

A Text Corpus Analysis Approach to the Conjunction Fallacy

Yuichiro Wajima (wajima@nm.hum.titech.ac.jp)

Kuninori Nakamura (knaka@ky.hum.titech.ac.jp)

Asuka Terai (asuka@nm.hum.titech.ac.jp)

Kimihiko Yamagishi (kimihiko@ky.hum.titech.ac.jp)

Masanori Nakagawa (nakagawa@nm.hum.titech.ac.jp)

Graduate School of Decision Science Technology, Tokyo Institute of Technology
2-12-1, Ookayama, Meguro-Ku, Tokyo, 152-8552, Japan

Abstract

This study aims to explain the conjunction fallacy (Tversky & Kahneman, 1983) in terms of degree of confirmation (Crupi, Tentori, & Gonzalez, 2007) by employing corpus data analysis. To accomplish this, we calculated indexes of the degrees of confirmation from the British National Corpus and fitted them to data of the previous study (Shafir et al., 1990). The results show that a major index of the degree of confirmation (Crupi et al., 2007) can significantly predict the conjunction fallacy, indicating a relationship between the conjunction fallacy and degree of confirmation as well as the importance of corpus data to explain biases in judgment.

Keywords: Conjunction Fallacy, Degree of Confirmation, Corpus Data Analysis

Introduction

The conjunction fallacy (Tversky & Kahneman, 1983) is one of the most famous violations of the normative axiom of probability. A representative experiment of this phenomenon asked participants to think about Linda as she is described in the following sentences:

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 antinuclear demonstrations.

Then, participants were required to rank various statements about Linda "by their probability." Two of these statements were "B" and "B and F":

(B) Linda is a bank teller.

(B and F) Linda is a bank teller and is active in the feminist movement.

In Tversky and Kahneman (1983), 85% of the participants ranked "B and F" as more probable than "B." This judgment is in apparent violation of the conjunction law $Pr(X \text{ and } Y|Z) < Pr(X|Z)$ for any statements X, Y, and Z, with strict inequality for nontrivial cases such as the present example. Tversky and Kahneman refer to this violation as the conjunction fallacy, and many studies have replicated it under various experimental conditions (see

Hertwig & Chase, 1998, for a review of findings; the original report is Tversky & Kahneman, 1983).

Many theoretical explanations for the conjunction fallacy as well as the replications have been reported (for a review, see Costello, 2009; Sides, Osherson, Bonini, & Viale, 2002). Some researchers (e.g., Crupi, Fitelson, & Tentori, 2008; Tversky & Kahneman, 1983) suggest that participants' responses in the conjunction fallacy are not based solely on the probability of the conjunction itself but also on other factors such as the representative heuristics (Tversky & Kahneman, 1981) or degree of the confirmation given by the conjunctions. Other researchers assume that participants utilize their pragmatic knowledge to understand meanings of the experimental materials employed in studies on the conjunction fallacy. This approach focuses on assumptions about how participants answer in the experimental tasks (Chase, Hertwig, & Gigerenzer, 1998; Dulany & Hilton, 1991; Macdonald & Gilhooly, 1986; Morier & Borgida, 1984; Politzer & Noveck, 1991), saying that participants, correctly following the pragmatics of communication rather than the logical meaning in the experimental task, interpret the single statement B as meaning " $B \wedge \neg A$ " and so rightly prefer the conjunction $B \wedge A$ (Dulany & Hilton, 1991; Macdonald & Gilhooly, 1990; Politzer & Noveck, 1991). In addition, researchers have recently proposed that the conjunction fallacy is an artifact that occurs due to unnatural experimental settings (Hertwig & Gigerenzer, 2002), possibly in combination with measurement error (Costello, 2008).

We seek to propose another explanation of the conjunction fallacy that is inspired by a recent development of the concept of confirmation (Carnap, 1955; Fitelson, 1999; Crupi, Tentori, & Gonzalez, 2007) and utilizes statistical corpus analysis. In what follows, we will argue how theories of confirmation can explain the conjunction fallacy. In addition, we will also present a solution to these problems using statistical corpus analysis.

Degree of confirmation

Consider a situation in which a rational agent collects data for hypothesis confirmation. The strength of the agent's belief about the hypothesis after seeing the data can be described as the posterior probability of the hypothesis, defined by Bayes' theorem, shown in (1):

$$P(h | d) = \frac{P(h)P(d | h)}{P(d)} \quad (1)$$

As you can see, Bayes' theorem defines a relationship between the prior belief ($P(h)$) and the posterior belief. In this vein, *whether* the data support the hypothesis or not can be represented by the difference between the prior and posterior probabilities; when the posterior probability is higher than the prior probability, the data support the hypothesis, and when not, they disconfirm it. Thus, degree of confirmation $c(h, d)$ can be expressed as shown in (2):

$$c(h, d) = \begin{cases} > 0, P(h | d) > P(h) \\ = 0, P(h | d) = P(h) \\ < 0, P(h | d) < P(h) \end{cases} \quad (2)$$

We must mention that it is possible to define many kinds of criteria that can satisfy the equation in (2). This equation solely proposes that the data support the hypothesis if the posterior probability is higher than the prior probability; it does not determine the degree of support. In fact, there are various types of criteria for degree of confirmation that can satisfy (2). Table 2 shows representative indexes of the degrees of confirmation (also see Eells & Fitelson, 2002; Fitelson, 1999; Tentori, Crupi, Bonini, & Fitelson, 2007).

Tentori, Crupi, Bonini, and Fitelson (2007) report on experiments that compared the adequacy of several such measures as descriptions of confirmation judgment in a probabilistic context. In the experiments, they showed participants two opaque urns that contained different numbers of white and black balls, and the numbers of each were known to the participants. Then, participants randomly chose one of the two urns by a coin toss, but the outcome of the coin toss was unknown to them. After selecting an urn in this way, the participants were required to draw a ball and estimate the impact of the results of the ball drawing with various types of response scales in 10 successive trials. Based on the number of balls that the two urns contained, the participants calculated various types of the confirmation indexes and compared them to the estimated degree of confirmation. Results showed that some of the confirmation indexes significantly correlate to the participants' responses, indicating that human hypothesis confirmation follows rational thought as the Bayesian agent is assumed to do so.

Crupi, Fitelson, and Tentori (2008) suggest that the conjunction fallacy can be explained in terms of the Bayesian hypothesis confirmation. They insist that participants in the conjunction fallacy experiments do not consider probabilities for conjunctions or each proposition that composes the conjunction. Rather, they appear to estimate how the description supports the conjunction as data. Based on this idea, Crupi, Fitelson, and Tentori specify a condition in which the degree of confirmation for the conjunction becomes larger than that for each proposition. This condition is shown in (3):

$$c(h_1, e) \leq 0 \quad \text{and} \quad c(h_2, e | h_1) \geq 1, \quad (3)$$

then $c(h_1 \& h_2 | e) \geq c(h_1, e)$

Another important feature of the condition is its robustness; Crupi et al. (2008) also show that this condition holds any confirmation measures, including D , R , L , C , S , and Z (table 1). In other words, each of these confirmation measures can explain the conjunction fallacy if people's probability judgment about conjunction depends on the degree of confirmation. Thus, not only *whether* the degree of confirmation but also *what* measure can explain the conjunction fallacy is an interesting empirical question.

Table 1. variation of the degree of conformation

$$D(A, E) = p(A | E) - p(A)$$

(Carnap, 1950; Eells, 1982)

$$C(A, E) = p(A \& E) - p(A) \times p(E)$$

(Carnap, 1950)

$$S(A, E) = p(A | E) - p(A | \neg E)$$

(Christensen, 1999; Joyce, 1999)

$$Z(A, E) = \begin{cases} \frac{p(A | E) - p(A)}{1 - p(A)} & \text{if } p(A | E) \geq p(A) \\ \frac{p(A | E) - p(A)}{p(A)} & \text{otherwise} \end{cases}$$

(Crupi, Tentori & Gonzalez, 2007)

However, one methodological problem of Crupi et al.'s (2007) explanation lies in how to determine probabilities related to the conjunction. All the confirmation measures, although they are not probabilities themselves, require values of probabilities for h_1 , h_2 , and e . In the experimental paradigm of the conjunction fallacy, this requirement is very difficult to fill, because the conjunction fallacy demonstrates that peoples' probability judgments for these components deviate from the normative principle. In other words, we cannot define the values of the probabilities required to calculate the degree of confirmation from probability judgment data. Rather, it is desirable to define the probability values based on objective data, not on subjective probability judgments.

Statistical corpus analysis

Statistical corpus analysis can address the above problems. Statistical corpora contain large amounts of sentences that are collected from natural language communication such as literature (British National Corpus) or newspapers (Mainichi-shinbun database) and enable researchers to access information on word frequencies or co-occurrences.

These databases can be considered as reflecting everyday verbal communication and have assumed to be substitutions for human memory structure (e.g., Steyvers & Tenenbaum, 2006; Stewart, Chater, & Brown, 2006).

Recently, the number of studies that attempt to describe human reasoning with corpus statistical data has been increasing (e.g., Sakamoto & Nakagawa, 2007; Stewart, Chater, & Brown, 2006; Terai & Nakagawa, 2007; Utsumi, 2010). These studies assume that frequencies of appearance or co-occurrence among words reflect word affinities or the strength of associative connections among words and substitute them as parameters for statistical models. For example, Sakamoto and Nakagawa (2007) hypothesized cognitive processes of category-based induction (e.g., Osherson et al., 1990) as classification based on associations between premises and conclusions, and employed probabilities for co-occurrence among the words that are contained in the premise and conclusion. The results of their study supported the hypothesis, suggesting that probabilistic properties of corpus data reflect basic knowledge of human reasoning.

Additionally, applying the corpus data analysis approach to the issue of the conjunction fallacy appears to be very natural upon inspection of the experimental procedure. In the typical experiment, participants read sentences that describe features of characters, such as “outspoken” or “philosophical,” and are required to rank alternatives (e.g., “bank teller,” “feminist”) by their probability. In this situation, it is natural to assume that participants access their memory to solve the task and that this assessment to their memory affects their probability judgment for each proposition. If corpus data reflect memory strength of the words contained in the stimulus sentences, we can employ the corpus data to test Crupi et al.’s (2007) hypothesis.

The purpose of this study

From the above discussion, we derive the following hypothesis about the conjunction fallacy; in the experimental procedure of the conjunction fallacy, participants estimate the degree of confirmation for the conjunction by assessing their memory strength. To test this hypothesis, we report two studies. The first reanalyzed the data of Shafir, Smith, and Osherson (1990) that are adequate to test our approach. The second collected new data from Japanese participants. In both studies, we calculate parameters of the degrees of probability with any given assumption using a language corpus and examine the idea that the conjunction fallacy occurs as a result of the Bayesian updating of degree of the confirmation.

Study 1: Reanalysis of Shafir et al. (1990)

The purpose of Study 1 is to test our hypothesis by reanalyzing the data of Shafir et al (1990). Shafir et al (1990) required participants to rate probabilities of occurrence of the contents of both 15 conditional sentences and combinations of propositions. The sentences that were judged by participants had the frame: Every single in the group is . For example, they employed stimulus sentences such as those below in order to consider the conjunction fallacy by comparing single and conjunctive condition texts:

Single condition:

“Every single lawyer in the group is conservative.”

Conjunctive condition:

“Every single labor-union lawyer in the group is conservative.”

In this case, the data (e) is “conservative” and single hypothesis (h_1) and conjunctive hypothesis ($h_1 \& h_2$) are “lawyer” and “labor-union lawyer,” respectively. The participants were requested to rate the probability of each condition. They calculated the degree of fallacy by subtracting the probability of the single condition from the probability of the conjunctive condition.

In Shafir et al.’s (1990) experiment, the attributes of the data (e) and hypothesis (h_1, h_2) are comprised of one word. In such experiments on the conjunction fallacy using stimuli, it is natural to assume that participants’ judgments are dependent on the characteristics and relationships of each word; therefore, it is very appropriate to consider the connection of language statistics and the conjunction fallacy. Thus, we selected it as the subject of analysis of this study. This study compares the degree of conformation that was calculated from the occurrence and co-occurrence probabilities of words in a language corpus and the results of the conjunction fallacy experiment performed by Shafir et al. (1990).

We use the British National Corpus (BNC) as language statistical data. The BNC, launched in 1991, is the world’s largest English corpus, consisting of approximately 100 million words including both British written and colloquial words, and is designed to represent modern British English as broadly as possible. It consists of approximately 90% written words and 10% colloquial words. Data to be added are selected based on statistical value by area of the number of annual publications in the UK. We decided to use the BNC as the original data source for calculating the degree of probability for the study because it contains enough data, and the date on which the original BNC texts were prepared is close to, or prior to, the date Shafir et al.’s (1990) experiment was performed.

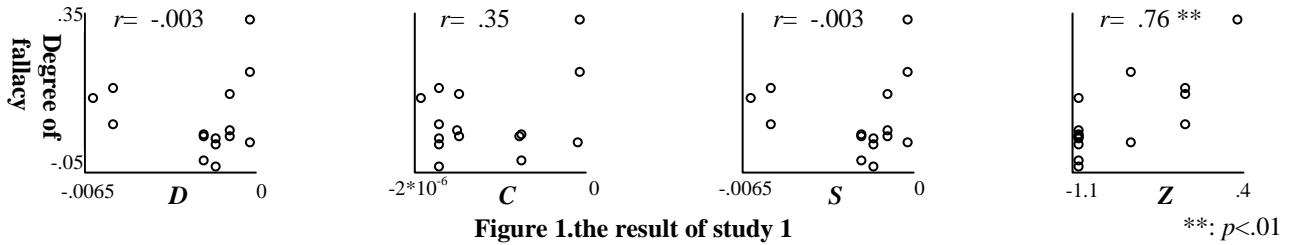


Figure 1.the result of study 1

The degree of probability was calculated through the following procedure.

First, we calculated occurrence and co-occurrence probabilities. We picked up 5,101,034 sentences by BNC. A frequency of a word (e.g. *h1*) was counted according to whether the word (*h1*) is included in each sentence. The occurrence probability of *h1* was calculated as follows,

$$p(h1) = n_{h1}/n_{all}$$

where the n_{all} is a number of all sentences ($n_{all} = 5,101,034$), and the n_{h1} is the frequency of the word (*h1*).

In the case of co-occurrence (e.g. *h1&h2*), we counted a co-frequency according to whether both words (*h1*) and (*h2*) are included in each sentence. The co-occurrence probability $p(h1 \& h2)$ was calculated as follows,

$$p(h1 \& h2) = n_{h1\&h2}/n_{all}$$

where the $n_{h1\&h2}$ is the co-frequency of words (*h1* & *h2*). According to the above procedure, all occurrence and co-occurrence probabilities for calculation of the confirmation-degree are calculated.

Then, after the calculating conditional probability based on Bayes' theorem, we calculated the degree of confirmation ($c(h1,e)$, and $c(h1\&h2,e)$ for each confirmation measures) to determine the degrees of probability of the conjunction and each proposition. Finally, we calculated the correlation coefficient between the degree of fallacy of Shafir et al. (1990) and $c(h1\&h2,e) - c(h1,e)$.

Scatter plots in Figure 1 demonstrate relationships among the data of Shafir et al. (1990) and values of the confirmation measures that are calculated from the BNC. Among the four confirmation measures, only *Z* correlated to the data significantly ($r = 0.76, p < 0.01$). These results support our proposition that the degree of confirmation calculated from the corpus data can explain the conjunction fallacy. Of course, robustness of the current results should be examined by another study, and Study 2 attempts to replicate this finding by using Japanese corpus in Japanese participants

Study 2 Examination in Japanese participants

In study 2, we conducted an experiment by using Japanese corpus for Japanese participants. The participants were 77 Japanese college students in a computer-literacy class. In the experiment, we used same contents that used in Study 1 and translated those 15 conditional sentences and combinations of propositions in Japanese. Participants were given the personality character and required to select one statement as most probable.

For example, participants were asked to think about a person A who is described in the following sentences:

“A is conservative.”

Then, participants were required to select one statement from three statements as most probable about person A. Three of these statements were as follows,

- i. **A is lawyer.**
- ii. **A is labor unionist.**
- iii. **A is labor-unionist and lawyer.**

In this case, the data (*e*) is “conservative,” and two single hypothesis (*h1*, *h2*) and conjunctive hypothesis (*h1&h2*) are “lawyer”, “labor unionist” and “labor-unionist and lawyer,” respectively. In study 2, a degree of fallacy is defined by selected ratio of conjunctive statement (iii).

In study 2, we use a Japanese corpus data that consists of approximately 650 million words, which extracted from 18-year span of a Japanese news paper (Mainichi Shinbun), 9706 books (by Aozora Bunko), two encyclopedias (Gakken Kokugo Daijiten, Nihon Hyakka Jiten) and Blog data. This Japanese corpus data consists of approximately 67% news papers, 8% books, 15% blogs and 10% encyclopedias. The degree of confirmation ($c(h1,e)$, $c(h2,e)$ and $c(h1\&h2,e)$ for each confirmation measures) was calculated through the same procedure in Study 1. We calculated the correlation coefficient between $c(h1\&h2,e) - c(h1,e)$ (or $c(h1\&h2,e) - c(h2,e)$) and the degree of fallacy of the result of experiment.

Figure. 2 show the relationships among the data of experimental result and values of the confirmation measures that are calculated from Japanese corpus. In the case of $c(h1\&h2,e) - c(h1,e)$, *Z* correlated to the data significantly ($r = 0.64, p < 0.05$), and the relation between $c(h1\&h2,e) - c(h2,e)$ and the degree of fallacy shows marginal significant positive correlations in the degree of *Z* ($r = .46, p < .1$).

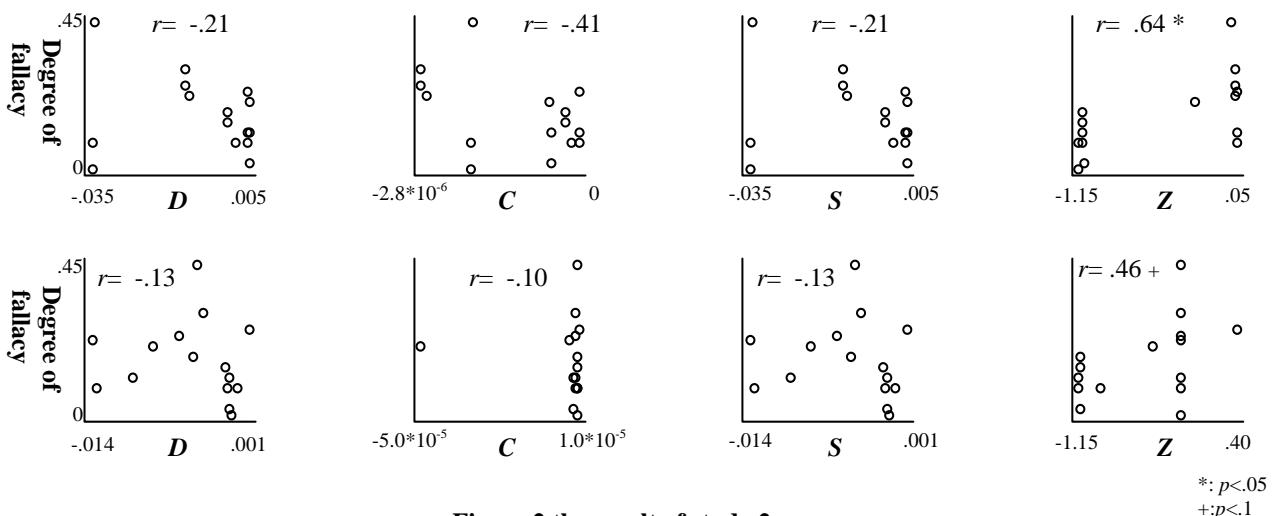


Figure 2.the result of study 2

Upperside : the relationship between $c(h1\&h2,e) - c(h1,e)$ and the degree of fallacy.
 Bottom side : the relationship between $c(h1\&h2,e) - c(h2,e)$ and the degree of fallacy

Discussion

The results of the two studies consistently showed that conjunction fallacy is explained by the degree of confirmation that can be defined by difference between prior and posterior probability that are reflected in the corpus data. These results are consistent with the hypotheses by Crupi et al. (2007), and can be positioned as another example that succeeded to explore cognitive processes of human reasoning by utilizing the corpus data.

We argue that the corpus data analysis is prospective approach to explore biases in human reasoning. Although the corpus data have been used to explore human reasoning such as category-based induction (Sakamoto & Nakagawa, 2007) or metaphor understanding (Terai & Nakagawa, 2007), application to the domain of cognitive biases very novel. We think that probabilistic aspects of the corpus data are useful for investigation to various agendas of decision-making research. We are planning to perform similar analyses and experiments on such decision making agendas to assess the probability of prediction of human decision-making through calculation of degree of probability based on language corpus.

Among the various types of the criteria of degree of confirmation, Z showed the highest fit to the data. This finding is meaningful because it suggests general cognitive processes for probability judgment including the conjunction fallacy. As the expression show in Table 2 indicates, this index defines degree of the confirmation as normalized difference between prior and posterior probability: when the target hypothesis is supported, sum of prior probabilities for hypothesis other than the target hypothesis becomes a normalizing term, and when disconfirmed, prior probability of the target hypothesis become the normalizing term. In other words, this index

uses set of the disconfirmed hypotheses as the normalizing term. In this vein, this index considers degree of confirmation as a result of competition between the target and the other hypothesis. This finding corresponds to recent findings that consider human reasoning as comparison between the target and alternative hypothesis (e. g., McKenzie & Amin, 2002; McKenzie & Mikkelsen, 2002, 2007). In addition superiority of Z suggests a coherency of human judgment because it satisfies the conditions needed to be a coherent index of confirmation (Crupi et al, 2007). In sum, the current results suggest a possibility that conjunction fallacy can be positioned as result of the rational cognitive processing. Although Costello (2008) made normative sense of the conjunction fallacy considering combination of measurement error terms, the current study proposed another explanation that can also rationalize the conjunction fallacy in terms of the normative sense.

One crucial assumption of this study is that probabilities for statements can be measured from the corpus data. Our rational for this assumption is mainly based on the findings in the previous studies that succeeded to explain human reasoning by utilizing the corpus data (e. g., Sakamoto & Nakagawa, 2007; Stewart, Chater, & Brown, 2006; Terai & Nakagawa, 2007; Utsumi, 2010). This rationale strongly depends on pragmatic arguments: that is, it solely states that this assumption is valid because of its ability to predict the data. Thus, this rationale does not appear to be strong because it does not explain *why* the corpus can be utilized to predict probability judgment. We believe that the current study add another evidence for utility of the usage of corpus data to investigate human reasoning. However, we must clarify our rational more precisely to justify using the corpus data for probability judgment.

References

British National Corpus (BNC) <http://www.natcorp.ox.ac.uk/>

Carnap, R. (1950). Logical foundations of probability. Chicago: University of Chicago Press.

Carnap, Rudolph, (1955), "Meaning and synonymy in natural languages", *Philosophical Studies*, 7: 33-47. Reprinted in Carnap 1956, pp. 233-247

Chase, V. M., Hertwig, R. and Gigerenzer, G. (1998), Visions of rationality, *Trends in Cognitive Sciences*, 2 206-214.

Christensen, D. (1999). Measuring confirmation. *Journal of Philosophy*, 96, 437- 461.

Costello, J. (2008) How probability theory explains the conjunction fallacy, *Journal of Behavioral Decision Making*, 22, 3, 213 – 234

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking & Reasoning*, 14, 182–199.

Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74, 229 – 252.

Dulany, D. E. and Hilton, D. J. (1991), Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition*, 9, 85–110.

Eells, E. (1982). Rational decision and causality. Cambridge, UK: Cambridge University Press.

Eells, E. & Fitelson, B. (2002). Symmetries and asymmetries in evidential support. *Philosophical Studies*, 107, 129- 142.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–129.

Joyce, J. (1999). The foundations of causal decision theory. Cambridge (UK): Cambridge University Press

Macdonald, R. and Gilhooly, K. (1986), More about Linda or conjunctions in context. *European Journal of Cognitive Psychology*, 2, 57–70.

McKenzie, C. R. M., & Amin, M. B. (2002). When wrong predictions provide more support than right ones. *Psychonomic Bulletin and Review*, 9, 821-828.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, 54, 33-61.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin and Review*, 7, 360- 366.

Morier, D. and Borgida, E. (1984), The conjunction fallacy: A task specific phenomenon? *Personality and Social Psychology Bulletin*, 10, 243–252.

Osherson, D., Smith, E. E., Wilkie, O., Lopez, A., and Shar, E. (1990). Category-based induction. *Psychological Review*, 97(2):185-200.

Politzer, G. and Noveck, I. A. (1991), Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research*, 20, 83–103.

Sakamoto, K., Nakagawa, M. (2007) "Risk Context Effects in Inductive Reasoning: An Experimental and Computational Modeling Study", B. Kokinov et al. (Eds.): CONTEXT 2007, LNAI 4635, Springer-Verlag Berlin Heidelberg, pp425-438

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory and Cognition*, 18, 229-239.

Sides, D., Osherson, N., Bonini, and R. Viale, (2002) On the reality of the conjunction fallacy. *Memory & Cognition*, 30(2):191-198.

Steyvers, M., & Tenenbaum, J. B. (2005). The large scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29, 41- 78.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53, 1-26.

Tentori, K., Bonini, N., and Osherson, D. (2004). The conjunction fallacy: a misunderstanding about conjunction? *Cognitive Science* 28 (3), pp. 467–477.

Tentori, Crupi, Bonini, & Osherson (2007). Probability, confirmation, and the conjunction fallacy, *Thinking & Reasoning*, Volume 14, Issue 2, pages 182 - 199 [3]

Terai, A., Nakagawa, M. (2007). A neural network model of metaphor understanding with dynamic interaction based on a statistical language analysis: targeting a human-like model, *International Journal of Neural Systems*, 17(4), 265-274.

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

Utsumi, A. (2010). Computational Exploration of Metaphor Comprehension Processes Using a Semantic Space Model. *Cognitive Science*, DOI: 10.1111/j.1551- 6709.2010.01144.x

Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus and problem type. *Cognition*, 107, 105–136.