Modeling the Conjunction Fallacy

Maryam Hedayati (hedayatim@carleton.edu)
CS328, Spring 2016, Carleton College

Huiji (Chelsea) Ying (yingh@carleton.edu)
CS328, Spring 2016, Carleton College

Abstract

This paper investigates the conjunction fallacy using the Linda problem, a well-known task in which people's answers tend to differ from the answer proposed by classical probability rules. We implemented the exemplar model to account for this fallacy, after comparing it to several other modeling approaches. The exemplar model is found to have results consistent with the conjunction fallacy. The robustness of the model is tested using several simulations in which assumptions that our model is based on are varied. In each of these simulations, the conjunction fallacy is found to hold.

Keywords: The Linda problem; Conjunction fallacy; Exemplar model; Representativeness heuristic; probability judgment

Introduction

The process through which humans make decisions has been of interest to cognitive psychologists for a long time. This includes the question of what role probability has in people's judgments. Until the 1980s, assumptions about decision making mostly presumed that human beings were rational, and that their judgments would be made based on rules of classical probability (Morris, 1977). However, since then, researchers have demonstrated that people's intuitions about probability are subject to various cognitive biases, and do not necessarily yield the same conclusions one would reach using mathematical rules of probability (Sieck & Yates, 2001; Kahneman & Frederick, 2002; Nilsson, Olsson, & Juslin, 2005; Costello & Watts, 2014; Lu, 2015). In particular, Tversky and Kahneman (1983) proposed the concept of the conjunction fallacy, an error in probability judgment that provides insight into how humans understand and use probability in their real-life decisions.

The conjunction fallacy is the idea that people rank specific conditions with typical statements as more probable than a single general atypical one, even when the specific typical category is a subset of the atypical general category. This probability judgment violates the extension rule of classical probability. According to the extension rule, if the extension of event A includes the extension of event B (i.e., $B \subseteq A$), then the probability of event A has to be no less than the probability of event B (i.e., $P(B) \leq P(A)$). Since the extension of the constituent includes the extension of the conjunction $A \wedge B \subseteq A$), the probability of the conjunction has to be less than or equal to the probability of any of its constituents (i.e., $P(A \wedge B) \subseteq P(A)$ and $P(A \wedge B) \subseteq P(B)$). Therefore, the misjudgment of this probability relationship by ranking the conjunctive event $A \wedge B$ (a subset of event A) as more probable than either event A or event B is a violation of the probability rule, and is thus called the conjunction fallacy.

Tversky and Kahneman (1983), in their landmark study on the conjunction fallacy, used a task referred to as "The Linda task" to test their proposed phenomenon. This task is now the primary apparatus used by researchers when studying the conjunction fallacy. In the Linda task, participants are given a personality sketch of a fictitious woman:

> 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.

Following this description, participants are given several statements, including "Linda is a bank teller" , "Linda is active in the feminist movement" , and "Linda is a bank teller and is active in the feminist movement" (Tversky & Kahneman, 1983). Participants are either asked to order these statements from most to least likely (direct test), or to rate the probability of each statement separately (indirect test). Because feminist bank tellers are a subset of bank tellers, Linda is at least as likely to be a bank teller as she is to be a bank teller who is active in the feminist movement. However, people usually estimated the probability of the latter conjunctive statement to be higher than the probability of the former constituent statement. This phenomenon was prevalent across people with differing levels of probability knowledge and statistical sophistication (Tversky & Kahneman, 1983), demonstrates that the conjunction fallacy is not based on a lack of knowledge, but rather on cognitive biases that may influence judgment.

Tversky and Kahneman (1983) proposed that human brains are designed to judge representativeness, not probability. They argued that in decision making, people often use cognitive heuristics, which are mental shortcuts that allow humans to make judgments quickly using limited cognitive resources. In particular, the representativeness heuristic, the tendency to make decisions based on how similar a sample is to a population, plays an important role in people's probability judgments (Tversky & Kahneman, 1983). In the example of the Linda problem, Linda is more representative of a feminist than a bank teller, and therefore is more similar to a feminist bank teller than a regular bank teller. Therefore, despite feminist bank tellers being a subset of all bank tellers, because Linda is more similar to the former, the statement containing feminist bank teller is ranked as more probable.

We were particularly interested in how the conjunction fallacy could be replicated by a computational model of cogni-

tion because machines and humans generally approach probability differently. In this paper, we adopt the exemplar model, a hypothesis using a rules and symbols approach to cognition, to account for the conjunction fallacy. The exemplar model suggests that people store individual exemplars of each category in their memory and make judgments based on the exemplars they are able to retrieve from each category (Medin & Schaffer, 1978; Sieck & Yates, 2001; Nilsson et al., 2005). In our investigation, we ran several simulated tests on our model to investigate the question of to what extent the exemplar model accounts for the conjunction fallacy and representativeness, and provides us with insights into human behavior.

Background

Before presenting our model and investigation, we will present some important models from previous studies on heuristics or probability judgment fallacies using three different approaches of modeling cognition: rules and symbols, networks, and probability and statistics. In addition, we will provide our rationale for choosing the exemplar model over these alternative models in our investigation.

Rules and Symbols Approach

The rules and symbols approach conceives human cognition by representing information in the world with discrete symbols, and representing people's strategies and skills as a set of rules. This approach assumes that humans apply a set of rules to the symbols stored in their memory. Besides the exemplar model we implemented in this paper, another model within the rules and symbols approach uses the prototype hypothesis. Instead of assuming that individual exemplars of each category are stored in memory, the prototype hypothesis claims that for each category, a prototype is compiled based on the features of the category's members. Only the prototype of each category is stored in human memory. Based on the prototype hypothesis, Kahneman and Frederick (2002) argued that when people judge the category of a new stimulus, the prototype of that category is retrieved. According to their REP[P] (REP = an interpretation of the representativeness heuristic; P = Prototype abstraction) model, the probability that a stimulus t belongs to category A is given by

$$p(A) = \frac{\text{similarity}(t \mid prototype_A)}{\sum_{c \in categories} \text{similarity}(t \mid prototype_c)}$$

where $prototype_A$ is the prototype of category A and $\text{similarity}(t \mid prototype_A)$ is the similarity between the stimulus t and (Kahneman & Frederick, 2002). Taking the Linda task as an example, REP[P] assumes that people have one prototype stored for each category(bank teller, feminist and feminist bank teller). When assessing the stimulus (Linda), the judgments reflect how similar people think Linda is to the single prototypical feminist bank teller, relative to the single prototypical bank teller or the single prototypical feminist in their mind.

Network Approach

The network approach represents information in the world as a set of nodes linked by edges indicating their relationships. It represents human cognition by spreading activation between nodes in this structure. Nodes are excited or inhibited as a result of the activation of the nodes they are connected to and properties of the corresponding edges in between. People learn associations between available cues and categories using error propagation (McClelland & Rumelhart, 1985). One example of a model using this approach is the Activation Strength Model (ASM) proposed by Sieck and Yates (2001). Using a two-layer neural network, the model accounts for the fallacy of overconfidence in probability judgments. The ASM assumes that people make probability judgments based on the total degree of activation in a probabilistic fashion. Features are encoded as nodes and the total activation of a category is given by the weighted sum of activations of related single nodes, denoted by $A = \sum_{i=1}^{m} w_i a_i$, where w_i indicates the strength of the associations between features and output category and a_i is the binary activation of each feature. The output activation, which generally falls within $[-1, 1]$, is then normalized into probability judgments that range from 0 to 1, given by $P_T = \frac{A+1}{2}$. To display the overconfidence effect, the model proposes that people's confidence for probability output $P_con = P_T$ only when $P_T \geq 0.5$, but $P_con = 0.5 + kP_T$ otherwise, where $0 \leq k \leq 1$ is a parameter indexing the person's willingness to give an extreme judgment. Learning then happens using the Rescoria-Wagner learning rule. Here, the new weights are given by

$$w_{(j+1)i} = w_{ji} + \beta a_i (T - A)$$

where β denotes the learning rate and T denotes the target output of total activation .

Probabilistic and Statistical Approach

The probabilistic and statistical approach of modeling cognition assumes that people are rational agents and provides language for describing constraints on human inductive inferences. Costello and Watts (2014) argue that people commit single probabilistic fallacies due to the random variation and noise in the reasoning process but their overall judgments still follow classical probability rules as deviations in probabilistic estimates generally cancel each other out. In particular, they claim that the deviation in probabilistic estimates for conjunction statements cancels out with the deviations for disjunctive statements. According to the probability addition law, they predict that human probability judgments follow the expression

$$X_E(A,B) = P_E(A) + P_E(B) - P_E(A \wedge B) - P_E(A \vee B) \approx 0$$

where E denotes estimation.

Rationale for choosing the Exemplar Model

We chose to use the exemplar model over the aforementioned models for several reasons. Firstly, we found the

rules and symbols approach to be best for the problem we were investigating. Networks lack the intuitive appeal of the rules and symbols approach. They require a lot of computational power, which does not account for the fast response time when people make particular simple judgments. Their weights are also often difficult to interpret. In addition, according to studies by Sieck and Yates (2001), the ASM model fails to predict the major quantitative phenomenon it was built to predict (probability overconfidence). The probability and statistics approach, on the other hand, fails to explain the regularity of the conjunction fallacy, the idea that it always occurs when the subset is a typical example of the stimulus, but that it does not occur otherwise.

Within the rules and symbols approach, we chose the exemplar model because the prototype model always predicts the fallacy, even when people only know one feminist bank teller, which does not match the actual human behaviors. The prototype model treats categories of various sizes as the same, since only one prototype is stored for each category. However, we thought the size of each category should play a role in the extent to which the conjunction fallacy holds. Finally, many papers comparing various models of the conjunction fallacy favored the exemplar model. Nilsson et al. (2005) demonstrated that the exemplar model is better than the prototype model for this particular problem, while Sieck and Yates (2001) found it to be superior to the ASM.

Implementation of the Exemplar Model

The exemplar model we implemented was a modified version of the Exemplar Retrieval Model (ERM) described by Sieck and Yates (2001). It elaborated Medin and Schaffer (1978) original model of category learning in a probabilistic fashion. According to the exemplar hypothesis, people mentally represent a category using all its member exemplars. Each exemplar is encoded individually in one's memory. In our model, exemplars were represented as vectors with binary values in each feature dimension. The exemplars that were already stored in memory were all labelled with their corresponding discrete categories. Our model assumed that given a new stimulus j, the probability that a previously stored exemplar k is retrieved is given by

$$p(k \mid j) = \frac{\text{similarity}(j,k)}{\sum_k \text{similarity}(j,k)}$$

The similarity function here multiplicatively accounts for the weights in each feature dimension using the expression

$$\text{similarity}(j,k) = \prod_{i=0}^{n} (1-w_i)^{d_i}$$

where w_i represents how much weight people put on each feature when judging similarity and d_i denotes the number of mismatching features in the ith dimension. Since we only had binary features for the exemplars in our model, $d_i = 0$ when the exemplars share the ith feature, and $d_i = 1$ otherwise. Based on the additive property of probability, the total probability that any previously stored exemplar from a category T is retrieved, given the new stimulus j, is observed is given by

$$p(k \in T \mid j) = \frac{\sum_{k \in T} \text{similarity}(j,k)}{\sum_{k \in T} \text{similarity}(j,k) + \sum_{k \notin T} \text{similarity}(j,k)}$$

We proposed that people make probability judgments about categorization based on the proportion of exemplars they are able to retrieve from each existing category, and thus $p(k \in T \mid new\_stimulus\_J)$ directly gave us the probability rating people would make when they were asked to categorize the new stimulus j to category T.

In the case of the Linda task, our model suggested that people judge the statement that Linda is active in feminist movement with higher probability than the statement that Linda is a bank teller because the given personality sketch produced a stimulus that shared more features with feminists and fewer features with bank tellers. In terms of the conjunctive statement, as the feminist bank teller category is the intersection of the feminist and the bank teller categories, it is very likely that this conjunctive category consists of exemplars having features both typical to bank tellers and to feminists, and thus the exemplars within would share more features with the Linda stimulus than the exemplars within the atypical bank teller category would. Since higher similarity would be assigned between the Linda stimulus and the feminist bank teller exemplars, the model would predict higher likelihood in retrieving previously stored exemplars from the conjunctive category than from the atypical category, and thus would assign a higher probability rating to the conjunctive statement that Linda is a bank teller who is active in feminist movement.

Simulation

The purpose of the simulation was to test whether our model makes the conjunction fallacy in situations where human do (i.e.: in both direct tests where people rank statements by likelihood and indirect tests where people give probability ratings to each statement separately), and to investigate when our model predicts the fallacy to occur.

Data Generation

The first step of our simulation was to generate exemplars. We started by separating the textual description of Linda into different phrases describing her characteristics. Then, these specific phrases were generalized to 8 different feature dimensions, so 31 years old, for example, became around 30 years old, and philosophy major became humanities major. To represent the exemplars in each category using binary features, we created a matrix of feature population distribution (FPD) where, for each category, a value between 0 and 1 was chosen to represent the distribution of people in that category that have that feature. The complete FPD is presented in Table 1. Here, *BT* represents the category of bank tellers; *F* represents the category of the feminists; *NBNF* represents the category of common people who are

not bank tellers and are not feminists; and FB represents the category of feminist bank tellers. For the conjunctive category of feminist bank tellers, we used the higher value from the bank teller and feminist categories, because we assumed that feminist bank tellers would have both the characteristics of bank tellers and feminists. The values in the FPD matrix were initialized based on: a) our own intuitive estimations of the population distribution, b) estimations we received by interviewing people around us, and c) strategies such as searching the internet for pictures of bank tellers, and estimating what percentage of them were female. Based on the resulting values, we confirmed that the abstracted description $x_{Linda} =< 1,1,1,1,1,1,1,1 >$ was a typical example of a feminist and an atypical example of a banker, as it this was an assumption of the original experiment.

Table 1: Feature Population Distribution (FPD).

Feature Dimension	BT	F	NBNF	FB
Around 30 years old	0.7	0.4	0.3	0.7
Single	0.4	0.8	0.5	0.8
Outspoken	0.3	0.8	0.4	0.8
Intelligent	0.5	0.7	0.5	0.7
Humanities Major	0.2	0.8	0.5	0.8
Concerned with discrimination and social justice	0.4	0.9	0.4	0.9
Participated in demonstrations	0.2	0.8	0.2	0.8
Female	0.8	0.9	0.5	0.9

Following this, exemplars were generated for each category in the form of vectors $x_i \in \mathbb{R}^8$ with a binary value on each feature dimension. The binary values were randomly chosen with 1 having the probability stated in the corresponding entry of the FPD matrix. These simulated exemplars were used as test sets for our implementation of the exemplar model, as described in the previous section.

Direct and Indirect Tests

10 exemplars from category BT were generated, along with 10 from F, 3 from FB and 30 from $NBNF$. A direct test was simulated by feeding all exemplars from the three categories BT, F, FB to our model at the same time so that our model would directly compare the probability of exemplar retrieval between the three categories, and the sum of those probabilities would add up to 1, as shown in Figure 1 below. The indirect test was simulated by feeding exemplars from 2 categories at a time to our model, B and $NBNF$, F and $NBNF$, FB and $NBNF$ respectively, to reflect the fact that human participants rate the statements independently of each other. As shown in Figure 2 below, the probabilities could add up to more than 1, in this case around 1.35 in the indirect test. We ran 100 iterations for both tests to simulate 100 different participants in human experiments. The results in the bar graphs are the average probability ratings for each category given the abstracted description of Linda over 100 iterations. The error bars show the standard deviation of the ratings. The simulation results indicate that in both tests, our model would be more likely to categorize the given stimulus, Linda, as a feminist bank teller than as a bank teller. This successfully mimicked the results of human data, and replicates the phenomenon of the conjunction fallacy.

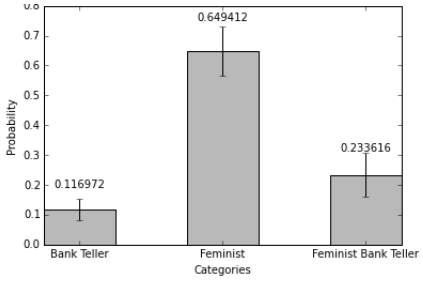

Figure 1: Average probability judgments for direct test over 100 iterations.

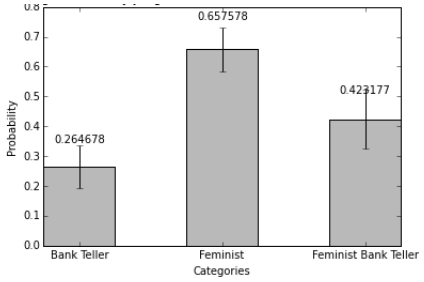

Figure 2: Average probability judgments for indirect test over 100 iterations.

Test with Atypical Feminist

In this test, we intended to establish a baseline for our model to show that it would not commit the fallacy in situations where humans do not. We had our model test a stimulus that was the exact opposite of Linda ($x_{\neg Linda} =< 0,0,0,0,0,0,0,0 >$): not around 30 years old, not single, not outspoken, not intelligent etc. We hypothesized that this stimulus would not cause the conjunction fallacy because this description is general and not necessarily typical to a bank teller, but definitely atypical to a feminist. Thus, according to human judgments, banker should be rated as most likely, followed by feminist, followed by feminist bank teller. Classical probability rules are likely to be followed in this condition. Our model's simulated results matched the expected human judgments, as shown in Figure 3 for a direct test simulation and Figure 4 for an indirect test simulation. Note that in the direct test, bank teller was rated with a very high probability (0.83), as direct comparison requires the probabilities to add up to 1. However, in the indirect test, bank teller was similarly probable to when Linda was used as the stimulus. This

result matched the fact that since people judge the probability of each statement independently in the indirect test, stimuli that are neither typical or atypical to a particular statement would not differ much in their probability judgments.

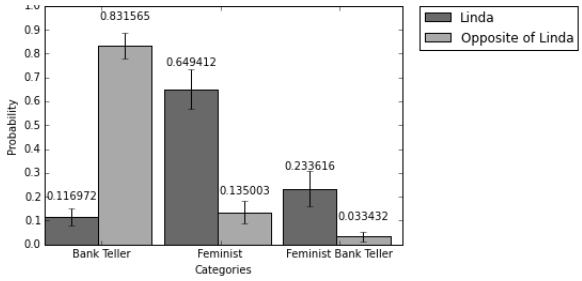

Figure 3: Average probability judgments for direct test over 100 iterations.

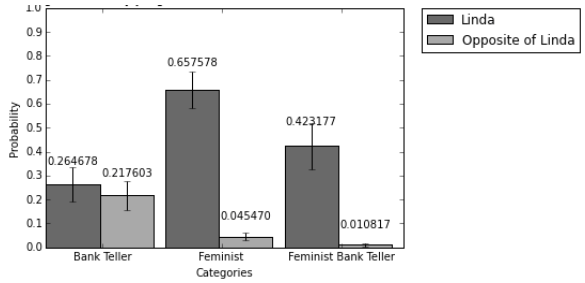

Figure 4: Average probability judgments for indirect test over 100 iterations.

Varying the Number of Exemplars Stored

The model that we built was founded on several assumptions. In order to test the robustness of our model, we challenged each of these assumptions to see if the model was still able to represent the conjunction fallacy when the assumptions were slightly tweaked. First, we investigated whether the number of exemplars stored by the model would make a difference in its performance. Instead of generating 10 bank teller exemplars, 10 feminist exemplars, 3 feminist bank teller exemplars, and 30 common people exemplars. In this test, we generated 10 times as many exemplars for each category while keeping the exemplar proportions constant, and simulated both direct and indirect tests using the bigger exemplar set over 100 iterations. In doing so, as shown in Figure 5, we found that average probability judgments were consistent relative to the exemplar set size. However, the variance was lower when more exemplars were used. This result was similar for both indirect and direct tests. The results of this test reflect human cognition in that the more previous experience or prior knowledge you have had, the more stable your judgment may be, although not necessarily more correct.

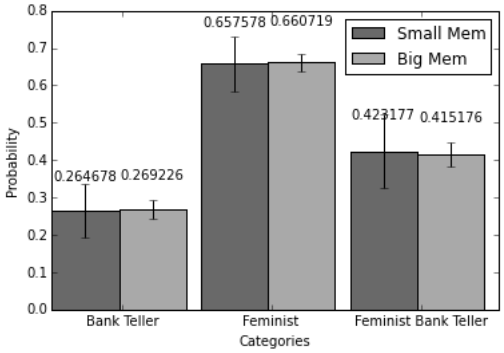

Figure 5: Average probability judgments for indirect test over 100 iterations.

Varying the Proportion of Exemplars Stored

The second assumption we tested was the proportion of exemplars in each category used by the model. Specifically, we changed the number of exemplars in one category while keeping the number of exemplars in the other categories constant. The same process was done for both direct and indirect tests, and we investigated the behavior of the model as the size of each category changed.

We started by varying the number of bank tellers we generated. In the direct test, shown in Figure 6 below, as well as the indirect test, the conjunction fallacy holds until a certain point. In the direct test, the conjunction fallacy holds until there are around 5 times as many bank tellers as feminist bank tellers. In the indirect test, it holds until around $\frac{2}{3}$ of common people are bank tellers. The results matched human intuition that when most of the people one knows are bank tellers, the conjunction fallacy will not hold. When the person is asked to make categorizations, even random people will more likely be categorized as bank tellers.

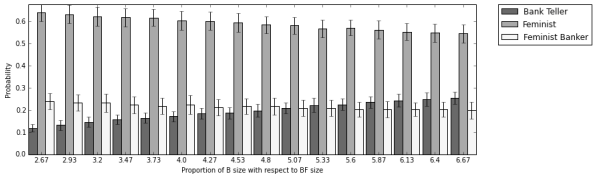

Figure 6: Varying B proportion while keeping F & FB constant (direct test over 100 iterations).

We then varied the number of feminists while keeping the other category sizes constant. We found that in both the direct and indirect test cases, the conjunction fallacy held, as long as there were more feminists than feminist bank tellers. The probability judgments for a direct test are shown in Figure 7.

We also varied the number of feminist bank tellers while keeping feminists and bank tellers at a constant equal value. As shown in Figure 8, we found that in the direct test, the conjunction fallacy held when the number of feminists was at least around 20% of the number of feminist bank tellers. In

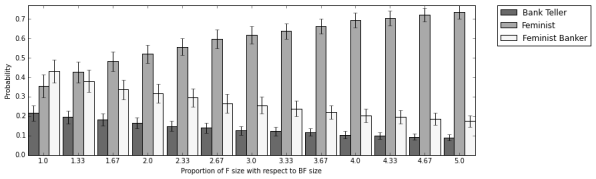

Figure 7: Varying F proportion while keeping B & FB constant (direct test over 100 iterations).

the indirect test, the conjunction fallacy held when the number of feminist bank tellers was at least 6% of the number of common people. As in the first case when the number of bank tellers was varied, this matches human intuition because knowing very few feminist bank tellers would make it unlikely for any stimulus to be categorized as a feminist bank teller.

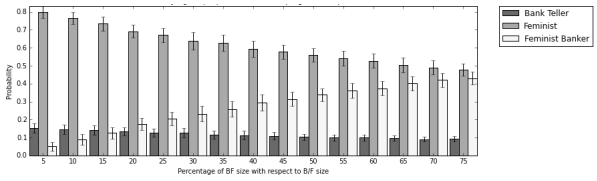

Figure 8: Varying FB proportion while keeping B & F constant (direct test over 100 iterations).

Finally, we varied the number of common people. We only did so in the indirect test, as the direct test does not use this as a parameter. As shown in Figure 9, we found that the conjunction fallacy held, regardless of the number of common people. With a higher number of common people, each probability is lower, but the categories are ranked in the order that is associated with the conjunction fallacy.

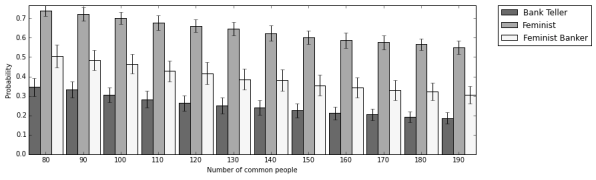

Figure 9: Varying CP proportion while keeping B, FB & F constant (direct test over 100 iterations).

Varying feature weights

Another assumption we tested was the weight of each feature. In our basic test to replicate the conjunction fallacy with our model, we set a uniform distribution for each feature weight ($w_i = 0.5$). In order to test this, we changed the weight of three different features: around 30 years old (more typical to bank tellers), participated in demonstrations (more typical to feminists), and female (similarly typical to both feminists and bank tellers). As shown in Figure 10, we found the conjunction fallacy to still hold in each of these three cases. When a feature more typical to a particular group is weighted more heavily, the probability for that group goes up slightly. Because the conjunction fallacy still held when changing these three features, we concluded that it would also hold for any other features, given that the ones with the largest difference between categories and the most neutral feature were all included here.

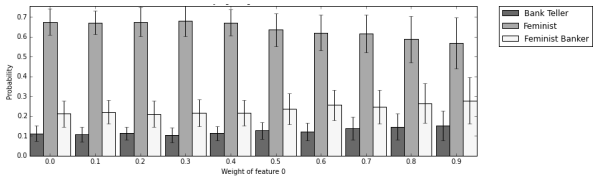

Figure 10: Varying feature 0 (Around 30 years old), a feature more typical to bank tellers than feminists.

Varying feature population distribution

We tested our two final assumptions simultaneously. The first of these was the feature population distribution (FPD). As mentioned before, these values were chosen somewhat arbitrarily, so we investigated if our model would still commit the conjunction fallacy when slightly different, but not unreasonable, values were given. Since humans can have varying understandings for the population distribution of each feature for each category, we also intended to challenge our assumption that everyone has the same mental models. We carried out a test of both assumptions by placing the distribution value of each feature for each category in the FPD over a normal distribution with a spread of 0.3, and did so for each iteration so that different values would be sampled from the normal distribution as entries for the FPD. This way, each iteration would represent a different person with a unique mental model. As shown in Figure 11, we found the conjunction fallacy to still hold when we varied the FPD values. However, there was much higher variance than when these values were kept constant, which was expected since randomly selecting values from a distribution could lead to more extreme outputs.

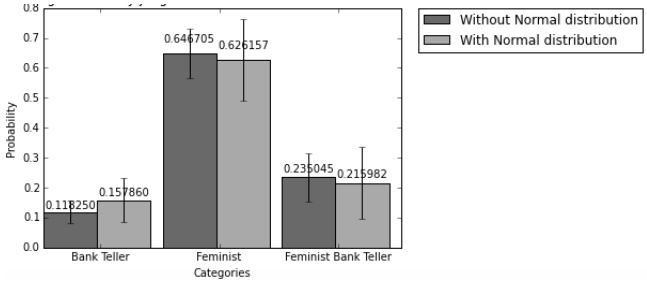

Figure 11: Average probability judgments for direct test over 100 iterations

Discussion

Overall, the exemplar model that we implemented was able to replicate the conjunction fallacy. Regardless of the assumptions that we varied, we found this phenomenon to be present. However, we do have suggestions for future research based on the weaknesses of our model that we were unable to implement given our limited time and resources.

The exemplar model itself has several limitations. One question of the model is whether it is generalizable to other decision making processes, as our current model is only used to predict answers to a very specific problem. Another limitation is that the exemplar model only allows for the categorization of new stimuli into pre-existing categories, and does not allow for the generation of new categories. The model also requires lots of parameters that we control for. Therefore we would like future research to further compare the exemplar model to models using other approaches. Although our review of the literature found a consensus that the exemplar model was superior, there is the possibility that because these other models approach the problem from a very different perspective, they could provide insight into what the conjunction fallacy demonstrates about human cognition.

Another weakness of our model is that we were unable to compare it to specific human data, as we did not have enough information from other papers testing the conjunction fallacy using the same measurements to do so. We were also worried about overfitting with the limited information we had, and did not have the resources to run a human experiment using the same measurements as our model, which would be a good measure of success. Along the same lines, many of our values were arbitrarily selected. In particular, the values in the FPD, the feature weights, and the proportions of exemplars that were chosen were selected based on our own mental models and our conversations with others. It would be interesting to survey more people in order to get values that are closer to what a wider range of people think about these categories. Ideally, a future area of research could be to run an experiment where the same people who complete the Linda problem also describe their mental models of feminists and bank tellers. This way, we could use their mental models in our implementation, and better see whether the results are similar.

Finally, the scope of our model, in terms of the feature dimensions and stimuli we use, is quite limited. In future research, we would recommend expanding the feature dimensions in the FPD matrix, as well as the number of statements ranked by the model. We chose to limit our model to the features listed in the description of Linda and the statements that were crucial to the conjunction fallacy. However, adding more feature dimensions would be useful for comparing different conjunctions of atypical, typical, and neutral irrelevant statements. The findings of our investigation would also be more robust if they were applied to more stimuli than just Linda and the opposite of Linda.

References

Costello, F., & Watts, P. (2014). Surprisingly rational: Probability theory plus noise explains biases in judgment. *Psychological Review*, *121*, 463–480.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and biases: The psychology of intuitive judgment*, 49–81.

Lu, Y. (2015). The conjunction and disjunction fallacies: Explanations of the linda problem by the equate-to-differentiate model. *Integrative Psychological and Behavioral Science*, 1–25.

McClelland, J., & Rumelhart, D. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.

Medin, D., & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.

Morris, P. (1977). Combining expert judgments: A bayesian approach. *Manag. Sci*, *23*, 679–693.

Nilsson, H., Juslin, P., & Olsson, H. (2008). Exemplars in the mist: The cognitive substrate of the representativeness heuristic. *Scandinavian Journal of Psychology*, *49*, 201–212.

Nilsson, H., Olsson, H., & Juslin, P. (2005). The cognitive substrate of subjective probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 600–620.

Rescoria, R., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. *Classical Conditioning II: Current Research and Theory*, 64–99.

Sieck, W., & Yates, J. (2001). Overconfidence effects in category learning: A comparison of connectionist and exemplar memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1003–1021.

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