Caverni, J. P., Rossi, S., & Péris, J. L. (2000). The alternatives taken into account in hypotheses testing: Two new paradigms for investigating strategies In J.A. Garcia-Madruga, N. Carriedo, & M.J. Gonzalez-Labra (Eds.), pp. 133-141. *Mental Models in Reasoning.* Madrid: UNED.

The alternatives taking into account in hypothesis testing:
Two new strategies tracing paradigms

Jean-Paul Caverni, Sandrine Rossi and Jean-Luc Péris

Centre de Recherche en Psychologie Cognitive
Cnrs & University of Provence, France[1]

---

[1] Correspondence should be adressed to Jean-Paul Caverni, Crepco, Université de Provence, 29 Av. R. Schuman, F-13621 Aix-en-Provence Cedex, France. E-mail: caverni@newsup.univ-mrs.fr

One of the main problem in the study of reasoning by human beings is to explain difficulties they have to consider alternatives. Numerous studies are concerned with the fact that people tend to focus on favoured hypotheses (e.g. Legrenzi, Girotto & Johnson-Laird, 1993) . This phenomenon, studied in cognitive and social psychology, has been referred to the preservation of beliefs and confirmation bias. The confirmation bias is defined by a tendency people exhibit to seek confirmation rather than disconfirmation of their hypotheses (Wason, 1960, 1962, 1966, 1968). They failed to produce counterexamples and to consider alternative hypotheses. This definition is linked to Karl Popper 's philosophy of science and to the refutation's criterion (Popper, 1959).

Hypothesis testing is an essential component of mental activities. Nevertheless, there is no consensus about cognitive processes involved In this paper, we report studies whose the aim is to examine a quite provocative question. Is the failure to consider alternatives a phenomenon really related to cognitive processes properties or does it come from any inefficiency of the tracing cognitive processes classical paradigm? To answer this question, we used the well-known 2-4-6 problem (Wason, 1960).

The Wason 2-4-6 problem: Classical results and interpretations

Let us first remember the 2-4-6 problem. The experimenter tells the participant that s/he has a rule in mind regarding number triples (i.e. " all sequences of increasing numbers "). The triple 2-4-6 is then presented as satisfying this rule. The participant is to attempt to discover the rule by proposing other triples.

For each proposed triple, the experimenter says to her/him whether or nor it follows the rule (yes-no). The participant is provided with a sheet on which s/he writes down (1) the proposed triple, (2) the tested hypothesis, and (3) the experimenter's feedback. Whenever her/him thinks s/he has discovered the rule, s/he states it. Then, the experimenter says whether it is correct or not. If incorrect, the participant continues to propose triples until s/he finds the rule or gives up.

Conventionally, the strategies implemented by participants have been traced by comparing the current triple the participant proposes and the current hypothesis s/he claims to test. Typically, participants tend to adopt hypotheses that are a subset of the rule, for example "increasing by two", and propose triples like "10-12-14" or "20-22-24". With the process tracing paradigm presented above, it is shown that in general participants propose triples that are positive examples of the hypotheses they claim to have tested. Even if participants discovered the rule, a few of them discovered it on the first announcement.

The prevailing interpretation for quite some time was that participants are prone to a *confirmation bias*: they would rather verify (confirm) their hypotheses than falsify (disconfirm) them. Such a behaviour has been claimed to be ineffective. In fact, because the to-be-discovered rule is "all sequences of increasing numbers", when giving triples as "10-12-14" or "20-22-24" participants are told that these triples satisfy the rule and receive an ambiguous verification of their (wrong) hypothesis (such as "increasing by two"). A good triple to receive a conclusive refutation would be a negative example (a counterexample) of the tested hypothesis, as for instance "10-20-30".

Two competitive interpretations have been proposed. For Evans (1983, 1989), if the phenomenon is really proceeding from a bias, it is not caused by a participants' deliberate choice, but by a cognitive difficulty in processing negative information: "It is not that participants do not wish to falsify, it is simply that they cannot think of the way to do it" (Evans, 1989, p. 143). According to Evans, the confirmation bias was not a motivational bias but the result of a set of cognitive failures. The failure is caused by a pre-conscious selective processing, that is, a bias to think about positive rather than negative information. This *positivity bias*, in most cases, leads participants to propose positive examples rather than negative examples of their current hypotheses. For Klayman and Ha (1987, 1989), positive hypothesis testing is not necessarily a bias. They considered participants' behaviour as a manifestation of a *positive test strategy*. This strategy entails testing hypotheses by considering the cases in which the property is expected to occur or it is known to have occurred. It's not incompatible with a deliberate search for falsification and it's a very good heuristic in many situations. It is only because of the particular relationship between the participant's hypothesis (specific) and the rule (general) that such a strategy turns out to be ineffective in Wason's original task. That is, the 2-4-6 problem, by its very nature, focuses the participants' attention on the specific property of the triple and therefore draws his attention away from the more general property that might apply also (that is, simply a series of increasing numbers). A hypothesis consistent with the triple's specific property like increasing by two, necessarily leads to positive feedback but does not allow the participant to discover the rule. All triples of numbers increasing by two also fit the experimenter's rule. In fact, as said before, in this particular situation, it's only by using a negative example (or test), such as "1-2-3", that the participant is in a position to falsify his hypothesis and thus eventually arrive at the solution. But let us suppose that the to-be-discovered rule was more specific than the hypothesis being tested, the latter could be rejected only by testing triples that verify it, but falsify the to-be-discovered rule. For instance, if the rule was "three consecutive even numbers" and the participant tested the hypothesis "three increasing even numbers" using the triple "4-8-10", then s/he would receive a negative feedback from the experimenter which would disconfirm her/his hypothesis. Klayman and Ha suggest that people use the positive test strategy as a general default heuristic in which the consequences vary with the characteristics of the task. This strategy is one that people use in absence of specific information that identifies some tests as more relevant than others, or when the cognitive demands of the task preclude a more carefully designed strategy.

A first problem with the usual strategies tracing

Is a negative example of the current hypothesis the one and only way to reject this hypothesis in the Wason's problem? Let us consider the following piece of protocol (Table 1), in which the participant first tests a "even numbers" hypothesis by a "8-10-12" triple, and then a "odd numbers" hypothesis by a "3-5-7" triple.

(Table 1 about here)

As in both cases the participant tests her/his hypotheses by proposing examples rather than counterexamples, the usual explanation argued that s/he is exhibiting a confirmatory (or positive) strategy. But a competitive explanation can be proposed. In particular, there are no theoretical arguments to assume that people test their hypotheses (1) by considering only the current triple they give, and (2) by failing to take into consideration both the previous and present hypotheses. It can be argued that in the second step of the indicated sample, the participant's goal is to disconfirm the previous "even number" hypothesis. To this aim, s/he uses a positive example of the "odd number" alternative, which is a negative example (a counterexample) of the "even number" hypothesis.

Given that the participant will be informed that both triples satisfy the rule, s/he will be able to reject both properties "even" and "odd" as belonging to the to-be-discovered rule. Thus, one might conclude that the participant exhibited not a confirming but a disconfirming strategy.

Tracing such a strategy involves comparing the current triple, not with the current hypothesis (as has been the usual procedure), but with the previous hypothesis that the participant has proposed (i.e. the hypothesis s/he proposed on the preceding trial t-1). Every time the current triple is a negative example of the previous hypothesis, it can be assumed that participants use a disconfirmative strategy, which can be done with either a positive or a negative test.

Is such a strategy used in solving the 2-4-6 problem? In order to answer this question 67 participants were presented with the usual instructions of the 2-4-6 problem (Wason, 1960). Participants were undergraduate students (in various disciplines, excluding psychology) at the Faculty of Literature and Humanities at the University of Provence in Aix-en-Provence, France.

For each participant, we computed the proportion of confirmations of the total number of triples, by comparing each triple (1) on the one hand with its current hypothesis (*current tracing*), (2) and on the other hand with the hypothesis held for the triple considered immediately prior (*shifted tracing*). Table 2 gives a sample of protocol coded with each tracing paradigm.

(Table 2 about here)

We compared the mean percentages of confirmative strategy for these two tracing ways (current tracing vs. shifted tracing), considering first the whole sample of participants [89.46 vs. 48.76, $F(1,66) = 138.40$, $p < .0001$], and then separately those (n=56) who found the to-be-discovered rule [88.52 vs. 45.30, $F(1,55) = 154.13$, $p < .0001$], and those (n=18/56) who found the rule at their first attempt [89.17 vs. 33.78, $F(1,17) = 87.83$, $p < .0001$].

In all cases the mean for the current tracing using the current hypothesis was significantly higher than the mean for the shifted tracing using the previous hypothesis. Let us point that for the successful participants who found the rule at the first announcement this mean was 33.78%, so in other words 66.22% of their hypothesis tests were disconfirmative.

We ran a final analysis, comparing among the successful participants those who found the rule at their first attempt (n=18/56) and those who did not (n=38/56), again using both tracing ways. Although the means were not significantly different for the current coding [89.17 vs. 88.21, $F < 1$], there was a significant difference when using the shifted coding method considering previous hypothesis in favour of successful participants who were correct on their first attempt [33.78 vs. 50.76, $F(1,54) = 6.51$, $p < .01$].

This first study showed that the positive test strategy is also effective in the 2-4-6 problem, that is it also led to the rejection of hypotheses. It seems that participants were able to test a hypothesis and its alternative. Whereas it would have been usual to expect them to do so with only a single step, they did so with two successive steps. We have seen that the most successful participants were the most disconfirmative. A possible interpretation is that expresses the manifestation of a deliberate strategy (Caverni & Rossi, 1997). However, a way to be sure of the alternative hypotheses' consideration is to take into account the participants' expected response to the experimenter's feedback. This leads us to present a second problem.

A second problem with the usual strategies tracing

Although we just have shown the contrary, let us postulate, as Wason did, strategies can be traced comparing the current triple the participant proposes and the current hypothesis s/he claims to test. The simple fact that participant try a triple that follows the hypothesis s/he claims s/he is testing is not a proof that s/he is using a confirmative (or positive) strategy.

As long as the participant expects a positive feedback from the experimenter, then there is a sense in calling a trial made with an hypothesis and a triple verifying it a confirmatory trial, and in calling a trial made with an hypothesis and a triple falsifying it a disconfirmative trial. But if the participant expects a negative feedback, then the opposite is true (Table 3): a trial made with an hypothesis and a triple verifying it must be called a disconfirmative trial, while a trial made with an hypothesis and a triple falsifying it must be called a confirmatory trial. So, our assumption is that to trace the participant's strategies, their expected feedback must be taken into account.

(Table 3 about here)

We ran an experiment in which forty-eight participants participated. They were undergraduate students (in various disciplines, excluding psychology) at the Faculty of Literature and Humanities at the University of Provence in Aix-en-Provence, France. Three between participants experimental conditions were compared which differed only by the instructions. Participants were randomly assigned to one of three experimental conditions (n=16 per condition). The possibility was given to participants to inform us or not about their feedback they expected to receive from the experimenter. They were asked to use a provided sheet of paper to write down: (1) in the "usual" condition, each proposed triple, the corresponding hypothesis, and the actual experimenter's feedback; (2) in the "feedback" condition, each proposed triple, the corresponding hypothesis, the feedback they were expecting from the experimenter, and the actual experimenter's feedback.

Nevertheless, we have used a third instruction less directive than the last to make sure that the participants didn't inform us about something that they didn't think. In a "feedback if" condition, they were asked to write down each proposed triple, the corresponding hypothesis, the feedback they were expecting from the experimenter if any, and the actual experimenter's feedback.

First, we have computed the proportion of expected negative feedback on the basis of the total number of steps made to solve the problem. This rate is 47% in the condition "feedback" and 47.1% in the condition "feedback if" [$F < 1$].

Considering the percentages of positive hypothesis tests, the positive hypothesis rate is significantly greater in the Usual condition than in the condition "feedback" [93.2 vs. 73.7, $F(2,45) = 6.83$, $p < .01$], and than the condition "feedback if" [93.2 vs. 76.7, $F(2,45) = 4.85$, $p < .03$]. However, the difference between the condition "feedback" and the condition "feedback if" is not significant [76.7 vs. 73.7, $F < 1$].

Finally, we computed the proportion of the four trial types when expected feedback is taken into account (Table 4). The type of hypothesis test the most employed is a positive hypothesis testing associated with a positive expected feedback [condition "feedback", $F(1,60) = 22.16$, $p < .0001$; condition "feedback if", $F(1,60) = 5.67$, $p < .02$]. This type of trial refers at the classical confirmatory strategy that Wason has pointed out.

(Table 4 about here)

Negative hypothesis tests are lower but most associated with a negative feedback. This result is significant only in the condition "feedback" [23.9 vs. 2.4, $F(1,60) = 13.73$, $p < .0005$]. The most interesting finding is that negative hypothesis tests associated with a positive feedback – which refers to the classical disconfirmative strategy defined by Wason – are less employed whatever the condition [condition "feedback", $F(1,60) = 40.5$, $p < .0001$; condition "feedback if", $F(1,60) = 15.96$, $p < .0002$]. In other words, when a participant does try to falsify his/her hypothesis, s/he does it by expecting a negative answer rather than by testing a triple that violates his/her hypothesis.

Conclusion

We have shown that the classical tracing of hypothesis testing strategies is insufficient. Beside the classically considered strategy that consists in the proposition of a counterexample of the hypothesis when the former is conceived, we have shown that two other possible strategies exist. One consists in testing the hypothesis at the next step, by proposing an example of the former alternative hypothesis formulated at the previous step. This strategy allows participant to make only positive tests of its current hypothesis, by mobilising a disconfirmative strategy. To the extent of the classic task puts no constraint neither temporal, nor in number of possible tests, this strategy is efficient. One has shown that participants' protocols succeeding at the first attempt, exhibit an important disconfirmation rate when its are coded with the shifted tracing paradigm.

We have shown a second strategy allowing to falsify hypotheses by proposing a positive example of the former: this strategy consists in expecting a negative feedback. Bring to the fore this strategy, which takes into account the participant's expected feedback from the experimenter, allows no longer to confuse positive test strategy and confirmation strategy on the one hand, negative test strategy and disconfirmative strategy on the other hand. It allows to distinguish on the one hand the search of a confirmative or disconfirmative strategy, independent on the one hand of a positive or negative test strategy, on the other hand of the confirming or disconfirming test result made by the participant.

These two new possible strategies, altering one and the other classic results, does not invalidate for as much the plausibility that participants behave according to Wason's presuppositions. Nevertheless, its show that classic results suppose that these presuppositions are satisfied. If its were not, classic interpretations would have to question, as for the incapacity to produce counterexamples (that it's turn out to be systematic with the shifted tracing paradigm) as well as for the incapacity to treat the negation (half of the expected feedback are negative). At hand, these two new strategies confirm that the main strategy is well the positive test strategy.

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. (1983). Selective processes in reasoning In J. St B. T. (Ed.), *Thinking and Reasoning: Psychological Approaches.* London: Routledge & Kegan Paul.

Evans, J. St. B. T. (1989). Confirmation Bias In *Bias in Human Reasoning. Causes and Consequences.* Hove, UK : Lawrence Erlbaum Associates.

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

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

Klayman, J., & Ha, Y. W. (1989). Hypothesis testing in rule discovery: Strategy, Structure and Content. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15,* 596-604.

Popper, K.R. (1959). *The logic of scientific discovery.* London: Hutchinson.

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. (1962). A reply to Wetherick. *Quarterly Journal of Experimental Psychology, 14,* 250.

Wason, P. C. (1966). Reasoning In B. M. Foss (Ed.), *New Horizons in Psychology*. Harmondsworth: England Penguin.

Wason, P. C. (1968). On the failure to eliminate hypotheses-A second look In P. C. Wason & P. N. Johnson-Laird. (Eds.), *Thinking and reasoning*. Harmondsworth, England : Penguin.

Table 1 - Piece of protocol

Proposed Triples	Tested Hypotheses
8-10-12	even numbers
3-5-7	odd numbers

Table 2 - Sample of Participant's protocol (n°20) coded with (1) current and (2) shifted tracing paradigms (Adapted from Caverni & Rossi, 1997).

Triples Proposed	Tested Hypotheses	Feedback	Current Tracing (1)	Shifted Tracing (2)
4-6-8	Even numbers increasing by two	YES	C	/
6-8-10	Even numbers increasing by two	YES	C	C
10-14-18	Increasing even numbers	YES	C	D
14-18-10	Even numbers independent of order and interval	NO	C	D
120-122-128	Increasing even numbers	YES	C	D
rule proposed: Even increasing numbers independent of interval				
18-14-10	Even numbers decreasing by four	NO	C	D
18-10-06	Decreasing even numbers	NO	C	D
18-16-14	Even numbers decreasing by two	NO	C	D
14-16-18	Even numbers increasing by two	YES	C	C
14-18-22	Even numbers increasing by four	YES	C	D
01-02-03	Increasing even and odd numbers	YES	C	D
rule proposed: Increasing numbers				

(1) Usual tracing takes into account the triple and the hypothesis at the same step (e.g. step n°2: 6-8-10; even numbers increasing by two: confirmation).

(2) The proposed shifted tracing takes into account the triple and the previous hypothesis only at previous step (e.g. step n°3: 10-14-18; even numbers increasing by two: disconfirmation).

C = Confirmation
D = Disconfirmation

Table 3 - The four trials types when expected feedback is taken into account

Triple Proposed	Expected Feedback	Strategy
satisfying the hypothesis	yes	confirmation
falsifying the hypothesis	no	confirmation
satisfying the hypothesis	no	disconfirmation
falsifying the hypothesis	yes	disconfirmation

Table 4 – Proportion of the four trial types when expected feedback is taken into account (%)

	Condition "feedback if"	Condition "feedback"
satisfying the hypothesis & "yes" expected	48.1	50.5
falsifying the hypothesis & "no" expected	18.4	23.9
Confirmative tests	*66.5*	*74.4*
satisfying the hypothesis & "no" expected	28.6	23.2
falsifying the hypothesis & "yes" expected	4.9	2.4
Disconfirmative tests	*33.5*	*25.6*