

The use of consistent and inconsistent evidence with a set of 2-4-6 problems

WILLIAM A. STOCK and ERIC FREITAG
Arizona State University, Tempe, Arizona

Subjects solved four problems in a modified form of the 2-4-6 paradigm (Wason, 1960). Rather than generating triads in pursuit of a solution, subjects were shown four series of triads containing specified amounts of rule-inconsistent information. With respect to rule-generation performance, for three rules, increasing amounts of rule-inconsistent information adversely affected subjects' performance. However, for one rule, the reverse was true.

The falsification principle (Popper, 1962) states that scientific explanation is strengthened by empirical attempts to falsify hypotheses. What hypothesis-testing strategies people actually employ first received empirical scrutiny by Wason (1960, 1968), who demonstrated that subjects prefer corroboration to falsification. In these studies, subjects were given a three-number sequence (e.g., 2-4-6, hereafter labeled a triad), told that the sequence was an instance of a relational rule (e.g., three ascending numbers), and asked to identify the rule. The subjects repeatedly generated a triad; stated a reason for their choice; and recorded whether the triad was rule-consistent or not, as indicated by feedback. When the subjects believed they knew the rule, they wrote it out. The subjects seldom spontaneously used a falsification strategy.

Neither training (Mynatt, Doherty, & Tweney, 1978) nor instructions to use falsification (Tweney et al., 1980) uniformly facilitate performance. Gorman and his associates (e.g., Gorman, 1986; Gorman & Gorman, 1984; Gorman, Stafford, & Gorman, 1987) found mixed results for instructions—with some facilitation occurring when subjects generated two alternative rules at the same time.

In studies based on the 2-4-6 task, subjects generate triads—a good procedure for identifying the strategies that the subjects elect to use. However, this paradigm is not as effective for determining whether or not subjects can use disconfirming evidence, because the amount of disconfirming evidence available depends on the subject. Nor do studies of the effects of instructions directly address this issue, because subjects vary in the ability to generate triads falsifying a particular hypothesis. Only studies controlling the amount of rule-consistent/rule-inconsistent information directly address this issue.

Here, across different sets of subjects, we manipulated the number of triads consistent with a rule. All subjects

were exposed to four problems. On the basis of previous research, we expected subjects receiving higher rates of confirming evidence to perform better.

METHOD

Subjects and Design

Fifty-eight undergraduates participated for course credit. The design had a between-subjects factor, amount of rule-consistent information, with four levels: 100% (24), 83% (20), 66% (16), or 50% (12) of the triads associated with a rule were consistent. Sample sizes for these four conditions were 15, 15, 16, and 12, respectively. There were two within-subjects factors, rule (described below) and trials with 24 levels.

Materials

Using the 2-4-6 paradigm, we chose four different rules: (1) Wason's original three ascending numbers, (2) first and second number less than or equal to the third, (3) first and second number sum to the third, and (4) two even and one odd number in any order. For each rule, 60 rule-consistent and 60 rule-inconsistent triads were created. Using a table of random numbers, numbers (1-99) were selected until either a rule-consistent or a rule-inconsistent triad was constructed. Each triad was typed on a piece of paper 120 × 190 mm, preceded by a sentence stating, "The three numbers below are *consistent (inconsistent)* with the rule." The statement "Write down on this page what you think the rule is," followed by three ruled lines, was typed below the triad.

Test booklets were constructed as follows: For each rule, a series of 24 triads was constructed from the 60 consistent and 60 inconsistent triads. At each level of the between-subjects factor, series were constructed of 24 consistent (and 0 inconsistent), 20 consistent (4), 16 consistent (8), or 12 consistent (12) triads. Triads within a series, orders of presentation of triads, and order of rules were all separately randomized for each subject. Before each series, a sheet was inserted on which the instructions were briefly reviewed. After each series, a post-test was inserted, which contained three consistent and three inconsistent triads, each followed by the words *consistent* and *inconsistent*.

Procedure

Subjects were run individually. A subject read typed instructions silently as the investigator read them aloud. The instructions stated that a single-sentence rule governed the creation of the triads; that subjects would see a triad and be informed whether it was consistent or inconsistent with the rule; and that subjects were to examine the triad as long as they wished (but not look back to previous triads) and then write out the rule. After these instructions, all questions were allowed and answered. There followed a practice session with a series of four triads (two consistent and two inconsistent) for the rule "three even numbers all the same." After instructions and practice, each subject was asked if he or she understood the task (all indicated "yes"). Thereafter, each subject was handed the experimental materials to complete.

This research was supported by an Arts, Social Sciences, and Humanities Grant from the Office of the Vice President for Research, Arizona State University. Requests for reprints should be addressed to William Stock, Learning and Instructional Technology Program, Arizona State University, Tempe, AZ 85287-0611.

Table 1
Means and Standard Deviations of Proportion of Correctly Generated Rules as a Function of the Number of Rule-Consistent Triads and the Type of Rule

Number of Rule-Consistent Triads	Rule								Overall	
	1		2		3		4			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
24	.51	.40	.00	.00	.84	.24	.42	.42	.44	.14
20	.46	.35	.02	.06	.51	.41	.48	.40	.37	.17
16	.22	.33	.05	.14	.53	.45	.34	.45	.28	.23
12	.38	.33	.11	.26	.40	.42	.17	.32	.27	.16
Overall	.39	.36	.04	.14	.58	.41	.36	.41		

RESULTS

The rules generated by subjects were scored as correct if they were an accurate paraphrase of the actual rule. Sums of corrects were computed for each block of six triads in a series. Also computed were sums of correct selections of consistent and inconsistent triads on the posttest for each rule. Rule generation and posttest scores were analyzed in separate analyses of variance.

Rule Generation Scores

Main effects were observed for amount of consistent information [$F(3,54) = 2.96, p < .05$], rule [$F(3,162) = 24.13, p < .001$], and trial blocks [$F(3,486) = 69.29, p < .001$]. The interactions of rule with amount of consistent information [$F(9,162) = 1.96, p < .05$] and of rule with trial blocks [$F(9,486) = 5.66, p < .001$] were also significant.

Table 1 contains means and standard deviations for the data classified by rule and amount of consistent information.

One degree of freedom tests confirmed that the group receiving 24 consistent triads did not differ from the group receiving 20 consistent triads, but that it did perform better than the groups receiving 16 [$F(1,54) = 6.06, p < .02$] and 12 [$F(1,54) = 6.54, p < .02$] consistent triads.

A comparison of the marginal means for rules indicates that Rule 2 (first and second number less than or equal to the third) was considerably more difficult than the remaining three rules. Rules 1 (three ascending numbers) and 4 (two even and one odd number in any order) were about the same in difficulty, but Rule 3 (first and second number sum to the third) was quite easy.

Individual contrasts indicated that the interaction between rule and amount of consistent information was restricted to differences in performance between subjects receiving 24 and 12 consistent triads [$F(1,162) = 3.30, p < .03$]. Figure 1 depicts the performance profiles of these two groups. There is one exception to the general performance decrement associated with receiving just 12 consistent triads. For Rule 2, no subject receiving 24 consistent triads ever generated a correct expression of the rule, while subjects receiving just 12 consistent triads generated a correct expression on about 11% of the trials. An individual comparison of Rule 2 performance between these two groups was statistically significant [$F(1,54) = 4.40, p < .05$].

Figure 2 depicts the interaction of rule and trial block. An examination of this figure reveals that there is a general tendency for correct rule generation to increase across blocks.

Relative to Rules 1 and 3, Rules 2 and 4 have less steep slopes across trial blocks, thereby accounting for the rule \times block interaction.

Posttest Scores

Analysis of the posttest corrects revealed main effects for the amount of consistent information [$F(3, 54) = 3.28, p < .03$] and for rule [$F(3,162) = 20.67, p < .001$]. The interaction of these effects was not significant.

Individual contrasts indicated that the group receiving 24 consistent triads performed better than the group receiving 12 consistent triads [$F(1,54) = 7.74, p < .01$], but that it did not differ from the remaining two groups. Posttest means for the groups receiving 24, 20, 16, and

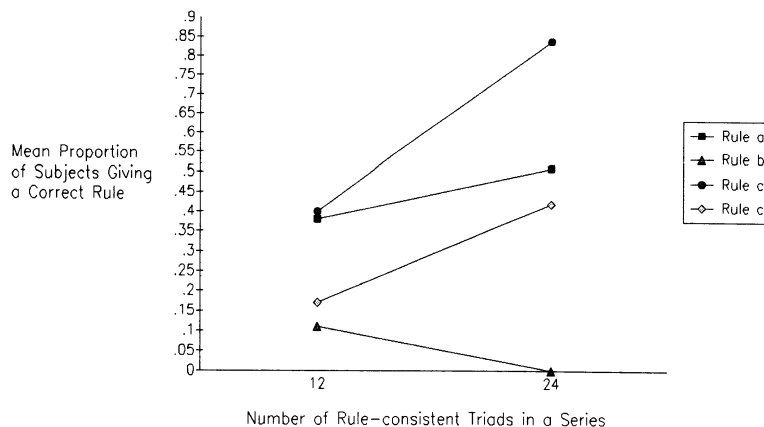


Figure 1. Mean proportion of subjects in the 12 and 24 rule-consistent triads conditions who generated a correct expression of a rule as a function of type of rule.

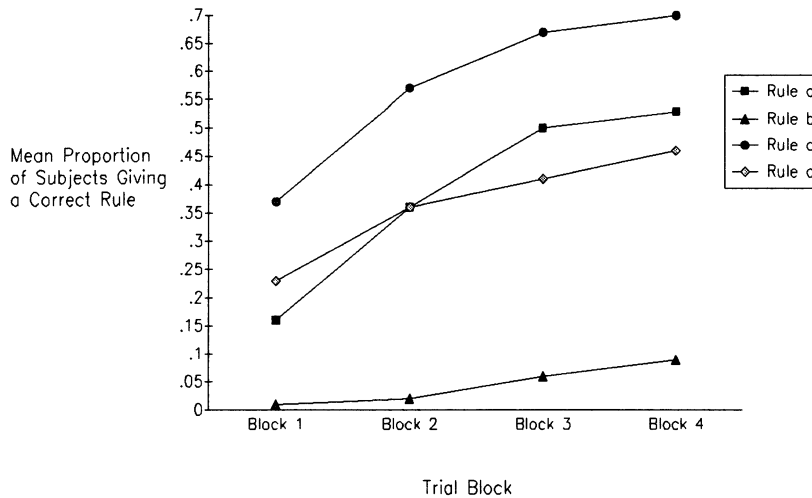


Figure 2. Mean proportion of subjects who generated a correct expression of a rule as a function of type of rule and trial blocks.

12 consistent triads were 4.75 ($SD = .55$), 4.68 ($SD = .54$), 4.34 ($SD = .75$), and 4.00 ($SD = .93$), respectively.

Posttest means for Rules 1, 2, 3, and 4 were 4.86 ($SD = 1.59$), 3.08 ($SD = 1.47$), 5.24 ($SD = 1.43$), and 4.67 ($SD = 1.67$), respectively. This ordering of means is identical to that obtained for rule-generation scores (see Table 1). Taken together, these results confirm that rules vary in their difficulty.

DISCUSSION

Previous research on falsification strategies has had two distinct phases. In the first phase, typified by Wason's research, experimental results clearly established that subjects seldom spontaneously adopt a strategy of falsifying hypotheses. Rather, they elect strategies with an intent to confirm a currently held hypothesis. The present results are consistent with this line of research, in that higher levels of rule identification are obtained when the evidence provided contains more rule-consistent than rule-inconsistent information.

Indeed, one might argue that subjects focus their attention on the rule-consistent triads and ignore rule-inconsistent triads. Performance for Rules 1, 3, and 4 is consistent with this argument, but performance for Rule 2 is not. On the average, subjects in the 24, 20, 16, and 12 consistent triad groups generated correct paraphrases for Rule 2 0%, 2%, 5%, and 11% of the time, respectively—with the difference between the 24 and 12 consistent triad groups being statistically significant. Under all conditions, Rule 2 was the most difficult, but some subjects receiving rule-inconsistent triads were solving the problem.

In the second phase of study of falsification strategies, typified by research by Mynatt and his associates (1978) and by Gorman and his associates (Gorman, 1986; Gorman & Gorman, 1984; Gorman et al., 1987), a variety of environmental manipulations (e.g., instructions in the use of different strategies, working alone vs. in groups, feedback veridicality, type of problem) have been examined, with equivocal results. In particular, instructions to use falsification strategies have not uniformly led to superior performance. Our results address instructional issues in an indirect, but relevant way.

The overall decrements in performance associated with increasing numbers of rule-inconsistent triads lead us to believe that subjects are less effective in their use of such information. If this is so, then instructions to use a falsification strategy will not necessarily benefit the inquiry

process. From a design point of view, we suggest that subsequent studies of falsification strategies incorporate preexperimental training for all subjects, with the object of training to bring all subjects to a standard level of proficiency on the use of rule-inconsistent information. Subsequently, any combination of strategies involving confirmation and falsification processes could be evaluated.

Offered for speculation is the thought that there exist some rules for which rule-consistent, relative to rule-inconsistent, triads facilitate identification, and vice versa. Our results suggest that Rule 2 may be an instance of a rule where rule-inconsistent information actually facilitates discovery. If a detailed logical analysis of rules were to establish that these distinct classes of rules did exist, then the use of rules that are neutral with respect to the consistency status of triads would represent an improvement in experimental methodology.

REFERENCES

- GORMAN, M. E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*, *77*, 85-96.
- GORMAN, M. E., & GORMAN, M. E. (1984). A comparison of disconfirmatory, confirmatory, and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, *36A*, 629-648.
- GORMAN, M. E., STAFFORD, A., & GORMAN, M. E. (1987). Disconfirmation and dual hypotheses on a more difficult version of Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, *39A*, 1-28.
- MYNATT, C. R., DOHERTY, M. E., & TWENEY, R. D. (1978). Consequences of confirmation and disconfirmation in a simulated research environment. *Quarterly Journal of Experimental Psychology*, *30*, 395-406.
- POPPER, K. R. (1962). *Conjectures and refutations*. New York: Basic Books.
- TWENEY, R. D., DOHERTY, M. E., WORNOR, W. J., PLISKE, D. B., MYNATT, C. R., GROSS, K. A., & ARKKELIN, D. L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, *32*, 109-123.
- WASON, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.
- WASON, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273-281.