Nonindependence of selections on the Wason selection task

P. POLLARD
Plymouth Polytechnic, Plymouth, Devon, England

(John T. E. Richardson, *sponsor*)

Card selections from selection tasks presented by Evans and Lynch (1973) and Manktelow and Evans (1979) were tested for independence. A strong positive association was found between selection (or nonselection) of the two mismatching cards, and a strong negative association was found between selection of the two antecedent cards. In general, selections were associated on the basis of their matching status, although these associations were often weak (with the exception of the above) and were consistently significant only for tasks using a doubly negated conditional rule. These analyses are compatible with recent explanations of matching bias, but are incompatible with the suggestion of an (additional) verification tendency. Some logical consistency between tasks is also displayed.

On a typical form of the "Wason selection task" (Wason, 1966, 1968), the subject is informed that there is a deck of cards showing letters on one side and numbers on the other, from which four cards have been taken. The subject is required to test the truth or falsity of a conditional statement about these four cards, of the general form "if P then Q" (e.g., "if the letter is an A, then the number is a 5"). The four cards are chosen such that they represent one instance each of TA and TC ("true," i.e., confirming, instances of the antecedent and consequent of the rule, e.g., "A" and "5") and one instance each of FA and FC ("false," i.e., nonconfirming, instances of the antecedent and consequent, e.g., "B" and "7"). The subject is asked to select those cards that it would be necessary to investigate (turn over) in order to test the rule. The rule can (only) be tested by searching for falsifying instances, and thus the appropriate decision is to select the potentially falsifying TA and FC cards, a selection combination that subjects rarely adopt. Most subjects select TA only or TA and TC (for recent reviews, see Evans, 1982, and Wason, 1983).

The typical TA and TC selection suggests a verification bias, because both cards can yield a positive instance of the rule (e.g., Johnson-Laird & Wason, 1970). However, Evans and Lynch (1973) suggested that subjects may simply be selecting the values mentioned in the rule, a "matching bias" previously noted on a different task using conditional rules (Evans, 1972). By systematic insertion of negatives into the rule (e.g., "if not P then not Q"), Evans and Lynch demonstrated that selections were indeed strongly affected by matching (although there

was also a tendency to select TA, whether or not it matched). For instance, most subjects select P and Q on both the rule "if P then Q" and on the rule "if P then *not* Q" (see also Wason & Evans, 1975). *Mismatching* cards (i.e., cards showing values not mentioned in the rule) are designated as $\bar{P}$ and $\bar{Q}$. On affirmative rules, TA and TC are clearly P and Q, and FA and FC are $\bar{P}$ and $\bar{Q}$, but these relationships change when negatives are introduced. For instance, for the rule "if A then *not* 5," a card showing "7" could be referred to as TC or as $\bar{Q}$. Any card may thus be uniquely designated in terms of either its logical or its matching status.

This paper is concerned with the presence, or absence, of associations between selections of pairs of cards. Evans (1977) concluded that it is "reasonable to regard subjects choices *as if* they were independent, until we have evidence to the contrary." A second conclusion that follows from this was that *combinations* of selections depending on a common logical principle have no psychological validity (i.e., such logical principles do not mediate subjects' selections). It follows from this, in turn, that the formerly common practice of analyzing selections in terms of such combinations was based on an erroneous assumption. A third conclusion drawn by Evans (1977) was that subjects' selection behavior could be adequately modeled in terms of a *stochastic* model, assuming probabilistic behavior *within*, rather than between, subjects, as opposed to the deterministic "insight" or strategy models that had previously been proposed.

Evans (1977) pointed out that previous studies, with the exception of Evans and Lynch (1973), had used only affirmative rules on which TA (P) is almost always selected and FA ($\bar{P}$) is rarely selected. Such ceiling effects make tests of association virtually impossible, and therefore Evans confined his analyses to tests of association between Q and $\bar{Q}$. The evidence for independence thus rests en-

tirely on the nonsignificant association between selection of these cards on affirmative rules. Fortunately, ceiling effects are less in evidence when negated rule forms are used, and a much broader test of null hypotheses concerning associations can be conducted on the data from these. Specifically, an analysis is conducted here on the data of Evans and Lynch (1973) and Manktelow and Evans (1979), two studies that used all four forms of conditional rule formed by systematic negation of each component. That is: "if P then Q" (AA), "if P then not Q" (AN), "if not P then Q" (NA), and "if not P then not Q" (NN).

Tests of all six possible associations were conducted, on each of the four rules, in the following manner: (1) Tests of association within each study were conducted by use of Fisher's exact tests (in accordance with Evans, 1977). Because the authors found no significant differences between the groups, data from abstract and thematic groups in the two experiments of Manktelow and Evans (1979) were pooled, producing N=48·in each case (as opposed to N=24 for the Evans & Lynch, 1973, data). (2) For this purpose, Fisher's exact tests were conducted in the direction of the association present, and a positive or negative Z value was assigned to each result. Z was set equal to zero whenever the tests yielded $p > .5$ in either direction (because such a result indicates a two-tailed probability of 1.0). (3) For each of the 24 tests (six pairs of associations on each of four rules), the Z values derived from each of the three sets of data were used to produce a combined estimate of Z for that association by the Stouffer method (see Rosenthal, 1978). The (two-tailed) significance of the association was then derived directly from the Z value. The results of these tests are shown in Table 1. For each pair of selections, the Z value obtained from each study on each rule form is shown, together with the combined estimate of Z.

Although the associations are often nonsignificant, except in the case of the NN rule, Table 1 shows noticeable consistency in the signs of the combined Z estimates. For each comparison, the sign is the same for each rule. Furthermore, these signs are all consistent with the idea that selections are positively associated when they have similar matching status and negatively associated when they have dissimilar matching status. In particular, decisions about the two mismatching cards ($\bar{P}$ and $\bar{Q}$) are significantly associated on all four rules. This result suggests that matching bias should more properly be regarded as a *mismatching bias*. However, these data show no evidence of consistent associations on the basis of logical status. In the important case of TA and FC selection (the logically correct solution), the association is significantly positive only on the NA rule, on which TA and FC are $\bar{P}$ and $\bar{Q}$, which are always strongly associated. The association between TA and FC is nonsignificantly positive on the AN rule, but is significantly *negative* on both the AA and NN rules.

Regarding tests within antecedent selections, it seems reasonable to conclude that these are negatively associated. All signs are negative within each study, and

Table 1
Z Values Derived From Fisher's Exact Tests of Association Between Pairs of Card Selections on the Four Rules: Data From Evans and Lynch (1973) and Manktelow and Evans (1979, Experiments 1 and 2)

Comparison	Rule Form	Study From Which Z is Derived			Combined Z Estimate
		E & L	M & E1	M & E2	
Q vs. $\bar{Q}$	AA	0.00	−0.21	−1.28	−0.86
	AN	−0.98	0.00	+0.29	−0.40
	NA	+0.55	−2.19	0.00	−0.95
	NN	0.00	−3.11	−1.94	−2.92**
P vs. $\bar{P}$	AA	−0.70	−0.60	−1.84	−1.81
	AN	−1.38	0.00	−0.34	−0.99
	NA	−2.50	−2.94	−2.69	−4.69***
	NN	−2.06	−4.00	−4.00	−5.81***
P vs. Q	AA	0.00	0.00	+0.87	+0.50
	AN	+0.98	0.00	0.00	+0.57
	NA	−0.15	+2.09	+0.32	+1.30
	NN	0.00	+2.41	+2.65	+2.92**
P vs. $\bar{Q}$	AA	−0.67	−1.17	−2.28	−2.38*
	AN	−2.68	−0.50	+0.30	−1.66
	NA	−0.72	−1.17	+1.01	−0.51
	NN	−0.89	−2.38	−2.91	−3.57***
$\bar{P}$ vs. Q	AA	0.00	−0.12	−1.24	−0.79
	AN	−0.21	−0.65	−1.07	−1.11
	NA	0.00	−1.40	0.00	−0.81
	NN	0.00	−2.46	−2.29	−2.74**
$\bar{P}$ vs. $\bar{Q}$	AA	+1.27	+0.87	+2.66	+2.77**
	AN	+1.38	+2.26	+1.60	+3.03**
	NA	+0.89	+1.59	+1.67	+2.40*
	NN	+1.82	+1.35	+3.98	+4.13***

Note— E & L = Evans and Lynch (1973); M & E1 = Manktelow and Evans (1979, Experiment 1); M & E2 = Manktelow and Evans (1979, Experiment 2). *$p < .05$. **$p < .01$. ***$p < .001$.

the lack of significance of the combined Z estimate on rules having affirmative antecedents may fairly be seen as a function of almost universal P, and negligible $\bar{P}$, selection.

It is worth looking at another type of association that may exist—sequential associations between selection of the same card on different rules. Using a conditional inference task, Pollard and Evans (1980) produced data revealing "no evidence of consistent MT [modus tollens] affirmation across all four rules." Pollard and Evans interpreted this lack of consistency as showing that subjects have little appreciation of the logical validity of modus tollens, even when accepting it. The analyses reported in Table 1 are consistent with their findings. However, associations on the basis of matching, but not logical, status would also be expected for sequential presentations of the task.

It was decided to test this on the most important selection in each category. In terms of logical status, FC selection appears to be most important for two reasons. First, although there is a strong tendency for subjects to select TA, there is little evidence that FC selections are generally made for other than nonlogical reasons. Second, unlike TA selections (which may provide either a verifying or a falsifying instance), FC selection can produce only a falsifying instance. Thus, if consistency in the selec-

Table 2
Z Values Derived From Fisher's Exact Tests of Association Between Selection of $\bar{Q}$
or FC Across Different Pairs of Rules: Data From Evans and Lynch (1973) and
Manktelow and Evans (1979, Experiments 1 and 2)

Comparison		Study From Which Z is Derived			Combined Z Estimate
		E & L	M & E1	M & E2	
AA vs. AN	(FC)	0.00	+2.00	+0.18	+1.26
	(Q)	0.00	−0.24	0.00	−0.14
AA vs. NA	(FC and $\bar{Q}$)	+1.02	+1.86	+0.94	+2.21*
AA vs. NN	(FC)	−0.18	+0.95	−1.28	−0.29
	(Q)	0.00	−0.56	+1.09	+0.31
AN vs. NA	(FC)	0.00	+2.63	+1.51	+2.39*
	(Q)	0.00	0.00	+2.35	+1.36
AN vs. NN	(FC)	0.00	+0.19	+1.67	+1.07
	(Q)	0.00	+0.01	+1.27	+0.74
NA vs. NN	(FC)	+1.30	+0.56	+0.50	+1.36
	(Q)	0.00	+0.18	+2.04	+1.28

Note—See Table 1 for definitions of E & L, M & E1, and M & E2. *$p < .05$.

tion of FC were observed, this would be unequivocally related to a falsification tendency. In the case of matching status, the previous analysis indicates that selection of the *mismatching* $\bar{Q}$ card is appropriate for investigation of consistent susceptibility to the bias.

Table 2 shows, for each of the six possible pairs of rule forms, the Z values obtained from Fischer's exact tests of associations between FC selection, and $\bar{Q}$ selection, on the two rules. It should be noted that, FC is $\bar{Q}$ on AA and NA rules, and thus only one set of tests is given for the comparison between these rules. It can be seen that there is a tendency for both FC and $\bar{Q}$ selections to be positively correlated across rules, although these associations are generally much weaker than those shown in Table 1.

DISCUSSION

With regard to the mathematical model proposed by Evans (1977), some modification of the assumptions of the original fit has been shown to be necessary by Pollard (1979) (see Evans, 1982, pp. 178-179). An alternative approach was adopted by Krauth (1982), who coped with the problems by proposing a slightly different model, which allows for existence of a verification bias. The lack of independence between selections, however, may be considered a challenge to both these models, which assume that probabilities for the selection of each card may be computed independently. It should be noted additionally that this nonindependence is observed here in data actually used in both the Evans (1977) and Krauth (1982) analyses. However, the present analyses have shown that Evans was correct in arguing that combinations of selections depending on a common *logical* principle have no psychological validity, at least on "abstract" tasks.

In that these analyses show clear consistencies on the basis of matching bias, they are compatible with theoretical explanations of reasoning that have seen matching bias

as an example of a genuine and generalizable psychological phenomenon (e.g., Evans, 1983, 1984; Pollard, 1982). One suggestion that they are certainly not compatible with, however, is the proposal (e.g., Krauth, 1982) that "abstract" selection task data are best accounted for by assuming a verification tendency in addition to matching and falsification (i.e., "logical") tendencies. From Table 1, it can be seen that the verifying selections tend to be positively associated on two rules but also tend to be negatively associated on the other two rules. Thus, there is no evidence of consistency in terms of verification.

REFERENCES

EVANS, J. ST. B. T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, **24**, 193-199.

EVANS, J. ST. B. T. (1977). Toward a statistical theory of reasoning. *Quarterly Journal of Experimental Psychology*, **29**, 621-635.

EVANS, J. ST. B. T. (1982). *The psychology of deductive reasoning.* London: Routledge & Kegan Paul.

EVANS, J. ST. B. T. (1983). Selective attention in reasoning. In J. St. B. T. Evans (Ed.), *Thinking and Reasoning: Psychological approaches.* London: Routledge & Kegan Paul.

EVANS, J. ST. B. T. (1984). Heuristic and analytic processes in reasoning. *British Journal of Psychology*, **75**, 451-468.

EVANS, J. ST. B. T., & LYNCH, J. S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391-397.

JOHNSON-LAIRD, P. N., & WASON, P. C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, **1**, 134-148.

KRAUTH, J. (1982). Formulation and experimental verification of models in propositional reasoning. *Quarterly Journal of Experimental Psychology*, **34A**, 285-298.

MANKTELOW, K. I., & EVANS, J. ST. B. T. (1979). Facilitation of reasoning by realism: Effect or non-effect? *British Journal of Psychology*, **70**, 477-488.

POLLARD, P. (1979, December). *Evans' stochastic reasoning model: A re-evaluation.* Paper presented at the London Conference of the British Psychological Society.

POLLARD, P. (1982). Human reasoning: Some possible effects of availability. *Cognition*, **12**, 65-96.

POLLARD, P., & EVANS, J. ST. B. T. (1980). The influence of logic on conditional reasoning performance. *Quarterly Journal of Experimental Psychology*, **32**, 605-624.

ROSENTHAL, R. (1978). Combining results of independent studies. *Psychological Bulletin*, **85**, 185-193.

WASON, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology*. London: Penguin.

WASON, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273-281.

WASON, P. C. (1983). Realism and rationality in the selection task. In J. St. B. T. Evans (Ed.), *Thinking and reasoning: Psychological approaches*. London: Routledge & Kegan Paul.

WASON, P. C., & EVANS, J. ST. B. T. (1975). Dual process in reasoning? *Cognition*, **3**, 141-154.