The Power of the Representativeness Heuristic

Sudeep Bhatia (s.bhatia@warwick.ac.uk)
Behavioral Science Group, University of Warwick
Coventry, United Kingdom

Abstract

We present a computational model of the representativeness heuristic. This model is trained on the entire English language Wikipedia corpus, and is able to use representativeness to answer questions spanning a very large domain of knowledge. Our trained model mimics human behavior by generating the probabilistic fallacies associated with the representativeness heuristic. It also, however, achieves a high rate of accuracy on unstructured judgment problems, obtained from large quiz databases and from the popular game show Who Wants to be a Millionaire?. Our results show how highly simplistic cognitive processes, known to be responsible for some of the most robust and pervasive judgment biases, can be used to generate the type of flexible, sophisticated, high-level cognition observed in human decision makers.

Keywords: Heuristic judgment, Representativeness, Conjunction fallacy, Adaptive rationality, Latent semantic analysis

Introduction

Human judgment and decision making is guided by the use of heuristics. Heuristics are short cuts for solving problems. They specify simple strategies for accessing and manipulating information, and are often able to provide quick and effortless responses in everyday judgment tasks (Gigerenzer & Todd, 1999; Gilovich et al., 2002; Tversky & Kahneman, 1974).

Despite the long history of heuristics research in psychology and cognitive science, there are two aspects of heuristic processing that are still the topic of considerable debate. Firstly, it is not clear how some heuristics, such as the representativeness heuristic (Kahneman & Tversky, 1973), can be formally defined. Although many scholars have specified the main properties of this heuristic, others have criticized these specifications for being too imprecise, and for not being able to provide clear, quantitative predictions regarding human judgment (e.g. Gigerenzer, 1991). It is certainly the case that there are currently no formal models that are able to take in as inputs the judgment problems offered to the decision maker, and produce as outputs the predictions of the representativeness heuristic (or for that matter, other similar heuristics) for these problems (but see e.g. Jenny et al., 2014; Tenenbaum & Griffiths, 2001).

Secondly it is not clear whether the use of heuristics like the representativeness heuristic should be considered detrimental for the decision maker. Some approaches to heuristic judgment have emphasized the fact that these heuristics lead to irrational biases, such as logical and probabilistic fallacies and violations of the tenets of economic rationality (Gilovich et al., 2002; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974, 1983). Other approaches have, however, stressed the usefulness of heuristics: they are easy to apply and can generate accurate responses across a variety of settings. In other words, they are adaptively rational (Gigerenzer & Todd, 1999). This debate is partially a product of the issue discussed above. The absence of formal models for important heuristics has made it impossible to test the accuracy of these heuristics in novel decision domains.

We attempt to address the above two issues with regards to the representativeness heuristic- one of the most prominent judgment strategies in decision making research, and a cornerstone of Kahneman and Tversky's heuristics and biases framework (Gilovich et al., 2002; Tversky & Kahneman, 1974). We begin by specifying a computational model that formalizes the cognitive processes assumed to be involved in generating judgments using this heuristic. These processes operate on similarity, assessed through latent semantic analysis (Landauer & Dumais, 1997), and are nearly identical to processes used to understand similarity-based cognition in lower level domains. We then train our model on the entire English language Wikipedia dataset, in order to allow it to judge the similarity, or representativeness, of various everyday objects and their descriptions. The result is a general model of heuristic judgment that is able to use representativeness to provide responses across a wide array of decision problems.

We apply our model to choice problems used in prior experiments on the representativeness heuristic. We find that the model is able to mimic human judgments on a number of classical tasks, such as the Linda problem (Tversky & Kahneman, 1983). Specifically the model generates similarity-based conjunction fallacies, which are typically attributed to the representativeness heuristic. After verifying that the model provides a satisfactory model of the biases generated by the representativeness heuristic, we test the accuracy of the model in novel judgment tasks. Particularly, we apply the model to a series of multiple choice trivia problems. These problems are obtained from the Who Wants to be a Millionaire? game show, and from a popular online geography quiz database. Overall, we find that the model is able to achieve an accuracy rate of 40-50% for four-option multiple choice problems, which is almost twice the accuracy of a random-choice model. Although this is far from perfect, it nonetheless showcases the power of judgments from representativeness. The mechanisms that violate the fundamental laws of probability are also able to

use a rich and complex information database to solve difficult and highly unstructured decision problems about an extremely wide range of topics. This suggests that heuristics, such as representativeness, do not only lead to biases in judgment: They may also be responsible for the types of quick, accurate and flexible judgments observed in human decision makers.

A Model

The Representativeness Heuristic

In their classic 1974 paper, Tversky and Kahneman described the representativeness heuristic as a way to answer questions of the following type: What is the probability that A belongs to/originates from/generates B? According to Tversky and Kahneman, decision makers do not consider probabilistic or logical relationships between A and B when answering these types of questions. Rather they make their judgment based on whether A is representative of, that is, similar to, B. Similarity is an important feature of cognition (Medin et al., 1993), and judgments using similarity can be made with relative ease. Indeed Tversky and Kahneman found that the representativeness heuristic could predict human responses in a range of decision problems of the above type, including problems in which the heuristic generated an incorrect response (see also Kahneman & Tversky, 1973; Tversky & Kahneman, 1983).

Since Tversky and Kahneman's groundbreaking work, a number of researchers have established the ubiquity of the representativeness heuristic and the biases that it generates (Gilovich et al., 2002). Indeed, the representativeness heuristic is the best-known and most-studied heuristic to emerge from Tversky and Kahneman's heuristic and biases framework. Despite this, this heuristic has not yet been formally modeled: We do not have a computational or mathematical specification of the representativeness heuristic that can provide precise predictions for the types of questions outlined in the first paragraph of this section. This is understandable. These types of questions can span a very large domain, and specifying a model that is able to apply the representativeness heuristic almost universally seems to be a highly complex task. That said, the absence of a formal model impedes theoretical development. By not being able to specify the representativeness heuristic's predicted responses in an apriori manner, we lose the ability to apply the heuristic in new settings. These sorts of tests cannot only examine the descriptive power of the representativeness heuristic (that is, its ability to explain human behavior) but also the desirability of this heuristic as a judgment strategy.

Latent Semantic Analysis

In this paper we provide a solution to this problem. Representativeness relies fundamentally on similarity, and similarity is a topic that has received attention not only from psychologists, but also from computer scientists and related researchers. There are, by now, a number of tools that can be used to establish the semantic and conceptual relatedness of natural language descriptions. One such tool is latent semantic analysis (LSA), which judges words to be similar in meaning if they occur in similar pieces of text (Landauer & Dumais, 1997). Formally LSA involves performing a singular value decomposition on a matrix of word counts per text in the text corpus on which the LSA model is being trained. The singular value decomposition uncovers the latent dimensions that characterize the structure of word concurrence in the different texts. Two phrases, descriptions, or texts are judged to be similar by LSA if their component words are characterized by the same latent dimensions, that is, if the cosine of the angle of their vector-word count representations on these latent dimensions is small.

LSA has a very appealing cognitive representation. Particularly, an LSA model can be represented as a locally-coded three-layer neural network, with the outer layers corresponding to the texts in the corpus and the individual words contained in the corpus respectively, and the middle layer corresponding to the latent dimensions that describe the structure of the corpus. Similarity is judged based on the overlap of activation on this hidden layer. As backpropagation has been shown to asymptotically implement singular value decomposition (Saxe et al., 2013), the LSA model can be trained using standard connectionist techniques.

Formal Model

LSA has been applied across wide variety of theoretical and applied domains (Landauer et al., 2013). Here we use it to study knowledge representation and manipulation in high-level judgment tasks typically answered using the representativeness heuristic. Particularly we train our model on the entire English language Wikipedia corpus to recover 1000 latent dimensions. Each article in this corpus is considered to be a separate text, and two words are judged to be semantically or conceptually related if they co-occur in the same Wikipedia article. Thus our analysis amounts to performing a singular value decomposition of the word co-occurrence matrix across the Wikipedia corpus. Due to computational limitations we consider only 300,000 unique word stems in our analysis (stems that are present in moderate frequency on Wikipedia). Also, prior to performing the singular value decomposition we apply a tf-idf weighting scheme to the matrix of word counts. The final LSA model uses 300,000 word stems across approximately 3.2 million Wikipedia articles. Our analysis is performed with the aid of the Gensim toolbox (Řehůřek & Sojka, 2010). An outline of the model is provided in Figure 1.

The article topics in Wikipedia correspond to the objects in the world that may be the topic of a judgment, the words used in these articles correspond to the descriptions of the different objects, and the 1000 latent dimensions capture the conceptual structure of the objects described in Wikipedia articles. Due to the scope of the Wikipedia corpus, our model can be seen as encoding a low-dimensional

representation of the structure of an extremely large domain of knowledge, and using assessments of similarity on this low-dimensional representation to make judgments from representativeness. Implicit in this exercise is that the assumption that the conceptual structure of human knowledge (which guides human judgments of representativeness) resembles that of the knowledge obtained from Wikipedia.

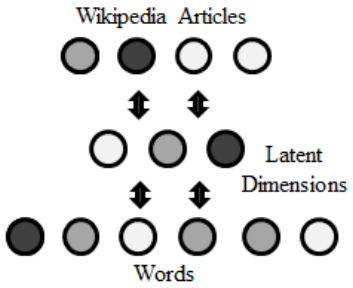

Figure 1: Outlines of LSA model trained on the Wikipedia corpus. Note that activation can flow in both directions. A representativeness score is generated based on activation overlap on the middle layer.

It is easy to see now how our model can be used to generate responses to questions of the type: what is the probability that A belongs to/originates from/generates B? A and B are either individual words (usually nouns), or extended descriptions, composed of a set of words. Using the structure of word co-occurrence Wikipedia, the model is able to quantify the conceptual similarity between A and B. This similarity is, in essence, a *representativeness score* for A and B, and can be used in the place of an actual probability judgment when answering the above question. This technique can then be applied by the model to provide responses in closed-end multiple-choice questions, where the response option with the highest similarity to the text in the question is selected as the model's answer.

The Conjunction Fallacy

The representativeness heuristic substitutes similarity for more complex probabilistic and logical relationships. This can lead to judgment fallacies in settings where response options that are highly similar to the object that is the topic of the judgment, cannot be more likely to be correct than their competitors. Consider, for example, the famous Linda problem (Tversky & Kahneman, 1983). In this problem decision makers are given the following description: "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." They are then asked whether she is more likely to be a bank teller or a feminist bank teller. Decision makers typically believe that Linda is more likely to be a feminist bank teller than a bank teller, despite the fact that the set of all feminist bank tellers is a subset of the set of bank tellers, making it impossible that Linda is a feminist bank teller but not a bank teller.

The representativeness-based model that we propose in this paper is able to make the same mistakes as decision makers, and thus is more likely to believe that Linda is a feminist bank teller and not a bank teller. Indeed, when we give our trained model the above question, it rates feminist bank teller as having a representativeness score of 0.031 to the description of Linda but bank teller as having a representativeness score of only 0.003. If the probability of selecting one response over another is given by the Luce choice rule, which applies a logistic transform to the difference between the representativeness scores of the two response options, then, like human decision makers, our model would be more likely to give the incorrect response in this question.

The Linda problem asks decision makers to judge whether a description A (Linda) is more likely to be B (bank teller) or B and C (bank teller and feminist). This problem is designed to elicit the conjunction fallacy, and is able to do so especially well when C is more similar to A then B. The conjunction fallacy weakens when both B and C are highly similar to A. Thus if asked to judge whether Linda is a social worker or a feminist social worker, decision makers are less likely to incorrectly choose feminist social worker as their response, relative to when they are given the bank teller version of the problem (though a majority of participants still make the conjunction fallacy) (Shafir et al., 1990). Our proposed model mimics this pattern, and ascribes social worker a representativeness score of 0.050 and feminist social worker a representativeness score of 0.065. This is a difference of only 0.015, less than 0.028, generated above. Subsequently our model is less likely to make a conjunction fallacy in the social worker version of the Linda problem compared to the bank teller version of the problem.

Shafir et al. (1990) do not only show how the conjunction fallacy depends on the similarity between the various components of a judgment problem. They also provide a more conclusive demonstration of the conjunction fallacy by replicating it in 28 different problems. The word stems in 22 out of the 28 problems are present in the 300,000 word stems that our model was trained on, implying that our model can be tested on these 22 problems. Overall, the model made fallacies in 67% of the problems in which fallacies were observed in decision makers, and there was a correlation of 0.29 between the conjunction fallacies generated by our model and those generated by the human participants in Shafir et al.

Testing Model Accuracy

Factual Judgments

The model is capable of answering more than just the description based probability questions outlined in the above section. It can also make general factual judgments regarding a wide array of topics and presentation formats. The model makes these judgments based on the similarity

between the text used in a question and the various response options offered to the decision maker. In essence it applies the representativeness heuristic on the mental representations of the objects and events that are the focus of the judgment.

For example, we asked the model "what is the capital of Kenya?", and offer it a choice between A: Tanzania, B: Nairobi, C: Kampala, and D: Mombasa. The model produced a representativeness score for the four response options based on their similarity with the words in the question, and chose the option with the highest score. In this case, the correct response, response B, was chosen. Since Nairobi is the capital of Kenya, the word "Nairobi" occurs very frequently with the words "Kenya" and "capital" in the Wikipedia corpus. Thus the trained model judges "Nairobi" to be most conceptually similar to the words in the question text, and assigns it a high representativeness score of 0.94. Note that Tanzania is a neighboring country of Kenya but is not a capital city, Kampala is a capital city, but not of Kenya, and Mombasa is a city in Kenya but is not its capital. Thus though these responses are considered somewhat similar to the text in the question (with scores of 0.79, 0.69 and 0.89 respectively), they are nonetheless less similar than the correct response.

Using this approach we can now test the general ability of the representativeness heuristic to provide accurate factual judgments in more general settings. Examining this is important. It can tell us whether the cognitive mechanisms responsible for the conjunction fallacy are beneficial for decision makers, that is, whether they are adaptively rational. If they are rational in this manner then the use of the representativeness heuristic can be justified, despite its tendency to systematically violate the laws of probability. If these strategies are not adaptively rational then we would be forced to ask why people continue to use this heuristic to make choices, and whether or not representativeness even plays a role in most everyday decisions.

Finding the representativeness heuristic is adaptively rational may also shed light on how sophisticated behavior can emerge from basic cognitive processes. Despite operating on an almost universal domain of knowledge, the model outlined in this paper is highly simplistic. It uses only similarity --that is, overlap in activation-- to generate responses, and can be implemented in the most basic type of neural network. Indeed it is this simplicity that makes the model computationally tractable: it would be impossible to train a more complex judgment model on such a rich data set. If the representativeness heuristic does manage to attain a high level of accuracy in general factual judgments, then it could present a part of the solution to one of the most fundamental questions in cognitive science.

Geography Quizzes

We first test the ability of the model to provide accurate responses using a set of geography quizzes obtained from the website About.com. The geography portal of this website has been posting multiple-choice quizzes since 1997, and describes itself as "the Internet's best geography quiz". As of 2014, there were over 200 geography quizzes on the website. These quizzes offer five multiple choice questions, with four responses each. Importantly for our purpose, they are in the public domain and are easy to access, and cover a diverse array of geography topics.

We used these questions to test the accuracy of the model in making factual judgments, in a manner similar to the Kenya capital question outlined in the above section. Particularly, each question was decomposed in to five pieces of text: the question text and the four response texts. The conceptual similarity between the words in the four responses and the words in the question, as assessed by the model, was then use to generate a representativeness score for each of the four responses. The response with the highest score was selected as the model's final answer.

Note that some of the quiz questions that the model was applied to involved choosing a response that did not satisfy a particular condition. For example, one of the questions in the geography quiz database asked which of a set of four countries did not border the Gulf of Aqaba. Responses to these types of questions were generated based on the lowest representativeness score. Thus response options whose text was least similar to the text in the question were chosen by the model for these questions. There were 85 questions in the geography dataset that had this property. Also note that there were some 345 questions in the geography dataset whose correct responses were composed entirely of word stems absent from our model's memory (i.e. word stems not part of the 300,000 stems that the model was trained on). As it is impossible for the model to make responses for these questions, they are excluded from subsequent analysis. This leaves a total of 836 questions for testing our model. .

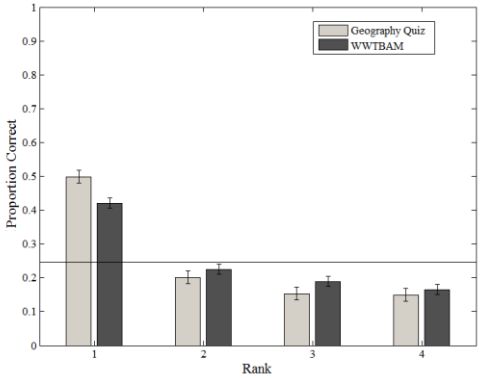

Figure 1*:* Accuracy of responses in the Geography Quiz and WWTBAM datasets, as a function of the rank ascribed to them by the model. Note that a random model would generate an accuracy of 25%. Error bars represent 95% confidence intervals.

We found that the model achieved fairly high accuracy rates. Particularly, the model was able to give the correct response 49.81% of the time, and was able to select the correct response as one of its top two choices 69.79% of the time. Both of these are statistically different from accuracy rates of 25% and 50%, which are what would be expected if

the model was choosing randomly ($p < 0.01$ using a binomial test). Figure 1 outlines the accuracy of the model responses. The bars represent the proportion of the times the model's most favored, second favored, third favored, and least favored responses were the correct responses.

We also examined the settings in which the model was most likely to give a correct response. Particularly we defined a new variable, discriminability, which was equal to the difference in the representativeness score of the most favored response relative to the average representativeness score of the remaining three responses. The discriminability of a problem captures the degree to which the HLM's favored response in the problem stands out relative to its competitors, and can be seen as a measure of the intuitive strength of the model's favored response for the problem.

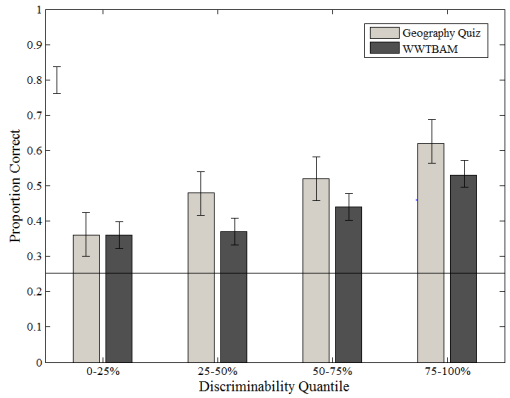

Figure 2: Accuracy of model responses in the Geography Quiz and WWTBAM questions, by discriminability quantile. Note that a random model would generate an accuracy of 25% for all quantiles. Error bars represent 95% confidence intervals.

We regressed the choice of the correct option in a problem on the discriminability of that problem to see if an increase in the intuitive strength of the most favored response led to a higher accuracy in the quiz problems. Our regression revealed a significantly positive coefficient (β = 2.29, z = 5.56, 95% CI = [1.48, 3.11], $p < 0.01$), indicating that model is more likely to be correct in problems where the intuitive strength of the favored response is higher. Figure 2 describes the average model accuracy for quiz problems in each quantile of the discriminability distribution. Thus we can see that the model achieved an accuracy of about 60% for the problems that were above the 75th percentile in terms of their discriminability, compared to the rest of the dataset.

Who Wants to be a Millionaire?

We also tested the ability of the model to provide accurate responses in a more general domain: one involving trivia questions on the popular television game show Who Wants to be a Millionaire? (WWTBAM). WWTBAM is a game show that offers contestants four-option multiple-choice questions spanning a very large range of topics, including history, current affairs, and popular culture.

The popularity of WWTBAM has spanned a number of fan-sites. One of these is wwtbam.com, where viewers post transcripts of the US game show's numerous episodes. We scraped 359 show transcripts from wwtbam.com, starting from 2007 (the earliest transcripts available on the website) and going up to 2010 (when the show's rules were changed). These transcripts generated a total of 2502 different questions that were used on the US television version of the WWTBAM game show.

As with the geography quizzes discussed above, each question was decomposed in to five pieces of text: the question text and the four response texts, with the conceptual similarity between the words in the four responses and the words in the question being used to generate a representativeness score for each of the four responses. Additionally, as above, many of the questions used on this show involved choosing a response that did not satisfy some condition. Questions of this form were answered by selecting the response with the lowest representativeness score. Finally, a total of 305 questions in the WWTBAM had correct responses that were composed entirely of word stems absent from our model's memory. These questions are not used in the subsequent analysis. This leaves a total of 2197 questions for testing our model.

Overall, we found that the model was able to provide the correct response 42.01% of the time, and was able to select the correct response as one of its top two choices 64.51% of the time. Although the accuracy of the model is a bit worse on this dataset, relative to the geography quiz dataset, it nonetheless far higher than that generated by a perfectly random model which would choose each response with a 25% chance ($p < 0.01$ using a binomial test). The proportion of the times the model's most favored, second favored, third favored, and least favored responses were the correct responses, are shown in Figure 1.

Once again we considered the discriminability of the model's favored response in each question. As above, this variable is defined as the difference in the representativeness score of the most favored response relative to the average representativeness score of the remaining three responses. We regressed the choice of the correct option on the discriminability of the problem and found a significantly positive coefficient (β = 1.90, z = 5.88, 95% CI = [1.26, 2.53], $p < 0.01$), indicating that an increase in the intuitive strength of the most favored response leads to a higher accuracy in the WWTBAM dataset. Figure 2 describes the average accuracy of the model for problems in each quantile of the discriminability distribution.

Finally, we were able to examine whether the model was more likely to make correct responses in easier questions. Each WWTBAM question in the US television version of the game show is accompanied a monetary value, and questions with a lower monetary value are typically easier. We found that the model had a roughly equal success rate for questions of all monetary values, indicating that the accuracy of the representativeness heuristic does not

typically vary with the difficulty of the problems that it is applied to.

Discussion and Conclusion

The representativeness heuristic is perhaps the best known and most studied heuristic in decision making research. The fallacies it generates are robust and systematic, and have, over the past four decades, shed light on an important limitation of human judgment. In this paper we have presented a formal model of the representativeness heuristic and have shown that it can both mimic human behavior by generating conjunction fallacies, and generate accurate responses in a wide array of factual judgment problems. The adaptive rationality of our model of representativeness explains why people are likely to use this heuristic despite the biases that it generates, and additionally, how they are able to achieve relatively high accuracy when making everyday judgments.

The model of representativeness that we have proposed relies fundamentally on stored representations in memory. Memory processes have been known to code information in a manner that reflects the structure of the environment, and one that is beneficial to the decision maker (Anderson & Schooler, 1991). These insights have led to the generation of heuristics that are able to make judgments based on assessments of recognition and familiarity (Goldstein & Gigerenzer, 2002). Our paper complements this work by studying heuristics that use similarity assessments on semantic memory. The ability to process semantic similarity is an important feature of human cognition, and similarity more generally forms one of the central theoretical constructs in cognitive psychology (Medin et al., 1993). We show that this important construct can provide accurate but flexible judgments across a very large range of topics.

The results in this paper also shed light on the settings in which similarity-based judgment processes are able to obtain the highest accuracy. Particularly we found that our model was most likely to give a correct response when only one option was strongly supported by the representativeness heuristic, that is, when the intuitive strength of the model's favored response was highest. It may be the case that decision makers use the representativeness heuristic to make judgments in these settings, but recruit higher-level deliberative processes in settings where the representativeness heuristic supports multiple options or doesn't support any option. Such a strategy would be able to achieve high accuracy rates without unnecessarily sacrificing speed and effort. Indeed the use of this strategy would be compatible with a dual-systems framework that stresses the primacy of intuitive heuristic processes over deliberate controlled processes (see e.g. Kahneman & Frederick, 2002).

Finally note that the model proposed in this paper differs greatly from previous heuristic models. It not only formalizes the mechanisms responsible for the representativeness heuristic, but trains these mechanisms on a very large knowledge database, Wikipedia. The model can thus make representativeness-based judgments for an extremely diverse range of judgment problems. This model is, in essence, a simulation of human judgments of representativeness that is able to both mimic human-like errors but also answer difficult, unstructured judgment questions with relatively high accuracy. Its ability to do this represents a heightened degree of formalism and theoretical rigor in decision modelling, and illustrates how the insights from multiple sub-fields within psychology can be combined in order to build a new class of powerful, flexible, domain-general models of everyday judgment.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6).

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, 2(1).

Gigerenzer, G., & Todd, P. M. (1999). *Simple Heuristics that Make us Smart*. Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review*, 109(1)

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4).

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2).

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013). *Handbook of Latent Semantic Analysis*. Psychology Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254.

Řehůřek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Learning hierarchical category structure in deep neural networks. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition*, 18(3).

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157)

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90(4), 293.