# Hypothesis testing in a rule discovery problem: When a focused procedure is effective

Sandrine Rossi
*University of Caen, Caen, France*
Jean Paul Caverni
*CNRS and University of Provence, Aix-en-Provence, France*
Vittorio Girotto
*CNRS and University of Provence, Aix-en-Provence, France and University of Trieste, Trieste, Italy*

We investigated individuals' ability to use negative evidence in hypothesis testing. We compared performance in two versions of Wason's (1960) rule discovery problem. In the original version, a triple of numbers—{2, 4, 6}—was presented as an example of a rule that the experimenter had in mind (i.e., "increasing numbers"). Participants had to discover the rule by proposing new triples. In the other version, the same triple was presented as a counter-example to the experimenter's rule (i.e., "decreasing numbers"). We predicted that, in both conditions, participants would form hypotheses based on the features of the triple, and test only instances of the hypothesized rule. However, in the counter-example condition, such focused testing would invariably produce negative evidence. As a consequence, participants would be forced to revise their hypotheses. The reported results corroborated our predictions: Participants solved the counter-example version significantly better than the original problem.

Can naive individuals efficiently use negative evidence in testing their hypotheses? Limitations in human hypothesis testing have been consistently found in several laboratory situations, in particular in "rule discovery" tasks, the best known of which is Wason's (1960) 2–4–6 problem: Participants are given a triple of numbers {2, 4, 6}, which fits a rule that the experimenter has in mind, and they have to discover the rule by proposing new triples of numbers. For each triple, the experimenter indicates whether or not it fits the correct rule. Participants continue to propose triples and receive the experimenter's

Requests for reprints should be sent to Sandrine Rossi, Laboratoire de Psychologie Cognitive et Pathologique, UPRES EA 1774, Université de Caen, Esplanade de la Paix, 14032 Caen Cedex, France.
Email: rossi@scvie.unicaen.fr
We thank Michel Gonzalez, Alice McEleney, and two anonymous referees for their helpful comments.


http://www.tandf.co.uk/journals/pp/02724987.html

feedback until they feel "highly confident" that they have discovered the correct rule. Typically, participants test rules such as "increasing even numbers", and do so by proposing only instances that fit these rules (e.g., {6, 10, 18}). The point is that the experimenter's rule (i.e., "increasing numbers") is more general than those hypothesized by participants. Given that all instances of the hypothesized rules are instances of the correct one, the testing procedure invariably produces confirming evidence. Thus, only very few participants discover the experimenter's rule at their first attempt.

In our view, there are two main sources of error in the 2–4–6 problem. On the one hand, individuals have a general tendency to focus on what they expect to be true rather than what they expect to be false (e.g., Evans, 1989; Legrenzi, Girotto, & Johnson-Laird, 1993). On the other hand, there are pragmatic reasons to assume that the salient features of the {2, 4, 6} sequence (i.e., even numbers increasing by two) are relevant (Girotto & Politzer, 1990). As a consequence, individuals are likely to form narrow hypotheses based on these features (e.g., "increasing even numbers", "increasing by two") and to conduct focused testing by proposing only instances that fit the hypothesized rule. In some cases, focused testing, or a "positive test strategy", may be effective (Klayman & Ha, 1987). However, in the 2–4–6 problem, this testing procedure yields erroneous solutions, given that it could never produce a conclusive falsification of the hypothesized rule.

The aim of this paper is to establish whether individuals can solve the 2–4–6 problem, despite their tendency to form narrow hypotheses, and to test them in a focused way. Let us suppose that the initial triple {2, 4, 6} is presented as a *counter-example* to, rather than as an example of, the experimenter's rule. In this case too, participants are likely to form hypotheses based on the salient features of the initial counter-example, such as "increasing *odd* numbers". Now, if they conduct a focused test of these hypotheses (i.e., if they test triples such as {3, 5, 7}), and if the experimenter's rule is "decreasing numbers", their procedure will invariably produce negative evidence. Indeed, no instance of the correct rule is an instance of the hypothesized rule. As a consequence, participants will subsequently eliminate their hypotheses about increasing triples, enhancing the probability of discovering the experimenter's rule. In other words, we predict that individuals may solve the 2–4–6 problem if they are led to test hypotheses, examples of which are counter-examples to the correct rule (and vice versa). Our predictions have been tested in the following experiment.

## Method

### *Participants*

Forty-eight undergraduates at the University of Provence participated in the experiment in order to fulfil a class requirement. They were randomly assigned to one of two equal-sized conditions: example and counter-example ($N$ = 24 in each condition).

### *Procedure*

In both conditions, participants were tested individually. The procedure was based on the classical one (Wason, 1960). Participants were given the triple {2, 4, 6}, presented as an example ("this triple follows the rule"), or as a counter-example ("this triple does not follow the rule") to a rule that they had to discover ("increasing numbers" vs. "decreasing numbers", respectively). The rest of the problem was the same in the two conditions. Participants had to generate new triples, receiving the experimenter's feedback. When confident of knowing the rule, they had to announce it. If they announced the correct rule, the experiment ended. Otherwise, they had to generate more triples and continue with attempts to discover it. Participants were provided with a lined response sheet, with three columns, labelled (from left to right): *triple* (in which participants had to record the proposed triple), *hypothesis* (in which they had to record the tested hypothesis), and *test result* (in which they had to record the experimenter's response: yes or no).

## Results

As predicted (see Table 1), there were more first-attempt solvers in the counter-example condition than in the example condition, $\chi^2(1) = 14.2$, $p < .001$. Participants in the former condition tested more triples before announcing their first rule than did participants in the latter, $t(46) = 3.22$, $p < .002$, and obtained more negative feedback, $t(46) = 7.05$, $p < .001$.

In both conditions, most participants tested an initial rule that concerned the "increasing numbers" feature. That is, they focused on the relevant features of the {2, 4, 6} triple. The most frequent initial rule was "ascending even numbers" in the example condition (58%), and "ascending odd numbers" in the counter-example condition (30%). This finding corroborates our prediction that participants focus on the salient features of the initial triple, regardless of its logical status.

One might argue that the counter-example condition is easier than the example condition because the "decreasing numbers" rule could be more accessible than the usual "increasing numbers" rule. In order to rule out this interpretation, a further group of 24 participants (from the same pool as that in the other conditions) was given a {6, 4, 2} triple presented as an example of the experimenter's rule (i.e., "decreasing numbers").

In this condition, the performance turned out to be similar to that produced in the example condition, and significantly inferior to that produced in the counter-example condition, $\chi^2(1) = 7.36$, $p < .01$. Compared to the latter condition, participants in the {6, 4, 2} condition tested fewer triples before announcing their first rule, $t(46) = 3.84$, $p < .001$, and obtained less negative feedback, $t(46) = 6.66$, $p < .001$.

In sum, the counter-example procedure (rather than the specific nature of the experimenter's rule) improved performance. Participants in the counter-example condition

TABLE 1
Performance of participants in the three conditions

| *Conditions*[a] | *Percentage of first-attempt solvers* | *Percentage of initial focused testing* | *Mean number of tested triples*[b] | *Mean number of negative feedback received*[b] |
|---|---|---|---|---|
| {2, 4, 6} counter-example | 54 | 88 | 8.7 | 5.5 |
| {2, 4, 6} example | 4 | 92 | 4.8 | 0.5 |
| {6, 4, 2} example | 16 | 83 | 4.9 | 0.6 |

[a]*N* = 24 per condition. [b]Before announcing the first rule.

received more negative feedback than those in the other conditions, and they were more likely to revise their hypotheses and therefore more likely to solve the problem than were participants in the other conditions.

## Discussion

In this paper, we have presented a procedure for improving performance in a rule discovery problem, without instructing individuals to use a disconfirmatory strategy. Unlike the latter procedure, which elicited rather mixed results (Evans, 1989), our procedure appears to elicit genuine solutions (Caverni & Rossi, 1997). These results show that individuals may correctly use falsifying evidence, gathered from experience, in their inductive hypothesis testing.

In the 2–4–6 problem, participants are likely to focus on the salient features of the {2, 4, 6} triple, both when it represents an example and when it represents a counter-example of the correct rule ("three numbers in increasing order" vs. "three numbers in decreasing order", respectively). Participants are likely to form hypotheses based on these features, and to conduct a focused test on them. However, the {2, 4, 6} triple presented as a counter-example to the correct rule is likely to elicit hypotheses (e.g., "three odd numbers in increasing order"), instances of which are not instances of the correct rule. Hence, if participants conduct a positive test of these hypotheses, they will invariably obtain negative evidence. The systematic falsification of their hypotheses leads them to discover the correct rule.

Klayman and Ha (1987, p. 215) noted that when the hypothesized rule and the correct one are "disjunct" (i.e., no instance of the former is an instance of the latter, and vice versa), individuals may obtain conclusive falsification by conducting focused testing. However, they concluded that this situation "is not likely in the 2–4–6 task, because you are given one known target instance to begin with". We concur with these authors. However, our results show that the 2–4–6 task can also produce such a situation, provided that the initial triple is presented as a counter-example to the experimenter's rule and, of course, provided that the rule is "decreasing numbers". (Suppose that the experimenter's rule were "the sum of the three numbers is an odd number". In that case, as in the example condition, focused testing would not provide negative evidence, given that all instances of the hypothesized rules would be instances of the correct one.)

The counter-example procedure may be related to the most successful manipulation reported in the literature. Tweney et al. (1980) found that individuals solved a version of the 2–4–6 problem, in which they had to discover two rules: One was Wason's original "increasing numbers" rule. The triples that fitted it were named DAX. The alternative rule generated all other triples, named MED. In the dual-task version, as in the original problem, participants are likely to form hypotheses based on the features of the initial triple (e.g., "increasing even numbers" for the DAX rule vs. "increasing odd numbers" for the MED rule), and to conduct focused testing (e.g., proposing triples such as {6, 8, 10} vs. {3, 5, 7}, respectively). Focused testing of the DAX rule will produce the usual confirming evidence (all instances of the hypothesized DAX rule are instances of the correct DAX rule, i.e., "increasing numbers"). However, given that the instances of the hypothesized MED rule are actually instances of the correct DAX rule, focused testing of the MED rule will falsify participants' hypotheses. That is, they will discover that a hypothesized MED triple such as {3, 5, 7} is actually a DAX triple. As a consequence, participants are forced to eliminate their initial hypotheses for both rules, increasing the probability of discovering the experimenter's one. Unlike the counter-example version, the dual task has a different structure from the original problem. However, in both cases, a focused test of incorrect hypotheses leads participants to solve the 2–4–6 problem.

The present findings corroborate some recent results on hypothesis testing, which showed that naive individuals solve problems requiring them to establish the probability of a hypothesis, on the basis of a given datum (e.g., Girotto & Gonzalez, 1999). In those cases, individuals produced the correct solution when they were forced to represent an alternative hypothesis right from the start. In the present case, despite the initial, incorrect representation, individuals produced the correct solution when the negative results of their testing procedure forced them to eliminate the initial hypothesis and to consider an alternative one.

## REFERENCES

Caverni, J.P., & Rossi, S. (1997). A nice bit of scandal: About a disconfirmation bias in the Wason's 2–4–6 problem. *Swiss Journal of Psychology*, *56*, 239–242.

Evans, J.St.B.T. (1989). *Bias in human reasoning. Causes and consequences*. Hove, U.K.: Lawrence Erlbaum Associates Ltd.

Girotto, V., & Gonzalez, M. (1999). Strategies and models in statistical reasoning. In W. Schaeken, G. De Vooght, A. Vandierendonck, & G. d'Ydewalle (Eds.), *Strategies in deductive reasoning*. Mahwah, NJ, Lawrence Erlbaum Associates, Inc.

Girotto, V., & Politzer, G. (1990). Conversational and world-knowledge constraints in deductive reasoning. In J.P. Caverni, J.M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases*. Amsterdam: North-Holland.

Klayman, J., & Ha, Y.W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Legrenzi, P., Girotto, V., & Johnson-Laird, P.H. (1993). Focussing in reasoning and decision making. *Cognition*, *49*, 37–66.

Tweney, R.D., Doherty, M.E., Worner, W.J., Gross, K.A., & Arkkelin, D. (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.

*Original manuscript received 15 February 1999*
*Accepted revision received 29 October 1999*