Journal of Economic Behavior & Organization
Vol. 41 (2000) 263–276

JOURNAL OF
Economic Behavior
& Organization

www.elsevier.com/locate/econbase

A violation of the monotonicity axiom: experimental evidence on the conjunction fallacy

Daniel John Zizzo [a,*], Stephanie Stolarz-Fantino [b], Julie Wen [b],
Edmund Fantino [b]

[a] *Hertford College, Oxford University, Oxford OX1 3BW, UK*
[b] *Department of Psychology, University of California at San Diego, San Diego, USA*

Received 23 June 1998; received in revised form 30 April 1999; accepted 11 May 1999

Abstract

The conjunction fallacy is an anomaly in human reasoning for which the conjunction of two events is rated more likely to occur than one of the events alone. In the context of decision under uncertainty, this violates the monotonicity axiom of probability, and consequentially also Bayes' Rule and the monotonicity axiom of preferences. Our experiments show how dynamic feedback and monetary incentives affect the fallacy rate, and how the complexity (and possibly the presentation) of the decision problem and an averaging heuristic might determine outcomes and reasoning. ©2000 Elsevier Science B.V. All rights reserved.

JEL classification: D81; D83

Keywords: Monotonicity axiom; Conjunction fallacy; Learning; Bounded rationality

---

1. Introduction

The conjunction fallacy occurs when a subject judges the conjunction of two events to be more rather than less (or at least equally) likely to occur than one of the events alone. The fallacy represents a violation of the basic monotonicity probability axiom (i.e. axiom 3 in the list by Hoel, 1984). This may lead to a violation of the monotonicity of preferences in the context of decision under risk, in the case, for example, of a lottery in which the joint occurrence of two events is preferred to the occurrence of only one of the events. It may also

---

[*] Corresponding author.
E-mail address: daniel.zizzo@hertford.ox.ac.uk (D.J. Zizzo).

PII: S0167-2681(99)00076-1

violate Bayes' Rule. Although psychologists have analyzed this case of bounded rationality to some extent since Tversky and Kahneman (1982, 1983) seminal work, economists have hardly taken notice. While some economists, such as Grether (1992), have analyzed the impact of another failure of Bayesian rationality, namely the base-rate error, the conjunction fallacy has mostly been neglected.

This paper analyzes experimental evidence on the conjunction fallacy. In Section 2 we review the literature. Section 3 describes the design of our experiments, which incorporated nontrivial monetary incentives and the possibility for learning. Section 4 reports the results, of which we offer our interpretation in Section 5. Section 6 concludes.

2. Review

The conjunction fallacy was first found in a considerable proportion of subjects by Tversky and Kahneman. The typical questionnaire included the description of a character, such as:

> Bill is 34 years old. He is intelligent but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

plus various statements on the character (say, 'Bill is an accountant' and 'Bill plays jazz for a hobby'), including a conjunction-of-two-events statement (say, 'Bill is an accountant who plays jazz for a hobby'). Subjects had to assess how likely, given the profile of the character (Bill), it is for the various statements to be true. The conjunction fallacy occurred when, for example, one found that 'Bill is an accountant who plays jazz for a hobby' was ranked more likely than 'Bill plays jazz for a hobby'.

A feature of Kahneman and Tversky, as well as of later psychological work, was that decisions were one-shot, i.e. either only one profile was given to the subjects or, less commonly, if more than one conjunction fallacy probability assessment was made, no feedback on whether the previous decision was correct or incorrect was given before the subject's next decision. It was difficult to assess the role of learning and feedback. Moreover, monetary incentives were not provided. On both grounds, such work appears vulnerable to objections by economists. Nevertheless, the robustness of the fallacy was verified in different ways.

In the Kahneman and Tversky data, the proportion of people committing the fallacy varied widely, from small to more than 80%, according to whether a description of the character (the 'framing description') was included. This pointed not only to the importance of framing effects, but also to the fact that the conjunction might or might not be a robust phenomenon. Further experimental work was aimed to verify both the robustness of the fallacy and the type of rule of thumb according to which subjects reason in assessing conjuncted events probabilities (e.g., among others, Yates and Carlson, 1986; Gavanski and Roskos-Ewoldson, 1991; Fantino et al., 1997). Overall, the conjunction fallacy seems an extremely robust phenomenon. While Gigerenzer (1991) argued that the fallacy disappears in judgments of frequency rather than probability, the data reviewed by Tversky and Koehler (1994) suggests a decrease but not a disappearance of the fallacy. Mulford and Dawes (1999) find significant fallacy committal in the recollection of the frequency for personal events.

One might be tempted to think that, if instructions provided hints favoring logical responses, the conjunction fallacy would be eliminated. However, in unpublished experimental work by the second and fourth authors of this paper, the instruction that ‘Your judgments should be made in terms of their probability and not simply in terms of whatever intuitive appeal is generated by the description above’ actually resulted in a greater proportion of University of California at San Diego (UCSD) students evincing the fallacy than under the standard instructions. Moreover, the fallacy does not disappear with training in logic. In Experiment 2 of Stolarz-Fantino et al. (1996), UCSD logic students were given a standard conjunction problem just prior to completing their course in logic. The logic professor introduced the task as one involving ‘reasoning’. Despite this context, 43% of the students committed the conjunction fallacy. This is comparable to other studies utilizing a brief training period (e.g., Benassi and Knoth, 1993). The experiments reported in Stolarz-Fantino et al. (1996) demonstrate that the conjunction fallacy is determined in a complex way and not simply due to framing effects. Indeed, a minimum of one quarter of the subjects still committed the fallacy even when the framing description was eliminated and the question was worded so as to facilitate a logical approach. When the framing description was eliminated, 41% of UCSD subjects from an experimental psychology class committed the fallacy, and this percentage rose to 78% with a framing description. The fraction of subjects committing the fallacy did vary according to the sample (class) used, but nevertheless a lower bound of about 25% of the subjects committing the fallacy underscores the robustness of the fallacy.

3. Design

3.1. Experiment 1 design

Forty-eight UCSD undergraduates, assigned to four experimental conditions of 12 subjects each, participated in this experiment. Subjects did the experiment individually, at different times, with only an experimenter present in the room. The experimenter gave each subject a booklet with seven pages. All the experimental material can be found in the Appendix A. The first page contained general instructions, which varied according to the experimental condition.

Subjects in the first experimental condition (Group 1) were asked to stop at the end of each page of the questionnaire. The experimenter would then tell them whether their answers on that particular page were correct, thus providing them individual feedback and the possibility for learning. For each single page answered correctly, they would receive US$ 3 (paid at the end of the experiment). Considering that the overall experiment did not last more than 15–20 min for each subject, the subject had the opportunity to earn US$ 18 for just one-third of an hour of work, the equivalent of US$ 54 on a hourly basis. Such monetary incentives seem nontrivial, especially for the undergraduate students that comprised the sample.

Subjects in Group 2 were also told to stop at the end of each page. The experimenter would then tell them whether their answers on that particular page were correct. However, while individual feedback was provided, no monetary incentives were mentioned. Subjects in Group 3 and 4 were just told to follow the instructions on the questionnaire, but no feedback (nor monetary incentive) was mentioned; however, subjects in Group 3 got a hint (‘there is a correct answer’) while subjects in Group 4 did not.

Pages 2 through 7 were identical for all groups. Each page contained a profile: the profile of ‘Ralph’ was on pages 2 through 5, while those of ‘Bill’ and ‘Linda’ were on pages 6 and 7, respectively (see the Appendix A). Subjects were asked to rank the likelihood of each of 7 statements on each page by entering a number from 1 to 7 on the left of the statement – where ‘1’ would be the least likely and ‘7’ the most likely. If, in this probability ordering, the conjuncted event number was greater than the number assigned to either of the two events being conjuncted, the conjunction fallacy was detected – and, in the case of Groups 1 and 2, would bring the experimenter to say that there was a mistake. The Ralph pages were identical to one another – except that a different conjuncted event statement was presented on each page.

The conjuncted event statements represented different combinations of ‘high’ and ‘low’ probability events. For the Ralph pages, the items had been determined to be ‘high’ or ‘low’ probability on the basis of pre-testing with UCSD students (different from those used in both experiments). For the Linda and Bill questions, we relied on Tversky and Kahneman’s subjects ratings of the likelihood of the various items: they each combined a high-likelihood event with a low-likelihood event.

In the case of Group 1, the payment stage followed, with US$ 3 assigned for any answer in which the conjunction fallacy was not detected.

3.2. Experiment 2 design

Forty-eight UCSD undergraduates, different from those in Experiment 1, were assigned to four experimental conditions of 12 subjects each. Experiment 2 aimed to reduce the *complexity* of the decision problem. There is evidence showing that, if problems are complex enough, little learning occurs even with dynamic feedback (e.g., Brehmer, 1992; Diehl and Sterman, 1995). Although the conjuncted events probability assessment seems easy enough a decision problem compared to those usually discussed in the dynamic feedback literature, there were ways to simplify it further and see whether this would significantly reduce the fallacy and improve the effectiveness of dynamic feedback. Experiment 2 was identical to Experiment 1 except in three respects:

1. for all groups, the only statements on each page were those of the two events to be conjuncted and of the conjuncted events. This should reduce confusion on the nature of the problem. Moreover, when – in the groups with feedback – the experiment said that a ‘mistake’ had been committed, subjects could spot it much more easily than if they had to decide which of seven statements was wrong, as in Experiment 1;
2. for all groups, instead of providing the likelihood ordering of the statements, subjects were asked to give probabilities between 0 and 100 for each statement. Again, this should help in simplifying the problem, and may also change the way subjects address it[1];
3. the hint used for Group 3 was stronger. Instead of just suggesting that a solution existed, a formula (‘$\Pr(A\&B) = \Pr(A) \times \Pr(B)$’) was suggested. While the formula is actually

---

[1] In this paper, we do not try to isolate this possible framing effect as such, from the reduction in complexity of the decision problem produced by our manipulations. Clearly, this should be object of further research.

true only for independent events, it was obviously sufficient for subjects not to commit the fallacy.

As in Experiment 1, Group 1 had both monetary incentives and feedback (but no hints): US$ 3 were given per correct answer, and feedback on whether an error had been committed was provided every round. Group 2 had the feedback but no monetary incentives nor hints. Group 3 had the stronger hint, but no monetary incentives nor feedback. Group 4 did not present monetary incentives, feedback or hints. The profiles were the same as in Experiment 1 (Ralph in the first four, then Bill, finally Linda), and so were the conjuncted event statements.

4. Results

4.1. *Definition of the null hypotheses*

We shall discuss the results of the experiments by considering the data first by subject, e.g. analyzing how the percentage of conjunction fallacies made by the subjects varied across conditions. We shall then analyze the data by round: we shall see how subjects changed their decisions across rounds in each condition, and how this affected the occurrence of the conjunction fallacy.

Before analyzing results, it may be worth briefly considering exactly what a null hypothesis may look like. One could actually think of *two* null hypotheses.

According to the *perfect rationality* null, we would expect the conjunction fallacy never to occur. This null also predicts that the complexity of the decision problem does not matter (since agents are perfectly rational).

According to the *zero rationality* null, subjects are confused and simply choose randomly: since there are six ways of ordering the probabilities of three events (A, B and A&B) and only in two of them the probability of the conjuncted events will be lower than either A or B, subjects behaving randomly will commit the error two thirds of the time[2]. An implication of the zero rationality null is that we would expect the conjunction fallacy rate to be approximately the same across rounds[3], e.g. not to follow any kind of pattern, since the choice is random whatever the specific features of the decision problem or the eventual feedback provided in previous rounds. One might perhaps suggest that some form of downward trend in conjunction fallacy rate could be possible from early to late rounds: perhaps, if subjects are confused, repetition will help them to be less confused. However, no other pattern is likely to be compatible with the zero rationality null. The same holds, of course, for the perfect rationality null. Indeed, we would not expect any pattern in the data also in a third case – namely, if subjects were rational but had occasional, random lapses[4].

---

[2] We thank a referee for pointing out to us this second null hypothesis.

[3] The same is, of course, trivially true for the perfect rationality null, since we would always expect a conjunction fallacy rate of 0%.

[4] If this were the case, we could perhaps try to estimate a common error rate (there is not enough data to estimate individual error rates).

Table 1
Percentage of conjunction fallacies made by subjects in each condition

Experiment 1 Group 1	Mean	0.6389
Money + Feedback	Std. Deviation	0.3469
Experiment 1 Group 2	Mean	0.5694
Feedback	Std. Deviation	0.3214
Experiment 1 Group 3	Mean	0.7778
Weak Hind	Std. Deviation	0.2171
Experiment 1 Group 4	Mean	0.7083
Nothing	Std. Deviation	0.3188
Experiment 2 Group 1	Mean	0.1250
Money + Feedback	Std. Deviation	0.2475
Experiment 2 Group 2	Mean	0.4444
Feedback	Std. Deviation	0.3205
Experiment 2 Group 3	Mean	0.2639
Strong Hind	Std. Deviation	0.3292
Experiment 2 Group 4	Mean	0.4583
Nothing	Std. Deviation	0.3343
Total	Mean	0.4983
	Std. Deviation	0.3623

4.2. Data by subject

Table 1 represents the average percentage of conjunction fallacies made by each group of subjects participating in our two experiments. In experiment 1, the average percentage of fallacies was 63.89, 56.94, 77.78 and 70.83 in Groups 1 (money + feedback), 2 (feedback), 3 (weak hint) and 4 (control), respectively. In experiment 2, the percentages drop to 12.50 for Group 1 (money + feedback), 44.44 for Group 2 (feedback), 26.39 for Group 3 (strong hint) and 45.83 for Group 4 (control). It is apparent that the most striking feature of the data is the improvement in performance in Experiment 2 relative to Experiment 1. A between-experiments F test proves this to be 1% significant (d.f. = 1, $P = 19.258$, $P < 0.0005$).

The mean number of fallacies made by a subject in Experiment 1 was 4.04 (out of a maximum of 6). This is comparable to previous experiments, and almost identical (67.36% vs. 66.67%) to the prediction of the zero rationality null hypothesis: a binomial test is non-significant ($P = 0.86$). Conversely, the perfect rationality hypothesis of a fallacy rate of 0% is clearly rejected[5].

In contrast, the mean fallacy rate in Experiment 2 was 1.94. While the perfect rationality null is still rejected[6], the rate is low enough that the zero rationality null of 2/3 is rejected as well ($P < 0.0005$ in a binomial test), in favor of some rationality.

---

[5] Under the null hypothesis, the probability of committing the fallacy is zero. The variance is also zero, so any positive percentage of fallacy committal would yield to a rejection of the null. If alternatively we consider the observed fallacy rate as the null, and compute the confidence intervals at conventional levels from there, we find they are far away from zero.

[6] At conventional levels, the confidence intervals considering the observed fallacy rate as the null are still well away from zero.

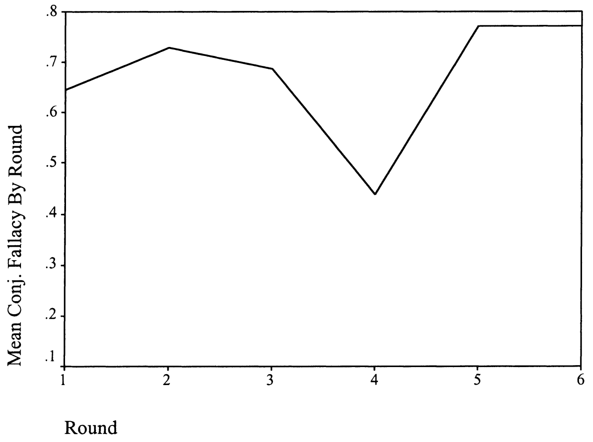

Fig. 1. All experiment 1 groups.

An F test on the average fallacy rates across Experiment 1 groups is non-significant (d.f. = 3, $F = 1.036$, $P = 0.386$). Thus, the monetary incentives, feedback and hint were ineffective. An F test on the average fallacy rates across Experiment 2 groups is significant, though (d.f. = 3, $F = 3.156$, $P < 0.05$). We did Tukey post hoc tests[7] to verify which between-groups differences are significant in Experiment 2: in one-tailed tests, group 1's (money + feedback) mean conjunction fallacy rate was found significantly lower than groups 2 (feedback only; d.f. = 46; $P < 0.05$) and 4 (control; d.f. = 46; $P < 0.05$). No other post hoc test was (even 10%) significant. So, in Experiment 2, providing money plus feedback did not seem significantly better than providing a strong hint, but yielded a significant decrease in the fallacy rate relative to providing nothing or feedback only.

It is worth noting that these results are not sensitive to the usage of parametric tests, for we were able to replicate them using nonparametrics. For example, a Kruskal–Wallis test on the between-groups mean fallacy rate in Experiment 1 is non-significant (d.f. = 3, $\chi^2 = 2.830$, $P = 0.419$), while it is significant for Experiment 2 (d.f. = 3, $\chi^2 = 10.535$, $P < 0.05$).

4.3. Data by round

Figs. 1 and 2 show the mean fallacy rate from round 1 to round 6 in Experiments 1 and 2, respectively. Figs. 3 and 4 display the mean fallacy rate in the groups where feedback was provided (Groups 1 and 2), again respectively for Experiments 1 and 2. The tables show no downward trend in the conjunction fallacy. With the partial exception of Fig. 4, the only regularity is a dip in the conjunction fallacy rate in Rounds 1 and 4. The average fallacy rate in Rounds 1 and 4 was 36.46%, against a 50% average fallacy rate in the other rounds.

---

[7] Post hoc tests are just t tests with stricter significance levels to control for the fact that, if – within the context of a significant F test – many t tests are done at the same time, the likelihood of finding something 'significant' increases (e.g., if 10 t tests are made, the likelihood of finding at least one correlation 5% 'significant' is $1 - 0.95^{10} = 40.13\%$). For further details, see Keppel (1991).

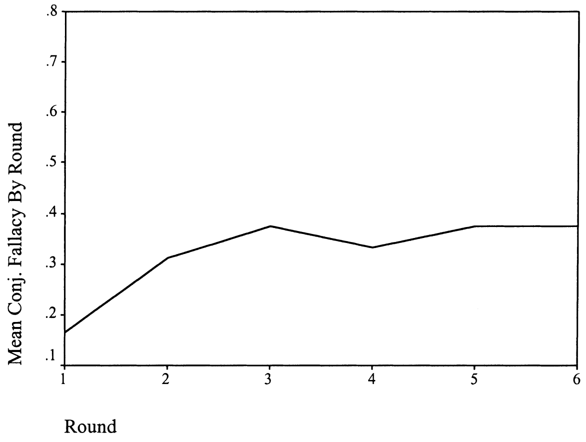

Fig. 2. All experiment 2 groups.

We tried to test the presence of any inter-round effect by using four variables. The variable Round has a value of 1 for the first round, 2 for the second, and so on until the sixth. The variable Learn is identical to Round, but considers only groups with feedback, so allowing the possibility of learning from past errors. The variable Last simply assigns a value of 0 to the first three rounds, and 1 to the last 3. Finally, the variable BothHighOrLow has a value of 1 for Rounds 1 and 4, and 0 otherwise; the reason for the variable name is that the first and fourth stages (Ralph, cases 1 and 4) presented conjunctions combining two high probability events and two low probability events, respectively, while the others (Ralph, cases 2 and 3, plus Bill and Linda) presented combinations of one high and one low probability event.

No matter what parametric or nonparametric tests are used, and not surprisingly given the graphs, Last, Round and Learn are non-significant. For example, *F* tests on Learn, Round and Last, with 5 d.f., give *F* values of 0.289, 0.856 and 0.111, respectively (the

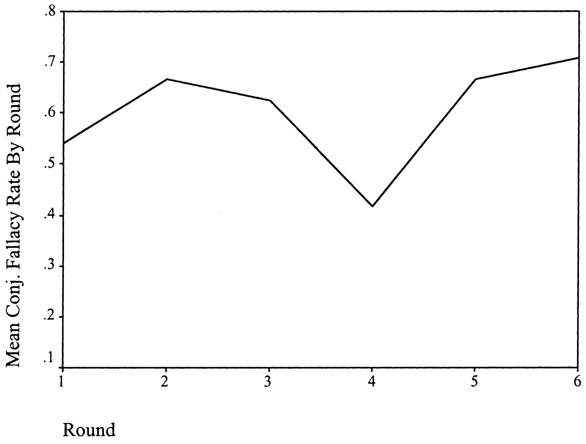

Fig. 3. Experiment 1 groups with feedback (1 and 2).

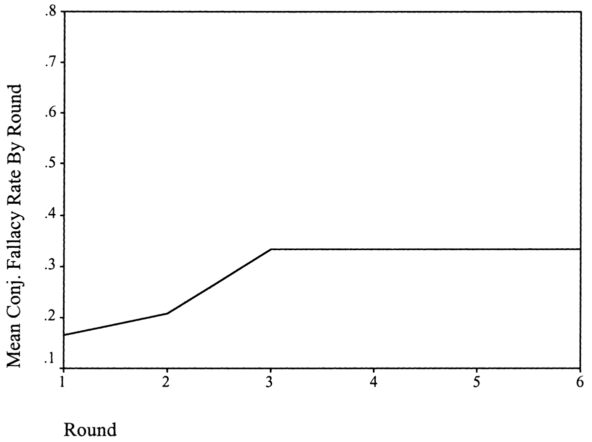

Fig. 4. Experiment 2 groups with feedback (1 and 2).

corresponding probabilities are 0.913, 0.519 and 0.74), and all Spearman's ρ correlations[8] with the mean round conjunction fallacy rate are non-significant.

Again no matter the choice of test, BothHighOrLow is significantly and negatively correlated with the mean round conjunction fallacy rate. An F test yields a value of 4.313 ($P < 0.05$). The Spearman's ρ correlation coefficient is -0.302 ($P < 0.05$).

5. Interpretation

This section contains our interpretation of the data presented in Section 4. We have tried to separate to a large extent such interpretation from the presentation of the data in order to allow readers to make their own judgment. We will first present reasons why we believe the perfect rationality and the zero rationality nulls fail. We will then suggest how a boundedly rational perspective which allows for the usage of decision rules beyond Bayes' Rule, such as the averaging heuristic, may be the most consistent with the data.

5.1. The perfect rationality null fails

As we saw in Section 4.2, the mean fallacy rate by the subjects (67.33% in Experiment 1, 32.33% in Experiment 2) is too high to be explained by the perfect rationality null. Surely the Experiment 2 results are 'in the direction of' the perfectly rational outcome, but it is unclear why the slight complexity differential between Experiment 1 and Experiment 2 should matter so much, if agents are perfectly rational. While a combination of feedback and monetary incentives does produce a further improvement, it is unlikely that many

---

[8] This is a standard nonparametric measure of correlation. A slightly different one, the Kendall tau-b, also shows a non-significant correlation (whereas that with BothHighOrLow is 5% significant).

decision problems of interest to economists present themselves as cleanly and simply as in Experiment 2.

The finding of Section 4.3 casts further doubt on the perfect rationality hypothesis: subjects performed significantly better in tasks where two highly probable or improbable events were joined (rounds 1 and 4). This pattern is left unexplained by the perfect rationality hypothesis, while – as we shall see below – it is consistent with the usage of a boundedly rational heuristic (the averaging heuristic).

5.2. *The zero rationality null fails*

The zero rationality hypothesis fares better than the other null, if nothing else because it is able to predict the mean fallacy rate in Experiment 1 (2/3 predicted, 67.33% observed). The lack of a learning trend (see Section 4.3) is also good news for the hypothesis that subjects behave randomly. Nevertheless, we believe that the weight of the evidence suggests that (at least some) subjects were not behaving randomly.

First, as we saw in Section 4.2, the zero rationality null is rejected in relation to the mean fallacy rate by the subjects in Experiment 2 (32.33% observed). Second, there is significant variability across the Experiment 2 conditions: monetary incentives plus feedback were able to improve subjects' performance significantly. Third, the zero rationality null cannot explain why subjects perform significantly better when faced with some kind of problems (conjunction of two highly probable or improbable events) rather than others (conjunction of one high and one low probability event).

For these reasons, we believe that a better explanation of the findings of our experiment has to be found elsewhere.

5.3. *Bounded rationality and the averaging heuristic*

The experimental data are consistent with the idea that agents are boundedly rational. They tend to commit the conjunction fallacy, but implementing ways to simplify the problem helps to improve their performance. This suggests that complexity may significantly reduce learning, even with dynamic feedback; the different performance in our two experiments might also be driven by differences in the ways subjects perceive the decision problem, according to how numerical assessments are elicited. In the simplified environment of Experiment 2, the provision of feedback plus monetary incentives may help to increase or to direct better the cognitive effort put out by the subjects. The fact that no learning was observed and that money incentives had a significant effect only in Experiment 2 would suggest that the observed deviations from the perfect rationality paradigm should not be undervalued.

In a bounded rationality perspective, it makes sense to ask which heuristics subjects are using in making choices in conjunction of events problems. If subjects use an *averaging* rule of thumb, i.e. they evaluate the probability of a conjuncted event as the average of the probabilities of the component events, the frequency of the conjunction fallacy should be lower when two high-probability or low-probability events are conjuncted than when a high and a low probability event are conjuncted (see Fantino et al., 1997). In the averaging heuristic model, if the conjunction consists of components rated about equally likely, then the incidence of the conjunction fallacy should be less than in the case of divergent likelihood of the individual events, since in the latter case the average likelihood falls more clearly above the less likely of the two components[9]. Conversely, there is no reason to expect this if subjects followed either Bayes' Rule or the representativeness heuristic rule of thumb of Tversky and Kahneman (1982), or if they had occasional lapses and only made random mistakes.

Research by Anderson (1981) and Shanteau (1975) supports the view that subjects may process perception judgments according to an 'information averaging' model. The conjunction fallacy evidence by Abelson et al. (1987), Gavanski and Roskos-Ewoldson (1991), Wells (1985) and Fantino et al. (1997) are consistent with the averaging model. The first and fourth rounds (Ralph 1 and 4) presented conjunctions combining two high probability events and two low probability events, respectively, while the others (Ralph 2 and 3, Bill and Linda) presented combinations of one high and one low probability event. Hence, it is a natural prediction of the averaging model that we would expect a lower average fallacy rate in rounds 1 and 4 (BothHighOrLow) than in the other rounds: as we saw in Section 4.3, such a prediction is supported by the data. We can conclude that there is evidence for an averaging heuristic, and lack of support for alternative bounded rationality hypotheses, such as purely random lapses[10].

6. Conclusions

The conjunction fallacy is an anomaly in human reasoning for which the conjunction of two events is rated more likely to occur than one of the events alone. In the context of decision under uncertainty, this violates the monotonicity axiom of probability, and consequentially also Bayes' Rule and the monotonicity axiom of preferences. Our experiments verified the robustness of the fallacy to the introduction of monetary incentives and of dynamic feedback. We found that, unless the decision problem is simplified, neither ameliorates the problem. A reduction of the general complexity of the decision problem improves the subjects' performance by a large extent, and is such as to make a joint introduction of monetary incentives and dynamic feedback have a further impact in reducing the fallacy rate. It is also possible that the different methods for eliciting numerical assessments, used in the two experiments, may change the way subjects approach the task.

Our overall interpretation of the evidence is that subjects are neither perfectly nor zero rational, but, rather, boundedly rational. An averaging heuristic seems best to explain the conjunction fallacy patterns in the data.

---

[9] Assume that $P(\mathrm{A}) = 20\%$ and $P(\mathrm{B}) = 30\%$. Then the average is 25%, and any slight random deviation (by 5%) may make the subject not commit the fallacy (by choosing 20% or less). Assume now that $P(\mathrm{A}) = 20\%$ but $P(\mathrm{B}) = 80\%$. Then the average is 50%, and one requires a random tremble of 30% in order for the conjunction fallacy not to occur. While the example is primitive, it makes the point that in an averaging model the fallacy is more likely to occur when events of similar likelihood are conjuncted.

[10] This explains why it is not useful to try to estimate a common error rate (see footnote 4).

Acknowledgements

We thank David Grether, Mark Machina, Lars Vik and an anonymous referee for valuable advice, and Matthew Mulford and the referee for useful references. The usual disclaimer applies. The research was conducted in accordance with human subjects guidelines at UCSD. Informed consent was obtained from subjects. For the statistical analysis, we used SPSS for Windows 7.5. The research was supported by NIMH Grant MH 57127 and NSF Grant IBN 94-07206 to the University of California, San Diego.

Appendix A

First page:

(*All Groups:*) Please follow the instructions on the questionnaire and evaluate each statement by writing your answer on the left line of each statement. (*Group 3 Exp. 1 only:*) *There is a correct answer*. (*All Groups:*) If you have any problem understanding the instructions, please ask the experimenter for further explanations.

(*Groups 1 and 2 both Exp.:*) After completion of EACH of the pages (Stop at the end of every page), inform the experimenter. The experimenter will tell you whether your answers on that particular page are correct.

(*Groups 1 only:*) For every single page answered correctly, you will receive US$ 3.

Second, third, fourth and fifth page: the same Ralph profile:

Ralph is 34 years old. He is intelligent though not especially creative, and his friends describe him as somewhat compulsive and dull. In college, he did well in the physical sciences but was weak in the humanities and social sciences.

(*Experiment 1 Likelihood Instructions:*) Please rank the likelihood of each of the following statements about Ralph by entering a number from 1 to 7 on the line to the left of the statement – where ‘1’ would be the least likely, and ‘7’ the most likely.

Do not use the same number twice.

(*Experiment 2 Likelihood Instructions:*) Please indicate the likelihood of each of the following statements about Ralph by entering a percentage on the line to the left of the statement – for example, ‘0’ would be virtually impossible, and ‘100’ virtually certain. You can think of the continuum of likelihood as looking like this:

0	00
virtually impossible	virtually certain

Since the statements are not mutually exclusive, the numbers (each from 0 to 100) need not sum to 100.

(*Exp. 2 Group 3 only:*) Hint: Pr(A&B) = Pr(A) × Pr(B).

____ Ralph builds radio-controlled gliders for a hobby.
____ Ralph is a building inspector.
____ Ralph collects stamps for a hobby.
____ Ralph plays in a heavy-metal band for a hobby.
____ Ralph is a park ranger.
____ Ralph is a kindergarten teacher.

Second page only:

____ Ralph is a building inspector who builds radio-controlled gliders for a hobby.

Third page only:

____ Ralph is a building inspector who plays in a heavy-metal band for a hobby.

Fourth page only:

____ Ralph is a kindergarten teacher who builds radio-controlled gliders for a hobby.

Fifth page only:

____ Ralph is a kindergarten teacher who plays in a heavy metal band for a hobby.

Sixth page:

Bill is 34 years old. He is intelligent but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Experiment 1 Likelihood Instructions: same as above, with word 'Bill' replacing 'Ralph'.

Experiment 2 Likelihood Instructions: same as above, with word 'Bill' replacing 'Ralph'.

(*Exp. 3 Group 3 only:*) Hint: Pr(A&B) = Pr(A) × Pr(B).

____ Bill is an architect.
____ Bill is an accountant.
____ Bill plays jazz for a hobby.
____ Bill surfs for a hobby.
____ Bill is a reporter.
____ Bill climbs mountains for a hobby.
____ Bill is an accountant who plays jazz for a hobby.

Seventh page:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Experiment 1 Only Likelihood Instructions: same as above, with word 'Linda' replacing 'Ralph'.

Experiment 2 Likelihood Instructions: same as above, with word 'Linda' replacing 'Ralph'.

(*Exp. 3 Group 3 only:*) Hint: Pr(A&B) = Pr(A) × Pr(B).

____ Linda is a teacher in a elementary school.
____ Linda is active in the feminist movement.
____ Linda is a psychiatric social worker.
____ Linda is a member of the League of Women Voters.
____ Linda is a bank teller.
____ Linda is an insurance salesperson.
____ Linda is a bank teller and is active in the feminist movement.

References

Abelson, R.P., Leddo, J., Gross, P.H., 1987. The strength of conjunctive explanations. Personality and Social Psychology Bulletin 13, 141–155.

Anderson, N.H., 1981. Foundations of Information Integration Theory. Academic Press, New York.

Benassi, V.A., Knoth, R.L., 1993. The intractable conjunction fallacy: statistical sophistication, instructional set, and training. Journal of Social Behavior and Personality 8, 83–96.

Brehmer, B., 1992. Dynamic decision making: human control of complex systems. Acta Psychologica 81, 211–241.

Diehl, E., Sterman, J.D., 1995. Effects of feedback complexity on dynamic decision making, organizational behavior. Organizational Behavior and Human Decision Processes 62, 198–215.

Fantino, E., Kulik, J., Stolarz-Fantino, S., Wright, W., 1997. The conjunction fallacy: a test of averaging hypotheses. Psychonomic Bulletin and Review 4, 96–101.

Gavanski, I., Roskos-Ewoldson, D.R., 1991. Representativeness and conjoint probability. Journal of Personality and Social Psychology 61, 181–194.

Gigerenzer, G., 1991. How to make cognitive illusions disappear: beyond heuristics and biases. European Review of Social Psychology 2, 83–115.

Grether, D.M., 1992. Testing Bayes rule and the representativeness heuristic: some experimental evidence. Journal of Economic Behavior and Organization 17, 31–57.

Hoel, P.G., 1984. Introduction to Mathematical Statistics. Wiley, New York.

Keppel, G., 1991. Design and Analysis: a Researcher's Handbook. Prentice Hall, Englewood Cliffs.

Mulford, M., Dawes, R.M., 1999. Subadditivity in memory for personal events. Psychological Science 10, 47–51.

Shanteau, J., 1975. Averaging versus multiplying combination rules of inference judgments. Acta Psychologica 39, 83–89.

Stolarz-Fantino, S., Fantino, E., Kulik, J., 1996. The conjunction fallacy: differential incidence as a function of descriptive frames and educational context. Contemporary Educational Psychology 21, 208–218.

Tversky, A., D. Kahneman, 1982. Judgements of and by representativeness, In: Kahneman, D., Slovic, P., Tversky, A. (Eds.), Judgement Under Uncertainty: Heuristics and Biases, Cambridge University Press, Cambridge, pp. 84–98.

Tversky, A., Kahneman, D., 1983. Extensional versus intuitive reasoning: the conjunction fallacy in probability judgment. Psychological Review 90, 293–315.

Tversky, A., Koehler, D.J., 1994. Support theory: a nonextensional representation of subjective probability. Psychological Review 101, 547–567.

Wells, G.L., 1985. The conjunction error and the representativeness heuristic. Social Cognition 3, 266–279.

Yates, J.F., Carlson, B.W., 1986. Conjunction errors: evidence for multiple judgment, procedures, including signed summation. Organizational Behavior and Human Decision Processes 37, 230–253.