

## Marketing: Selected Doctoral Theses

### **“Essays on Understanding and Combating Misinformation at Scale”**

**Author:** Jennifer Allen (2024)

**Committee:** David Rand (chair), Abdullah Almaatouq, Adam Berinsky (MIT Political Science)

**Abstract:**

In Chapter 1, I explore the use of crowdsourcing as a potential solution to the misinformation problem at scale. Perhaps the most prominent approach to combating misinformation is the use of professional fact-checkers. This approach, however, is not scalable: Professional fact-checkers cannot possibly keep up with the volume of misinformation produced every day. Furthermore, many people see fact-checkers as having a liberal bias and thus distrust them. Here, we explore a potential solution to both of these problems: leveraging the “wisdom of crowds” to make fact-checking possible at scale using politically-balanced groups of laypeople. Our results indicate that crowdsourcing is a promising approach for helping to identify misinformation at scale.

In Chapter 2, joint with David Rand and Cameron Martel, I extend work on crowdsourced fact-checking to assess the viability of crowdsourcing in an opt-in, polarized environment. We leverage data from Birdwatch, Twitter’s crowdsourced fact-checking pilot program, to examine how shared partisanship affects participation in crowdsourced fact-checking. Our findings provide clear evidence that Birdwatch users preferentially challenge content from those with whom they disagree politically. While not necessarily indicating that Birdwatch is ineffective for identifying misleading content, these results demonstrate the important role that partisanship can play in content evaluation. Platform designers must consider the ramifications of partisanship when implementing crowdsourcing programs.

In Chapter 3, I examine the role of online (mis)information on US vaccine hesitancy. I combine survey experimental estimates of persuasion with exposure data from Facebook to estimate the extent to which (mis)information content on Facebook reduces COVID vaccine acceptance. Contrary to popular belief, I find that factually-accurate vaccine-skeptical content was approximately 50X more impactful than outright false misinformation. Although outright misinformation had a larger negative effect per exposure on vaccination intentions than factually accurate content, it was rarely seen on social media. In contrast, mainstream media articles reporting on rare deaths following vaccination garnered hundreds of millions of views. While this work suggests that limiting the spread of misinformation has important public health benefits, it highlights the need to scrutinize accurate-but-misleading content published by mainstream sources.

### **“Nothing to See, Nothing to Say, and Nothing How Much: Three Essays on Information and Behavior”**

**Author:** Matthew Cashman (2024)

**Committee:** Drazen Prelec (chair), David Rand, Rahul Bhui

**Abstract:**

I present three essays that examine information flows and behavior. The first examines the effect of sequential play in Public Goods Games in cases where players move one after another but do not see each others’ moves. Even with no information flow—when there is nothing to see of others’ decisions—order of play affects contributions to the public good. The second essay considers pre-play socializing and its effects on coordination games and hold-up games. Pre-play small-talk results in better outcomes even when players talk before they know they will be playing a game—before they have anything of strategic relevance to say. The third essay presents a novel quantitative, empirical means of measuring the flow of memes through minds. Most ways of learning what other people know rely on strong commitments to what the right question to ask is. Using cloze completion tasks I outline a principled, content-agnostic method of estimating how much information from a given text is stored in a reader’s mind.

## **“Essays on Understanding Marketing Innovations”**

**Author:** Keyan Li (2024)

**Committee:** Ducan Simester (co-chair), Juanjuan Zhang (co-chair), T. Tony Ke

### **Abstract:**

The dissertation consists of three chapters on understanding marketing innovations, including targeted marketing and new product development. The first chapter studies a novel targeting problem that many firms face and develops a new method for targeting experimentation. Adaptive learning policies that guide how firms trade off acquiring new information to improve a current targeting policy, versus exploiting the current policy to harvest, typically focus on settings in which customers arrive individually, in a frequent sequence. However, in practice, firms often conduct marketing campaigns in batches, in which they target a large group of customers with personalized marketing actions together. This has an important implication for how firms resolve the tradeoff between acquiring new information and exploiting the current policy. The large number of customers in each batch (campaign) introduces an information externality: the incremental information contributed by a single customer depends upon the assignment decisions for other customers in the batch. We investigate how to optimally acquire and coordinate information in these settings. The algorithm we propose uses Gaussian processes to estimate the value of incremental information, while accounting for the information externality between customers in the same batch. Findings are validated using data from a field experiment.

The second chapter studies customer demand in a non-market-oriented economy. The economics and marketing literature has primarily focused on market economies and studied factors such as price and advertising when analyzing customer demand. However, in non-market-oriented economies, social factors like corruption can have a significant influence on customer decisions. In particular, this paper focuses on the demand for luxury products, which are widely used for gift-giving and even bribery in emerging markets. One possible mechanism is that when the relative size of non-market-oriented sectors in the local economy increases, luxury products can be used to identify those who have a higher willingness to pay for scarce resources. As a result, the demand for luxury products moves together with the degree of corruption. By leveraging natural experiments of top-down anti-corruption campaigns that temporarily halt this channel, an empirical study is performed using a comprehensive dataset that covers the sales of all cigarette brands and the local social environment in China. The results suggest that these social factors can have an unanticipated impact on the demand for luxury products.

The third chapter studies how customer search can stifle product innovations. Conventional wisdom suggests that when an incumbent fails to innovate, there is a greater risk to the incumbent of competition from other innovators. I show conditions in which the opposite is true: by delaying innovation, an incumbent can create entry barriers that deter innovation by competitors. Consequently, both competition and innovation are suppressed. The key insight driving these outcomes is that customer search is endogenous, and absence of innovation today can disincentivize customers from searching in the future. Since customers need to search to discover innovations, when they search less, it both creates entry barriers for competitors, and reduces the competitors' incentives to innovate. Postponing innovation can benefit incumbents if it motivates customers to search less, and thus competitors to innovate less. Notably, I show that searching less is a rational customer response.

## **“Essays on Understanding and Combating Misinformation at Scale”**

**Author:** Yifei Wang (2024)

**Committee:** Catherine Tucker (co-chair), Duncan Simester (co-chair), Birger Wernerfelt

### **Abstract:**

This dissertation consists of three chapters on understanding the opportunities and challenges of using individualized data in marketing, in the context of mobile economy, retailing, and education.

The first chapter investigates the impact of market structure on the use and abuse of individualized data. In particular, I study how a shift in competition affects the nature and type of a firm's intrusion on consumers' privacy. I study this question in the context of Android app markets in China, and measure privacy by examining apps' permission requests. I investigate a 2017 regulation that reduced competition in censored app categories by prohibiting censorship-circumvention tools commonly used to access apps banned by the

government. This regulation made banned apps less accessible and reduced the competition faced by permitted apps in censored categories but did not affect apps in uncensored categories. I use a synthetic differences-in-differences approach to Android permission requests by apps in censored and uncensored categories before and after the regulatory change. I show that reducing competition led to a significant increase in the permission requests by apps in censored categories. Empirically, I show that this increase in privacy-invasive behavior is due to treated apps' efforts to improve consumer engagement and monetize attention.

The second chapter investigates the correlations in an individual customer's willingness to engage in search across different decisions and contexts. I show empirically show that the amount of search a customer engages in is correlated across seemingly unrelated tasks. I prove theoretically that this leads to correlations in customer decisions in seemingly unrelated product categories and contexts. I use these theoretical and empirical findings to explain the 'harbingers of failure' phenomenon documented in the recent literature, which is a series of findings showing there exist customers who systematically buy new products that fail across product categories and decision contexts. In this paper I argue that one latent characteristic that could contribute to the effect is a customer's willingness to engage in search and show how the theoretical and empirical findings on interrelated search of individual customers can explain the harbingers of failure phenomenon.

The third chapter studies how access to digital educational content affects inequality in education. In particular, our analysis uses individualized data on children's reading behavior from an eBook app to trace out both the short-run and long-run treatment effects of providing free access to digital reading resources to children with different socio-demographic backgrounds. We find that free access to digital content leads to a dramatic and immediate increase in reading time for treated children, and that this immediate effect is much large for children from less developed regions with fewer educational resources. However, children's reading activities decline quickly after the start of their free access, and this decline is much faster for children from less developed regions. Further evidence suggests that children from more developed regions benefit more from the free access in the long run. Our mechanism analysis further reveals a nuanced complementarity between digital and non-digital education.

## **“Product Returns Management in Online Retail”**

**Author:** Marat Ibragimov (2023)

**Committee:** John R. Hauser (chair), Duncan Simester, Artem Timoshenko (Kellogg School of Management, Northwestern University)

### **Abstract:**

In Chapter 1, I and coauthors study the problem of predicting the product return rate using the products' visual information. In online channels, products are returned at high rates. Shipping, processing, and refurbishing are so costly that a retailer's profit is extremely sensitive to return rates. Using a large dataset from a European apparel retailer, we observe that return rates for fashion items bought online range from 13% to 96%, with an average of 53% – many items are not profitable. Because fashion seasons are over before sufficient data on return rates are observed, retailers need to anticipate each item's return rate prior to launch. We use product images and traditional measures available prelaunch to predict individual item return rates and decide whether to include the item in the retailer's assortment. We complement machine-based prediction with automatically extracted image-based interpretable features. Insights suggest how to select and design fashion items that are less likely to be returned. Our illustrative machine-learning models predict well and provide face-valid interpretations – the focal retailer can improve profit by 8.3% and identify items with features less likely to be returned. We demonstrate that other machine-learning models do almost as well, reinforcing the value of using prelaunch images to manage returns.

In Chapter 2, I consider customer search and product returns on the individual level. Previous research has focused on linking customers' purchase and return decisions. However, online retailers have access to the information which precedes the purchase decision – customer search. I demonstrate that customer search

information provides important insights about product returns. Using data from a large European apparel retailer, I propose and estimate a joint model of customer search, purchase, and return decisions. I then provide theory and data indicating that using search filters, viewing multiple colors of a product, spending more time, and purchasing the last item searched are negatively associated with the probability of a return. Finally, I use the proposed model to optimize the product display order on the retailer's website.

Chapter 3 extends and reinforces the results obtained from previous Chapters. In the paper, I study the assortment planning problem in presence of frequent product returns. I develop a deep-learning model of customer search, purchase, and return. The model is based on a transformer framework and allows the recovery of important relations in the data. I use the estimated model to demonstrate that retailers could identify successful and unsuccessful products and modify the assortment. The modified assortment would increase the retailer's sales and at the same time decrease returns. Lastly, I provide qualitative insights on which products are most likely to be unsuccessful in online retail.

### **“Augmented Machine Learning and Optimization for Marketing”**

**Author:** Yuting Zhu (2022)

**Committee:** Juanjuan Zhang (co-chair), Duncan Simester, Tony T. Ke (The Chinese University of Hong Kong)

#### **Abstract:**

This dissertation consists of three essays exploring how to augment machine learning and optimization methods for marketing management. The first essay considers an augmentation of deep-learning-based recommender system for sales force management. Helping new salespeople succeed is critical for many organizations. We develop a deep-learning-based recommender system to help new salespeople recognize suitable customers, leveraging historical sales records of experienced salespeople. One challenge is how to learn from experienced salespeople's own failures, which are prevalent but often do not show up in sales records. We develop a parsimonious model to capture these “missing by choice” sales records and incorporate the model into a neural network to form an augmented, deep-learning based recommender system. We validate our method using sales force transaction data from a large insurance company. Our method outperforms common benchmarks in prediction accuracy and recommendation quality, while being simple, interpretable, and flexible. We demonstrate the value of our method in improving sales force productivity.

The second essay explores an augmentation of large-scale linear programming optimization method for targeting with constraints. Personalization, which aims to target different marketing actions to different customers, has attracted broad attention in both academia and industry. While most research has focused on training personalization policies without constraints, in practice, many firms face constraints when implementing these policies. For example, firms may face volume constraints on the maximum or minimum number of actions they can take, or on the minimum acceptable outcomes for different customer segments. They may also face fairness constraints that require similar actions with different groups of customers. These constraints can introduce difficult optimization challenges, particularly when the firm intends to implement personalization policies at scale. Traditional optimization methods face challenges solving large-scale problems that contain either many customers or many constraints. We show how recent advances in linear programming can be adapted to the personalization of marketing actions. We provide a new theoretical guarantee comparing how the proposed method scales compared to state-of-the-art benchmarks (primal simplex, dual simplex and barrier methods). We also extend existing guarantees on optimality and computation speed, by adapting them to accommodate the characteristics of personalization problems. We implement the proposed method, and compare it with these benchmark methods on feasibility, computation speed, and profit. We conclude that, volume and similarity (fairness) constraints should not prevent firms from optimizing and implementing personalization policies at scale.

The third essay studies collective search in an organization. In this paper, we build a two-member two-period model to show that when a group of people with different preferences conduct search and make a decision together, they can benefit from making a commitment on the number of products to search ex ante when the search cost is very small or relatively large. The underlying mechanism is that, because of the preference divergence between group members, they tend to search fewer products and thus have lower

expected utility in group search than in single agent search, and making a commitment on the number of products to search can help mitigate the preference divergence problem in group search. If consumers can observe product prices before search and the firm sets product prices endogenously, the firm can benefit from letting consumers commit to the number of products to search ex ante if consumers search as a group and their search cost is small. We also consider several extensions to show the robustness and boundary conditions of our findings.

## **“Essays on MarTech: Learning to Design, Deliver, and Diffuse Interventions”**

**Author:** Jeremy Yang (2021)

**Committee:** Juanjuan Zhang (co-chair), Sinan Aral (co-chair), Dean Eckles

### **Abstract:**

Chapter one develops an algorithm to predict the causal effect of influencer video advertising on product sales. A summary statistic, motion-score, or m-score, is proposed to capture the extent to which a product is advertised in the most engaging parts of a video. Pixel-level product placement is located with an object detection algorithm and pixel-level engagement is estimated as a saliency map by fine-tuning a deep 3D convolutional neural network on video-level engagement data. M-score is then defined as pixel-level engagement-weighted advertising intensity of a video. The algorithm is constructed and evaluated with influencer video ads on TikTok. Causal effects of video ads on product sales are identified by exploiting variation in video posting time. Videos of higher m-score indeed lift more sales. This effect is sizable, robust, and more pronounced among impulsive, hedonic, or inexpensive products. The mechanism can be partially traced to influencers’ incentives to promote themselves rather than the product. How various stakeholders in entertainment commerce can use m-score in a scalable way to optimize content, align incentives, and improve efficiency are discussed.

Chapter two proposes a method to optimize a targeting policy that maximizes an outcome observed only in the long term. Traditionally, this typically requires delaying decisions until the outcome is observed or relying on simple short-term proxies for the long-term outcome. The method builds on the statistical surrogacy and off-policy learning literature to first impute the missing long-term outcomes and then approximate the optimal targeting policy on the imputed outcomes via a doubly robust approach. It is applied in large-scale proactive churn management experiments at The Boston Globe by targeting optimal discounts to its digital subscribers to maximize their long-term revenue. It is shown that conditions for the validity of average treatment effect estimation with imputed outcomes are also sufficient for valid policy evaluation and optimization; furthermore, these conditions can be somewhat relaxed for policy optimization. The method is also validated empirically by comparing it with a policy learned on the ground truth long-term outcomes, they are shown to be statistically indistinguishable. It also outperforms a policy learned on short-term proxies for the long-term outcome.

Chapter three investigates how network embeddings can be applied to the study of product diffusion. Three sets of results are documented using a combination of real and simulated datasets: First, node embeddings can predict adoption decisions and timing better than standard centrality-based summary statistics. Second, node embeddings as control variables reduce the bias in the estimation of peer effect, especially when tie formation depends on unobservables. Third, graph embeddings based on the diffusion process as a whole reveal meaningful similarities between different diffusion processes such as simple vs. complex contagion.