Differential difficulty of categorical syllogisms

LOUIS S. DICKSTEIN
Wellesley College, Wellesley, Massachusetts 02181

Previous research with categorical syllogisms indicates that subject performance is better on valid syllogisms than on syllogisms which are indeterminate. The present study tested the hypothesis that the poorer performance on invalid syllogisms is a function of the disproportionate number of invalid syllogisms in the traditional syllogism task. Subjects may not expect so many invalid syllogisms and may be reluctant to draw so many nonpropositional conclusions. One group of subjects received the standard set of 19 valid and 45 invalid syllogisms, while a second group received a set of 35 valid and 29 invalid syllogisms. There were no significant differences between the groups on valid or invalid syllogisms, disconfirming the hypothesis. Several previous findings with categorical syllogisms were replicated.

Previous research utilizing both valid and invalid categorical syllogisms has found that performance is considerably better for syllogisms which have a specific propositional conclusion than for syllogisms which are indeterminate. Thus, Roberge (1970) has presented data for 59 syllogisms in which the percentage correct for 14 valid syllogisms was 51.2%, while the percentage correct for 45 invalid syllogisms was 35.8%. Similarly, Dickstein (1975) reported the percentage correct for all 19 possible valid syllogisms to be 72.6%, as compared to 58.2% for the 45 invalid syllogisms.

These data indicate that subjects are more willing to endorse specific propositions as conclusions than they are to endorse the alternative that no valid conclusion may be deduced from the premises. Revlis (1975a) has argued that subjects possess a response bias against nonpropositional conclusions, and he attributes this response bias to the universe of problems which constitute the traditional syllogistic reasoning task. A complete set of categorical syllogisms consists of 64 different premise combinations. Of these, only 19 yield propositional conclusions while 45 have no valid conclusion. Revlis argues that this imbalance between valid and invalid syllogisms biases the subject toward propositional conclusions because the subject does not expect so many invalid syllogisms. According to this explanation, the poorer performance on invalid syllogisms does not necessarily reflect some fundamental aspect of the reasoning process but, rather, is an artifact of the task situation.

The primary purpose of the present experiment was to test the hypothesis that performance on invalid syllogisms is a function of the proportion of valid and

invalid syllogisms constituting the task set. If the relatively poor performance on invalid syllogisms is due to their disproportionate representation in the task set, then performance on such syllogisms should be improved by utilizing a modified set of syllogisms in which the proportion of valid syllogisms is increased while the proportion of invalid syllogisms is reduced. A smaller proportion of invalid syllogisms should make subjects less reluctant to endorse nonpropositional conclusions.

A secondary purpose of the present experiment was to replicate two previous findings. Chapman and Chapman (1959) proposed that invalid syllogisms might be subdivided into two categories based upon the reasoning processes leading subjects to erroneous conclusions. One category consisted of syllogisms where erroneous acceptance of the converse of universal affirmative and/or particular negative premises could account for the predominant subject errors. The second category consisted of syllogisms in which the predominant errors could be attributed to probabilistic inference. Dickstein (1975) presented evidence in support of this distinction based upon the differential difficulty of the two categories of invalid syllogisms. In that study the percentage correct for the probabilistic inference syllogisms was 62.2%, while the percentage correct for conversion syllogisms was only 25.9%. A second finding in the study was that performance on the valid syllogisms was significantly correlated with performance on the invalid syllogisms ($r = .640$). A correlation of this magnitude suggests that similar processes are involved in the solution of both valid and invalid syllogisms. The present study was conducted, in part, to see whether these findings could be replicated.

METHOD

Subjects were 32 male and female undergraduates enrolled in an introductory psychology course and participating in the experiment to fulfill a course requirement. Sixteen subjects were assigned to a traditional group which received the standard set of 64 syllogisms containing 19 valid syllogisms and 45 invalid syllogisms. The remaining subjects constituted a modified

group. This group also received 64 syllogisms, but the set consisted of 35 valid and 29 invalid syllogisms.

The standard syllogism task consists of 16 syllogisms for each of four figures, where "figure" refers to the ordering of terms in the major and minor premises. The number of valid syllogisms varies in the different figures. Thus, there are four valid syllogisms in Figure 1, four in Figure 2, six in Figure 3, and five in Figure 4. The modified set was constructed by repeating 16 of the valid syllogisms, four for each figure. The only syllogisms which were not repeated were AA and OA in Figure 3 and IA in Figure 4. (The letters used to name the syllogisms constitute the standard nomenclature for the premises of categorical syllogisms. An "A" premise is a universal affirmative premise, an "E" premise is a universal negative, an "I" premise is a particular affirmative, while an "O" premise is a particular negative. Thus, an OA syllogism is one in which the major premise is a particular negative proposition, while the minor premise is a universal affirmative proposition.) The 16 additional valid syllogisms replaced 16 invalid syllogisms. Since there are only 13 conversion invalid syllogisms, while there are 32 probabilistic inference invalid syllogisms, it was decided to omit only probabilistic inference syllogisms. Again, four syllogisms were deleted from each figure. The 32 probabilistic inference syllogisms consist of eight different premise combinations, and two syllogisms for each premise combination were deleted. The syllogisms deleted were IE, IO, EO, and OE in Figure 1, II, IO, EE, and OE in Figure 2, IE, EO, OI, and OO in Figure 3, and II, EE, OI, and OO in Figure 4.

All subjects received problem booklets with instructions which clarified the rules of the syllogistic task. Thus, subjects were told that the word "some" means "at least some" and does not mean "only some." In addition, subjects were told that in order to draw a substantive conclusion the premises had to compel that conclusion. The instructions were part of the booklet and subjects could refer back to them at any time. The booklet consisted of 16 pages of syllogisms, with four syllogisms on a page. Each premise combination was followed by five response alternatives, and the subject was required to select one of them. The alternatives were "All S are P," "Some S are P," "No S are P," "Some S are not P," and "No valid conclusion." These alternatives were always presented in the same order to avoid confusing the subjects. Similarly, the letters S, M, and P were always used to represent the subject, middle term, and predicate to avoid confusion. In the modified set, care was taken that two identical syllogisms did not appear on the same page nor in consecutive succession. Subjects completed all of the syllogisms at a single session and no time limit was imposed.

RESULTS

The means and standard deviations for the number correct for both the traditional and modified groups for valid, invalid, conversion, and probabilistic inference syllogisms are presented in Table 1. For the traditional group the mean percentage correct for valid syllogisms was 64.8%, while the mean for the modified group was 59.7%. On invalid syllogisms the mean was 39.9% for the traditional group and 38.3% for the modified group. Within the set of invalid syllogisms, the mean for conversion syllogisms was 17.3% for the traditional group and 22.6% for the modified group. Finally, for the probabilistic inference syllogisms the mean was 49.0% for the traditional group and 51.2% for the modified group.[1] Since the two groups received different numbers of the different types of syllogisms, statistical analyses were conducted on percentage correct scores. Four t tests

Table 1
Means and Standard Deviations for Number Correct on Valid
and Invalid Syllogisms

Syllogisms	Traditional		Modified	
	M	SD	M	SD
Valid	12.31	3.03	20.88	4.83
Invalid	17.94	10.63	11.12	8.21
Conversion	2.25	2.38	2.94	3.49
Probabilistic Inference	15.69	10.46	8.19	5.31

were conducted and none of the comparisons reached significance, with the highest t value only 1.02.

In the traditional group, performance on valid syllogisms was significantly better than performance on invalid syllogisms $[t(15) = 5.52, p < .01,$ two-tailed]. In addition, performance on probabilistic inference syllogisms was significantly better than performance on conversion syllogisms $[t(15) = 5.01, p < .01,$ two-tailed]. Similarly, for the modified group, performance was significantly better on valid than on invalid syllogisms $[t(15) = 2.67, p < .05,$ two-tailed] and performance on probabilistic inference syllogisms was significantly better than on conversion syllogisms $[t(15) = 4.99, p < .01,$ two tailed].

The correlation between percentage valid correct and percentage invalid correct was computed for the two groups separately. For the traditional group, $r(14) = .642, p < .01,$ two-tailed, while for the modified group, $r(14) = -.002.$ Finally, since subjects in the modified group received repetitions of 16 valid syllogisms, performance on the first presentation of these 16 premise combinations was compared with performance on the second presentation. There was no significant difference in performance, with subjects achieving 60.2% correct on the first presentation and 60.6% on the second presentation.

DISCUSSION

The present study confirms the previous finding that performance on valid syllogisms is significantly better than performance on invalid syllogisms. For both the traditional and modified groups a differential of more than 20% was obtained between the valid and invalid syllogisms. However, the present study does not support the proposed explanation advanced by Revlis (1975a). Performance on invalid syllogisms was unaffected by the relative proportions of valid and invalid syllogisms. Thus, it appears untenable to maintain that subjects avoid nonpropositional conclusions because there are too many of them in the set.

These findings do not rule out the possibility that subjects may indeed have a response bias against nonpropositional conclusions. Rather, they demonstrate that such a bias cannot be explained as an artifact of the task structure. Perhaps subjects are simply uncomfortable with the conclusion that the two premises do not provide any definitive information. Subjects may wish to avoid the ambiguity of a nonpropositional conclusion and may, therefore, endorse conclusions which are possible even though they are not required by the premises.

An alternative possibility is that the errors subjects make on invalid syllogisms may result from the systematic operations

which subjects perform when solving reasoning problems. Subjects may systematically misinterpret the premises, or they may make systematic combination errors which then result in erroneous propositional conclusions. The processes of illicit conversion and probabilistic inference suggested by Chapman and Chapman (1959) illustrate such errors. Erickson's (1974) proposal of various interpretation, combination, and labeling processes provides another possible explanation. Although a completely satisfactory specification of the processes involved in erroneous reasoning has not yet been provided, it appears premature to conclude that performance must be due to response bias.

The present study replicates the finding (Dickstein, 1975) that subjects do better on probabilistic inference invalid syllogisms than on conversion invalid syllogisms. Thus, the distinction proposed by Chapman and Chapman appears well supported. The present study also replicates the significant correlation obtained previously between performance on valid and invalid syllogisms. These findings support the notion that there are common interpretation or combination processes occurring on both valid and invalid syllogisms, although, again, complete specification of these processes remains a task for the future.

The absence of a significant correlation between performance on valid and invalid syllogisms for the modified group was unanticipated. One possible explanation is that, since 16 invalid syllogisms were deleted for this group, the invalid syllogisms data did not provide as accurate an assessment of individual differences as is provided by the full set of invalid syllogisms.

Finally, the absence of any difference in performance between the first and second presentations of the 16 repeated valid syllogisms for the modified group replicates a finding by Revlis (1975b). It is not surprising since, in the absence of feedback about the correct answer, there is no reason to expect improvement simply on the basis of repeated exposure to the problems.

REFERENCES

CHAPMAN, L., & CHAPMAN, J. Atmosphere effect reexamined. *Journal of Experimental Psychology*, 1959, **58**, 220-226.

DICKSTEIN, L. S. Effects of instructions and premise order on errors in syllogistic reasoning. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **1**, 376-384.

ERICKSON, J. R. A set analysis theory of behavior in formal syllogistic reasoning tasks. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium.* Hillsdale: Lawrence Erlbaum, 1974.

REVLIS, R. Syllogistic reasoning: Logical decisions from a complex data base. In R. J. Falmagne (Ed.), *Reasoning: Representation and process.* Hillsdale: Lawrence Erlbaum, 1975. (a)

REVLIS, R. Two models of syllogistic reasoning: Feature selection and conversion. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 180-195. (b)

ROBERGE, J. A reexamination of the interpretation of errors in formal syllogistic reasoning. *Psychonomic Science*, 1970, **19**, 331-333.

NOTE

1. The percentage correct for the modified group for all invalid syllogisms combined is lower than that for the traditional group, even though the percentages correct for the modified group for both subsets of invalid syllogisms separately are higher than for the traditional group. This occurs because the two subsets are of differential difficulty and the two groups received different proportions of the two subsets.