UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

## Title
Attenuation of Belief Perserverance In a Covariation Judgement Task

## Permalink
https://escholarship.org/uc/item/5xj424kb

## Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 15(0)

## Authors
Vallee-Tourangeau, Frederic
Baker, A.G.

## Publication Date
1993

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

# Attenuation of Belief Perseverance In a Covariation Judgment Task[1]

Frédéric Vallée-Tourangeau and A.G. Baker
Department of Psychology
McGill University
Montréal, Québec H3A 1B1

## Abstract

A wide variety of judgment tasks have shown that once a reasoner favors a hypothesis, encountering evidence which contradicts it might not, in and of itself, dislodge that hypothesis. The interaction of prior belief and new evidence was studied in a covariation judgment task where subjects monitored multiple predictor-outcome relationships. Each relationship was programmed to reflect a strong positive contingency in a first phase, but in the second phase the contingency was negative, disconfirming the acquired expectation. For two of these relationships, the negative evidence was framed as positive evidence for alternative relationships, while in a third relationship, the negative evidence was not presented as supporting alternative explanations. Subjective contingency estimates indicated that the negative contingency was recognized in all three conditions. Belief perseverance, as measured by the likelihood of predicting the outcome on trials where the original predictor variable was present, was the strongest in the condition without alternatives. These results support the notion that belief change is a function of the negative evidence pertaining to that belief and the presence of alternative explanations which seek their support from that same evidence.

## Introduction

The evaluation of new evidence in the light of a preexistent belief has often been shown to depart from the normative implication of that evidence. For example, in belief-updating tasks, subjects show an "inertia effect" in that revisions of the probability that a hypothesis is true given new evidence are much more conservative than the predictions of a Bayesian analysis (e.g., Peterson & DuCharme, 1967; Pitz, Downing & Reinhold, 1967). Research in areas such as rule discovery, scientific and causal reasoning, correlation judgment, impression formation and stereotype maintenance has repeatedly found that once a hypothesis is favored, new evidence is asymmetrically evaluated such that supporting evidence is attributed greater probative weight than opposing evidence. This project sought to determine the factors which attenuate the tendency to maintain a belief or a hypothesis when the reasoner faces sustained disconfirming evidence.

The work on illusory correlation carried out by Chapman and Chapman (1967,1969) and by Hamilton and Rose (1980) has suggested that reasoners attribute more importance to cases which confirm their correlational expectation and that, consequently, relationships are often inferred when in fact none exist in the data. Both teams of researchers attempted to attenuate the illusory correlations by presenting their subjects with data which strongly opposed their expectations. Flooding the belief with disconfirming data reduced the magnitude of the illusory correlations but failed to change substantially the nature of the covariation expectation. These results lend credence to Nisbett and Ross (1980): "...people who subscribe to a given covariation theory before encountering evidence that should have served to overturn the theory, often emerge from such encounters with the theory unblemished and unscathed." (p. 169)

Anderson and others (Anderson, 1982; Anderson, Lepper, & Ross, 1980) have further demonstrated that a plausible hypothesis may survive the complete discreditation of the data base from

[1]This research was supported in part by a J.W. McConnel Memorial Fellowship to Vallée-Tourangeau and a NSERC of Canada grant to Baker.

which it was formulated. In these studies, subjects are shown case histories from an occupational group (e.g., fire fighters) which suggest that job effectiveness is a strong function of a certain trait (e.g., risk preference; another group of subjects is led to believe that risk aversion determines fire fighting success). The subjects are asked to formulate a plausible explanation for that relationship (e.g., "fire fighting is a risky occupation: one must take risks to be a good fire fighter"). The subjects are then thoroughly debriefed: the case histories are discredited and they are told that another group of subjects were shown successful risk-averse fire fighters. Yet, upon presentation of new case histories, subjects still use risk preference as a good predictor of fire fighting performance: the plausible explanation has survived the invalidation of the data and "(has) remained as an available and reasonable way of understanding how the various (variables) could be related." (Anderson, 1982, p. 128).

A hypothesis or a theory can accumulate an impressive amount of negative evidence against it, but if its plausibility is particularly appealing then it will survive as a *heuristic structure* with which new data are interpreted and new predictions are generated. Dawes (1988) offers a number of examples where hypotheses were shielded from disconfirmation by their plausibility. He argues, for instance, that the unassailable plausibility of the premise that clinical patients will project aspects of their personality in the interpretation of ambiguous stimuli, has protected the Rorschach, for the last sixty years, against the onslaught of psychometric disconfirmation. After reviewing other cases of hypothesis perseverance, Dawes conjectures that the best way of accentuating a reasoner's sensitivity to negative evidence is not so much to amass empirical anomalies but rather to offer a new plausible hypothesis which can make sense of the new evidence as well as the old data. A number of researchers have converged on the same conclusion (e.g., Anderson, 1982; Einhorn & Hogarth, 1986; D. Kuhn, 1989): old beliefs yield to negative evidence only after the adoption of a new plausible hypothesis; the brute impact of negative data is often insufficient to fuel theory change (this notion borrows heavily from T. Kuhn's, 1970, analysis of theory change in science).

The present study uses a covariation judgment task in which a reasoner first acquires an expectation about the relatedness of a predictor variable and an outcome variable; in a second phase, this expectation is contrasted with evidence that strongly opposes it. The procedure is designed to create a conflict between a belief and new data and to offer a window on the interaction between "theory" and evidence. The design of the experiment is within-subjects. The task involves the simultaneous monitoring of different predictor-outcome contingencies, a subset of which all show the same positive contingency in the first half. For two of these contingencies, the negative evidence in the second half is presented in terms of *alternative* predictor-outcome relationships. In one case, the negative evidence for the $\text{Predictor}_1$ - $\text{Outcome}_1$ relationship is framed, in part, as positive evidence for a $\textbf{Predictor}_2$ - $\text{Outcome}_1$ relationship (i.e., an alternative predictor now signals the outcome), and in the second case, the negative evidence is framed, in part, as positive evidence for a $\text{Predictor}_1$ - $\textbf{Outcome}_2$ relationship (i.e., the old predictor now signals an alternative outcome). In the third contingency, the negative evidence is not framed in terms of alternative relationships. A comparison of the magnitude of the adjustment to the negative evidence between these three conditions will assess the impact of disconfirming evidence when it is supplemented with alternative conceptions with which to interpret the disconfirmation data, and will assess its impact when it is not.

This project offers two new methodological features which enhances its importance. The first is the within-subjects design. Some have argued (e.g., Baron, 1988) that perseverance is a cognitive trait and that some reasoners are simply more open-minded and will thereby show faster adaptation to negative evidence. While this point may deserve consideration, a method which elicits adaptation is some conditions and not others for the same subjects, affords a more specific understanding of belief perseverance. Second, as will be described shortly, the measures of adaptation to the negative evidence will attempt to separate a) the recognition that the second half contingency is negative and b) the extent to which the subjects still rely on the disconfirmed hypothesis to generate positive predictions about the presence of the outcome; i.e., the extent to which the heuristic value of the predictor variable survives the recognition that the correlational evidence is now negative.

## Method

### Subjects

Twelve McGill University undergraduates were paid $5 to participate in this experiment.

## Procedure

The predictor-outcome contingencies were presented in a scenario involving teams of astronomers (the predictors) and "the detection of comets in a neighboring galaxy" (the outcomes). Subjects evaluated the effectiveness of different teams at detecting different comets. The evaluations were made on the basis of a series of trials which reported the presence or absence of a team at an "observation station" and whether the comet was detected or not. For each trial subjects predicted whether or not the comet would be observed. The frequency of events for the three experimental conditions are shown in Table 1. The contingency between the presence of the team and the comet was moderately positive over the first 24 trials ($\Delta P = p(\text{Comet}|\text{Team}) - p(\text{Comet}|\text{no Team})$; $\Delta P = .67$) and was moderately negative for the last 24 trials ($\Delta P = -.67$). The cumulative contingency was zero, since the sum of the two contingency tables in Table 1 produces a contingency table with all four cell frequencies at 12. The subjects were not informed that the contingency would change over trials. (See Yates and Curley, 1986, for a similar design used to investigate primacy effects in covariation judgments.)

Two types of trials provide evidence that is inconsistent with a positive contingency: trials where the absence of the team is paired with a comet detection (*no predictor* → *outcome* trials) and trials where the presence of the team is paired with the absence of a comet detection (*predictor* → *no outcome* trials). In the contingency for the second half (see Table 1) there are 10 negative trials of each type. In the ALTERNATIVE PREDICTOR condition, a second team was present on the 10 *no predictor* → *outcome* trials. In the ALTERNATIVE OUTCOME condition a second comet occurred on the 10 *predictor* → *no outcome* trials. In the SINGLE PREDICTOR conditions, these negative trials were not coded in terms of either the presence of an alternative team or an alternative comet.

To mask the manipulation, and to increase the complexity of the monitoring task, three additional team-comet relationships were designed, one in which the contingency between the presence of the team and the detection of the comet was moderately negative, a second one which showed a zero team-comet contingency, and a third one which started out as a zero contingency, but where over the last 24 reports the team's effectiveness increased. With six team-comet contingencies, defined over 48 trials, subjects were presented with a total of 288 trials. Trial order within each half was randomized for each subject.

Table 1. Frequencies of the four types of event pairings in all three experimental conditions

**FIRST HALF**

| COMET | ASTRONOMERS present | ASTRONOMERS absent |
|---|---|---|
| present | 10 | 2 |
| absent | 2 | 10 |

**SECOND HALF**

| COMET | ASTRONOMERS present | ASTRONOMERS absent |
|---|---|---|
| present | 2 | 10 |
| absent | 10 | 2 |

## Measures

An effort was made to separate the subjects' evaluation of the evidence in the second half from their reliance on the presence of the predictor variables to make new outcome predictions. Initial acquisition of the positive contingencies and subsequent adjustment to the negative evidence were thus measured in two ways.

1. After each block of 72 trials, subjects were asked to evaluate the effectiveness of each of the six teams using a scale ranging from -100 to 100. These ratings were estimates of the contingency between the presence of the team and the detection of a comet.

2. In order to assess how the subjects used the presence of the team to predict the comet, the proportion of comet predictions in the presence of the team was calculated over blocks of 12 trials for each relationship.

## Results

The mean contingency judgments for the three experimental conditions are shown in the left panel of Figure 1. These estimates indicate that the positive contingency was judged with reasonable accuracy in all three conditions over the first two blocks. Adjustment to the rapidly decreasing correlation in the second phase occurred in all three conditions, although estimates in the SINGLE PREDICTOR condition were slightly less negative over the last block. A two-factor repeated measures analysis of variance (ANOVA), with Condition and Block as the two factors, was performed on the mean estimates for Blocks 2, 3, and 4 (the rejection level was .05 for this and all subsequent analyses): only the main effect for Block was reliable ($F(2,22) = 31.5$; maximum nonreliable $F(4,44) = 2.0$). This analysis indicates that the decrease in mean contingency estimates paralleled the decrease in actual contingencies in all three conditions: the presence of alternative relationships did not produce greater sensitivity to the negative evidence.

The mean proportions of comet predictions on trials in which the appropriate team was present are shown in the right panel of Figure 1. Acquisition of the predictive value of each team was similar in all three conditions over the first two blocks. When the team's predictive value decreased suddenly over Blocks 3 and 4, adjustment in prediction responses was most pronounced in the two conditions which framed the negative evidence in terms of alternative relationships. In fact, in the SINGLE PREDICTOR condition, comet detections were still predicted on more than 75% of the trials in which the team was present, even though the teams had *failed* to produce

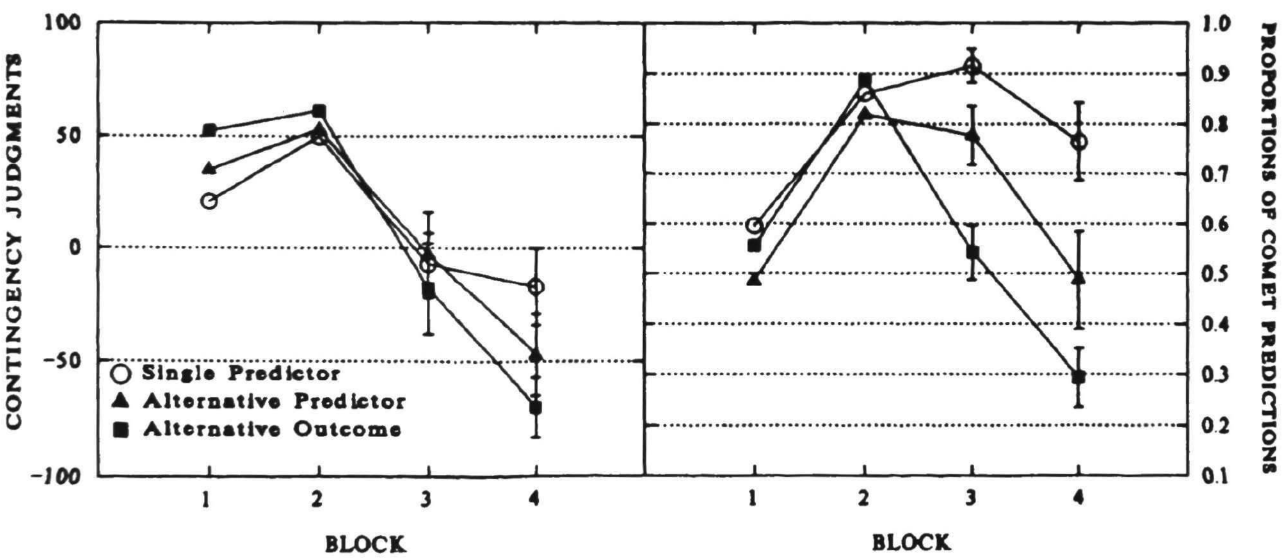


Figure 1. Mean contingency judgments for the three experimental conditions, left panel. Mean proportions of comet predictions in the presence of the astronomers, for the experimental conditions, right panel.

a comet detection in over 80% of the trials in the second half in all three conditions. The two-way repeated measures ANOVA on the mean comet predictions support these observations: over Blocks 2, 3, and 4, both main effects, for Block ($F(2,22) = 17.0$) and Condition ($F(2,22) = 17.5$), were significant, as was the interaction ($F(4,44) = 6.15$). Even under sustained disconfirmation, the correlation expectation in the SINGLE PREDICTOR condition is still employed to make predictions about future comet detections.

## Discussion

The impact of continuous disconfirmation on an acquired correlational expectation was investigated using a within-subjects design. All subjects learned, in a first phase, that different predictor variables (teams of astronomers) were informative signals of the presence of different outcome variables (detection of comets). The acquisition of the correlational expectation was reflected in both estimates of contingency and proportions of comet predictions in the presence of the team of astronomers. Over the second half of the observation reports, the team-comet contingency was suddenly reversed to be in direct opposition to the original contingency. Prediction responses were adjusted best in the conditions where the negative contingency was expressed as a positive contingency between an alternative team and the original comet or as a positive contingency between the original team and an alternative comet. In the condition which offered no alternative conceptions with which to make sense of the negative evidence, the presence of the team was still deemed a good predictor of the detection of the comet. This difference cannot have arisen out of a different recognition of the negative contingency in the second half because, although there were slight differences in the mean contingency estimates in the second half, these were not significant. This fact does not support an attention decrement explanation for the lack of adjustment in the SINGLE PREDICTOR condition. Rather, it suggests that subjects recognized the negative evidence in all three conditions but did not abandon their correlational expectation, as expressed in their prediction responses, unless alternative conceptions were offered.

Einhorn and Hogarth's (1986: Equation 6) belief-updating model captures some aspects of the present data. Their equation is an attempt to formalize the idea that the judged importance of an explanation (X) for event (Y) is a function of the evidence for that relationship as well as the discounting impact of the recognition of alternative explanations. Similarly, in the present study, the heuristic value of the team of astronomers in the second phase was a function of the team's predictive value (i.e., its contingency with comet detection) *and* of the discounting effect of the presence of alternative predictors which signalled the outcome. Without the presence of alternative hypotheses, subjects were less likely to abandon their correlational expectation, and would persevere in predicting the presence of the outcome. The rapid decrease in the ALTERNATIVE OUTCOME condition simply showed that subjects quickly learned to associate the predictor with the new outcome. The difference in predictive responses between the SINGLE PREDICTOR and the ALTERNATIVE PREDICTOR conditions is the most interesting theoretically and suggests that the heuristic value of a predictor variable is a function of its predictive value minus the weighed predictive value of alternative predictors.

The event frequencies of Table 1 are amenable to a Bayesian analysis. The prior odds that the comet will be detected on any given trial are 1:1, that is, the comet is as likely to be detected as it is to be undetected prior to knowing about the presence of the team of astronomers. These priors stay the same throughout the task. As well, the probability that a team is present on any given trial is also .5. Consequently, the posterior odds always equal the likelihood ratio. The diagnosticity of the team of astronomers is very high in the first half: the odds that a comet is detected in their presence are 5 to 1. The astronomers' diagnosticity, in all three experimental conditions, decreases sharply during the second half, to reach a terminal point where the comet is as likely to be detected in their presence as it is in their absence. Bayes's Theorem thus predicts swift adjustment to the negative evidence in the second half. Interestingly, the fact that an alternative team is present on the ten trials in which the comet is detected and the original team is absent, does not affect the Bayesian calculation of the original team's diagnosticity. In effect, Bayes's Theorem predicts as much adjustment in the ALTERNATIVE PREDICTOR condition as in the SINGLE PREDICTOR condition, but *not more*. It may predict more adjustment in the ALTERNATIVE OUTCOME condition if at the start of the second half the likelihood ratio is not calculated cumulatively, but is rather reset at zero. Otherwise, the predictions are the same in the three experimental conditions.

Thus, a Bayesian analysis of belief updating is sensitive to the diagnosticity of the predictor variable but is indifferent to whether the absence of that predictor variable is coded in terms of alternatives or not. This is because Bayesian predictions are a direct function of the probability of the outcome given the predictor which is the same in both the SINGLE PREDICTOR and ALTERNATIVE PREDICTOR conditions. As a normative model of belief updating, a Bayesian analysis falls short of capturing the psychological importance of the presence of an alternative predictor which supplements the mere accretion of evidence that undermines the diagnosticity of the original predictor. The data point to the importance of alternatives in favoring adjustment to negative evidence. Thus, any model which seeks to capture belief perseverance and adjustment to negative evidence must not only specify the normative implication of the evidence for the current predictors, but must also capture the impact of alternative predictors.

The tendency to rely on a disconfirmed hypothesis can be hard to dislodge; subjects may prefer to err in their predictions rather than abandoning the only theory they have. These results support the contention that hypotheses do not simply yield to the accumulation of negative evidence: rather, the negative evidence must be interpreted from the perspective of alternative hypotheses to motivate a change in belief.

## References


Anderson, C. A. (1980). Inoculation and counterexplanation: debiasing techniques in the perseverance of social theories. *Social Cognition* 1:126-139.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology* 39:1037-1049.

Baron, J. (1988). *Thinking and deciding*. Cambridge University Press.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology* 72:193-204.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology* 74:271-280.

Dawes, R. M. (1988). *Rational Choice in an Uncertain World*. HBJ, Publishers.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin* 99:3-19.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality and Social Psychology* 39:832-845.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review* 96:674-689.

Kuhn, T. (1970). *The structure of scientific revolutions*. Second Edition. The University of Chicago Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall.

Peterson, C. R., & DuCharme, W. M. (1967). A primacy effect in subjective probability revision. *Journal of Experimental Psychology* 73:61-65.

Pitz, G. F., Downing, L., & Reinhold, H. (1967). Sequential effects in the revision of subjective probabilities. *Canadian Journal of Psychology* 21:381-393.

Yates, J. F., & Curley, S. P. (1986). Contingency judgment: Primacy effects and attention decrement. *Acta Psychologica* 62:293-302.