# Dual Processes in Relational Judgment: A Computational Framework

**Sudeep Bhatia, Russell Richie, Wenjia Joyce Zhao**
Department of Psychology, University of Pennsylvania
{bhatiasu, drrichie, zhaowenj} @sas.upenn.edu

**Keywords: Vector Semantics; Computational Model; Drift Diffusion Model; Bayesian Analogy with Relational Transformation Model; Heuristics**

People have the ability to accurately judge the truth-values of propositions involving relations, though are sometimes prone to errors which arise due to the use of heuristics and cognitive shortcuts. For example, while judging whether or not the *PartOf* relation holds between the concepts *sparrow* and *bird*, people may use structured symbolic knowledge representations that directly evaluate the truth or falsehood of the proposition *PartOf*(*sparrow*,*bird*), but may also query unstructured representations that merely consider the association between *sparrow* and *bird*. The latter are often more accessible and can be used to make quick and effortless judgments. However, they can lead to errors when the association between two concepts does not correspond to whether or not the target relationship holds between the concepts, such as in the example considered here.

This interplay between structured and unstructured relational judgment is often understood through the lens of dual process theories (see Evans & Stanovich, 2013 for a review), which propose that people's responses are the product of two types of cognitive processes: Type 1 processes are automatic and rely on heuristics of association, contiguity, and relatedness, whereas Type 2 processes are deliberative and rely on symbolic and rule-based computations. The automaticity of Type 1 processes implies that their responses serve as intuitive defaults, which may be overridden by Type 2 processes with effort and time.

Dual process theories have provided a useful organizational framework for theorizing about judgment and reasoning. That said, nearly all dual process theories of high-level cognition are expressed verbally and are capable of making only qualitative (and not quantitative) predictions regarding participant responses. There are two reasons for this. First, Type 1 and Type 2 processes involve complex dynamic interactions, which are difficult to describe using mathematical models. Second, the representations over which Type 1 and Type 2 processes operate are often highly complex. It is impossible to build a formal model of Type 1 and 2 processes for common judgment and reasoning tasks, without specifying unstructured (associative) and structured (symbolic) representations for the thousands of concepts that could be evaluated in these tasks.

We present a computational framework for modeling relational judgment for pairs of words, that addresses these two problems. Our framework is based on our prior work on decision making (Bhatia, 2013; Bhatia & Mullett, 2016; Bhatia & Pleskac, 2019; Golman et al., in press; and Zhao et al., 2019) and on semantic representation (Bhatia, 2017; Bhatia et al., 2019; Bhatia & Stewart, 2018). In the former, we examine mathematical models capable of capturing complex dynamic and stochastic aspects of two-alternative forced choice. Although such models are typically applied to simple perceptual, lexical, or preferential choice, here we show that they can also be used to model the interplay between Type 1 and 2 processes in high-level relational judgments, such as those asking participants to judge whether *PartOf*(*sparrow*,*bird*) is true or false. Specifically, we use the diffusion decision model (DDM) (see Ratcliff & Smith, 2004 for a review; also see Zhao et al., 2019 for a related application). We specify the signals generated by Type 1 processes as starting point effects that automatically bias the individual in favor of one of the two responses at the beginning of the decision process, and the signals generated by Type 2 processes as drift rate effects that gradually drive the deliberation towards the correct response.

The specific starting points and drift rates in our framework are obtained from word vector models of semantic representation. Vector semantic models describe words and concepts as points in high dimensional spaces, which are derived from word-co-occurrence statistics in large-scale natural language data. Prior work has found that similarity in these spaces corresponds to word association, and is thus able to predict associative bias in high-level judgment tasks (see Bhatia et al., 2019 for a review). Following this research, we specify the starting point bias in our DDM model as a linear transformation of the cosine similarity between the word vectors for the words involved in the relational judgment (e.g. *sparrow* and *bird*). For the purposes of this paper we use pretrained GloVe vectors (Pennington et al., 2014).

Vector semantic models are also useful for predicting whether or not a given relation holds between a pair of words. Specifically, the Bayesian Analogy with Relational Transformation (BART) model (Lu et al., 2019) passes word vectors for entities in a proposition (e.g. *sparrow* and *bird*) through a relation-specific non-linear function, to generate a continuous measure of the degree to which the proposition being judged is true or false. Such functions can be learnt from sufficient relational data and can thus be used to make predictions for novel propositions (not explicitly hard-coded by the modeler). Following this work, we specify the drift rate in our model as a linear transformation of the BART prediction for the proposition.

We build and test our framework using relations between word pairs obtained from ConceptNET, a large open source knowledge base (Speer et al., 2017). Specifically, we extract over 12,000 word pairs that are linked to each other by at least one of ten different relations (*IsA*, *PartOf*, *UsedFor*, *AtLocation*, *Causes*, *HasProperty*, *Synonym*, *Antonym*, *DistinctFrom*, and *MannerOf*). These word pairs and relations represent common sense knowledge about the world obtained from numerous sources, including crowd-sourced data. By training our models on this data we are able to quantify Type 1 and 2 signals for propositions that apply the ten ConceptNET relations to nearly any English word pair.

We find that the average cosine similarity of word pairs that make up the ConceptNET propositions is 0.31, whereas the average cosine similarity of two randomly selected unrelated words is 0.07. Consequently, Type 1 processes, that rely entirely on association, are able to successfully predict whether or not two words are related to each other. This result indicates that Type 1 processes (as specified by our framework) are adaptive. If these processes provide immediately accessible and cost efficient signals, then these signals serve as good priors for further deliberation. Indeed, the DDM provides a particularly compelling interpretation of this result. According to the sequential probability ratio test, priors in optimal sequential decision making should enter as starting points in a DDM-based dynamic process. Thus, using simple association to determine the DDM starting point in a relation judgment task is (boundedly) rational if this association is immediately and effortlessly available.

Of course, the association between a pair of word (e.g. cosine similarity of *sparrow* and *bird*) is not enough to judge whether or not a specific relation (e.g. *PartOf*) holds between the words. Our BART-based specification of Type 2 processes, which provides a relation-specific prediction for the words, is able to address this problem. We find that the signal generated by this model achieves an out-of-sample accuracy rate of 73% (significantly higher than a random accuracy rate of 50%) in making relation predictions on the ConceptNET data. This shows that BART-based Type 2 processes can be used to substantially improve the response tendencies generated by Type 1 associative processes.

In order to test whether our framework is able to successfully describe participant responses, we ran an experiment in which 42 subjects were asked to judge the truth or falsehood of 300 propositions made up of 30 word pairs for each of the ten ConceptNET relations. Half the propositions were true, and the false propositions were chosen to match the true propositions in terms of the strength of association of their word pairs. Our dual process DDM specification predicts that subjects should be more likely to respond correctly in congruent trials (in which Type 1 and Type 2 processes both support the correct response) than in incongruent trials (in which only Type 2 processes support the correct response). It also predicts that correct responses should be quicker in congruent trials than in congruent trails. Both these patterns emerge in our data. We also fit our DDM model to choice and response time data using hierarchical Bayesian model fitting, and find that cosine similarity plays a significant role in the starting point (with starting points closer to the "true" threshold for highly associated word pairs) and that the BART rating plays a significant role in the drift rate (with drift rates favoring the "true" response when BART judges the proposition to be true). These results are consistent with the theoretical structure proposed above.

Overall, our framework offers novel insights that significantly advance our understanding of reasoning and judgment processes. By combining technical and theoretical ideas from two important subfields of cognitive science it is able to make precise choice and response time predictions for nearly any English language word pair on the ten ConceptNET relations. This makes it unique amongst computational models of reasoning and judgment, which seldom make quantitative predictions, and are usually limited to a small set of stimuli that are hand-coded by the modelers. We are currently in the process of running additional experiments to test our framework. These involve relations obtained from other datasets (e.g. SemEval-2012) as well as experimental manipulations that alter Type 1 responses with additional participant training, and restrict Type 2 processing with working memory load. We also plan to use our framework to test for more complex interactions between Type 1 and 2 processes. We are excited about these tests and look forward to presenting their results at the Cognitive Science conference in July.

## References

**Bhatia, S.** (2013). Associations and the accumulation of preference. *Psychological Review*, 120(3), 522-543.

**Bhatia, S.** (2017). Associative judgment and vector space semantics. *Psychological Review*, 124(1), 1-20.

**Bhatia, S.** & Mullett, T. (2016). The dynamics of deferred decision. *Cognitive Psychology*, 86(7), 112-151.

**Bhatia, S.** & Pleskac, T. (2019). Preference accumulation as a process model of desirability ratings. *Cognitive Psychology*, 109, 47-67.

**Bhatia, S.,** Richie, R, & Zou, W. (2019). Distributed semantic representations for modelling human judgment. *Current Opinion in Behavioral Sciences*, 29, 31-36.

**Bhatia, S**. & Stewart, N. (2018). Naturalistic multiattribute choice. *Cognition*, 179, 71-88.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223.

Golman, R., **Bhatia, S.** & Kane, P. (in press). The dual accumulator model of strategic deliberation and decision making. *Psychological Review*.

Lu, H., Wu, Y. N., & Holyoak, K. J. (2019). Emergence of analogy from relation learning. *Proceedings of the National Academy of Sciences*, 116(10), 4176.

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. *EMNLP*.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333.

Speer, R., Chin, J., & Havasi, C. (2017, February). Conceptnet 5.5: An open multilingual graph of general knowledge. *AAAI*.

Zhao, W. J., Diederich, A., Trueblood, J. S., & **Bhatia, S.** (2019). Automatic biases in intertemporal choices. *Psychonomic Bulletin and Review*, 26(2), 661-668.