

# Hierarchical Drift-Diffusion Model for Moral Dilemma: Understanding Reaction Times and Choices

Richard Kim<sup>1</sup> (kimr@mit.edu) Niccolo Pescetelli<sup>1</sup> (niccolop@mit.edu) Max Kleiman-Weiner<sup>2</sup> (maxkw@mit.edu),  
Edmond Awad<sup>1</sup> (awad@mit.edu), Sohan Dsouza<sup>1</sup> (awad@mit.edu),  
Josh Tenenbaum<sup>2</sup> (jbt@mit.edu), Iyad Rahwan<sup>1</sup> (irahwan@mit.edu)

<sup>1</sup>Media Lab / <sup>2</sup>Brain and Cognitive Sciences, Massachusetts Institute of Technology  
Cambridge, MA 02139

## Abstract

Discrete choice models (e.g. logistic regression) are popular models in the economics literature that describe choices between two or more discrete alternatives. These models have been successfully used to model value-based decisions, e.g. decisions in moral dilemmas, although temporal components of a decision, such as reaction times and changes of mind are not included. In cognitive sciences, another class of decision models, namely sequential-sampling models, has gained popularity in modelling choice accuracy, reaction time and decision uncertainty (e.g. confidence judgments). Here, we model decisions in moral dilemmas using a variant of a hierarchical drift-diffusion model, factor drift diffusion, that combines the value-based approach with that of evidence accumulation mechanism by sequential-sampling. Specifically, we model the evidence accumulation process as resulting from a subjective weighting of abstract moral dimensions (factors). We train our model on a data set of 6500 moral decisions by 500 respondents on a popular web platform (MoralMachine.mit.edu) and separately infer different sources of uncertainty in moral decisions. We show that the model successfully predicts reaction times and choices in moral dilemmas, while also leading to unexpected results.