



Avoiding the conjunction fallacy: Who can take a hint?

Simon Klein

Spring 2017
Master's thesis, 30 ECTS
Master's programme in Cognitive Science, 120 ECTS
Supervisor: Linnea Karlsson Wirebring

Acknowledgments: The author would like to thank the participants for enduring the test session with challenging questions and thereby making the study possible, his supervisor Linnea Karlsson Wirebring for invaluable guidance and good questions during the thesis work, and his fiancée Amanda Arnö for much needed mental support during the entire process.

AVOIDING THE CONJUNCTION FALLACY: WHO CAN TAKE A HINT?

Simon Klein

Humans repeatedly commit the so called “conjunction fallacy”, erroneously judging the probability of two events occurring together as higher than the probability of one of the events. Certain hints have been shown to mitigate this tendency. The present thesis investigated the relations between three psychological factors and performance on conjunction tasks after reading such a hint. The factors represent the understanding of probability and statistics (statistical numeracy), the ability to resist intuitive but incorrect conclusions (cognitive reflection), and the willingness to engage in, and enjoyment of, analytical thinking (need-for-cognition). Participants ($n = 50$) answered 30 short conjunction tasks and three psychological scales. A bimodal response distribution motivated dichotomization of performance scores. Need-for-cognition was significantly, positively correlated with performance, while numeracy and cognitive reflection were not. The results suggest that the willingness to engage in, and enjoyment of, analytical thinking plays an important role for the capacity to avoid the conjunction fallacy after taking a hint. The hint further seems to neutralize differences in performance otherwise predicted by statistical numeracy and cognitive reflection.

Människor begår ofta det så kallade ”konjunktionsfelslutet”, genom att felaktigt bedöma sannolikheten för sammanträffandet av två händelser som större än sannolikheten för en av händelserna. Vissa typer av ledtrådar har visat sig mildra denna tendens. Denna uppsats undersökte relationerna mellan tre psykologiska faktorer och prestation på konjunktionsuppgifter efter att ha läst en sådan ledtråd. Faktorerna motsvarade förståelsen för sannolikhet och statistik (statistisk räknefärdighet, *eng.*, statistical numeracy), förmågan att motstå intuitiva men felaktiga slutsatser, (kognitiv reflektion, *eng.*, cognitive reflection), samt viljan och lusten till analytiskt tänkande (behov-av-tänkande, *eng.*, need-for-cognition). Deltagare ($n = 50$) besvarade 30 korta konjunktionsuppgifter och tre psykologiska mätskalor. En bimodal svarsfördelning motiverade dikotomisering av resultaten. Behov-av-tänkande var signifikant, positivt korrelerat med prestation, vilket varken räknefärdighet eller kognitiv reflektion var. Resultaten tyder på att viljan och lusten till analytiskt tänkande spelar en viktig roll i förmågan att undvika konjunktionsfelslutet efter att ha fått en ledtråd. Ledtråden verkar också neutralisera skillnader i prestation som annars uppstår på grund av räknefärdighet och kognitiv reflektion.

One of the main trends in cognitive psychology has been to compare human judgments and decisions against some form of normative standard. This perspective gained traction through the *heuristics-and-biases program* initiated by Amos Tversky and Daniel Kahneman (1971, 1974). The fundamental perspective taken in this research program is that human judgments and decisions under some circumstances deviate systematically (i.e., are biased) from normative theories of probability and decision-making. This deviation is hypothesized to be due to the use of mental shortcuts or “rules of thumb” known as heuristics. As Gilovich and Griffin (2002) point out, the *cognitive revolution* was an important factor in the development of the heuristics-and-biases program, since it allowed for viewing the mind as dependent on subroutines assessing adjustment, availability, and similarity. Sloman (1996) gave a computational account of heuristics and biases, as well as the related concept of *dual-systems*. Sloman reviewed the empirical case for an often-made distinction between *associative* and *rule-based* thoughts. While associative thoughts are based on similarities and dependent on temporal factors, rule-based thoughts can operate on symbolic structures in a logical and coherent manner. It is

commonly suggested that heuristics, and the biases and fallacies that they give rise to, are the result of associative cognitive processes being used in a situation where logical reasoning is needed to reach the normative solution. Evans and Stanovich (2013) gave a closer description of the relations between dual-system (alternatively *dual-process*) theories and research on heuristics and biases, nuancing the rather simplistic claim that intuition is always to blame, while still maintaining the usefulness of the distinction and its relation to biased behavior and cognition.

From the viewpoint of this thesis, the most important heuristic described by Tversky and Kahneman (1974) is the *representativeness heuristic*, wherein probability estimates concerning two or more events, in a broad sense, are hypothesized to be based on how representative, or similar, the events are of each other. This heuristic could for example be used when answering the question “what is the probability that object A belongs to class B?”. Tversky and Kahneman gave the example of a fictive person Steve who, among other traits, is shy, but helpful, has a need for order and structure, and a passion for detail. Since the description of Steve is representative of a librarian, judgments made concerning Steve’s occupation will often portray him as highly likely to be a librarian. However, from a normative standpoint, similarity or representativeness is not the only relevant factor in making such judgments. For example, if the question concerns whether it is more probable for Steve to be a librarian or a farmer, the base rate of librarians and farmers in the population should have a prominent role in the judgment process. A person who relies too heavily on the representativeness heuristic will however respond in a biased manner, since other factors are disregarded. Later accounts of the representativeness heuristic can be found in for example Kahneman and Tversky (1982) and Kahneman and Frederick (2002).

One of the most robust findings in research on heuristics and biases is the *conjunction fallacy*, or *conjunction effect*, introduced by Kahneman and Tversky (1982) and Tversky and Kahneman (1983). It is committed whenever the probability of the co-occurrence of several events is judged as being higher than the probability of one of the events, regardless of whether the other events occur. For two events A and B , the fallacy can be expressed as the claim that $p(A\&B) > p(B)$. This is a clear violation of the *conjunction rule* (cf. Tversky & Kahneman, 1983) in probability theory, which states that $p(A\&B) \leq p(B)$. In plain English, the conjunction rule says that the probability of A and B occurring together must be smaller or equal to the probability of B occurring, regardless of whether A occurs.

As a trivial example of a conjunction, imagine a standard deck of cards. Denoting queens as Q , and red cards (i.e., hearts or diamonds) as R , we write the probability of drawing a queen as $p(Q)$, and the probability of drawing a red card as $p(R)$. Consequently, the probability of drawing a queen of a red suit is written as $p(Q\&R)$. In this expression, Q and R are the two constituents, while $Q\&R$ is a conjunction of the two constituents. Since all instances of $Q\&R$ are also instances of Q and R separately, it is impossible for the probability of the conjunction, $p(Q\&R)$, to be greater than that of either of its constituents, $p(Q)$ and $p(R)$, by themselves.

The most famous example of a task where the conjunction fallacy is often committed is the Linda problem, first described by Kahneman and Tversky (1982) and Tversky and Kahneman (1983). In this task, a participant is given a description of a person, Linda, and is asked to rank order statements concerning Linda according to how probable they are. The full description of Linda given to participants by Tversky and Kahneman, with alternatives, is given below:

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.

Linda is a teacher in elementary school.

Linda works in a bookstore and takes Yoga classes.

Linda is active in the feminist movement. (F)

Linda is a psychiatric social worker.

Linda is a member of the League of Women Voters.

Linda is a bank teller. (T)

Linda is an insurance salesperson.

Linda is a bank teller and is active in the feminist movement.

(T&F). (1983)

Naturally, the version shown to the participants did not include the parenthesized *T*, *F*, and *T&F*. These were included in the article by Tversky and Kahneman (1983) to help the reader see the relevant alternatives and understand the set-theoretical circumstances, where *T&F* marks the conjunction of the two constituents *F* and *T*. In the study by Tversky and Kahneman, 85% of the 88 participants ranked *bank teller and feminist* as more probable than *bank teller*, thereby committing the conjunction fallacy. In a similar task concerning Bill, who's description was representative of a stereotypical jazz musician, 87% of the participants ranked *accountant and jazz musician* as more probable than the constituent *accountant*.

The explanation put forward by Tversky and Kahneman (1983) as to why participants commit the fallacy is that they, presumably, make use of the representativeness heuristic, rather than a strict mathematical notion of probability, when ranking the alternatives. In the terms of Sloman (1996), participants are using their associative system, when they should have been using their rule-based system. The conjunction fallacy illustrates a common theme for the representativeness heuristic, namely that people overestimate the probability of specific, but more representative, events in comparison to more general, but less representative, events (Kahneman & Tversky, 1982). Kahneman and Tversky discussed the potential real-world implications of the conjunction effect/fallacy on the judgments of professionals such as judges, politicians or physicians, where the probability of highly unlikely, but representative, scenarios might be overestimated. More recent empirical findings have shown that a tendency to commit the conjunction fallacy when judging possible outcomes of short vignettes is positively linked to belief in paranormal phenomena (Rogers, Davis, & Fisk, 2009; Rogers, Fisk,

& Wiltshire, 2011) and conspiracy theories (Brotherton & French, 2014). Given the real-world effects of these beliefs, and other effects of the conjunction fallacy, there is a need to understand when, how, and why human judgment overestimates the probability of conjoined events. Such a focus on boundary conditions of the conjunction fallacy was taken by for example Wedell and Moro (2008).

While the conjunction fallacy is one of the more robust findings in decision-making research, it has long been known that the way in which the conjunction tasks are constructed has an effect on the responses of the participants (Fiedler, 1988; Tversky & Kahneman, 1983; Wedell & Moro, 2008). In standard conjunction tasks, such as the formulation of the Linda problem above, most, but not all, participants commit the fallacy. A common critique to the standard formulation of the conjunction fallacy is that fewer fallacious answers are given if the problem is stated in frequencies, which is claimed to be a more natural setting (Fiedler, 1988; Hertwig & Gigerenzer, 1999). However, Sloman, Over, Slovak, and Stibel (2003) showed that a majority of participants can avoid the fallacy even without a frequency formulation, as long as they are aware of the *nested-sets* structure of the situation. The notion of nested sets can be illustrated with the card deck described above. All elements in the set *cards that are queens of a red suit* are also elements in the set *cards that are of a red suit*.

Stergiadis (2015) showed that participants are more likely to avoid the conjunction fallacy, even for standard Linda-like conjunction tasks, if they are presented with a nested-sets hint prior to the tasks. The hint used by Stergiadis described that even with a candy bag mostly containing black candies, it is more probable to randomly pick *a candy*, than it is to pick *a black candy*. It further provided two corollary examples, stating that, for the same reasons, it is more likely that a person is *tall*, compared to *tall and thin*, as well as *blond*, compared to *blond and has blue eyes*. While not explicitly stating the conjunction rule, the hint reminds the participants of the nature of nested-sets and constraints on the probabilities of subsets (e.g., black candies) in relation to superordinate sets (e.g., candies). Kidane and Saghai (2017) used a shortened version of the same nested-sets hint (in the present thesis termed the *black candy hint*, see Appendix 1), excluding the corollary examples, and found that responses to conjunction tasks in this condition were almost evenly split between avoiding and committing the conjunction fallacy. Both above cited studies used 30 Linda-like tasks to assess the participants' tendency to commit or avoid the conjunction fallacy. Kidane and Saghai noted that participants were relatively consistent in their responses, some choosing mostly conjunction alternatives (*A&B*) and some choosing mostly constituent alternatives (*A*).

As described above, even participants given the same tasks or experimental manipulations may differ in their responses. Research focusing not only on the most common response to psychological tasks, but on the *individual differences* between those who give different responses, is useful in trying to fill the knowledge gap on the causes and conditions for a psychological effect (Stanovich & West, 2000). In the case of the conjunction fallacy, the question therefore becomes what the psychological differences are between participants who avoid or commit the fallacy

under a certain set of circumstances. The present study builds upon these findings and ideas by investigating the individual differences between those who commit versus avoid the conjunction fallacy after having read a nested-sets hint. In other words, who benefits from receiving a black-candy hint? The present study used the one-example black-candy hint also used by Kidane and Saghai (2017).

An important question when studying individual differences is what factors to investigate, and how to measure them. Several studies have been performed measuring different psychological factors in relation to responses to conjunction tasks (e.g., Liberali, Reyna, Furlan, Stein, & Pardo, 2012; Oechssler, Roeder, & Schmitz, 2009; Wedell, 2011; Winman, Juslin, Lindskog, Nilsson, & Kerimi, 2014). Through a review of the literature, three psychological factors could be identified as most relevant to responses to conjunction tasks. They were *numeracy* (Paulos, 1988), *cognitive reflection* (Frederick, 2005), and *need-for-cognition* (NFC, Cacioppo & Petty, 1982). These factors, and their relations to conjunction tasks, will now be presented.

The construct of numeracy was popularized by Paulos (1988), and has later been defined as “the ability to understand and use numbers” (Reyna, Nelson, Han, & Dieckmann, 2009, p. 945). Paulos argued that poor numeracy is a major reason for faulty decision-making, both on an institutional and personal level. Much research on the effects of poor numeracy has focused on risk perception and medical decision making, albeit with some theoretical problems within the field (see Reyna et al., 2009, for a review). Peters et al. (2006) studied correlations between numeracy and several decision-making tasks and found support for the hypothesized correlations, such that individuals lower in numeracy were more likely to be affected by irrelevant information and consequently make non-normative decisions.

Theoretically, it can be argued that having a high level of numeracy should correlate with avoidance of the conjunction fallacy, based on the relative ease with which people with high levels of numeracy should be able to use mathematical and probabilistic concepts in their reasoning. For conjunction tasks, this would make it easier for participants to see the nested-sets structure of the situation (Slooman et al., 2003).

A commonly used scale for measuring numeracy is the 11-item *numeracy scale* (*NS*) by Lipkus, Samsa, and Rimer (2001), which is an expansion of a 3-item scale by Schwartz, Woloshin, Black, and Welch (1997). Significant positive correlations have indeed been found between scores on the *NS* and avoiding the conjunction fallacy by for example Liberali et al. (2012) and Winman et al. (2014). However, Wedell (2011) did not find a significant correlation between numeracy and responses to standard conjunction tasks. In the factor analysis by Liberali et al. (2012) for scores on the *NS* and the *Cognitive Reflection Test* (*CRT*, Frederick, 2005, see below), three out of the four factors correlated positively with avoiding the conjunction fallacy on a non-standard conjunction task, indicating that multiple parts of the numeracy concept are involved in increasing avoidance of the conjunction fallacy.

Even though the *NS* by Lipkus et al. (2001) is widely used, it has been criticized for exhibiting ceiling effects when used on highly educated participants, due to most of the questions being relatively easy (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012). Cokely et al. introduced the 4-item *Berlin Numeracy Test (BNT)*, *Cronbach's* $\alpha = .59$) as a test for *statistical numeracy* and *risk literacy* with better psychometrical properties, such as better discriminant validity, in highly-educated samples. Statistical numeracy was described by Cokely et al. as "(...) an understanding of the operations of probabilistic and statistical computation (...)" (2012). For risk literacy, Cokely et al. (2012) used "risk" as it is commonly used in economics and psychology, to refer to situations which involve "known chances" (Rakow, 2010). Since the sample in the present study was expected to be largely composed of university students, the *BNT* was chosen as the measure of numeracy to avoid ceiling effects.

Cognitive reflection (Frederick, 2005; Thomson & Oppenheimer, 2016; Toplak, West, & Stanovich, 2011) was also identified as a factor of interest for the present study. It has been defined as "(...) the ability or disposition to resist reporting the response that first comes to mind." (Frederick, 2005, p. 35), and was developed within a dual-system theoretical setting. Relevant reviews of dual-system, or dual-process, theories can be found in, for example, Evans and Frankish (2009), Evans and Stanovich (2013), and Sloman (1996). As the definition would suggest, tests of cognitive reflection consist of questions to which most participants will think of an intuitive, but incorrect, answer that they must suppress to reach the correct answer. For example, one of the questions in the original 3-item *CRT* (Frederick, 2005, p. 26) is "A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? ____ cents.", a question originally reported by Kahneman and Frederick (2002). Most participants respond "10 cents", seemingly having simply subtracted \$1.00 from \$1.10, when the correct answer is "5 cents", implying the cost of the bat to be \$1.05 and the sum of the costs being the wanted \$1.10.

Since cognitive reflection represents the tendency to resist intuitive answers, it can be seen as an indicator of how often and well the rational/analytical system/processes override the automatic responses given by the more automatic, intuitive and associative system/processes. A related theoretical concept is *default-interventionism* (Evans, 2007; Evans & Stanovich, 2013), which describes the intuitive system/processes as always active and responsible for default responses, while the rational system/processes can intervene and override the default responses. In a conjunction task, cognitive reflection could therefore manifest itself through the realization that probability does not only rely on similarity. A higher level of cognitive reflection could also lead to remembering, and applying, the conjunction rule if one has learned it previously. The theoretical role of cognitive reflection as the inhibition of intuitive responses and promotion of rational cognition, as well as the empirical findings of correlation to normative responses on conjunction tasks, made it a variable of interest in the present study.

Results on the *CRT* have been shown to correlate with normative responses for rational thinking tasks in general (Toplak et al., 2011), as well as for conjunction

tasks specifically (Liberali et al., 2012; Oechssler et al., 2009). While the standard, 3-item *CRT* (Frederick, 2005) has been widely used, critical questions have been raised concerning its validity for use in present-day research (Baron, Scott, Fincher, & Metz, 2015; Thomson & Oppenheimer, 2016; Toplak, West, & Stanovich, 2014). One such concern is that the questions in the *CRT*, especially the bat-and-ball problem cited above, have been spread in popular science literature, most famously Kahneman (2011), and taught in for example psychology and cognitive science classes at universities. This leads to prior exposure to the questions and misleading scores on the test (Chandler, Mueller, & Paolacci, 2014; Haigh, 2016; Thomson & Oppenheimer, 2016). The *CRT* has also shown floor effects even in some university student populations (Frederick, 2005), indicating that the questions may be too hard. Furthermore, several studies have shown significant correlations between scores on numeracy scales and the *CRT* (Liberali et al., 2012; Thomson & Oppenheimer, 2016). Since the *CRT* questions all involve some form of arithmetic operations, this is to be expected. However, it provides an obstacle in interpretation of the results from studies using the *CRT*, due to collinearity. Finally, a gender effect is often observed, such that men on average score higher than women (Alos-Ferrer, Garagnani, & Hugelschafer, 2016; Frederick, 2005; Toplak et al., 2014). This effect might be due to the overlap in abilities required for achieving high scores on the *CRT* and numeracy scales, and the presence of a gender difference in numeracy (Thomson & Oppenheimer, 2016).

Due to the above issues with the standard *CRT*, the present study used the 4-item *CRT-2* presented by Thomson and Oppenheimer (2016), which was developed to have a smaller overlap with numeracy and low prior exposure to the questions in common study populations. For example, one question of the *CRT-2* is “If you are running a race and you pass the person in second place, what place are you in?” (Thomson & Oppenheimer, 2016, p. 101), where the intuitive answer is “first place” but the correct answer is “second place”. While the answer is numerical, there is no need for arithmetic or probabilistic reasoning involved in finding the correct answer. Thomson and Oppenheimer also found that the gender difference normally observed when using *CRT* was not present in their study using *CRT-2*.

The third and final factor in the study, *NFC*, was first defined by Cohen, Stotland, and Wolfe (1955) as “(...) a need to structure relevant situations in meaningful, integrated ways. It is a need to understand and make reasonable the experiential world” (p. 291). However, few studies used the construct prior to the construction of an easy-to-administer *NFC* scale by Cacioppo and Petty (1982). The construct of *NFC* was further incorporated into *Cognitive self-experiential theory* (Epstein, 1994; Epstein, Pacini, Denes-Raj, & Heier, 1996; Pacini & Epstein, 1999). Cognitive self-experiential theory is one of many dual-systems theories to stipulate the existence of two independent, but interacting, systems, in this case relying on rationality and experientiality (i.e., intuition), respectively. Within cognitive self-experiential theory, the *NFC* scale is seen as measuring “engagement in and enjoyment of cognitive activities” (Pacini & Epstein, 1999, p. 973).

Previous studies investigating the relation between NFC and responses to conjunction tasks have provided mixed results. Wedell (2011) found a positive correlation between NFC and avoiding the conjunction fallacy under standard formulations of the task. Path analyses performed by Wedell indicated a direct, albeit weak, correlation between NFC and conjunction scores. Wedell explained this effect with the possibility that individuals with higher levels of NFC would be more likely to analyze the structure of the problem and therefore make less errors. On the other hand, West, Toplak, and Stanovich (2008) used a composite thinking dispositions score which combined an 18-item *NFC* scale with the 41-item *Actively Open-minded Thinking scale* (Stanovich & West, 2007) and found no significant correlation between a rational thinking disposition and responses to a conjunction task.

Findings by (Simon, Fagley, & Halleran, 2004), who studied framing effects (i.e., when decisions differ based on normatively irrelevant features of how the problem is presented), suggest an important role for NFC in debiasing interventions, such as the hint used in the present study. Simon et al. used a level-of-processing manipulation where some participants were asked to motivate their decisions, with the hypothesis that this would mitigate the framing effect. They identified an interaction between NFC and the level-of-processing manipulation, such that the manipulation was more effective in participants higher in NFC.

Theoretically, the three variables described above, that have been seen to correlate with (i.e., predict) normative responses to conjunction tasks, can be described as having distinct and important roles in avoiding the conjunction fallacy. Statistical numeracy and risk literacy represents the ease of understanding and using probabilistic concepts (e.g., nested sets), cognitive reflection the ability to resist settling for intuitive, but wrong answers, and NFC the will and ability to engage in cognitive activities. Applied to the conjunction task, numeracy allows the participant to see the nested-sets structure of the problem and act according to the conjunction rule, regardless of whether the participant has explicit knowledge of the rule, cognitive reflection makes the participant doubt the intuitive answer, and NFC makes the participant try to understand the situation in a more analytical manner.

Tying together the advances, methods, and psychological constructs described above, the present study applied the individual-differences perspective to a subtle hint designed to help participants avoid the conjunction fallacy. By doing so, the present study aims to gain a deeper understanding of why the hint works for some, but not all, participants.

In conclusion, the purpose of the study was to investigate the individual differences between those who commit versus avoid the conjunction fallacy after having read a hint on the structure of nested-sets. Despite the fact that numeracy, cognitive reflection, and NFC have all been shown to correlate with avoiding the conjunction fallacy, the present study is, to the knowledge of the present author, the first to analyze all three correlations in the same sample. The present study also used newer and, in some aspects, better scales of measurement for numeracy and cognitive

reflection than previous studies. Knowledge concerning the psychological characteristics of participants who commit the fallacy despite having read the hint can be used to produce other types of hints that may be more effective in participants that do not respond to the present hint. The three hypotheses of the study were that numeracy, cognitive reflection, and NFC would be positively correlated with the tendency to avoid the conjunction fallacy after having been given a nested-sets hint.

Methods

Participants

A convenience sample of 52 participants (31 female, 21 male, $M_{Age} = 24.96$, $SD = 4.14$) was recruited through advertisements on Umeå University campus and by way of personal contact. To avoid difficulties in the interpretation and answering of questions used in the study, as well as effects of cognitive aging, participants were required to be native speakers of Swedish and be between 18 to 40 years of age. The actual age span in the sample was from 18 to 38 years of age. Two participants (both female) were excluded from the sample after completion of the study due to prior knowledge of the conjunction fallacy. The sample used for data analysis therefore totaled 50 participants, consisting of 29 females ($M_{Age} = 25.00$, $SD = 4.51$) and 21 males ($M_{Age} = 24.90$, $SD = 3.67$). Participants read and signed an informed consent form prior to participation, and received 150 SEK (approximately \$17) each as a compensation for their time and effort.

Data collection

This section will describe the materials, study design, and procedure used in the present study.

Materials and design

Data collection was performed digitally, using two different platforms. The first block of the study was implemented in E-Prime® 2 and consisted of the black candy hint (see *Appendix 1*), a test trial (for familiarization with the answering procedure), and the same 30 Linda-like conjunction tasks used by for example Carlberg (2017), Kidane and Saghai (2017) and Stergiadis (2015). The conjunction tasks used consisted of one-sentence descriptions of fictive persons, followed by the question “Which of these alternatives is the most probable?” and four alternatives (two targets and two distractors). For example, one of the tasks given to the participants was:

Fabian, 33, has studied musicology and likes jazz.

Is it more probable that the person described is a:

1. Taxi driver and record collector (*A&B*)
2. Ostrich farmer (*D*)
3. Taxi driver (*A*)
4. Taxi driver and orienteers (*A&C*)

The target alternatives were in principle similar to “bank teller” (*A*) and “bank teller and active in the feminist movement” (*A&B*) in the Linda problem (Tversky & Kahneman, 1983), in that the description was representative of the hobby *B*, while *A* was an occupation of which the description was not representative. Hence, choosing *A&B* is an instance of the conjunction fallacy, while choosing *A* is an instance of avoiding the fallacy. For each task, the two distractor alternatives included a rare occupation (*D*) and a conjunction of *A* and a non-representative hobby (*A&C*).

The second block of the study was implemented in Websurvey by Textalk (<http://www.textalk.se/websurvey/>). In this block, participants answered five blocks of questions in the following order: (1) questions concerning their answering method in the first block and demographic questions (age, gender, and field of study), (2) *CRT-2*, (3) *BNT*, (4), *NFC*, and (5) control questions concerning previous knowledge of the conjunction fallacy and the Linda-problem. Participants were asked to, in their own words, describe what method they used when answering the questions in part 1 of the study. Questions concerning answering method and demographic questions were the same as in Carlberg (2017). Translation of *CRT-2* (Thomson & Oppenheimer, 2016) was done by the present author and approved by the thesis supervisor. Previously validated Swedish translations of the *BNT* (Lindskog, Kerimi, Winman, & Juslin, 2015) and *NFC* (Bjorklund & Backstrom, 2008) were used. The final control questions included those used by Carlberg (2017) as well as more detailed questions constructed by the present author.

Based on Kidane and Saghai (2017), it was assumed that there would be an approximate 50/50 split between the number of participants who avoided the conjunction fallacy and those who committed it. Therefore, the participants were all given the same treatment, but with the expectation that they would have differing outcomes. The aim for an even split in outcomes was done to maximize the power of the study, since it allowed for a larger number of participants in the smallest outcome group than if an uneven split had been achieved.

Procedure

Data collection sessions were performed in computer rooms at Umeå University with one to four participants per session. At the beginning of a session, participants were given instructions on the session and informed of their rights as participants, after which they signed an informed consent form. Subsequently, they began the conjunction task block of the session. Upon completion of the first block, participants contacted the research leader for initiation of the second block. The order of presentation of the scales and items within the second block was kept constant across participants. After the session, participants were given a debriefing, wherein the conjunction fallacy and the purpose of the study was explained.

Data analysis

Data analysis was performed using IBM SPSS Statistics 24. For the conjunction task, the proportion of correct responses was calculated for each individual participant. Since it was expected that most participants would have a clear tendency to mostly commit or avoid the fallacy, the 30 trials for each participant were not independent

measurements. Consequently, it was expected that the distribution of avoidance rates would be bimodal with many participants nearly always committing (i.e., avoidance rate near 0) or avoiding (avoidance rate near 1), as in Kidane and Saghai (2017). This was confirmed through both visual inspection of the data and statistical tests for non-normality. Since the distribution of avoidance rates did not allow for linear regression, some form of transformation was necessary. Based on the expected and observed distribution, avoidance rates for each participant were dichotomized as “0” or “1”, depending on whether the proportion of correct answers was smaller than 0.5 or not. Dichotomizing the avoidance rates made it possible to perform a logistic regression with the outcome variable *performance*, where a “1” corresponded to “avoiding the conjunction fallacy at least 50% of the time”.

Answers to psychometric scales were scored according to the sources of the respective scales. For the *BNT*, the 4-item “paper-and-pencil” version was used, albeit in a digital distribution format. Since the *BNT* and *CRT-2*, scores are calculated as the number of correct answers (Cokely et al., 2012; Thomson & Oppenheimer, 2016), and both scales have 4 questions each, the maximum score on each of the scales is four. However, items in the 20-item *NFC* scale are answered on a 5-point Likert scale (Pacini & Epstein, 1999). This gives a range of possible scores from 20 to 100, with a higher score indicating a higher level of *NFC*. Since the range of possible scores differed markedly between the scales, the scores were standardized to z-scores prior to analysis, to enhance the interpretability of the coefficients in the logistic regression models.

After the data collection sessions, some participants reported that they had changed their response pattern during the session. One reason given was that they had realized that it was wrong to choose the conjunction alternative (i.e., *A&B*), sometimes explicitly stating that they after a while realized the meaning of the hint given before the tasks. A *post hoc* comparison of performance during the first and second half of the sessions showed that participants avoided the fallacy significantly more during the latter half. Performance from the first 15 (*early trials*) and last 15 (*late trials*) trials was therefore analyzed with separate logistic regression models in a secondary analysis.

Results

The data analysis served to test the three hypotheses that scores on the *BNT*, *CRT-2* and *NFC* scales would correlate positively with avoiding the conjunction fallacy after the participants had received a subtle hint. This section will therefore describe the responses of the participants, descriptive statistics for, and correlations between, the variables in the study, and the logistic regression models used to test the three hypotheses.

Responses and performance

The mean number of correct responses per participant (i.e., avoidance rate) over 30 trials was 11.42, meaning that 38% of all responses in the study avoided the conjunction fallacy. The observed ratio of avoidance of the fallacy is similar to the

13.50 (45%) rate of avoidance responses found by Kidane and Saghai (2017) in their “one example” setting.

Descriptive statistics

Descriptive statistics for the raw, undichotomized response data and the independent variables used in the analysis are given in Table 1. The table illustrates several important features of the data. For example, the mean and standard deviations of scores on the *NFC* differ markedly in size from those of the *BNT* and *CRT-2*. To enhance the interpretability of the logistic regression analyses, the scores for these scales were standardized prior to inclusion in the models. Furthermore, the difference in responses between the early and late trials is apparent. A paired-samples *t*-test revealed a significant difference between the means for early ($M = .32$) and late ($M = .43$) trials, $t(49) = 3.35, p = .002$.

Table 1: *Descriptive statistics for the raw, undichotomized avoidance rates and independent variables used in the study. Minimum and maximum values indicate the observed extreme values.*

	Min	Max	Mean	Std. Dev	Skewness	Std. Error	Kurtosis	Std. Error
Avoidance rate (All trials)	0	1	.38	.39	.53	.34	-1.39	.66
Avoidance rate (Early trials)	0	1	.33	.39	.84	.34	-.94	.66
Avoidance rate (Late trials)	0	1	.43	.43	.30	.34	-1.72	.66
BNT	0	4	1.94	1.10	.12	.34	-.70	.66
CRT-2	0	4	2.38	1.12	-.27	.34	-.84	.66
NFC	42	90	71.66	11.28	-.45	.34	-.25	.66

Note a: Avoidance rate: Proportion of instances in which participants avoided the conjunction fallacy (range: 0 – 1); BNT: Berlin numeracy test (range: 0 – 4); CRT-2: Cognitive reflection test-2 (range: 0 – 4); NFC: Need-for-cognition scale (range: 20 – 100).

The reliability of the three psychometric scales was assessed through Cronbach’s α . For the two scales with four items each and binary response coding relatively modest reliability statistics were observed; *BNT* ($\alpha = .44$), *CRT-2* ($\alpha = .46$). For the 20-item, 5-point Likert scale, *NFC*, a higher level of reliability was estimated ($\alpha = .88$).

The distribution of avoidance rates for all 30 conjunction trials is shown in Figure 1. It is apparent from the figure that the distribution of responses is not normal, which was also confirmed through Kolmogorov-Smirnov ($p < .001$) and Shapiro-Wilk ($p < .001$) tests for normality (Ghasemi & Zahediasl, 2012). Furthermore, most participants had avoidance rates near the extreme points, indicating always

committing or avoiding the fallacy. These results demonstrate the need for, and viability of, dichotomization of the avoidance rates for each individual participant, as described in *Methods*.

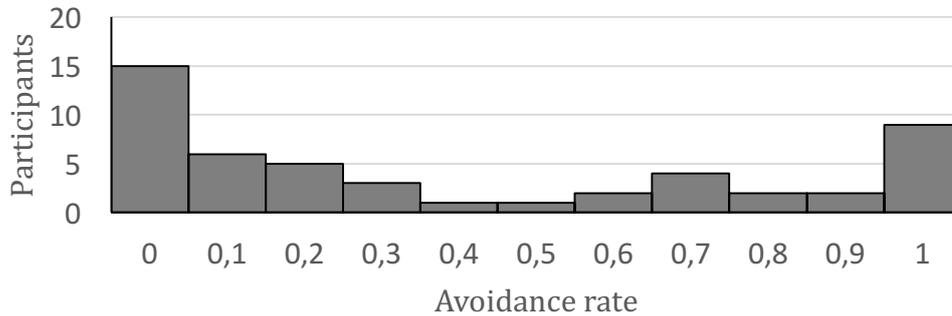


Figure 1: Distribution of avoidance rates per participant over 30 trials. Avoidance rates range from 0 (never avoiding the fallacy), to 1 (always avoiding the fallacy). Labels on the x-axis indicate the upper bounds of the respective bins.

Correlations

To further understand the characteristics of the independent variables, a Pearson correlation analysis was performed. The results can be seen in Table 2. A significant correlation was found between *BNT* and *CRT-2*. This indicates that some of the logistic regression models reported below should be taken with some caution, since multicollinearity is present in the models.

Table 2: Correlations table for the independent variables.

		<i>BNT</i>	<i>CRT-2</i>	<i>NFC</i>
<i>BNT</i>	Pearson	1	.33*	.25
	<i>p</i> (2-tailed)		.02	.08
<i>CRT-2</i>	Pearson	.33*	1	.12
	<i>p</i> (2-tailed)	.02		.43
<i>NFC</i>	Pearson	.25	.12	1
	<i>p</i> (2-tailed)	.08	.43	

Note b: *BNT:* Berlin numeracy test; *CRT-2:* Cognitive reflection test-2; *NFC:* Need-for-cognition scale. * = $p < .05$.

Logistic regressions

To test for correlations between the outcome variable, performance, and the three independent variables of interest, logistic regression models were constructed.

Primary analysis: 30 trials

The primary analysis was done using performance on all 30 trials in the session as the dependent variable. 30 participants had avoidance rates lower than .5, and were therefore given a performance score of 0. Consequently, 20 participants avoided the fallacy in at least 15 of the trials (i.e., an avoidance rate of .5 or more), and were

given a performance score of 1. The model correctly predicted 22 of the cases with a performance of 0 (73.3%) and eight of the cases with a performance of 1 (40%), with a total percentage of correct predictions at 60%. Table 3 shows the logistic regression model using the standardized scores for the three independent variables. *NFC* was positively correlated with performance. Neither *BNT* nor *CRT-2* were correlated with performance.

Table 3: *Logistic regression model for three variables with performance as the outcome variable.*

	<i>B</i>	<i>S.E.</i>	<i>Wald</i>	<i>Df</i>	<i>Sig.</i>	<i>OR</i>	<i>95% CI for OR</i>	
							<i>Lower</i>	<i>Upper</i>
<i>z(BNT)</i>	-.19	.36	.27	1	.61	.83	.41	1.68
<i>z(CRT-2)</i>	-.07	.33	.05	1	.83	.93	.49	1.77
<i>z(NFC)</i>	.78	.37	4.40	1	.04	2.17	1.05	4.48
<i>Constant</i>	-.44	.31	2.03	1	.16	.64		

Note c: z(BNT): Standardized score for Berlin numeracy test; z(CRT-2): Standardized score for Cognitive reflection test-2; z(NFC): Standardized score for Need-for-cognition scale.

In logistic regression, a common objective is to find the most parsimonious model (Hosmer, Lemeshow, & Sturdivant, 2013). Since *NFC* was the only significant predictor in the above model, a model using only *NFC* was made. In this model, *NFC* was still significantly correlated with performance ($B = .70$, $OR = 2.00$, 95% CI: 1.03 – 3.90, $p < .04$). The *NFC*-only model correctly predicted 24 out of the 30 (80%) performance scores of 0 and 9 out of 20 (45%) performance scores of 1, with the model thus totaling 66% correct predictions.

Secondary analysis: Early and late trials

The avoidance rates in the early and late trials were calculated for each participant. A new dichotomization into performance scores was made on these rates, again using a proportion of correct answers of .5 or more as the criterion for being coded as a “1” and any proportion smaller than that being coded as “0”. In the early trials, 15 participants had a performance of 1, and 35 a performance of 0, while the corresponding numbers for the late trials were 21 and 29, respectively.

For the early trials, *NFC* was significantly correlated with performance ($B = .86$ $OR = 2.37$, 95% CI: 1.04 – 5.39, $p < .04$), but this was once again not the case for *BNT* and *CRT-2*. The model correctly predicted 32 out of 35 (91.4%) cases where the fallacy was most often committed (i.e., a performance of 0) and 2 out of 15 (13.3%) cases where the fallacy was most often avoided (i.e., a performance of 1), giving a total accuracy of 68.0%.

In the three-variable model for the late trials, *NFC* was again significantly correlated with performance ($B = .84$, $OR = 2.32$, 95% CI: 1.11 – 4.84, $p < .03$), which was not the case for neither *BNT* nor *CRT-2*. However, the model for the late trials was better at predicting avoidance cases than the model for the early trials. This led to it being

correct for 19 of 29 (65.5%) cases with a performance of 0, and 10 out of 21 (47.6%) cases with a performance of 1, with a total percentage of 58.0%.

As with the three-variable model, the *NFC*-only model for the early trials predicted few avoidance outcomes. Of the two participants that the model predicted would avoid the fallacy, only one did. The total accuracy of the model was 70.0%, due to it being 97.1% accurate for cases with a performance of 0 and 6.7% accurate for cases with a performance of 1. Although this model only predicted one of the 15 avoidance cases, *NFC* was still significantly correlated with performance during early trials ($B = .75$, $OR = 2.11$, 95% CI: 1.02 – 4.39, $p < .05$).

For the *NFC*-only model of the late trials, *NFC* was once again significantly correlated with performance ($B = .76$, $OR = 2.13$, 95% CI: 1.09 – 4.17, $p < .03$). More predictions of avoidance cases were made than for the early trials, with a 75.9% accuracy for cases where the fallacy was committed, and 52.4% accuracy for cases where the fallacy was avoided, giving a total of 66.0% correct classifications.

Investigating the effect of collinearity

As a control for the possibility that the non-significant results for numeracy and cognitive reflection, a logistic regression for all trials using only *NFC* and *CRT-2* was performed. *CRT-2* was chosen since it had a lower degree of correlation with *NFC*, as compared to *BNT*. Despite the removal of *BNT* from the model, *CRT-2* was still not significantly correlated with performance ($B = .13$, $OR = .879$, 95% CI: .483 – 1.601, $p > .67$). *NFC* was still correlated with performance ($B = .715$, $OR = .2043$, 95% CI: 1.045 – 3.997, $p < .04$).

Discussion

The purpose of the present study was to investigate what it takes to avoid the conjunction fallacy. The use of a subtle nested-sets hint combined with an individual-differences approach showed that willingness to engage in, and enjoyment of, analytical thinking can be used to predict who will avoid the conjunction fallacy after receiving a hint. Interestingly, the correlations between normative responses and the factors of numeracy and cognitive reflection seen in previous studies not using such a hint were not seen, indicating that the hint is especially beneficial to participants with lower levels of these psychological traits. This part of the thesis will discuss possible interpretations of the present results and its relations to previous research, as well as implications for future research and limitations of the study. The discussion will end with summarizing conclusions.

Interpretation and relations to previous research

The present results corroborate the findings by Simon et al. (2004) that *NFC* plays a role in mediating so called *debiasing* interventions, meant to increase the tendencies for normative responses and behaviors. Since the present study, like Simon et al., used a manipulation to increase the number of normative responses, it is reasonable to believe that *NFC* would play an important part in whether a participant makes use of the hint.

The role of NFC in debiasing is also supported by the theoretical line of reasoning that individuals who find pleasure in analytical thinking and in learning new ways to reason would be more susceptible to information concerning how to think in a more rational manner. For example, Stanovich and West (2008) argued that thinking dispositions (e.g., NFC) are related to *override detection*, such that individuals exposed to classical heuristics-and-biases tasks will differ in their approach to a situation based on their thinking dispositions. Individuals higher in NFC will therefore be more likely to exercise caution and detect the need for overriding default (i.e., associative) processes.

The present study diverges from previous studies where avoidance of the conjunction fallacy has shown correlations to both numeracy and cognitive reflection (Liberali et al., 2012; Oechssler et al., 2009; Winman et al., 2014). However, Wedell (2011) investigated both factors and only found a significant correlation for cognitive reflection. The presence of a nested-sets hint in the present study is a plausible explanation for the divergence of the present results from previous findings. It is possible that the hint in some way eliminates, or at least decreases, the disadvantages normally present for participants with lower levels of these factors. As a metaphor, the correlations between multiplying skills and results on a math test would likely differ depending on whether the students are allowed to use a pocket calculator or not. In the same way that the calculator “levels the playing field” with regards to multiplication, the hint in the present study might aid participants low in numeracy and/or cognitive reflection in becoming aware of the relevant nested-sets structure and possibly even alter their intuitions concerning the probabilities of conjunctions. Such an effect could also be imagined as closing a *mindware gap* (Stanovich & West, 2008), a term indicating the absence of rules, strategies, or procedures necessary to reach the normative answer to a problem. In the present study, the hint could have been effective in participants higher in NFC since they would be more interested in the information. With the mindware gap on nested sets and conjunction tasks filled, these participants were better able to avoid the conjunction fallacy.

Alternative explanations to the null results for numeracy and cognitive reflection include the fact that the present study used different scales to measure the constructs of numeracy and cognitive reflection than those used in previous studies. While the scales were chosen to avoid collinearity between the variables, collinearity between numeracy and cognitive reflection was present in the analysis. The model including only *CRT-2* and *NFC* suggests that collinearity alone cannot explain the null results.

Furthermore, the *BNT* and *CRT-2* scales showed low reliability, as measured through Cronbach’s α . Since the *BNT* and the *CRT-2* were designed to show high discriminant validity, the relatively low α values observed in both the present and previous studies are not surprising. The items on these scales are scored in a binary way, making it hard to achieve both discriminant validity and reliability. While discrimination is achieved by making the items of a scale different enough so that floor and ceiling effects are avoided, reliability is a result of participants having

similar scores on all items. An alternative explanation for the low α values can be found in Cortina (1993), where it was demonstrated that low α values may be found in scales which have few items and/or measure multiple dimensions. The present scales do have few items, but they were not explicitly designed to measure multiple dimensions.

Implications for future research

The fact that NFC predicted avoidance of the fallacy in the present study raises the question of whether this is a necessary correlation for all hints, or whether the present hint was just not engaging or accessible enough for participants with lower levels of NFC. Research on boundary conditions can clarify the nature of a phenomenon, and some research of that nature has already been done on both the conjunction fallacy (Wedell & Moro, 2008) and the nested-sets hypothesis (Kidane & Saghai, 2017; Sloman et al., 2003; Stergiadis, 2015). While the nested-sets hint used in the present study was seen to benefit those lower in numeracy and cognitive reflection, it would be desirable to find hints which are accessible even to those low in NFC. For example, participants in Stergiadis (2015) who were given two corollary examples after the “black candy” text had a significantly higher rate of normative responses than the participants in Kidane and Saghai (2017). It would be interesting to know what the reason for this difference is, and if NFC is correlated with performance using Stergiadis more effective hint as well. Other possible means of designing a more accessible hint might be to use graphical material, such as Euler diagrams (Agnoli & Krantz, 1989), to more easily demonstrate the nested-sets structure of a situation. It is also possible that monetary incentives coupled to performance could increase the motivation of participants low in NFC enough to make use of the relevant and available information in a hint. A comparison of different forms of hints and their respective correlations with psychological factors is a possible way of gaining further knowledge on the workings of human probability judgments of conjunctions.

The present results could also motivate further research on the importance of NFC in the efficiency of debiasing methods in general. Experimental comparisons of the correlations between NFC and normative behavior in common heuristics-and-biases tasks after use of a debiasing intervention could be of interest. Such studies could compare the strengths of the correlations between NFC and normative behavior between an experimental group, which receives a debiasing method, and a control group.

Since the present study did not use a control group, it is hard to discern the direct effect of the hint and any standard effects of the psychological factors on performance. Future research with the possibility of a bigger sample size could therefore also provide a more direct comparison between the participants who avoid or commit the fallacy as a function of whether they were given a nested-sets hint.

The possibility of a “levelling of the playing field” effect for numeracy and cognitive reflection on standard conjunction tasks from a nested-sets hint (e.g., the black-

candy hint) invites further research on whether this effect generalizes to an individual's intuitions on conjunctive probabilities. This is especially important in relation to the findings by Rogers et al. (2009; 2011) and Brotherton and French (2014), connecting a tendency to commit the conjunction fallacy with increased belief in paranormal phenomena and conspiracy theories. Would the incidences of these anomalous beliefs decrease if people were made aware more often of nested-set relations?

Limitations of the present study

As with all scientific studies, the present study has limitations. This section will discuss matters concerning sample size, study design, and analysis methods. The point is raised that interpretation of the present study is complicated by the fact that it is difficult to know whether results compatible with a null hypothesis are due to a lack of true correlation or other factors, such as low power or collinearity between independent variables.

Sample size and event-per-variable

An important limitation of the study is the relatively small sample size ($n = 50$) in comparison to the number of analyzed variables. A slightly larger sample size would have been preferred, but data collection was ended at 50 valid participants due to time restraints of the thesis work. The small sample size led to a low *event-per-variable* (EPV) in the multivariable logistic regressions. EPV is calculated as the size of the smallest outcome group divided by the number of variables. A commonly used "rule-of-thumb" for logistic regression is to have at least 10 EPV (a.k.a. "the rule of ten"), but the results of simulation studies vary in their findings concerning the effects of low EPV (van Smeden et al., 2016). For example, Peduzzi, Concato, Kemper, Holford, and Feinstein (1996) showed a number of adverse effects for logistic regression models done with <10 EPV. On the other hand, Vittinghoff and McCulloch (2007) found that the adverse effects of a small EPV were considerably more pronounced for analyses with 2 – 4 EPV than those with 5 – 9 EPV, indicating that there may be conditions under which analyses with <10 EPV can still be considered informative. Vittinghoff and McCulloch drew the conclusion that "only a minor degree of extra caution is warranted" (p. 717) in the interpretation of significant correlations in analyses with 5 – 9 EPV, especially for *a priori* hypotheses where the correlations are plausible and highly significant.

The smallest EPV in the present study was 5, in the three-variable model for the early trials. A summary of the EPVs for all analyses is given in Table 4.

Table 4: *Event-per-variable summary for all analyses.*

<i>Trials</i>	<i>n(smallest group)</i>	<i>Variables</i>	<i>EPV</i>
<i>All</i>	20	<i>BNT, CRT-2, NFC</i>	6.67
		<i>NFC</i>	20
<i>Early</i>	15	<i>BNT, CRT-2, NFC</i>	5
		<i>NFC</i>	15
<i>Late</i>	21	<i>BNT, CRT-2, NFC</i>	7
		<i>NFC</i>	21

Note d: *BNT: Berlin numeracy test; CRT-2: Cognitive reflection test-2; NFC: Need-for-cognition scale; EPV: Event-per-variable.*

Furthermore, it is difficult to know how to interpret a result which is compatible with the null hypothesis. An advice given by Hoenig and Heisey (2001) is to look at the breadth of the confidence interval where the smaller the range of the interval, the more confidence one should have in the null result. The results of the present study indicate that any odds ratio between standardized scores on the *BNT* or *CRT-2* and performance, after controlling for the other independent variables, are likely not above 1.68 and 1.77, respectively (see Table 3). In the above analyses, there is however also a problem with collinearity between *BNT* and *CRT-2* (see Table 2), which is known to disturb the results of multiple regression models (De Veaux, Velleman, & Bock, 2012). This collinearity was surprising, given the fact that the *CRT-2* was developed as to not correlate with numeracy (Thomson & Oppenheimer, 2016).

A further problem caused by considerations of sample size and EPV was that the present study could not include more variables of interest than three without disturbing the analyses excessively. Emphasis in the literature review was put on both empirical and theoretical reasons which made the variables interesting for inclusion, but it is of course possible that there are important variables that were not included in the present study.

The above discussion indicates that some caution should be taken in the interpretation of the results of these analyses. However, the causes for caution apply mostly to the null results of numeracy (*BNT*) and cognitive reflection (*CRT-2*). *NFC* (*NFC* scale) was significantly correlated to performance in all models, regardless of whether potential effects from numeracy or cognitive reflection were controlled for. It is however worth noting that the confidence intervals for the odds ratio of *NFC* span from just over 1 to over 5, indicating that the present study cannot make precise estimations of the true effect size of *NFC*.

Lack of experimental design features

Considerations on sample size in relation to the number of variables to be analyzed also forced a compromise on experimental design features. It was deemed that it was not reasonable to include a control group that would not receive a hint, especially since the number of participants in such a group that would avoid the hint would be even lower.

Furthermore, since all participants performed the different psychological scales in the same order, with the same ordering of items, it is also impossible to rule out any order effects between the independent variables and the items within them. The order of the scales, as presented to participants, was *CRT-2*, *BNT*, and *NFC*. Other studies have chosen different scale answering orders. For example, Liberali et al. (2012) presented participants with the commonly used numeracy scale, *NS* by Lipkus et al. (2001), a subjective numeracy scale by Fagerlin et al. (2007), and finally the *CRT* (2005). Liberali et al. discussed whether the completion of an objective scale prior to a self-assessment scale could give rise to *fluency effects*, which is of course a concern for the *NFC* scale used in the present study. Wedell (2011) did not report on the order in which participants completed the numeracy and *NFC*-scales used.

The order of presentation in the present study was based on theoretical considerations of what scales should be most stable, and the probable effects of hypothetical scale orders. The line of reasoning behind the chosen order was that the *CRT-2* should be performed with as little clues as possible that the questions can be deceptive. The *BNT* was thought to be a more stable measure, which should depend more on the numerical ability of the participant than their alertness to the possibility of trick questions. Finally, *NFC* was put in last place to avoid the participants committing to a thinking style prior to answering any of the other scales, especially the *CRT-2*. This means that a fluency effect may be present, in that the participants' perception of how well they did on the *CRT-2* and *BNT* could have influenced their answers in the *NFC*. This concern is however not supported by the correlations between the independent variables, since *NFC* scores were not significantly correlated with *CRT-2* scores, and only correlated at a marginally significant level to *BNT* scores.

Dichotomization

The choice to dichotomize the avoidance rates can be considered controversial, since it lumps together participants with differing avoidance rates, for example treating a participant with 50% correct responses the same as one with 100% correct responses. Dichotomization of variables in psychology has been criticized (e.g., Babyak, 2004; MacCallum, Zhang, Preacher, & Rucker, 2002), but MacCallum et al. also note that one of the exceptions where it can be justified is when an extreme point has a large number of observations. This is similar to the situation of the present study, where multiple participants either always committed or always avoided the conjunction fallacy. Furthermore, the bimodal distribution of avoidance rates made many other transformations and analysis methods inappropriate.

It should also be noted that, due to the learning effect which was discovered during the present study, the individual avoidance rate of a participant would be likely to differ if the study would have used a different amount of trials than 30. A problem with dichotomization is that a participant who understood how to apply the hint to the conjunction tasks after 10 trials would likely get an avoidance rate of over 50% for all 30 trials, but not if only 15 trials had been used. Likewise, a participant who gained the same insight after 20 trials would have a total avoidance rate below 50%

for 30 trials, but would probably have an avoidance rate above 50% in a 60-trial session. In conclusion, while dichotomization does lead to a loss of information and is an imperfect transformation, it was deemed appropriate for the present study due to the relatively clear bimodality of responses, and problems with other forms of analysis.

Conclusions

The main conclusion of the present thesis, and the proposed answer to the title question of who can take a hint, is that willingness to, and enjoyment of, engagement in analytical thinking plays an important role in mediating the effect of such a manipulation. This conclusion is supported by the present finding that *NFC* scores predicted performance on conjunction tasks after having read the black-candy hint introduced by Stergiadis (2015). It is further supported by earlier empirical findings and theoretical arguments, while also motivating further research on the role of need-for-cognition in the efficiency of debiasing efforts. While the present study also investigated numeracy and cognitive reflection, through the *BNT* and *CRT-2* scales, these variables were not correlated with performance. The null results for these variables were difficult to interpret due to issues with a relatively small sample size and collinearity between the two variables. The results of the present study suggest that the black candy hint “levels the playing field” between participants with regards to numeracy and cognitive reflection, two factors which have repeatedly been demonstrated by previous research to correlate with avoidance of the conjunction fallacy. More research is however needed to confirm these findings.

References

- Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy [Abstract]. *Cognitive Psychology*, *21*(4), 515-550. doi:10.1016/0010-0285(89)90017-0
- Alos-Ferrer, C., Garagnani, M., & Hugelschafer, S. (2016). Cognitive Reflection, Decision Biases, and Response Times. *Frontiers in Psychology*, *7*. doi:10.3389/fpsyg.2016.01402
- Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, *66*(3), 411-421. doi:10.1097/01.psy.0000127692.23278.a9
- Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition*, *4*(3), 265-284. doi:10.1016/j.jarmac.2014.09.003
- Bjorklund, F., & Backstrom, M. (2008). Individual differences in processing styles: validity of the Rational-Experiential Inventory. *Scandinavian Journal of Psychology*, *49*(5), 439-446. doi:10.1111/j.1467-9450.2008.00652.x
- Brotherton, R., & French, C. C. (2014). Belief in Conspiracy Theories and Susceptibility to the Conjunction Fallacy. *Applied Cognitive Psychology*, *28*(2), 238-248. doi:10.1002/acp.2995
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116-131. doi:10.1037//0022-3514.42.1.116
- Carlberg, J. (2017). *Are normative probability judgments a system two-operation?* (Bachelor's thesis), Umeå University, Umeå.
- Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaivete among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112-130. doi:10.3758/s13428-013-0365-7
- Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *Journal of abnormal psychology*, *51*(2), 291.
- Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring Risk Literacy: The Berlin Numeracy Test. *Judgment and Decision Making*, *7*(1), 25-47.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98-104. doi:10.1037//0021-9010.78.1.98
- De Veaux, R. D., Velleman, P. F., & Bock, D. E. (2012). *Stats : data and models* (ed. 3. ed.). Boston: Addison-Wesley.
- Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709-724. doi:10.1037//0003-066x.49.8.709
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390-405. doi:10.1037/0022-3514.71.2.390
- Evans, J. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, *13*(4), 321-339. doi:10.1080/13546780601008825

- Evans, J. B. T., & Frankish, K. (2009). *In two minds : dual processes and beyond*. Oxford: Oxford University Press.
- Evans, J. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223-241. doi:10.1177/1745691612460685
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a Math test: Development of the subjective numeracy scale. *Medical Decision Making*, 27(5), 672-680. doi:10.1177/0272989x07304449
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research-Psychologische Forschung*, 50(2), 123-129. doi:10.1007/bf00309212
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), 25-42. doi:10.1257/089533005775196732
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*, 10(2), 486-489. doi:10.5812/ijem.3505
- Gilovich, T., & Griffin, D. (2002). Introduction - Heuristics and biases: Then and now. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of human judgment* (pp. 1-18). Cambridge, UK: Cambridge University Press.
- Haigh, M. (2016). Has the Standard Cognitive Reflection Test Become a Victim of Its Own Success? *Advances in Cognitive Psychology*, 12(3), 145-149. doi:10.5709/acp-0193-5
- Hertwig, R., & Gigerenzer, G. (1999). The 'conjunction fallacy' revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12(4), 275-305.
- Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *American Statistician*, 55(1), 19-24. doi:10.1198/000313001300339897
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed. ed.). Hoboken: Wiley.
- Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). Cambridge, UK: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84-98). Cambridge, UK: Cambridge University Press.
- Kidane, D., & Saghai, S. (2017). *Hur många exempel behöver du? : Om hur du undgår tankefel*. (Master's thesis), Umeå University, Umeå.
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, 25(4), 361-381. doi:10.1002/bdm.752
- Lindskog, M., Kerimi, N., Winman, A., & Juslin, P. (2015). A Swedish validation of the Berlin Numeracy Test. *Scandinavian Journal of Psychology*, 56(2), 132-139. doi:10.1111/sjop.12189

- Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical Decision Making*, 21(1), 37-44.
- MacCallum, R. C., Zhang, S. B., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7(1), 19-40. doi:10.1037//1082-989x.7.1.19
- Oechssler, J., Roeder, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72(1), 147-152. doi:10.1016/j.jebo.2009.04.018
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, 76(6), 972-987. doi:10.1037/0022-3514.76.6.972
- Paulos, J. A. (1988). *Innumeracy : mathematical illiteracy and its consequences*. London: Penguin, 1990.
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373-1379. doi:10.1016/s0895-4356(96)00236-3
- Peters, E., Vastfjall, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science*, 17(5), 407-413. doi:10.1111/j.1467-9280.2006.01720.x
- Rakow, T. (2010). Risk, uncertainty and prophet: The psychological insights of Frank H. Knight. *Judgment and Decision Making*, 5(6), 458-466.
- Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How Numeracy Influences Risk Comprehension and Medical Decision Making. *Psychological Bulletin*, 135(6), 943-973. doi:10.1037/a0017327
- Rogers, P., Davis, T., & Fisk, J. (2009). Paranormal Belief and Susceptibility to the Conjunction Fallacy. *Applied Cognitive Psychology*, 23(4), 524-542. doi:10.1002/acp.1472
- Rogers, P., Fisk, J. E., & Wiltshire, D. (2011). Paranormal Belief and the Conjunction Fallacy: Controlling for Temporal Relatedness and Potential Surprise Differentials in Component Events. *Applied Cognitive Psychology*, 25(5), 692-702. doi:10.1002/acp.1732
- Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine*, 127(11), 966-972.
- Simon, A. F., Fagley, N. S., & Halleran, J. G. (2004). Decision framing: Moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making*, 17(2), 77-93. doi:10.1002/bdm.463
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119(1), 3-22. doi:10.1037//0033-2909.119.1.3
- Slooman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, 91(2), 296-309. doi:10.1016/s0749-5978(03)00021-9
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(5), 645-+. doi:10.1017/s0140525x00003435

- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning*, 13(3), 225-247. doi:10.1080/13546780600780796
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, 94(4), 672-695. doi:10.1037/0022-3514.94.4.672
- Stergiadis, D. (2015). *Does providing a subtle reasoning hint remedy the conjunction fallacy?* (Bachelor's thesis), Umeå University, Umeå.
- Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11(1), 99-113.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7), 1275-1289. doi:10.3758/s13421-011-0104-1
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning*, 20(2), 147-168. doi:10.1080/13546783.2013.844729
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-&. doi:10.1037/h0031322
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty - Heuristics and biases. *Science*, 185(4157), 1124-1131. doi:10.1126/science.185.4157.1124
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293-315. doi:10.1037/0033-295x.90.4.293
- van Smeden, M., de Groot, J. A. H., Moons, K. G. M., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2016). No rationale for 1 variable per 10 events criterion for binary logistic regression analysis. *Bmc Medical Research Methodology*, 16. doi:10.1186/s12874-016-0267-3
- Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, 165(6), 710-718. doi:10.1093/aje/kwk052
- Wedell, D. H. (2011). Probabilistic Reasoning in Prediction and Diagnosis: Effects of Problem Type, Response Mode, and Individual Differences. *Journal of Behavioral Decision Making*, 24(2), 157-179. doi:10.1002/bdm.686
- Wedell, D. H., & Moro, R. (2008). Testing boundary conditions for the conjunction fallacy: Effects of response mode, conceptual focus, and problem type. *Cognition*, 107(1), 105-136. doi:10.1016/j.cognition.2007.08.003
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and Biases as Measures of Critical Thinking: Associations with Cognitive Ability and Thinking Dispositions. *Journal of Educational Psychology*, 100(4), 930-941. doi:10.1037/a0012842
- Winman, A., Juslin, P., Lindskog, M., Nilsson, H., & Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.00851

APPENDIX 1 – BLACK CANDY HINT

English version

Imagine buying a bag of wine gums with almost exclusively black candies – improbable, but not impossible. When picking a candy at random it is probable that you will get a *black candy*.

However, if someone would ask you whether it would be more probable to pick a *black candy* or a *candy*, the correct alternative would be *candy*. That alternative contains all subcategories – red candy, green candy, yellow candy and so on – but also *black candy*, which was the other alternative. As *black candy* is a subcategory to *candy*, it can never be more probable to randomly pick a *black candy* than to pick a *candy*.

(...) (*omitted in the present study to use the one example condition from Kidane & Saghai, 2017*)

In some, but not all of the tasks in this study, you will be able to utilize the information above. (Stergiadis, 2015, Appendix B)

Swedish version

Tänk dig att du köper en påse Gott & Blandat med nästan bara svarta godisar – osannolikt, men inte omöjligt. När du sticker ned handen och tar en godis på måfå är det mest sannolikt att du fiskar upp en svart godis. Om någon däremot skulle fråga dig om det är mer sannolikt att du får upp en *svart godis* eller en *godis* är det givna alternativet att välja *godis*. Det alternativet innehåller ju alla underkategorier – röd godis, grön godis, gul godis och så vidare – men även *svart godis*, som var det andra alternativet. Eftersom *svart godis* är en underkategori till *godis* kan det aldrig vara mer sannolikt att du plockar upp en *svart godis* än en *godis*.

(...) (*uteslutet i denna studie för att använda ett exempel-betingelsen från Kidane & Saghai, 2017*)

I vissa, men inte alla, av studiens uppgifter kommer du att kunna dra nytta av ovanstående bakgrundsinformation. (Kidane & Saghai, 2017, Appendix B)