Unreliable Sources and the Conjunction Fallacy

Andreas Jarvstad (jarvstad@cf.ac.uk)
School of Psychology, Cardiff University, Tower Building, Park Place
Cardiff, CF10 3AT, UK

Ulrike Hahn (hahnu@cf.ac.uk)
School of Psychology, Cardiff University, Tower Building, Park Place
Cardiff, CF10 3AT, UK

Abstract

We provide the first empirical test of a recent, normative account of the conjunction fallacy. According to Bovens and Hartman (2003), an unlikely statement from a partially reliable source is not necessarily more likely than a conjunction statement from another partially reliable source. Hence once information is considered to be coming from potentially not fully reliable sources, the conjunction fallacy is no longer at odds with probability theory. We provide here a simple experimental test of this account, and report comparisons of the Bovens and Hartmann model with Wyer's (1976) model and a simple averaging model. Wyer's model provided the best fit and the averaging model had the highest true positive rate in determining whether individual participants would commit the fallacy or not.

Keywords: conjunction fallacy, conjunction effect, probability judgment, number of components, Bayesian reasoning, normative models.

Introduction

The conjunction fallacy is arguably one of the best-known judgment errors in the cognitive literature. The fallacy consists of judging the conjunction of two statements, one likely (L) and one unlikely (U), as more likely than the least likely statement (P(L,U) > P(U)) (Tversky & Kahneman, 1983). At least for the original testing paradigm, the fallacy has proven to be robust (but see e.g., Hertwig, Benz & Krauss, 2008, on different versions of the problem). Yet, twenty-six years of extensive research[1] have failed to produce an adequate account of the phenomenon (Fisk, 2004).

Here a novel Bayesian account of the conjunction fallacy is assessed. Bovens and Hartman (2003) argue that it is natural to take into consideration the reliability of sources when judging the probability of statements. In particular, viewing experimenters as less than fully reliable sources is arguably not unreasonable (e.g., McKenzie, Wixted & Noelle, 2004), especially given the widespread use of deception in sub-fields of psychology (Nicks, Korn & Mainieri, 1997) and given that exposure to deception increases the expectation of being deceived in future experiments (Krupat & Garnonzik, 1994).

[1] A Google Scholar search with the term: ' "conjunction fallacy" OR "conjunction effect" ' yields 2050 hits (http://scholar.google.co.uk search performed 21.01.2009).

Bovens and Hartman (2003) show that, when source reliability is taken into account, it is sometimes *normative* to commit the conjunction fallacy. Specifically, receiving a report that matches our prior belief (a report of likely fact L), from a source who's reliability we are agnostic about, causes greater belief updating than receiving a report that seems improbable given our prior belief (a report of unlikely fact U). If we receive a statement that is the compound of a likely and an unlikely statement, we are justified in believing the compound statement (LU) more than only the unlikely fact – rendering the fallacy a "non-fallacy".

Bovens and Hartman's (2003) conjunction fallacy model is captured by the Bayesian Network in Figure 1. L and U are binary variables representing a likely and an unlikely claim respectively (an actual sample scenario follows below). Reliability (REL) is a binary variable that captures the reliability of the source; the prior degree of belief in the source's reliability is represented by ρ in the following. REP_L and REP_U, finally, are report variables representing whether or not we receive a report of facts L or U.

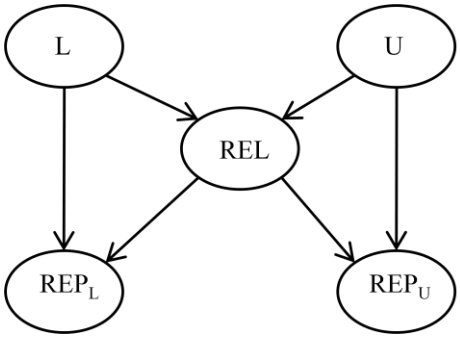

Figure 1: Bovens and Hartman's model of the conjunction fallacy as a Bayesian Network (adapted from Bovens and Hartman (2003)).

If a source is reliable it simply reports the truth. That is, it provides a positive report if the fact in question is true and no report if it is false. If the source is unreliable, however, the source decides at random whether or not to provide the report, independently of whether or not the fact in question is true. Specifically, the source will provide a positive report with probability *a*. Here it is assumed, throughout, that

unreliable sources are unbiased. That is, they are as likely to provide a report as to provide no report (a = 0.5).

Exploiting, the conditional dependencies in Figure 1, Bovens and Hartman derive the following equation for the posterior probability of the unlikely component statement:

$$P(U|REP_U) = (priorL(\rho + a\,\neg\rho)) / (priorU * \rho + a\,\neg\rho)) \quad (1)$$

and for the posterior probability of the conjunction:

$$P(L,U|\,REP_L, REP_U) = ((priorL * priorU * (\rho + a^2\,\neg\rho) / (priorL * priorU * \rho + a^2\,\neg\rho)) \quad (2)$$

They then show that, given a prior reliability of .5, there is a considerable range of priors for which the posterior degree of belief in the conjunction will be greater than for the unlikely component. That is:

$$\Delta P = P(L,U|REP_L, REP_U) - P(U|REP_U) > 0 \quad (3)$$

in which case it is normative to commit the "fallacy".

As the model assumes that the fit between perceived source reliability and the prior plausibility of the reported facts affects posteriors, it follows that adding a further likely or unlikely component to a standard conjunction (LU) can affect the incidence of the fallacy. For example, an additional likely component will increase the perceived reliability of the source and increase the belief in the conjunction relative to the classical LU conjunction. Hence, providing participants with an LLU conjunction should yield an even higher incidence of the fallacy (Bovens & Hartman, 2003). By contrast, adding a further unlikely component (LUU) should decrease its incidence.

To our knowledge, only three studies with three-component conjunctions have been conducted. Teigen, Martinussen and Lund (Experiment 1, 1996a) found no evidence that the addition of an extra likely component (LLU) increased the incidence of the fallacy relative to a two-component conjunction (LU) for a classical conjunction fallacy problem ('Linda, Tversky & Kahneman, 1983). In a second experiment, various outcomes in the 1994 football World Cup were estimated. When evaluating these outcomes, three-component conjunctions resulted in fewer conjunction fallacies compared to two-component conjunctions (see also Teigen, Martinussen & Lund, 1996b). Unfortunately, these results do not speak to the predictions of Bovens and Hartman's (2003) model. The frequency of two-component fallacies in the above studies was an aggregate of the incidence of the fallacy for three two-component conjunctions (e.g., L_1U, L_1L_2, L_2U). Hence it is impossible to determine, which two-component conjunction the three-component conjunction fallacy (e.g., L_1L_2U) frequency changed relative to.

A third study was conducted by Stolarz-Fantino, Fantino, Zizzo, and Wen (2003). They used a standard conjunction problem ('Bill', Tversky & Kahneman, 1983) but participants were required to estimate only the conjunction and were given the component probabilities (LUU = L(0.8), U(0.2), U(0.1)). The number of fallacies did not differ between the LUU condition (55% incidence) and the LU condition (52% incidence). However, it is unclear whether explicit probabilities are processed in the same way as internally generated probabilities. Overall, it is difficult to extrapolate from previous results to assess the predictions of Bovens and Hartman's (2003) model.

To investigate if an extra component does indeed affect the incidence of the conjunction fallacy, as predicted by Bovens and Hartman (2003), we manipulated the probability of an extra component in a classical conjunction fallacy scenario ('Bill', Tversky & Kahneman, 1983). After reading a personality description, participants' either rated an LLU, an LU or an LUU conjunction and their respective component probabilities. Bovens and Hartman's model predicts the following relationship for the incidence of the conjunction fallacy: LLU > LU > LUU.

To anticipate our results, there were no significant differences in the incidence of the fallacy. To further explore the accuracy of the model we compared the fit of the model with the fit of Wyer's (1976) model and a simple averaging model. Wyer's model produced the best quantitative fit, whilst the simple averaging model best predicted whether or not individual participants' committed the fallacy.

Method

Participants

Sixty undergraduates at Cardiff University participated receiving a chocolate bar as payment.

Materials

The material was presented in questionnaire format. Each questionnaire contained one modified Bill scenario (Tversky & Kahneman, 1983). There were three versions of the questionnaire corresponding to three conditions. The LU version contained one likely- (L) and one unlikely (U) component statement and their conjunction (LU). The LLU and the LUU version contained an additional likely (L) and an additional unlikely (U) component respectively. The LUU version is provided as an example:

> Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In college, he was strong in mathematics but weak in social studies and literature. You are told one of the following
>
> Bill is an accountant _____
>
> Bill surfs for a hobby _____

Bill plays jazz for a hobby _____

Bill is an accountant, plays jazz for a hobby and surfs for a hobby _____

How much would you believe the statement in each case? Please provide a rating between 0 (definitely untrue) and 100 (definitely true) for each statement.

Design and Procedure

A between subject design was used. The independent variable was conjunction type: LU, LLU or LUU. The dependent variable was the believability estimate.[2] Participants were approached in Cardiff University public areas. The task took approximately two minutes to complete.

Data Analysis

All responses were transformed by dividing each value with 100. A nominal conjunction fallacy incidence variable was created: estimates that conformed to (LU/LLU/LUU) > (U) were classified as exhibiting the fallacy and estimates that conformed to (LU/LLU/LUU) ≤ (U) as not exhibiting the fallacy (where U is the least likely component statement for each participant).

Results

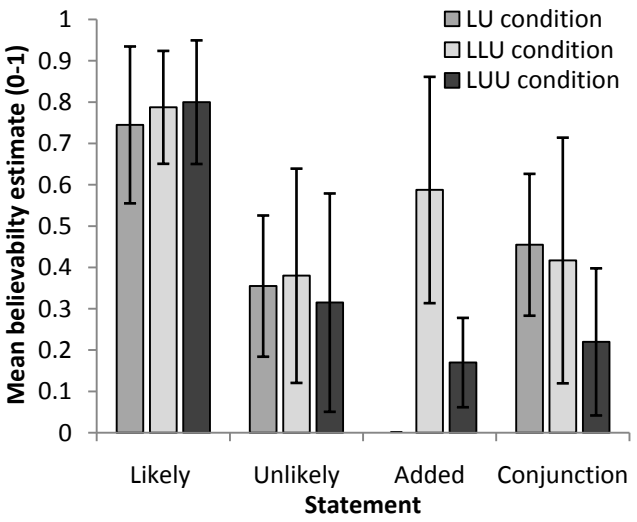

Figure 2: Mean believability estimate as a function of condition and statement type (error bars are ±1 standard deviation). Note that for the 'Added Statement' the light grey bar indicates a likely statement and the dark grey bar indicates an unlikely statement.

As can be seen in Figure 2, the likely statements and the unlikely statements were not rated differently across conditions (F (2, 59) = .65, p = .53, MSE= .026 and F (2, 59) = .39, p = .680, MSE = .055). The added likely component statement was rated as more likely than the added unlikely component statement (F (1, 39) = 40.23, p < .001, MSE= .043). In addition, there was an overall effect of condition on the conjunction statement probabilities (F (2, 44) = 6.39, p = .004, MSE = .05). The LUU conjunction was rated lower than the LU conjunction (Mean difference = -.235, p = .048, SEM = 0.08) and lower than the LLU conjunction (Mean difference = -.197, p = .019, SEM = 0.07. However, the LLU conjunction was not rated as more likely than the LU conjunction (Mean difference = 0.038, p = .95, SEM = 0.08). Also noticeable in Figure 2 is the relatively large variability in the LLU condition.

Although the mean estimates of the added likely and unlikely component statements differed significantly, the relatively large variability for the added likely component ('Added', LLU condition, Figure 2) may indicate a less than perfect likelihood manipulation. In support of this, the added likely component was not judged as significantly higher than .5 (one sample t-test, t(19) = 1.429, p > 0.05). The added unlikely component, by contrast, was judged to be significantly lower than .5 (t(19) = -13.65, p < 0.001).

Twelve participants (60%) in the LLU condition judged the conjunction as more likely than the unlikely component. Likewise, 12 participants (60%) in the LU condition committed the fallacy. Nine participants (45%) exhibited the conjunction fallacy in the LUU condition. The likelihood of a fallacy in the LU condition was used as a baseline empirical estimate of the likelihood of committing a fallacy. The number of participants committing the fallacy was not lower in the LUU condition than in any of the other conditions (one-tailed Binomial test (N = 20, p = .6) p = 0.13)).

Discussion

An equal number of fallacies were committed in the LLU and the LU condition. Thus, the addition of a likely component statement did not increase the frequency of the fallacy as predicted. There was a trend towards committing fewer fallacies in the LUU condition, but it was not significant. This trend in combination with the finding that the likely component was not judged as more likely than 0.5 makes it difficult to confidently refute the model. In the following we further assess the model by modeling.

Modeling

Modeling Methods

Participants' component estimates are, conceptually, posteriors[3]. However, Equation 2 above requires priors for

[2] Phrasing the problem in terms of either believability or probability has been found to yield equivalent results (Hertwig & Gigerenzer, 1999).

[3] There appears to be no straightforward way to empirically assess priors in classical conjunction problems given Bovens and Hartman's (2003) interpretation. The priors are formed as a result of reading the personality description. As soon as one is asked

the calculation of the conjunction posterior. These priors for the component statements can be eliminated by rearranging Equation 1 above, to derive the priors, and then using this to replace the priors in Equation 2:

$$\text{Prior} = (a - a\rho) / (((\rho + a - a\rho) / (\text{Posterior}) - \rho) \qquad (4)$$

Model fitting then simply involves finding the prior reliability (ρ) that minimizes the sum of the squared deviations between the model's predicted conjunction ratings and participants' conjunction ratings. In order to fit the model to the LLU and LUU conditions, Equation 2 (posterior probability for the conjunction) was extended to incorporate a third component statement:

$$P(L, A, U|REP_L, REP_A, REP_U) = ((\text{priorL} * \text{priorA} * \text{priorU} (\rho + a^3 \neg\rho) / (\text{priorL} * \text{priorA} * \text{priorU} * \rho + a^3 \neg\rho)) \qquad (5)$$

where A is the added likely or unlikely component.

The reliability parameter was allowed to vary between 0 and 1 and a single value of this parameter was estimated for the whole data set (i.e., it was assumed that all participants shared the same prior belief in the reliability of the source). Solver (Excel) was used to find the reliability parameter that minimized the sum of squared deviations.

The fit of Bovens and Hartman's (2003) model was assessed by several criteria. The quantitative fit of the model was compared to two other models in several ways. Firstly, models were compared on two indices of fit: the sum of the squared deviations and the r^2. Secondly, the models' ability to classify individual participants as committing the fallacy or not committing the fallacy (proportion of true positives and true negatives) was assessed. Finally, the models' ability to predict the average conjunction rating from the average component ratings for the standard conjunction condition (LU) was compared.

The first comparison model was Wyer's (1976) model of conjunction estimates. Wyer's model calculates the mean of a probability average and a multiplicative component:

$$P_{AB} = 1/2 ((P_A + P_B)/2 + P_A * P_B) \qquad (6)$$

and for the three conjunct case:

$$P_{ABC} = 1/2 ((P_A + P_B + P_C)/3 + P_A * P_B * P_C) \qquad (7)$$

The second model was a simple averaging model. It calculated the mean of participants' component ratings.

Modeling Results

Table 1: Model fits (sum of squared deviations & r^2) as a function of model type.

	$\sum (X_{Data} - X_{Model})^2$	r^2
Bovens & Hartman	3.25	0.47
Wyer	1.82	0.48
Averaging	3.50	0.42

As can be seen in Table 1, Wyer's (1976) model provides the best quantitative fit - regardless of whether fit is measured by the sum of squared deviations or r^2. This is impressive considering that it is has no free parameters. However, it is noteworthy that Bovens and Hartman's model (2003) explains approximately the same amount of variance as Wyer's model. The averaging model performs worse than the other models on both measures.

Inspection of the scatter plots in Figure 3, 4 and 5 reveals systematic deviations of the models from the data. Bovens and Hartman's model (Figure 3) consistently underestimates the conjunction ratings whereas Wyer's (1976) model overestimates low conjunction ratings and underestimates higher ratings (Figure 4).

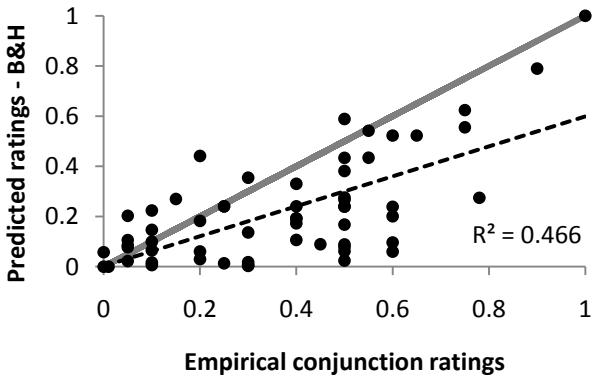

Figure 3: The relationship between predicted conjunction ratings from Bovens and Hartman's (2003) model and actual conjunction ratings. The dotted line is the line of best fit.

about some trait, one interprets the statement as a statement from a potentially less than fully reliable source.

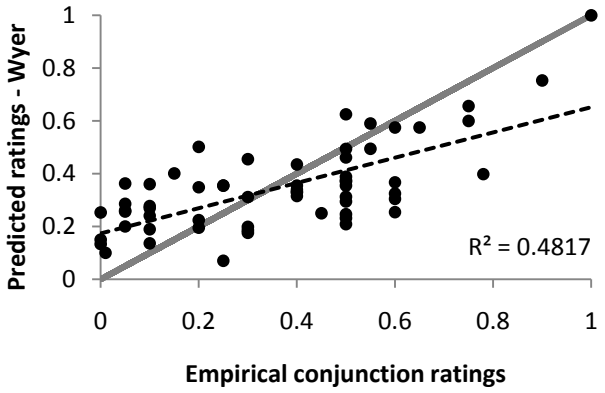

Figure 4: The relationship between predicted conjunction ratings from Wyer's (1976) model and actual conjunction ratings. The dotted line is the line of best fit.

The simple averaging model (Figure 5) substantially overestimates ratings below 0.5 and is quite accurate for estimates above 0.5.

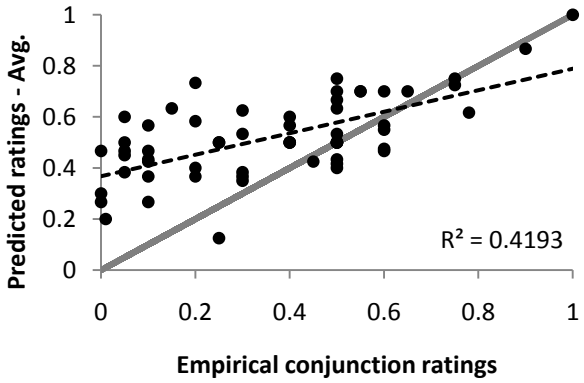

Figure 5: The relationship between predicted conjunction ratings from the simple averaging model and actual conjunction ratings. The dotted line is the line of best fit.

A quantitative model of the conjunction fallacy should arguably also be able to predict, from a participant's component ratings, whether or not that participant will commit the fallacy. The simple averaging model classified 62% (37/60) of the participants correctly, while Wyer's model (1976) classified 57% (34/60) of the participants correctly. Bovens and Hartman's model (2003) classified 50% (30/60) of the participants correctly. Hence, the averaging model appears best at discriminating between those who do and those who do not commit the fallacy. Although, as for the r^2 measures, no model vastly outperforms any of the others[4].

Yet another way to assess Bovens and Hartman's (2003) model is to ask how closely it predicts the average conjunction rating. In other words, how good is the model at capturing *average data*? The difference between the mean unlikely component rating and the mean conjunction rating in the LU condition was computed (Empirical – Figure 5). For each model, the difference between the model's conjunction estimate and the mean empirical unlikely component rating was computed. As can be seen in Figure 6, Wyer's (1976) model best matches the average empirical difference. The difference predicted by Bovens and Hartman's (2003) model is in the opposite direction to the empirical difference. In other words, given the mean ratings for the component statement in the data, the model predicts that the conjunction will be rated *lower* than the component probability.

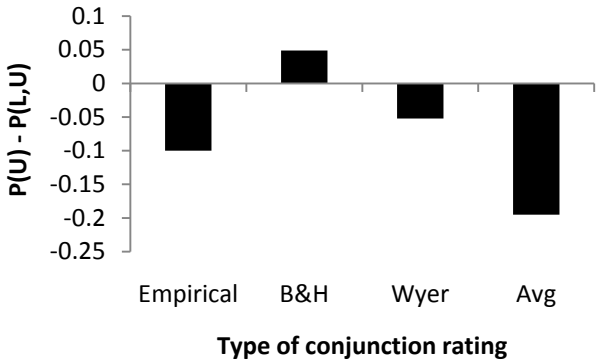

Figure 6: The difference between the mean empirical unlikely component rating and four different types of conjunction ratings.

General Discussion

Bovens and Hartman's (2003) prediction that the addition of an extra component in classical conjunction problems would affect the incidence of the conjunction fallacy was not confirmed. A trend toward a lower incidence of the fallacy in the LUU condition, and the fact that the extra likely component in the LLU condition was not rated as significantly higher than 0.5, however, means that it may be premature to refute the model on this alone.

Given that the likely component was not interpreted as likely and given the relatively small sample size (N=20 per

[4] Comparing the percentage of correctly classified responses using tests for differences in proportions between two samples, or comparing each sample to a reference proportion of .5 using a two-tailed one-sample binomial test produces n.s. results.

condition)[5] it is possible that adding extra components to conjunction problems could affect the fallacy. On the other hand, if a larger sample is required it would suggest that the effect is moderate at best. Furthermore, although Bovens and Hartmann's model (2003) fit the data reasonably well ($r^2 = 0.47$), other arguably more parsimonious models fit the data better.

The quantitative fit of Bovens and Hartman's model (2003) is poorer than a model without free parameters (Wyer, 1976), the ability of the model to predict whether a participant will commit the fallacy is poorer than both Wyer's model and the simple averaging model, and the model failed to qualitatively predict the difference between the average unlikely component rating and the average conjunction rating. It has also been noted that the model cannot predict certain conjunction fallacies (i.e., Björn Borg scenario, see Crupi, Fitelson & Tentori, 2008) and the task perspective that the model adopts has been criticized (Levi, 2004).

Hence, although participants may view experimenters as less than fully reliable sources of information and although some fallacies may be a result of this judgment it is unlikely to explain the majority of committed conjunction fallacies.

However, it should also be noted that although Bovens and Hartman's (2003) model contains a free parameter and the other models do not, it was "clamped" by assuming that it remained invariant across participants. The extent to which this might be an unduly severe restriction, is a matter for future empirical investigation.

It should also be noted that whilst the averaging model or a combined averaging-multiplying model fit the data in this study well, both models have other weaknesses. Neither model can, for example, predict the occurrence of double fallacies for LL conjunctions (Yates & Carlson, 1986).

Given that no account seems to be able to explain all aspects of the conjunction fallacy (Fisk, 2004), multi-component models of the fallacy may be one way forward. Future models should perhaps take into account individual differences in participants' problem solving strategies (e.g., see Yates & Carlson, 1986) or in participants' interpretation of the problems (e.g., Hertwig et al., 2008).

Acknowledgments

Andreas Jarvstad was supported by the Economic and Social Research Council. We thank the reviewers for their helpful comments.

References

Bovens Bovens, L., & Hartman, S. (2003). *Bayesian Epistemology.* Oxford, UK; Clarendon Press.

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, Confirmation, and the Conjunction Fallacy. *Thinking and Reasoning, 14,* 182-199.

Fisk, J. E. (2004). Conjunction fallacy. In R. F. Pohl, (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment, and memory*. London: Psychology Press.

Fisk, J. E., & Pidgeon, N. (1996). Component probabilities and the conjunction fallacy: Resolving signed summation and the low component model in a contingent approach. *Acta Psychologica, 94,* 1-20.

Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of and. *Cognition, 108,* 740-753.

Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12,* 275-305.

Krupat, E., & Garonzik, R. (1994). Subjects' expectations and the search for alternatives to deception in social psychology. *British Journal of Social Psychology,. 33,* 211–222.

Levi, I. (2004). Jaakko Hintikka. *Synthese, 140,* 37 – 41.

McKenzie, C. R. M., Wixted, J. T., & Noelle, D. C. (2004). Explaining purportedly irrational behavior by modeling skepticism in task parameters: An example examining confidence in forced-choice tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 947-959.

Nicks, S.D., Korn, J.H., & Mainieri, T. (1997). The rise and fall of deception in social psychology and personality research, 1921 to 1994. *Ethics and Behavior, 7,* 69–77.

Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., & Wen, J. (2003). The conjunction effect: New evidence for robustness. *American Journal of Psychology, 116,* 15-34.

Teigen, K. H., Martinussen, M. & Lund, T. (1996a). Linda versus world cup: Conjunctive probabilities in three-event fictional and real-life predictions. *Journal of Behavioral Decision Making, 9,* 77-93.

Teigen, K. H., Martinussen, M. & Lund, T. (1996b). Conjunction errors in the prediction of referendum outcomes: Effects of attitude and realism. *Acta Psychologica, 93,* 91-105.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Wyer, R. S., Jr. (1976). An investigation of the relations among probability estimates. *Organizational Behavior and Human Performance, 15,* 1-18.

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.

[5] Cf., Tversky & Kahneman (1983) N = ~ 80, but see Hertwig & Gigerenzer (1999) N = ~ 22.