My primary research interest is to develop statistical machine learning methods for human-centered applications, particularly ones which identify influences on human behavior. As these problems require interpretable results, I focus on Bayesian models, in which latent random variables capture the hidden structure in observed data. To discover the values of these hidden model parameters, I derive and implement scalable inference algorithms, then apply them to real-world data. I use statistical model checks, held-out prediction, visualization, and domain expert reviews to validate and refine these models.

Modeling text Written text is a rich and abundant source of data that can tell us about human behavior, relationships, and influences. Probabilistic topic models discover the underlying themes in collection of text documents; these themes can be used to summarize, organize, explore, and analyze the corpus.

Topic models, however, are high-level statistical tools—a user must scrutinize numerical distributions to understand their results. To make these results accessible, David Blei and I developed a method for visualizing topic models. Our method creates a navigator of the documents, allowing users to explore the hidden structure that a topic model discovers and understand the collection in new ways.

Historians face a related problem: they read many documents to identify significant events that influence individuals and agencies. Hanna Wallach, David Blei, and I developed methods to help historians identify possible events from diplomatic cables or similarly structured text (such as email). We built on topic modeling to distinguish between topics that describe "business-as-usual" and event topics that deviate from these patterns during particular periods of time. We developed scalable variational inference algorithms for these methods and are currently expanding our methods to incorporate network interactions.

Modeling user behavior Logged user actions, such as clicks on web posts, can also help us understand influences on behavior. Algorithmic recommendation systems use this data to uncover latent "preferences" for items and form personal recommendations based on the activity of others with similar tastes.

With David Blei and Tina Eliassi-Rad, I developed the *social Poisson factorization* (SPF) recommendation model. Prior work represents users only in terms of general preferences; these models do not capture that a user may like an item because her friend likes that item. SPF models both signals, discovering both latent preferences and unobserved influence between pairs of connected users; these learned parameters can then be used to explore data. We developed scalable algorithms for analyzing data with SPF and demonstrated that it outperforms competing methods on six real-world datasets.

With Mike Gartrell, Jake Hofman, and others, I explored how group settings influence users by performing a large-scale study of television viewing habits. Our analysis revealed how engagement in group viewing varies by viewer and content type, and how viewing patterns shift across various group contexts. We then constructed a simple model of how individual preferences are combined in group settings.

Current research With Young-suk Lee and Babara Engelhardt, I am developing *generalized nonparametric deconvolution models* (NDMs), a family of Bayesian nonparametric models for collections of data in which each observational unit (e.g., regional population) is comprised of unobserved heterogeneous particles (e.g., individuals). Like other models, NDMs describe the data in terms of latent factors or hidden patterns. Unlike existing models, NDMs recover the local fluctuations for each observation, describing how it deviates from global patterns and enabling us to analyze how it is impacted by external influences.

Future research Algorithmic recommendation systems conflate the concepts of user behavior and user preferences and use this amalgamation to potentially influence users. I plan to disentangle the concepts of user behavior and user preferences in the context of recommendation systems and estimate the causal impact of these systems on both behavior and preferences. I will use text-based domains for this project, integrating my previous work in visualizing and modeling text with my work in recommendation.