

© Copyright 2018

Aravinda Garimella

Essays on User Engagement in Online Collaborative Communities

Aravinda Garimella

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Ming Fan, Chair

Suresh Kotha

Yong Tan

Program Authorized to Offer Degree:

Business Administration

University of Washington

Abstract

Essays on User Engagement in Online Collaborative Communities

Aravinda Garimella

Chair of the Supervisory Committee:
Associate Professor Ming Fan
Department of Information Systems and Operations Management

In my dissertation, I study user engagement on online collaborative communities. I explore user interactions in two contexts, Crowdfunding (Collaborative funding) and Massive Open Online Courses (Collaborative learning) platforms. In the context of collaborative fundraising, using data from JD.com, China's leading crowdfunding platform, I study entrepreneurial activities before and after the launch of crowdfunding campaigns and how these activities affect the likelihood of campaign success. Specifically, I study the effect of raising awareness prior to the launch of a campaign on the likelihood of a campaign meeting its goal. In the context of collaborative learning, I study student engagement on MOOCs, an outcome of utmost interest to practitioners. Using a rich data set of daily student interactions over a period of forty months on Khan Academy, a pioneer MOOC platform in the K-12 space and text mining techniques, I study the effect of the display of empathy by students on the level of conversation on the platform .

TABLE OF CONTENTS

List of Figures	3
List of Tables	4
Dedication.....	5
Acknowledgements.....	6
Chapter 1. Introduction	7
Chapter 2. Launch on a High Note: How Prefunding Affects Crowdfunding Outcomes	10
2.1 Introduction.....	10
2.2 Literature Review.....	13
2.3 Theory and Hypotheses.....	16
2.4 Research Context and Data	21
2.5 Empirical Analyses	24
2.5.1 Identification Strategy.....	24
2.5.2 Propensity Score Matching.....	25
2.5.3 Econometric Analysis	26
2.5.4 Results.....	29
2.6 Discussion and Implications	31
2.6.1. Discussion.....	31
2.6.2 Implications.....	33
2.7 Conclusion	34
2.7.1. Contributions.....	34
2.7.2 Limitations	36

2.7.3	Concluding Remarks.....	36
Chapter 3.	Human Touch in the Online World: The Role of Empathy in Online Education.....	38
3.1	Introduction.....	38
3.2	Literature Review.....	41
3.3	Theory and Hypotheses.....	43
3.4	Research Context and Data.....	43
3.4.1.	Research Context	46
3.4.2.	Data.....	48
3.5	Model.....	53
3.5.1.	Assumptions.....	54
3.4.2.	Identification Challenges	56
3.4.3.	Estimation Strategy.....	56
3.4.4.	Model Specifications	57
3.4.5.	Results.....	58
3.6	Discussion and Conclusion.....	59
Chapter 4.	Conclusion.....	62
Bibliography.....		63
References for Chapter 1.....		63
References for Chapter 2.....		64
References for Chapter 3.....		67
Tables.....		69
Figures.....		77

LIST OF FIGURES

Figure 2.1. Sample Project in Prefunding Phase.....	91
Figure 2.2. Theoretical Model.	92
Figure 2.3. Propensity Score Distributions for Non-Prefunding and Prefunding Projects.	93
Figure 2.4. Number of Backers: Prefunding and Non-Prefunding Projects.	94
Figure 2.5. Contribution Size: Prefunding and Non-Prefunding Projects.	95
Figure 2.6. Distribution of Contribution Size on First Day of Fund-Raising	96
Figure 2.7. Distribution of Contribution Size on Last Day of Fund-Raising	97
Figure 3.1. Example of a micro-lecture on Khan Academy	98
Figure 3.2. Attenuation of Conversation on Khan Academy.....	99
Figure 3.3. CDF of Conversation Level Per Question.....	100

LIST OF TABLES

Table 2.1. Variable Descriptions.	78
Table 2.2. Summary Statistics.	79
Table 2.3. Logit Regression on Prefunding Decision.	80
Table 2.4. Average Treatment Effect of Prefunding on Project Success.....	81
Table 2.5. Effects of Prefunding Activities on Initial Performance	82
Table 3.1. Variable Descriptions.	86
Table 3.2. Summary Statistics.	89
Table 3.3. Effects of Expressed Empathy on Student Conversation on MOOCs.....	90

DEDICATION

To my parents, Ramesh Garimella and Annapurna Garimella for instilling in me the love for learning.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those who helped me to get where I am today. I want to thank Professor Ming Fan who has been a great advisor and mentor to me in the past few years. His support and guidance have shaped and will continue to shape my approach to research through my academic career. I also want to sincerely thank the members of my supervisory committee, Professors Abhishek Borah, Elina Hwang, Suresh Kotha, Yong Tan and Dongsheng Zang for their guidance, inputs and their support of me and my work since I formed my committee.

I would like to extend special thanks to Shawna Reimers, the ISOM Department Administrative Assistant, and Jaime Banaag, the Assistant Director of the PhD program for their kindness, support and consistently positive attitude. Also, my special thanks go to my friends in the doctoral program, Antino Kim, Haoyan Sun, Xue Tan, and Amirreza Fazli, who consistently supported me and made my PhD experience so much better. I feel fortunate to have had the opportunity to know these individuals and learn from them.

Last but not the least, I thank my family for their unconditional support. My parents always believed in me and gave me their unwavering love and support. My sister and brother-in-law were my pillars of support during the doctoral support. I will be eternally thankful to them.

Chapter 1: Introduction

In “The Origin of Wealth”, Eric Beinhocker (2006) differentiates between two types of technologies: physical and social. Physical technologies are “methods and designs for transforming matter, energy, and information from one state into another in pursuit of a goal”, while social technologies are “methods and designs for organizing people in pursuit of goals”. Modern advances in social technologies have enabled the democratization of traditionally centralized industries through a class of platforms that are collectively known as *online collaborative communities*. Online communities emphasize community collaboration to foster interaction frequency and intensity among members and leverage the team spirit of a community. Therefore, they heavily depend on the engagement level of the users in the community.

In this dissertation, I study user interactions in two distinct contexts within online collaboration: Crowdfunding (Collaborative fundraising) and Massive Open Online Courses (Collaborative learning). Crowdfunding has democratized access to capital and MOOCs have reduced geographic and socio-economic barriers to education. The first chapter of my dissertation focuses on how entrepreneurs can be more effective in fundraising on crowdfunding platforms. I study the activities entrepreneurs can engage in prior launching their fundraising campaigns and how these activities affect the likelihood of campaign success. The second chapter focuses on student interaction on MOOCs. I study factors that affect student engagement, an outcome of utmost interest to MOOC practitioners.

This manuscript is structured as follows. Chapter 2, which is the first essay is motivated by the fact that in an increasingly competitive crowdfunding market, entrepreneurs are seeking out new ways to improve their likelihood of meeting of their goals. This is one of the reasons for the

emergence of prefunding platforms, which help entrepreneurs to raise awareness of their forthcoming fund-raising projects. Using data from an online crowdfunding platform, I show that opting for prefunding prior to fund-raising significantly increases the likelihood of a campaign meeting its fund-raising goal. I also find that prefunding projects transition to the fund-raising stage with a higher number of backers, and a higher contribution size on the first day of funding when compared to non-prefunding projects. This strong head-start on the first day of the solicitation period then enhances a project's ability to garner funds over the fund-raising window, thus increasing the likelihood of eventual success. Moreover, when entrepreneurs proactively update information about their projects and actively engage with followers during the prefunding period, their campaigns are more likely to launch with a higher number of backers and higher contribution size.

This study has important implications for both entrepreneurs and crowdfunding platforms. Through this work, I complement the research stream of crowdfunding that highlights the importance of early traction by studying the means to achieve this early traction. For crowdfunding practitioners, both entrepreneurs and platforms, I provide a set of actionable insights on the effects of prefunding and on how to employ it effectively to gain an edge in the competitive crowdfunding market.

In Chapter 3, which is an essay on collaborative learning through MOOCs, I address one of the key concerns of MOOC providers: fostering student engagement. I analyze on a rich data set of daily student interactions over a period of forty months on Khan Academy, a pioneer MOOC platform in K-12 education. In particular, I study how the display of empathy by students impacts the evolution of student conversation on MOOCs. The first part of the analysis shows that empathy in both questions and answers, encourages conversation. Interestingly, the role of empathy

becomes more prominent as the difficulty level of the content increases. Finally, given that this study is in an educational setting, I examine the interplay of empathy and analytical thinking and find contrasting effects in questions and answers. I explain why this is the case and discuss implications. I use text mining techniques and employ a dynamic panel generalized method of moments (GMM) estimation technique to provide a rich set of insights on the pivotal role that emotions play in student conversation on MOOCs. Empowered with these insights, MOOC providers can plan and execute meaningful interventions to garner and improve student engagement on their platforms.

Chapter 4 provides a general conclusion of my dissertation. Online collaborative communities thrive on user engagement and the understanding of how to garner and sustain engagement is crucial to the sustenance of these platforms.

Chapter 2: Launch on a High Note: How Prefunding Affects

Crowdfunding Outcomes

2.1. Introduction

Crowdfunding is a radically novel approach to early stage financing as it enables entrepreneurs to raise small amounts of capital from a large network of people who pool their money to support new venture ideas that they resonate with and care to support (Mollick 2014). However, with increasing competition, entrepreneurs now have to compete for financial backing from a large, albeit, finite number of website visitors (Viotto 2015). As such, entrepreneurs are seeking out new ways to improve their likelihood of success in meeting their fund-raising goals. This is perhaps the reason for the emergence of *prefunding* platforms, which help entrepreneurs to raise awareness of their forthcoming fund-raising projects (Richardson 2014).

In look and feel, prefunding platforms are similar to crowdfunding platforms such as Kickstarter and Indiegogo, which enable entrepreneurs to create campaign webpages to share product descriptions, funding goals, and backing options. Prefunding platforms can be stand-alone sites, such as Prefundia or LaunchRock in the US, which are independent from crowdfunding platforms, or integrated as an optional feature into existing crowdfunding platforms, such as JingDong (JD) Crowdfunding, China's leading reward-based crowdfunding platform (McLaughlin, 2016). During the prefunding period, entrepreneurs employ images, videos, and detailed new-venture descriptions or narratives to showcase forthcoming new-venture projects. Entrepreneurs can also engage with potential backers via posting updates and participating in online discussions. However, prefunding differs from regular crowdfunding in that entrepreneurs are not permitted to raise funds during the prefunding time period. In other words, monetary

contributions can be received only after the project is formally launched as a fund-raising campaign (Lomas 2013).

Prefunding provides an opportunity to engage with potential backers and garner their feedback before the launch of fund-raising campaigns. However, going through the prefunding phase requires substantial time and effort by entrepreneurs. As such, it is of interest and importance to know whether prefunding has an impact on fund-raising outcomes in order for entrepreneurs to engage in prefunding. The significant difference between prefunding and crowdfunding portals motivates us to propose the following research questions in this study: *Does prefunding increase the likelihood of a new venture project achieving its funding goal on crowdfunding platforms?* To address this question, we examine the mechanisms through which prefunding affects the likelihood of campaign success and the effects of specific prefunding activities on fund-raising outcomes.

While extant literature has explored various determinants of crowdfunding success (Lin and Viswanathan 2015, Mollick 2014, Ordanini 2011, Burtch et al. 2013), the antecedents to crowdfunding that could potentially influence outcomes have yet to receive attention. We extend this literature stream by exploring the specific ways entrepreneurs can engage potential backers before the fund-raising stage. Additionally, while prior literature has shown that early traction during the funding period is a strong predictor of crowdfunding success¹ (e.g. Burtch et. al 2013, Kuppuswamy and Bayus 2015, Colombo et. al 2015), how entrepreneurs can garner this early traction remains an important and underexplored research question. We find evidence to show that informing and engaging potential backers through prefunding is an effective means to achieve this initial support. Further, we contribute to the broad literature in finance and economics on market mechanisms, such as book building in Initial Public Offerings (IPOs) and call exchange markets,

¹ Success has been operationalized using different measures in the literature: Whether the funding goal is met, how much is raised, and how fast the goal is met, among others.

that aggregate information prior to trade executions (Benveniste and Spindt 1989, Schnitzlein 1996). We find that prefunding can lead to more informed and engaged backers who contribute higher financial support to the campaigns. Finally, prefunding, as either a campaign choice or a stand-alone platform offering, has received scant attention in literature. To the best of our knowledge, our study is the first study in the area of crowdfunding that focuses on the role prefunding plays in crowdfunding campaign success.

We use data collected from JD Crowdfunding, a platform launched in 2014. It is China's leading crowdfunding platform with a market share of about 34%, and over 7,500 projects raised more than \$500 million in 2016. JD Crowdfunding has a prefunding feature built into its platform. Entrepreneurs can choose this option, and then seamlessly transition to fund-raising phase, during which project updates and discussions that happened during the prefunding phase are available to potential backers as entrepreneurs solicit funds. Figure 1 shows a sample project in the prefunding stage.

----- Insert Figure 1 about here -----

Our study yields the following interesting findings. We find that prefunding significantly increases the likelihood of a campaign meeting its fund-raising goal. Specifically, we find that prefunding projects transition to the fund-raising stage with a higher number of backers, and higher contribution size on the first day of funding when compared to non-prefunding projects. This strong head-start on the first day of solicitation then enhances a project's ability to garner funds over the fund-raising window, thus increasing the likelihood of eventual campaign success. Further, when entrepreneurs proactively update information about their projects and actively engage with followers during the prefunding period, their campaigns are more likely to launch with a higher number of backers and higher contribution size from such backers.

Our study has strong implications for both entrepreneurs and platform operators. First, in an increasingly competitive crowdfunding market, a better understanding of prefunding's impact on campaign outcomes can assist entrepreneurs to choose the right campaign option in order to engage effectively with the potential backers. Second, since platform owners receive a percentage of the funds entrepreneurs raise when projects meet funding goals, platforms should be motivated to offer features that can help entrepreneurs to successfully raise funds.

The rest of the chapter is organized as follows. In Section 2, we review prior crowdfunding literature in an effort to motivate our research. In Section 3, we propose theoretical arguments to develop our hypotheses. We describe our data in Section 4, and present our econometric model and results in Section 5. In Section 6, we discuss our findings. Finally, in Section 7, we highlight our contributions, discuss the limitations, and provide concluding remarks.

2.2. Literature Review

Prior crowdfunding studies can be broadly categorized into three research streams. First, prior studies have examined crowdfunding from the perspective of entrepreneurs who seek financial capital and have investigated how entrepreneurs' attributes and actions affect funding outcomes (Lin and Viswanathan 2015, Mollick 2014). A second stream has focused on backers' contribution behavior and the role it plays in crowdfunding campaigns (Ordanini 2011, Burtch et al. 2013). Third, researchers have examined how campaign design and attributes such as contribution policies and rules impact funding outcomes entrepreneurs experience (Wash and Solomon 2014, Chen et al. 2013). An important question addressed across all three research streams is: What determines fund-raising success in the context of crowdfunding?

At the entrepreneur level, researchers have primarily focused on highlighting the attributes and actions of entrepreneurs that impact fund-raising success. Attributes such as the entrepreneur's

geographic location can have an effect on the backers attracted due to home bias (Lin and Viswanathan 2016). A strong track record of past entrepreneurial success (Skirnevskiy et al. 2017) can positively influence fund-raising outcomes by signaling experience and prowess. Entrepreneurs' internal social capital within crowdfunding platforms built through supporting other entrepreneurs' projects (Colombo et al. 2015) and external social capital built outside the platform through other social networks (Ordanini et al. 2011) are also factors that positively affect campaign success. Additionally, entrepreneurial actions such as the presentation of typo-free textual content, the use of videos (Mollick 2014), and display of passion (Li et al. 2017) are associated with campaign success.

Backer-level heterogeneity and behavioral patterns also influence fund-raising dynamics and outcomes. A key factor that influences the behavior of backers is the information on prior contributions, including the amount and timing of others' contribution (Burtch et. al 2013). Hence, early-stage contributions to crowdfunding projects trigger a self-reinforcing mechanism that leads to eventual funding success (Kuppuswamy and Bayus 2015, Lehner 2014). Backing during the initial days of a project is essential to trigger a "success-breeds-success" process, which then leads to a campaign successfully meeting its funding goals (Colombo et. al 2015). Research carried out in peer-to-peer lending settings suggests that backers' herding behavior depends on their sophistication; backers with little or formal expertise in evaluating projects behave differently from institutional backers with significant investment experience (Lin, Sias and Wei, 2015), which then has important ramifications for project success.

At the campaign level, researchers have studied attributes and design choices that impact fund-raising outcomes. Research suggests that campaigns with shorter time durations and smaller funding goals are more likely to succeed (Frydrych et al. 2014). Funding rules such as the keep-it-

all (KIA) versus all-or-nothing (AON) have also been shown to affect funding outcomes (Cumming et al. 2014). The KIA model involves the entrepreneur setting a fund-raising goal and keeping the entire amount raised, regardless of whether she meets her goal, while the AON model involves the entrepreneur setting a fund-raising goal and keeping nothing unless the funding goal is achieved.

Our study bridges the following gaps in current literature. First, entrepreneurial actions prior to the launch of crowdfunding campaigns have received little attention. We examine prefunding activities by entrepreneurs and their effects on fund-raising dynamics upon launch. Second, while prior literature has shown that early traction plays an important role in crowdfunding success (Burtch et. al 2013, Kuppuswamy and Bayus 2015, Colombo et. al 2015), how entrepreneurs can attain this early support is a question is yet to be addressed. Finally, we study the effects of prefunding, which is a campaign design decision that has not been examined in the current literature to the best of our knowledge.

Our research is also related to the market microstructure literature in finance and economics (O'Hara 1995), especially the studies on book building in IPOs and on call markets in financial trading. By marketing the IPOs to potential investors in "road shows," book building can aggregate information by encouraging investors to reveal their beliefs about issue's value (Benveniste and Spindt 1989). Similarly, in call auctions, orders are batched for simultaneous execution, and this accumulation of orders can lead to lower adverse selection costs noise traders incur (Schnitzlein 1996). We contribute to the broad literature in finance and economics of information systems that examine the effects of new market mechanisms.

Finally, prefunding, as a distinct activity prior to fund-raising, has the potential to significantly impact outcomes. Prefunding activities are distinct from and should not be

confounded with early crowdfunding period activities, as prefunding can aggregate interest and information on projects prior to fund-raising. This distinction is important as prefunding provides opportunities for entrepreneurs to engage with audience, process feedback, and clarify or revise their narratives prior to the formal launch of a campaign. Through our study, we hope to examine prefunding as an important antecedent to early performance, which, subsequently, should affect the outcomes of fund-raising.

2.3. Theory and Hypotheses

Our intent is to examine how prefunding increases the likelihood of a new-venture project achieving its funding goal on a crowdfunding platform. To address our intent, we study three inter-related questions. First, we discuss whether prefunding should help or hurt the chances of a campaign meeting its goal. Second, we examine the mechanisms through which this effect manifests itself in impacting funding success. Third, and finally, we take a closer look to examine the effects of specific prefunding activities on crowdfunding campaigns.

Prefunding could have opposing effects on the fund-raising success of a campaign. On the one hand, prefunding activities should benefit fund-raising. It provides entrepreneurs with additional time to raise awareness and visibility around their new venture ideas. It also provides entrepreneurs an opportunity to refine their pitches based on feedback garnered during the prefunding phase. On the other, prefunding does induce a time gap between securing the “buy-in” from potential backers and the point in time when contribution solicitation becomes feasible. This time delay could result in the loss of potential backers. Below, arguing that prefunded projects are more likely to meet their funding goals, we offer our theoretical reasoning.

Startups present high levels of uncertainty and information asymmetry (Venkataraman 1997, Fischer and Reuber 2007, Li et al. 2017). On a crowdfunding platform, potential backers

who visit a campaign webpage, see the project pitch or narrative, which is usually comprised of text, pictures, and/or videos and is succinct by design. Upon the campaign launch, the narrative and video are the key sources of information about the new-venture idea. While the narrative does attempt to provide the important information about the product being offered, often uncertainty about the product idea in the minds of potential backers still remains.

During the prefunding phase, in addition to having access to the project pitch (i.e., video and descriptive narrative), potential backers can interact with the entrepreneur and with other like-minded backers interested in the supporting the campaign. Over time, the entrepreneur and campaign followers co-create a repository of information through such interactions. And when the project is launched (i.e., followers are allowed to contribute funds), this corpus of information is visible to all potential backers interested in funding the campaign. Information thus aggregated helps reduce campaign uncertainty. Under the condition of less information and high uncertainty, backers often hesitate to fund both, high and low caliber projects. Therefore, all projects perform sub-optimally. Evidence suggests that information aggregation about a campaign helps reduce the uncertainty regarding its quality (Griffin 2003, Loughran and McDonald 2014), which causes backers to support (at least) the high caliber projects more confidently. We, therefore, expect the net effect of information aggregation on fund-raising success to be positive.

The second effect of prefunding is that it can help entrepreneurs to engage potential backers even prior to launch. Users who participate in crowdfunding platforms do so because they like engaging in innovative behavior (Moritz et al., 2015). Backers see their role on crowdfunding platforms not only as financial resource providers but also as enablers and co-innovators. The literature on brand community suggests that consumers who are deeply involved with a brand can develop a common understanding of shared identity with the brand (Muniz and

O’Guinn 2001). Both, the sense of enabling and the sense of shared identity, are achieved through prefunding because such campaigns involve their potential backers even before launch.

Based on the two effects described above, namely the information aggregation effect and engagement effect, we hypothesize as follows:

H1: Prefunding a campaign increases the likelihood that it will meet its fund-raising target.

Prefunding campaigns differ from non-prefunding campaigns in that the information aggregation has occurred and the level of uncertainty reduced substantially prior to the launch of the campaign. Moreover, potential backers have been engaged through updates and discussions, making them vested and interested in the campaign. As a result, on the day of launch, we expect that prefunding campaigns, compared to non-prefunding ones, should have a higher number of backers, i.e., a higher level of participation. A high level of participation on the first day, regardless of the amount raised, is an observable indicator and a signal that the idea resonates with a broad cross section of the potential backers upon first exposure.

A high level of participation on a campaign’s launch day affects the subsequent course of the campaign in two ways. First, people frequently engage in “observational learning,” drawing quality inferences from observing peer choices (Zhang 2010). Second, even in the absence of learning, a high level of participation on the first day can trigger a bandwagon effect (Leibenstein 1950). The bandwagon effect refers to the tendency of individuals to demand more of a product because other individuals in the market also demand more of the product. Together, these two forces trigger a self-reinforcing mechanism that takes projects with a high participation on the first day all the way to the finish line (Kuppuswamy and Bayus 2015, Lehner 2014, Ordanini et al. 2011). Based on the above analyses, we hypothesize the following:

H2: The positive relationship between prefunding and campaign success is mediated by the initial number of backers for the project.

The funding a campaign receives comes in various sizes and patterns, and the wide range of contribution sizes a campaign garners on crowdfunding platforms is indicative of the heterogeneity of both the sophistication and the informedness of backers on crowdfunding platforms. In addition to the number of backers, prefunding should affect the pattern of contribution as well.

We argue that prefunding should lead to a higher average contribution size on the first day of a campaign. This could happen for two reasons. Prefunding makes users more informed, and more informed backers will be comfortable to contribute more. Conversely, backers who have the resources to contribute high amounts are more likely to invest time and effort acquiring information through the prefunding phase. This is consistent with findings in finance literature that large investors are usually more informed than the small ones (Bhattacharya 2001, Hirshleifer et. al 2008). Either way, projects that go through prefunding should kick off with more informed backers who invest more on average.

A higher average contribution size means informed and serious backers approve of the entrepreneur's idea. Therefore, while a large number of contributors indicates a general product appeal, a high contribution size serves as an early endorsement from informed backers, which can, subsequently, attract the less informed backers, who have a tendency to herd or mimic the informed backers (Scharfstein and Stein 1990). For this reason, we expect that prefunding projects, which launch with a high average contribution size, are more likely to attract subsequent funding required to eventually meet their fund-raising targets. Therefore, we hypothesize:

H3: The positive relationship between prefunding and campaign success is mediated by the initial contribution size for the project.

----- Insert Figure 2 about here -----

We now examine the prefunding stage closely to see what happens during prefunding can lead to the aforementioned strong performance on the first day. Specifically, we study the effects of two prefunding activities, discussions and updates, on the number of backers and contribution size upon launch.

Discussions could affect both, the participation level and contribution sizes on the day of campaign launch. First, discussions in the prefunding period represent the interactions between potential backers and the entrepreneur. They are usually initiated by potential backers, and then receive responses from others including the entrepreneur. Through such discussions, backers can seek clarification about the new venture idea and entrepreneurs often respond by providing additional information. Alternatively, backers can provide feedback and entrepreneurs can acknowledge or communicate intent to incorporate the feedback into their original ideas. Such discussions establish dialogues between the entrepreneur and potential backers before launch. Consequently, discussions will induce both the information aggregation effect as well as the engagement effect as discussed earlier. Information aggregation reduces the variance around the signal of quality of projects, while engagement heightens the sense of shared identity among potential backers. These two effects will lead to a higher-level participation from potential backers, and more informed backers, who on average contribute more. Thus, we hypothesize:

H4 (a): Campaigns are more likely to have higher initial number of backers when there is a higher level of discussion during the prefunding period.

H4 (b): Campaigns are more likely to have higher initial contribution sizes when there is a higher level of discussion during the prefunding period.

The prefunding period also allows for entrepreneurs to preemptively volunteer information through updates. Entrepreneurs can use updates to communicate the status of the project, design improvements, milestones reached and/or media coverage garnered during the time period leading up to the campaign launch. Updates can reduce project-related uncertainty by providing additional information unavailable in the entrepreneur's original pitch. While discussions on the campaign's webpage trigger both, the information and engagement effects, updates are primarily responsible for the former. Nevertheless, the updates entrepreneurs provide would still result in more participation as well as more informed backers, which leads to the following hypothesis:

H5(a): Campaigns are more likely to have higher initial number of backers when the entrepreneur provides more updates during the prefunding period.

H5(b): Campaigns are more likely to have higher initial contribution sizes when the entrepreneur provides more updates during the prefunding period.

2.4. Research Context and Data

We use data from JD Crowdfunding, a reward-based, AON platform similar to US-based crowdfunding platforms such as Kickstarter. Although inspired by Kickstarter, JD Crowdfunding is primarily a presale platform with many products ready-to-ship (Alois 2014). The platform lists projects under the following categories: technology, health, entertainment, design, household applications, journalism, and charity. The layout of project pages is also similar to Kickstarter; project pages display information such as project pitch (via description and pictures), funding goals, funds raised, percentage of funds raised and number of days left to meet the funding goal

set by entrepreneur. Entrepreneurs post project updates and engage backers and potential backers via online discussions.

JD Crowdfunding has a prefunding option built into its crowdfunding website. Entrepreneurs can choose whether to prefund their projects, and if they do, they can then seamlessly transition into the fund-raising stage within the platform. If an entrepreneur opts for prefunding, she can set up her prefunding page with a pitch, goal and perk descriptions and post updates. Potential backers can post questions or share feedback and entrepreneurs can respond. Once the campaign opens for monetary contributions all the information generated during the prefunding phase is available to all potential backers who visit the campaign webpage.

We have data of JD Crowdfunding projects for over a year starting from April 2015 to June 2016. We observe the projects from the initial launch all the way through to the end of the fund-raising period. If a project was prefunded, we observe it from the beginning of the prefunding until the end of the fund-raising period. We examined projects across all the categories highlighted earlier. After excluding outliers, e.g. projects with extremely low funding goals, we have a sample of 3,176 projects.

For each project, we collected campaign-related variables, such as the goal, duration, and category, and entrepreneur-related, such as prior experience. Table 1 lists and describes the variables in our data set. In Table 2, we provide summary statistics. Below, we describe the variables central to our analysis.

----- Insert Table 1 about here -----

----- Insert Table 2 about here -----

Success is an indicator of whether the fund-raising goal of the campaign was met. As JD Crowdfunding is an AON platform, we code *Success* as 1 if the campaign meets its fund-raising target, and 0 otherwise. It has a mean of 0.77 and the relatively high success rate on the platform is due to the presale model described earlier.

Prefunding is an indicator of whether the campaign used the prefunding option prior to fund-raising. We code it as 1 if the project went through prefunding and 0 otherwise. *Prefunding* has a mean of 0.67 and a standard deviation of 0.47.

Backers is the initial number of backers for a campaign. We measure it as the total number of people who contributed to the campaign by the end of the first day of fund-raising. *Backers* has a mean of 220.2 and a standard deviation of 777.9, suggesting that projects are highly heterogeneous in the number of backers they attract on the first day.

Amount is the average contribution amount per backer for a campaign at the end of the first day of fundraising. We also refer to it as contribution size throughout the paper. It has a mean of 204.4 Chinese Yuan (CNY), approximately \$30.5 and a standard deviation of 568.2 CNY (\$84.8), also indicating vast heterogeneity among projects in the amounts they attract on the first day.

Goal is the target funding amount and cannot be changed over the course of the campaign. The average goal for the projects in our data set is 122,984.3 CNY (\$18,355.9). The standard deviation of *Goal*_{*i*} is 197,899.9 CNY (\$29,537.3). Once again, the standard deviation is higher than the mean, indicating over dispersion.

Duration is the length of the fund-raising window in days. It is the period over which monetary contributions are possible. The mean duration of projects is 35 days, with a standard

deviation of 11 days. The most common durations are 30 days and 45 days. *Levels* represents the number of reward options offered. *Link* is a dummy variable that indicates where the campaign provides an external link with additional information.

PreDuration is the duration of the prefunding period in days. It has a mean of 2.4 and standard deviation of 2.9. *PreUpdates* is the number of updates by the entrepreneur during the prefunding stage, with a mean of 0.1412 and a standard deviation of 0.544. *PreDiscussions* is the number of discussions in the prefunding period and has a mean of 9.68 and a standard deviation of 34.46.

2.5. Empirical Analyses

2.5.1. Identification Strategy

In studying the effect of prefunding on funding success, the first econometric challenge arises due to the nature of observational data. In non-experimental settings, the estimate of a causal effect obtained by comparing treatment and comparison groups could be biased due to self-selection. Specifically, in our context, the factors that influence an entrepreneur's propensity to opt for prefunding might simultaneously affect the likelihood of the success of her projects, creating a self-selection bias.

To mitigate self-selection bias, we employ a matching technique based on propensity scores (Rosenbaum and Rubin 1983). The objective of propensity score matching is to assess the effect of a treatment by comparing observable outcomes between a sample of treated observations and a matched sample of untreated observations. We match a sample of projects that went through prefunding to a sample of projects that did not go through prefunding phase according to the propensity of being chosen for prefunding. Matching projects in this way should substantially reduce any remaining selection bias issues.

2.5.2. Propensity Score Matching

To implement the propensity matching technique, we need to identify a number of observed variables that might influence a project's propensity to be prefunded. To find the factors that are most likely to influence the decision to opt of prefunding, we ran a logistic regression with the prefunding decision as the dependent variable and several project and entrepreneur-specific characteristics as explanatory variables.

Table 3 summarizes the results of this regression. We find that the funding goal, project category, levels, pictures and links, all significantly affect the dependent variable. In addition, we include entrepreneur prior experience as an additional matching criterion to ensure that our control and treatment groups are not significantly different with respect to entrepreneur-specific observables.

----- Insert Table 3 about here -----

We then calculate the propensity score for each project, which is the probability that the project receives treatment given its covariates. In Figure 3, we plot the propensity scores for non-prefunding projects (control group) and prefunding projects (treatment group) and the distributions of propensity scores have a similar shape. We use the one-to-one nearest neighbor greedy matching without replacement approach (Austin 2009). Our matched sample consisted of 891 prefunded and 891 non-prefunded projects. We conducted our analysis using this matched sample.

----- Insert Figure 3 about here -----

Matching methods are based on the conditional independence or unconfoundedness assumption, which states that the researcher should observe all variables simultaneously influencing the participation decision and outcome variables. This is a strong identifying assumption. We cannot test the unconfoundedness assumption itself, because it would amount to testing that there are no unobserved variables that influence the selection into treatment. Therefore, we use a bounding approach proposed by Rosenbaum (2002) to estimate the degree to which our results hinge on this untestable assumption. In other words, we determine how strongly an unmeasured variable can influence the selection process in order to undermine the implications of our matching analysis. We compute the Mantel and Haenszel (MH, 1959) test statistic for binary outcomes (Aakvik 2001) following the procedure described in Becker and Caliendo (2007). Our bounds suggest that our study is insensitive to a bias that would as much as double the odds of a project being picked for prefunding by the entrepreneur. This suggests that the matching procedure is fairly robust to unobserved factors.

2.5.3. Econometric Model

Having obtained a matched sample of prefunding and non-prefunding projects, we first calculate the expected effect of prefunding on funding outcomes by estimating the average treatment effect (ATE) using a two-sample t-test with equal variances. In addition, to mitigate the bias from omitted variables, in Model 1, we estimate the effect of prefunding on success through the following logistic regression:

$$\text{logit}(\text{Success}_i) = \alpha_0 + \alpha_1 \text{Prefunding}_i + \mathbf{a}'\mathbf{X}_{1i} + \varepsilon_i, \quad (1)$$

where \mathbf{X}_{1i} is a vector of control variables including Goal, Duration, Level, and Link. The descriptions of those variables can be found in Table 2.

In the context of a treatment study, it is critical to highlight the mechanisms by which a treatment achieves its intended effect. To this end, we follow MacKinnon (2008) and conduct mediation analyses, in Model 2, using the following structural equations:

$$\begin{aligned} \text{logit}(\text{Success}_i) = & \beta_0 + \beta_1 \text{Prefunding}_i + \beta_2 \ln(\text{Backers}_i) + \beta_3 \ln(\text{Amount}_i) \\ & + \mathbf{b}' \mathbf{X}_{1i} + u_{1i}, \end{aligned} \quad (2)$$

$$\ln(\text{Backers}_i) = \gamma_0 + \gamma_1 \text{Prefunding}_i + u_{2i}, \quad (3)$$

$$\ln(\text{Amount}_i) = \delta_0 + \delta_1 \text{Prefunding}_i + u_{3i}, \quad (4)$$

where both *Backers* and *Amount* are mediators, and \mathbf{X}_{1i} is the same vector of control variables as in Equation (1).

Finally, we examine the effects of the number of updates (*PreUpdates*) and discussions (*PreDiscussions*) in the prefunding period on the initial number of backers (*Backers*) and the initial contribution size (*Amount*) in the following model:

$$y_i = \omega_0 + \omega_1 \text{PreUpdates}_i + \omega_2 \text{PreDiscussions}_i + \mathbf{w}' \mathbf{X}_{2i} + v_i, \quad (5)$$

where y_i can be either *Backers_i* or *Amount_i*, which are dependent variables, and \mathbf{X}_{2i} is a vector of control variables.

There are two econometric challenges here. First, there is potential endogeneity induced from unobservables that could simultaneously affect the outcome variables in Equation (5), and prefunding updates and discussions. Second, there is a self-selection issue as these two explanatory variables, namely, the number of updates (*PreUpdates*) and discussions (*PreDiscussions*) in the prefunding period, do not exist for non-prefunding projects.

We employ an estimation strategy that adjusts for biases from both self-selection and endogeneity following Wooldridge (2010). In this estimation approach, we need at least two

instruments that affect the prefunding selection and the endogenous covariates, *PreUpdates* and *PreDiscussions*, but not the outcomes in Equation (5).

We are able to identify the two instruments Z_1 and Z_2 . Z_1 is the recency of the prefunding option, i.e., the number of months between the introduction of the prefunding option by JD Crowdfunding (July 2014) and the launch of project i . The longer the lapse since the introduction of a feature, the higher we expect the awareness, and consequently, consumption of the feature to be. Therefore Z_1 affects the choice to opt for prefunding. Also, this increased level of awareness can lead to entrepreneurs exploiting prefunding features such as updates more, and visitors exploiting features such as the discussion board more. However, there is no apparent reason why the recency of prefunding should affect the number of backers or contribution size on the first day.

Z_2 is the percentage of projects that used prefunding in the month before project i was launched. The level of adoption of prefunding in the recent past can influence whether a project is selected for prefunding, because entrepreneurs are likely to take current trends into consideration when they make their campaign design decisions. Also, awareness of recent adoption of the prefunding option could motivate entrepreneurs and potential backers to explore and use features such as updates and discussions. But, once again, there is no rationale for why the level adoption of prefunding in the previous month should directly affect the number of backers or the amount raised on the first day of a campaign.

We briefly describe our two-step procedure here to estimate the effects of updates and discussions during the prefunding period. First, we estimate a Probit model for the selection of prefunding option. The independent variables include instruments Z_1 and Z_2 as well as other variables such as Goal, Duration, and Level. From this estimation, we calculate the Inverse Mills Ratio, which we then add to Equation (5).

Second, using only the sub-sample of projects that went through prefunding, we estimate the effect of the number of prefunding updates and discussions on the initial number of backers and initial contribution size using a 2SLS procedure, treating *PreUpdates* and *PreDiscussions* endogenous and using the inverse mills ratio, Z_1 , and Z_2 as instrumental variables.

2.5.4. Results

In order to estimate the expected effect of prefunding on funding outcomes, we construct a matched sample of prefunding projects and non-prefunding projects and first run a two-sample t-test with equal variances. Table 4 shows the results of the t-test. While non-prefunding projects have an average success rate of 67.2%, a comparable sample of prefunding projects have an average success rate of 82% ($p < 0.001$).

----- Insert Table 4 about here -----

In Model 1, we then estimate the effect of prefunding on success through a logistic regression, controlling for the funding goal, campaign duration, the number of perks, and the presence of an external link. Table 5 summarizes the results of the estimation of Model I. Once again, we find that prefunding has a positive and significant effect on success (0.148, $p < 0.001$).

In Model 2, we conduct mediation analysis to test hypotheses 2 and 3. Table 5 shows the results of our mediation analysis. We find evidence in support of both hypotheses; the relationship between prefunding and project success is mediated by both the initial number of backers as well as the initial contribution size. We know that a variable mediates the effect of a relationship, when (i) the effect of the independent variable (e.g. prefunding) on the mediator (e.g. number of backers or contribution size) is significant; (ii) the effect of mediator on the outcome (e.g. success) is significant; and (iii) the effect of the independent variable on the outcome loses significance upon the introduction of mediators. In Table 5, we can see the following results: (i) The effect of

prefunding on initial number of backers is positive and statistically significant (1.069, $p < 0.1$) and the effect of prefunding on initial contribution size is also positive and statistically significant (0.093, $p < 0.1$). (ii) The effect of initial number of backers on prefunding is positive and statistically significant (0.097, $p < 0.01$) and the effect of initial contribution size on prefunding is positive and statistically significant (0.077, $p < 0.01$). (iii) The direct effect of prefunding on success, which is positive and statistically significant (0.0148, $p < 0.01$), loses significance when the mediators are introduced (-0.029 , $p > 0.1$). The results support both Hypotheses 2 and 3 that the initial number of backers and initial contribution size serve as mediators of the relationship between prefunding and campaign success.

----- Insert Table 5 about here -----

A closer look at the trajectories of prefunding and non-prefunding projects shows that prefunding projects launch with a higher number of backers than non-prefunding projects and go on to attract even more backers, widening the gap over time. Prefunding projects kick off with more informed backers, who contribute more on average. Over time, less serious backers get involved. In contrast, non-prefunding projects start off with relatively non-serious backers. On average, prefunding projects reach the finish line with both, a higher number of backers and a higher contribution sizes on average. Figures 4 and 5 demonstrate this finding.

----- Insert Figure 4 about here -----

----- Insert Figure 5 about here -----

We estimate the effects of updates (H4) and discussions (H5) on the initial number of backers and initial contribution size. In addition to addressing the selection bias using selection criteria in a Probit selection model, we also include two instruments to address potential endogeneity

concerns. We present our results in Table 6. We find that a high level of discussion in the prefunding period has a positive and significant effect on the number of backers (1.491, $p < 0.01$), as well as on the contribution size (1.574, $p < 0.05$) on the first day of the campaign. A higher number of updates during prefunding leads to higher contribution size on average on the first day (7.899, $p < 0.026$), but does not necessarily lead to more backers (2.045, $p > 0.1$).

2.6. Discussion and Implications

6.1. Discussion

Using rich, granular data from a unique crowdfunding platform, we show that prefunding a campaign before fund-raising significantly increases the likelihood of a campaign meeting its funding goal. Prefunding affects funding outcomes by giving projects an early traction when compared to non-prefunding projects. This early traction is realized in two distinct ways: Prefunded projects launch with (i) a higher number of backers, and (ii) a pool of backers who contribute a higher amount on average.

Our results suggest that the initial number of backers is one of the mediating factors between prefunding and project success. As shown in Figure 4, prefunding projects begin with a higher number of backers than non-prefunding ones. This stronger start triggers a virtuous circle that attracts even more backers over time, a finding that is consistent with prior literature (Kuppuswamy and Bayus 2015, Lehner 2014, Colombo et. al 2015). Through the course of the campaign, the substantive gap on the first day widens over the fund-raising window with prefunding projects reaching the finish line with significantly higher number of backers on average compared to non-prefunding ones.

Not only does prefunding attract more backers on the first day, but also a distinctly different pool of backers. As shown in Figure 5, there is a significant difference between

prefunding and non-prefunding projects in terms of average contribution size on the first day of launch. The explanation is that prefunding can either attract more informed backers or make potential backers more informed, which leads to higher contribution size. After projects launch, even though the average contribution size increases for non-prefunding projects over time, it appears that the lack of initial momentum makes it difficult for non-prefunding projects to outperform prefunding campaigns on contribution size.

----- Insert Figure 5 about here -----

Figure 6 further illustrates the stark difference in first day contribution distributions between prefunding and non-prefunding campaigns. On the first day of fund-raising, non-prefunding projects are dominated by small backers, while prefunding projects have much higher contribution size. About 67% of the non-prefunding projects have an average contribution size below 50 CNY (\$7.5), and 15% between 50 – 99 CNY (\$7.5 - \$15). On the other hand, about 22% of prefunding campaigns have average contribution size of 250 CNY (\$37) or more, compared to around 9% for non-prefunding ones. Small-size backers also decrease for prefunding campaigns, with about 38% with an average contribution size below 50 CNY (\$7.5).

----- Insert Figure 6 about here -----

Over time, the distribution of contribution size has not changed much for prefunding campaigns. For example, as shown in Figure 7, the number of campaigns with contribution over 250 CNY have narrowly changed from the first day to the last day of the funding. However, contribution size increases significantly for non-prefunding campaigns, and the percentage of

campaigns with average contributions of 250 CNY or higher increased from about 9% to 17%. The results suggest that with information aggregation and engagement, backers are more comfortable to support non-prefunding campaigns with high contributions over time. Despite of this, non-prefunding projects are unable to bridge the funding gap with prefunding projects.

----- Insert Figure 7 about here -----

Further, we explored only the campaigns that employ prefunding and find that prefunding is effective when entrepreneurs keep their followers informed through updates and engaged via discussions. We find that discussions attract more backers as well as a higher contribution size, while updates only bring a higher contribution size, but do not have a significant effect on the number of backers. The results suggest that informational and engagement roles of discussions are relatively more pronounced compared to those of updates. This difference could be stemming from the difference in the nature of discussions, which are dialogues, versus updates, which are monologues. While discussions induce both the information aggregation effect as well as the engagement effect, updates are more helpful with the former and not directly responsible for the latter. Without the engagement effect, it is more difficult to generate high interests from the potential backers.

6.2. Implications

Our findings have direct implications for entrepreneurs and crowdfunding platforms. From an entrepreneur's perspective, prefunding is a valuable tool to improve the chances of fund-raising success. Entrepreneurs should view the prefunding option as more than just additional time on the platform. Prefunding can help aggregate information and mitigate uncertainty prior to fund-

raising. It can engage potential backers, helps garner feedback and serves as a “soft” launch through which entrepreneurs get to share and refine their entrepreneurial pitches. The significant role of discussions and updates in the effectiveness of prefunding shows that entrepreneurs who engage with potential backers prior to launch are much more likely to succeed than those who launch first and then attempt to engage later.

The implications from a platform perspective are twofold. First, any offering that makes entrepreneurs more likely to meet their fund-raising goals makes platform operators better off financially. Although prominent crowdfunding platforms charge a small, fixed fee to host campaigns, a majority of their revenues result from charging a percentage of the funds raised from successful campaigns. As such, platform operators are constantly searching for ways to improve the success rate of campaigns. Second, at the time of this study, JD Crowdfunding is one of the few crowdfunding portals worldwide to offer an integrated prefunding solution. Stand-alone platforms such Prefundia in the US were adopted enthusiastically in the beginning (Lomas 2013), but have not turned into the mainstream way of raising awareness before launching a crowdfunding campaign on Kickstarter or Indiegogo. This could largely be due to the disintegrated model, which does not allow for seamless flow of aggregated information from prefunding to fund-raising. It appears that integration is the key to reduction of uncertainty, which is central to the effectiveness of prefunding.

2.7. Conclusion

2.7.1. Contributions

Our contributions are three-fold. First, the literature on crowdfunding till date has explored various determinants of crowdfunding success, however, most of these determinants fall within the fundraising window. The antecedents to crowdfunding and entrepreneurial actions prior to

launch have received little attention as potential influencers of fund-raising outcomes. We examined the activities of entrepreneurs and explored the specific ways in which they can engage potential backers prior to fund-raising.

Second, we contribute to research that has focused on backers' herding behavior on crowdfunding platforms. Research has shown that early traction during the funding period is a strong predictor of crowdfunding success (e.g. Burtch et. al 2013, Kuppuswamy and Bayus 2015, Colombo et. al 2015) because support in the initial days of a campaign creates a self-reinforcing virtuous cycle. However, a less examined question is how entrepreneurs can acquire this initial support. We contribute to this stream of work by showing that engaging with potential backers through a prefunding stage can be an effective means to launch fundraising campaigns with substantially superior first day performance.

Third, we contribute to the broader literature in finance and economics on market mechanisms, such as book building in Initial Public Offerings (IPOs) (Derrien and Womack 2003). During the book building period, underwriters solicit demand and aggregate information prior to trade executions (Benveniste and Spindt 1989, Schnitalein 1996). We examine an analogous process in crowdfunding markets, and show that information aggregation and early engagement play an important role even in the less sophisticated reward-based crowdfunding context. We find that prefunding leads to more informed and engaged backers, who then contribute more to campaigns.

To the best of our knowledge, this is one of the first studies that focuses on the role prefunding plays in new venture project success. Prefunding is a recent phenomenon, and as such, has not yet been examined either as a campaign design choice or a stand-alone platform offering in current literature. Through our study, we hope to initiate a conversation on prefunding and on

the potential of pre-launch entrepreneurial actions to create and run effective fund-raising campaigns.

2.7.2. Limitations

Our study has several limitations. First, our findings are applicable to crowdfunding platforms that are currently offering or contemplating to offer prefunding as an integrated feature, but may or may not be directly extensible to stand-alone prefunding platforms. The seamless transitioning from prefunding to fund-raising plays an important role in engendering the information aggregation and engagement effects, and we cannot ascertain that the effects of prefunding described in our study will manifest in disjoint and disconnected settings.

Second, while we argue that engaging potential backers is the cornerstone to success through prefunding, we do not delve into the content or nature of updates or discussions. Updates and discussions could be used to offer information to backers proactively, provide clarification reactively or express appreciation for support. Studying the effects of different types of engagement can provide more nuanced insights with useful implications.

Finally, success could be viewed as a multi-dimensional outcome. Previous studies have operationalized crowdfunding success using outcomes such as goal attainment (Greenberg and Mollick 2017), amount raised (Calic and Mosakowski 2016), and number of backers (Ahlers et al. 2015). Since we analyze data from an AON platform, we operationalize campaign success as the meeting of the funding goal to keep our study focused. The effects of prefunding on other outcomes are worth examining and can create interesting opportunities for future research.

7.3. Concluding Remarks

As entrepreneurs compete for resources on crowdfunding platforms, prefunding has emerged as a novel option to raise visibility and awareness prior to launching in fund-raising campaign. Using

a rich data from a crowdfunding portal that offers a prefunding option, we examine the effect of prefunding on fund-raising outcomes. We find that prefunding a campaign significantly improves its likelihood of success. Prefunding prior to a campaign launch helps attract more backers and more informed backers who contribute more on average compared to non-prefunding projects. Prefunding helps entrepreneurs gain early traction when they engage potential backers through discussions and updates.

Our study has significant contributions to both, the crowdfunding literature and practitioners. To the best of our knowledge, this is one of the early studies to explore factors and strategies outside the fund-raising window that affect fund-raising outcomes. We complement the research stream that highlights the importance of early traction by studying the means to achieve this early traction. For crowdfunding practitioners, both entrepreneurs and platforms, we provide a set of actionable insights on the effects of prefunding and on how to employ it effectively to gain an edge in the competitive crowdfunding market.

Chapter 3: Human Touch in the Online World: The Role of Empathy in student conversations on MOOCS

3.1. Introduction

Massive Open Online Courses, or MOOCs, continue to attract staggering enrollments and have been altering the landscape of education over the last few years (Friedman 2013). The MOOC market is estimated to grow from USD 1.83 Billion in 2015 to USD 8.50 Billion by 2020, at an estimated compound annual growth rate (CAGR) of 36 % (Markets and Markets 2015). But, the idea of global access to education to anyone with an Internet connection is powerful even beyond these figures. MOOCs are reducing barriers to education, both geographic and socio-economic (Leber 2013). Tens of thousands of students from countries across the world join online classrooms on Coursera, Udacity, EdX and other platforms to study subjects ranging from Data Science and Statistical Inference to Greek and Roman Mythology.

While MOOCs have been credited with reducing geographical and socio-economic barriers to education, one of the biggest concerns around MOOCs has been a high level of attrition (Yan et. al 2013). Dropouts have been attributed to student disengagement (Ramesh et. al 2013). MOOCs are essentially platforms of collaborative learning, therefore their effectiveness heavily relies on student engagement. Engagement, therefore is a crucial intermediate outcome on online educational platforms. A recent quote from the office of MITx, the MOOC program at MIT, echoes the view of practitioners at large. *“Engagement is an essential part of the pedagogical experience on MOOCs. Without engagement, MOOCs might as well be (and have been compared to) the correspondence courses of the 1800s or your local public radio or TV station. It is just information transfer, not true knowledge development.”* As a result, recent conversation among

practitioners has shifted from attracting students towards keeping students engaged. Some of the early academic research in this area has also shown that increased engagement is a precursor to increased student persistence on MOOCs (Ramesh et. al 2013).

One of the salient means through which students engage is *conversation*. The role of content creators on MOOCs is usually limited to posting lecture videos. Once the content is disseminated, much of the knowledge is co-constructed by students through conversations in discussion forums. Student-student conversations are essential for building a *community of inquiry*, the hallmark of education within virtual learning environments (Garrison, Anderson and Archer 2001). Multiple works from socio-cognitive literature also show that conversation is a reflection of cognitive presence² on online learning environments.

But, social interaction is challenging even in the offline world, and much more challenging in the online context. Which personal thoughts and attitudes are communicated to others and the way they are expressed are different in digital versus traditional face-to-face settings. A significant portion of traditional face-to-face communication tends to be nonverbal (eg, body language, tone of voice), and without these cues in the online world, it is easy to miss out on the emotional and non-verbal subtext of what others are saying (Turkle 2015). Online users may also dissociate those at the other end of the communication by subconsciously viewing them merely as avatars or usernames instead of actual persons.

Such dissociation has led to reduced levels of expressed empathy in digital settings. An emerging stream of research across disciplines explores the construct of “digital empathy” (Terry and Cain 2015). In the context of education, empathy plays an especially crucial role. It strengthens the sense of community and helps foster a positive “classroom culture” that has direct effects on

² Cognitive presence is the extent to which learners are able to construct and confirm meaning through reflection and discourse in a critical community of inquiry (Garrison, Anderson and Archer 2000).

learning outcomes (Owen 2015). MOOCs and other new platforms of online education present challenges to the socio-communicative aspects of education and therefore require an understanding of digital empathy.

With this challenge in mind, in our study, we address the following questions.

- What is the role of empathy in online conversations among students?
- Does the effect of empathetic language vary based on the social role of the student in the online classroom? Does it vary based on content?
- Does being empathetic benefit or impair the underlying logical / analytical processes in an educational context?

We conduct our analysis on Khan Academy, a pioneer MOOC platform in the K-12 (Kindergarten through 12th Grade) space. In less than a year of its launch, the Forbes magazine called it one of those “*why-didn't-anyone-think-of-that*” stories that is rapidly becoming the most influential teaching organization on the planet. As of March 2017, forty million students use the platform each month. Using a rich data set of daily interactions between 47,000 students on Khan Academy over a period of forty months, our study provides insights on student conversational patterns in large-scale online educational discussion forums.

Our study leads to the following findings. First, we find that empathy in both questions and answers, encourages conversation among students. That is, questions phrased in an empathetic manner attract more answers, and questions attract more answers when the answers up to that point in time employ empathetic language. Interestingly, the role of empathy becomes more prominent as the difficulty level of the content increases. Finally, given that our study is in an educational setting, we examine the interplay of empathy and analytical thinking. We find that questions, when posed analytically or empathetically, attract more answers but these styles may

be complementary to each other. On the other hand, in answers, analytical language combined with empathetic language can have a synergistic effect and drive even more subsequent conversation than each of these elements alone. We offer reasons for these findings in the following sections.

The rest of this manuscript is organized as follows. In Section 2, we outline the literature streams that our work draws from and contributes to. In Section 3, we build our theory and hypotheses. In Section 4, we describe our research context and data, followed by sections on the econometric model and results. We conclude with a discussion on managerial implications.

3.2. Literature Review

Our work is at the intersection of three streams of literature. The first of these is the information systems research on user participation on online communities. The second is the nascent research on student behavior on MOOCs. Finally, our work builds on the emerging literature on digital empathy and is one of the first papers in Information Systems to explore engagement from this perspective.

Over the past decade, the significance of user participation under the broader umbrella of online communities has been studied extensively in the Information Systems literature. One of the early studies in this area finds that people contribute knowledge when they perceive that it enhances their reputation, when they have the experience to share, and when they are structurally embedded in the network (Wasko and Faraj 2005). Social ties, trust, norm of reciprocity, identification, shared vision and shared language are also factors that influence individuals' knowledge sharing in virtual communities (Chiu, Hsu and Wang 2006). Ma and Agarwal (2007) adopt an identity-based view and study how the use of IT-based features in online communities is associated with online knowledge contribution. Ray, Kim and Morris (2014) extend Ma and

Agarwal's (2007) framework by proposing that engagement is a central element that mediates beliefs of identity and ability, and promotes truly prosocial behavior. Another study finds that members may have psychological bonds to a particular online community based on need, affect, and/or obligation and that the form of commitment to a community affects the likelihood that a member will engage in particular behaviors (Bateman, Gray and Butler 2011). We contribute to this literature by (1) studying user participation in a novel and unique setting, a large-scale online educational community, and (2) adopting the lens of emotions to study participation, which to the best of our knowledge has not been done before.

MOOCs are a relatively recent phenomenon, so there is a small, but growing collection of studies in this area in both, information systems and computer science. Much of the extant literature on MOOCs focuses on student retention as an outcome. One of the early studies in information systems focuses on identifying the student, course, platform, and university characteristics that affect student retention (Adamopoulos 2013). Early studies in computer science also have focused on retention, or conversely, dropout rate as the outcome of interest. These studies analyze the impact of video production quality (Guo, Kim and Rubin 2014), social network positioning (Yang et al. 2013) and student attitude towards course (Wen, Yang and Rose 2014) on dropout rates.

Our work is distinct from the work on MOOCs so far in that we focus, not on persistence, but a more intermediate outcome of interest, student conversation. Student conversation levels have shown to be good predictors of student persistence on MOOCs (Ramesh et. al 2013, Carini, Kuh and Klein 2006). The theory of student involvement views student time and energy as finite institutional resources. According to the theory, the extent to which students can achieve particular developmental goals is a direct function of the time and effort they devote (Astin 1984).

How does involvement relate to persistence? If we conceive of involvement as occurring along a continuum, the act of dropping out can be viewed as the ultimate form of noninvolvement and dropping out depicts involvement at the lowest.

On Khan Academy, one of the forms of involvement is voluntary participation through conversation with other students by answering their questions. When students answer each other's questions, they are choosing to expend their physical and cognitive resources in thinking about the solution and articulating their approaches. But, student conversation as outcome has not been studied much in extant MOOC literature. Through our work, we hope to draw focus to this pivotal metric.

Finally, our work contributes to the emerging literature on digital empathy. The expression of empathy in online settings has been explored, albeit sparsely, across various disciplines (Cummings, Butler and Kraut 2000, Wolf 2000). Some of these studies predate the advent of social media and compare participants' interaction on email to that in face-to-face settings (Kraut et. al 1998). Our work is one of the first to explore digital empathy in the context of collaborative learning. It is important to adopt the lens of empathy when studying educational outcomes because of the crucial role that empathy plays in learning (Feshbach and Feshbach 2002, Cooper 2011).

3.3. Theory and Hypotheses

Empathy is an important psychological process that facilitates human communication and interaction (Xiao et. al 2016). Acquired during evolution (Preston and Waal, 2002), empathy serves to motivate sympathetic, helping, cooperative, and prosocial behaviors, and facilitates social communication. The term of empathy takes multiple interpretations. Hoffman defined it as

“an affective response more appropriate to another’s situation than one’s own” (Hoffman 2002), while Batson listed eight distinct phenomena that are all forms of empathy (Batson 2009). Despite conceptual variations, the understanding of empathy consists of three major sub processes: knowing what another person is feeling, feeling what another person is feeling, and responding compassionately to another person’s distress (Levenson and Ruef 1992).

Empathy is seen as a central phenomenon for social interaction because empathy is an important component to reach rapport between communicators and build a basis for trustworthy communication (Comfort 1984). However, empathy in online settings is challenging for a number of reasons. One of the theories proposed to explain this difference in empathy expression is the *online disinhibition effect* (Suler 2004).

The online disinhibition effect describes some subtle, but powerful underlying factors that contribute to the nature of online communication. First, the anonymity often associated with online settings allows people to take on an online identity and hide behind a pseudonym. This form of dissociative anonymity allows people to separate from their in-person identity and moral agency. Dissociating those at the other end of the communication can also stem from subconsciously viewing them as avatars and usernames instead of actual persons. Second, in order to understand the thoughts and feelings of the other in offline communication, people often observe facial expressions, and body movements. They keep eye contact, listen to the tone of the voice, and use other non-verbal cues to complement the verbal messages. It is estimated that about 90% of emotional expressions offline are conveyed non-verbally (Goleman 1995). This is a challenge for online communication as the mediating technology restricts the use of non-verbal cues.

As an increasing number of people spend time in online communities to make friends, develop relationships, exchange emotional support, and in our case receive education, it is of great importance to understand empathy in the online context; if we learn to understand what constitutes empathy in online communication, we can also find out how to nurture it and design online communities that support empathetic communication.

On online platforms of collaborative learning, social interaction plays a pivotal role. Engagement has repeatedly been shown to be a precursor to ultimate learning outcomes. But, engaging online by jumping into a conversation thread takes courage. Quite like in an offline classroom, fear and shyness are barriers to participation. In such a situation, empathy can play a very important role in cultivating a positive, relaxed and informal environment, which makes it easier for learners to voice their perspectives, thoughts and approaches. The empathetic language of peers makes students feel encouraged and supported, making them more likely to engage in a conversation. We hypothesize as follows.

H1 (a): Questions with empathetic language are more likely to attract a higher number of answers.

H1 (b): Questions to which previous answers contain empathetic language are more likely to attract a higher number of subsequent answers.

The barriers of shyness and lack of confidence increase manifold as the educational content gets more difficult and challenging. Intuitively, a student would require much more push to chime in on a discussion on integral calculus than she would on a discussion on linear algebra. The empathy of peers becomes even more critical in this situation to provide that push to go that extra mile. We find interesting analogies in different contexts. In an ethnographic research at Mount Everest in Nepal, Gulnar Tumbat (2008) studies the extreme context of high-altitude

mountaineering expeditions. Her study finds that the importance of emotional support is the highest when the expeditioners scale higher altitudes that are closest to the summit. Although in a very different setting, this allows us to make an important observation in our own context. We expect that the role of empathy becomes much more important as the difficulty of the subject matter increases.

H2: The effect of empathy on participation is higher for content of higher difficulty.

Finally, because the context is education, we expect the ability to demonstrate analytical thinking to also play an important role in driving participation. When both analytical thinking and empathy are displayed in a question or in its answers, we expect it to lead to synergies that drive even more participation.

H3: (a) Questions that display high empathy as well as analytical thinking attract a higher level of participation.

H3 (b) Questions to which previous answers display high empathy as well as analytical thinking attract a higher level of participation.

3.4. Research Context and Data

The source of our data is Khan Academy (www.khanacademy.org), one of the pioneer MOOC platforms. In this section, we describe our research context and details of the student-interaction data we collected. We then share some descriptive statistics before we discuss our model.

3.4.1. Research Context

At the time of our research, Khan Academy primarily catered to K-12 (kindergarten through twelfth grade) students via thousands of free micro-lectures in fifteen subjects such as Math, Science, History, Finance, Economics, etc. Micro-Lectures are embedded YouTube videos that explain a concept typically in less than ten minutes. After watching a video, students can post

questions in the discussion forum under each video. They can also view questions posted by other students and provide answers to these questions. The website also provides the option to up-vote or down-vote others questions and answers. The exchanges between students in these forums provide a symbiotic learning ground where much of the learning on MOOCs takes place. Figure 1 shows a sample conversation between students that one is likely to encounter under each video. It is this student conversation via questions and answers that we wish to understand better, and with this end in mind, we collect a rich data set of daily student interactions over a period of three years on Khan Academy.

We collected student interaction data at the highest level of granularity since the launch of the Khan Academy website in July, 2010 until October 11th, 2013. This amounted to exchange data between 106,145 students over a period of 40 months under the subjects: Math, Science, Economics and Finance. On October 11th, 2013, we collected the data using a Python script that sifted through the webpage of each lecture and scrapped the historic conversation up to that point including time stamps. From this data, we used the time stamps of videos, questions and answers to construct a daily level panel dataset.

On October 22nd, 2013, the Math section of Khan Academy was re-organized to map its existing videos to the US Grade system. Our data collection cut-off date is two weeks before this change was implemented, so we avoid any confounding effects that such content rearrangement might have on student behavior. Additionally, we dropped the data from the first three months in order to side-step the conversation around the stabilization period following the launch of a new platform.

Further, we decided to restrict our attention to the Math section and used only the Math subset of the data. The reasons for this were twofold. Firstly, this limits the heterogeneity of subject matter and allows our empirical model to focus on the effects of interest without having to control for this explicitly. Secondly, the Math section was Khan Academy's oldest, most complete and stable branch at the time of data collection. It was well-organized in a manner that a student could begin from the topic of Pre-Arithmetic and build her way through all the way to advanced topics such as Differential Equations. As our model will reflect, this allows us to control for the difficulty of the subject matter when we analyze engagement.

Our resulting data set consisted of 2400 math lectures with 93,927 questions and 127,488 answers among 47,000 students. We collected lecture-specific, question-specific, answer-specific and user-specific data from the website. We collected the upload date and the length of each video. Under each video, for every question asked, we scrapped the textual content, time stamp and student ID of the questioner. Likewise, for every question, we collected the textual content, timestamp and student ID of every answer given.

3.4.2. Data

We now describe our dependent, independent and control variables. Some of these were readily available in our data set, while others were constructed through text mining techniques, which we explain in detail.

4.2.1. Dependent Variable

As stated above, for every question asked on Khan Academy, we collected all the answers given. We define our dependent variable, $conversation_{it}$ as the number of answers attracted by each question, i at time t .

Conversation_{it}: Total number of answers attracted by question i in period t .

The length of our time period t , is one day. We found that 60% of conversations end in a week, 85% in two weeks and 95% in three weeks. Figure 2 illustrates this attenuation. So, we sampled the conversation around each question for thirty days from the date the question was posted to ensure that we captured nearly all of the dialogue. We deliberated on other choices for the time period length, such as one week, but chose one day. Disaggregate temporal analysis, whenever possible, should be leveraged to get deeper insights into dynamics and to avoid biased estimates (Tellis and Franses 2006).

The total number of answers a question receives varies substantially. Figure 3 shows the cumulative distribution function (CDF) of our dependent variable, excluding the top 0.1% outliers. The CDF shows that not all questions achieve the engagement they hope to evoke on the platform. In fact, 70% of the questions do not get any answers, and a very small percentage of questions get more than five answers. The engaging of peers seems to be a goal few students are able to achieve, which makes engagement an interesting and worthwhile dependent variable to study.

4.2.2. Independent Variables

Our interest is in the effect of empathy on engagement. But, empathy is in part an internal mental process, which is difficult to gauge directly by observation. Measurement of it relies on human perception and subjective assessment. However, spoken language encodes a multitude of information including a speaker's intent, emotions, desires as well as other physical, cognitive and mental state and traits (e.g., speaker age and gender). By analyzing the language transcripts of interactions we may infer the empathy processes that are driving, and reflected in, the language expressions. To this end, we used the Linguistic Inquiry and Word Count (LIWC)

program developed by Pennebaker et al. in 1993. We used the LIWC 2015 dictionary, which was the most up-to-date version of the program at the time of our analysis.

The LIWC program uses psychometrically pretested dictionaries to perform counts of words that represent a concept. The 2015 master dictionary is composed of almost 6,400 words and word stems. Each word or word stem is associated with one or more word categories. For example, the word *cried* is part of five word categories: Sadness, Negative Emotion, Overall Affect, Verb, and Past Focus. Hence, if the word *cried* was found in the target text, each of these five categories' scores would be incremented. The categories are designed such that they tap the basic emotional and cognitive dimensions often studied in social sciences, health and personality psychology.

The LIWC dictionaries offer strong, reliable convergence between the dimensions they extract and manual coding performed by human coders (Pennebaker and Francis 1996). It is, therefore, not surprising that the use of LIWC is increasing in recent years across disciplines (e.g., Bantum and Owen 2009; Pennebaker and Stone 2003), and specifically within Information Systems (Yin, Bond and Zhang 2014) and Marketing (Ludwig et. al 2013, Moore 2014, cite more here).

To construct a measure of empathy, we use the LIWC dimensions, second person pronouns and third person pronouns. We operationalize Empathy as $\text{Empathy} = (\text{Second Person Pronoun Use} - \text{Third Person Pronoun Use})$. Empathy is the speaker's identification, with varying degrees (ranging from degree 0 to 1), with a person who participates in the event that he describes in a sentence. Studies in linguistics show that the effect of using a second person pronoun is often the creation of empathy (cf. O'Connor, 1994, Stirling and Manderson, 2011). When such language is employed, the addressee is presupposed to empathize with the (set of)

protagonist(s) about which some statement is made and to imagine herself in the relevant situation (Rubenstein, 2010). The use of third person pronouns, on the other hand, reflect distancing and isolation and an inability to take the perspective of the addressee (Kuno and Kburaki 1997).

To capture the extent of analytical and logical thinking displayed, we use the *cognitive processing* dimension that measures and captures several sub-processes such as insight, causation, discrepancy, tentative, certainty, and differentiation (reflected by words such as cause, know, ought, think, know, because, effect, should, would, maybe, perhaps, always, never, but, else).

Below, we describe and explain our independent variables. We first describe the question-specific variables, which were derived by running LIWC on the content of the questions. These are time- invariant variables, as the content of a question does not change over time.

(1) *QuesEmpathy_i*: Level of empathetic language in question *i*.

(2) *QuesCognitive_i*: Level of cognitive processing reflected in question *i*.

Corresponding to each of the variables above, we have answer-level independent variables. A question might receive more than one answer in a period, so these variables are averages. These are time-variant variables, as the answers in each period are different.

(3) *AnsEmpathy_{it}*: Average level of empathy reflected in answers to question *i*

(4) *AnsCognitive_{it}*: Average level of cognitive processing reflected in in answers to question *i*

until and including period *t*.

4.2.3. Control Variables

We now describe some additional question-specific, video-specific and student-specific variables we collected. As we show later in the model section, we use these as control variables in our estimation.

(5) *QuesLength_i*: Number of characters in question *i*. The length of a question is a proxy for the effort invested by the questioner in framing, articulating and posting it. One might argue that longer questions, because they signal effort, attract more answers. To rule this effect out, we control for the length of the question.

(6) *QuestionerProfile_i*: Profile type (private or public) of the student asking the question, *i*. Khan Academy allows students to choose between one of two profile types: Public and Private. Enabling the public profile allows other students to view ones progress on the website, badges earned, questions asked and answers given. The Private profile options masks nearly all of this information and allows access to a much more restricted subset of the student history. It is possible that students' behavior towards an anonymous student is systematically different from that towards a student whose credentials and history are visible. To account for this, we control for the questioner's profile type in our model.

(7) *LectureLength_i*: Length (in seconds) of the lecture under which question *i* was asked. While most lectures on the platform are under ten minutes long, it is possible that students' inclination to engage in conversation depends on the length of the conversation.

(8) *LectureDifficulty_i*: Difficulty level of the topic under which question *i* was asked. Discussions on arithmetic could be systematically different from those around calculus. In order to control for this, we coded the difficulty level of each video on a scale of 0-8 based

on Khan Academy’s roadmap for its Math content. For instance, a lecture on Pre-Arithmetic is coded as 0, while one on Differential Equations is coded as 8.

For ease of reference, in Table I, we list and describe our variables concisely. In Table II, we provide summary statistics.

3.5. Model

In this section, we develop a dynamic model of student conversation on MOOCs. Our model is dynamic because we account for the fact that our y variable, namely, conversation in the current period, depends on its lagged values, i.e., conversation in the previous periods. Such state-dependence is important to consider while studying outcomes that build on an “accumulated stock” over time. Recent examples of such phenomena studied using dynamic models include mobile-user content consumption (Ghose and Han 2011), YouTube video views (Yoganarasimhan 2012), and firm sales (Sonier, McAlister and Rutz 2011), to name a few.

Let $y_{it} = \ln(\text{Conversation}_{it})$. We model y_{it} as follows:

$$y_{it} = c + \alpha y_{it-1} + \beta X_i + \gamma Z_{it-1} + \delta_i + \varepsilon_{it} \quad (1)$$

The composition of each of these terms is as follows:

- y_{it-1} is the lag of the dependent variable, i.e., $\ln(\text{Conversation}_{it-1})$.
- X_i is a vector of question-specific time-invariant covariates. It includes our explanatory variables, QuesEmpathy_i and QuesCognitive_i . In addition, it also has the control variables, QuesLength_i , $\text{QuestionerProfile}_i$, LectureLength_i and $\text{LectureDifficulty}_i$.

- Z_{it-1} is a vector of lagged time-variant covariates, namely $AnsEmpathy_{it-1}$, $AnsCognitive_{it-1}$.
- δ_i is a time-variant unobserved fixed effect. It captures other inherent characteristics of the question, lecture or the topic that might potentially affect the number of answers it attracts.
- ε_{it} captures random shocks that the question might experience in a period.

In other words, we model conversation as a function of three components: (1) Emotions in the question (captured in X_i), (2) Prior conversation level (captured in y_{it-1}), and (3) Emotions in prior conversation (captured in Z_{it-1}). Figure 4 illustrates our framework.

3.5.1. Assumptions

We make the following assumptions regarding the model:

1. $E(\varepsilon_{it}) = 0$, $E(\delta_i) = 0$, and $E(\varepsilon_{it}, \delta_i) = 0 \forall i, t$

ε_{it} and δ_i are mean-zero and uncorrelated for all i and t .

2. $E(\varepsilon_{it}, \varepsilon_{iu}) = 0$, if $t \neq u$, and σ_i^2 if $t = u$

ε_{it} s are allowed to be heteroskedastic across questions, but assumed to be serially uncorrelated. This assumption is required for reasons that we explain in our estimation strategy. We test the validity of this assumption using the Arellano-Bond (2) test.

3. $E(\delta_i, \delta_j) = 0$, if $i \neq j$, and σ_δ^2 if $i = j$

The unobserved fixed effect, δ_i is assumed to be an independent draw for each question, that is, it is not correlated across questions.

4. $E(Z_{it}, \varepsilon_{iu}) = 0$, if $u > t$, but $E(Z_{it}, \varepsilon_{iu}) \neq 0$ if $u \leq t$

We allow for correlation between the error-term, ε_{it} , and both future and current Z_{it} s. For instance, a positive shock to the number of answers in period t could, in some situations, mean that question i will receive more emotionally charged answers in and after period t . Hence, we cannot assume Z_{it} to be strictly exogenous. So we impose the weaker restriction that Z_{it} is exogenous only to future shocks. This is non-restrictive. We are saying that future shocks to the number of answers do not impact the emotion level of conversation that has already transpired.

5. $E(Z_{it}, \delta_i) \neq 0 \forall i, t$

Z_{it} s may also be correlated with δ_i s because the unobserved attributes that affect engagement may also affect the time varying covariates, that is, the emotions in engagement.

6. $E(X_i, \delta_i) \neq 0 \forall i$

We also allow for correlation between the time-invariant network properties X_i and δ_i because unobserved attributes of the question that affect engagement might be related to even stem from the observed time-invariant properties of the question.

6. The realizations of y_{it} in the initial periods are centered around its long-term mean and the deviations from the mean are uncorrelated to the mean itself. This is similar to the Initial Conditions assumption in Blundell and Bond (1998). This is a reasonable assumption in most settings, including ours, because it is essentially a form of stationarity assumption on the initial conditions. We clarify the need for this assumption in section 4.3.

3.5.2. Identification Challenges

In estimating the model above, we need to address several econometric challenges. Specifically, endogeneity manifests in three different ways in our model. First, the unobserved question-level fixed attributes (δ_i) that affect engagement (y_{it}) may also affect the time-varying covariates (Z_{it}). For example, a controversial question might receive a large number of answers, but it is also likely to receive more emotional answers. Second, time-variant question characteristics are correlated with shocks to past and current engagement levels. That is, the error-term, ε_{it} , may be correlated to and both future and current Z_{it-1} . For instance, a positive shock to the number of answers received in period t could also increase the probability of question i getting more emotional answers in that period and afterwards. Finally, even observed time-invariant question properties (X_i) can be correlated with the unobserved question fixed effect (δ_i). For instance, a question that is long (a fixed property that we observe) may be correlated to the quality of the question itself (a fixed property that we do not observe).

3.5.3. Estimation Strategy

Our strategy needs to be able to handle all three types of endogeneity listed above. Standard VAR-based estimation strategies require error terms to be uncorrelated with all the explanatory variables. Random-effects estimation and fixed-effects estimation cannot be used in a dynamic setting. External instruments for the endogenous variables would be ideal, but strong and valid instruments are challenging to find. Therefore, we use a GMM style estimator of dynamic panel data models that exploit the lags and lagged differences of explanatory variables as instruments. Anderson and Hsaio (1981) showed that in the absence of serial correlation in error-terms (Assumption 2), lags of explanatory variables

can be used to instrument for the endogenous explanatory variables in first-differenced equations of interest. Arellano and Bond (1991) provided a specification test, the Arellano-Bond test for serial correlation, to check the validity of this assumption of serially uncorrelated errors. We confirm the absence of serial correlation using this test.

While the Anderson and Hsaio approach allows us to address all three types of endogeneity, first-differencing all time invariant covariates. But, we are interested in recovering these coefficients. To address such requirements, Blundell and Bond (1998) proposed a system GMM approach that uses both first-differenced and level equations. Using this method, by imposing mild assumptions the initial deviations of the dependent variable are independent of its long-term average (Assumption 7), we address all three forms of endogeneity, as well as recover coefficients of time-invariant covariates.

3.5.4. Model Specifications

Adhering to the generic specification in Equation (1) above, we developed three variants of the model to address our research questions. In both models, y_{it} (dependent variable, *conversation_{it}*), δ_i (time-invariant fixed effect) and ε_{it} (random shocks) remain the same; the compositions of X_i and Z_{it-1} vary to address two different questions.

Model I is our baseline model, in which we focus on the effects of empathy on conversation. In this model, X_i contains *QuestionEmpathy_i* and control variables, while Z_{it-1} comprises of the *AnsEmpathy_{it-1}* (average empathy in the answers so far), and control variables.

In Model II, we extend the baseline model by interacting empathy with content difficulty. In this model, X_i constitutes *QuestionEmpathy_i*, *Difficulty_i* and *QuestionEmpathy_iDifficulty_i*,

and control variables. Z_{it-1} , in this model, is comprised of $AnsEmpathy_{it-1}$, $AnswerEmpathy_{it-1}XDifficulty_i$ and control variables.

In Model III, we extend the model further by also interacting empathy with cognitive processing. In this model, X_i constitutes $QuestionEmpathy_i$, $Difficulty_i$, $QuestionCognitive_i$, $QuestionEmpathy_iXDifficulty_i$, $QuestionEmpathy_iXQuestionCognitive_i$ and control variables. Z_{it-1} , in this model, is comprised of $AnsEmpathy_{it-1}$, $AnswerEmpathy_{it-1}XDifficulty_i$, $AnswerEmpathy_{it-1}XAnswerCognitive_{it-1}$ and control variables.

In all models, we control for the length of the question, the cognitive processing reflected in the question, the profile type of the questioner, the length of the lecture and the difficulty level of the topic of the lecture.

3.5.5. Results

Table 2 summarizes the results of the estimation of the two models described above.

In our first model, we find that the expression of empathy in both questions and answers makes it more likely for students to engage in a conversation. The coefficient for $QuestionEmpathy_i$ is positive and highly significant (0.089, $p < 0.01$), suggesting that a student posing a question with a positive tone is more likely to attract more answers. The coefficient for $AnswerEmpathy_{it-1}$ is also positive and highly significant (0.002, $p < 0.01$), suggesting that questions to which the answers up to a point are empathetic are also likely to attract more answers after that point.

In the second model, we extend the model to study how empathy interacts with content difficulty. Here, we find an interesting result. The role of empathy in attracting engagement becomes increasingly substantial as the subject matter becomes more difficult. This is true of both questions and answers. The coefficient for $QuestionEmpathy_iXDifficulty_i$ is positive

and highly significant (0.033, $p < 0.01$), and the coefficient for $AnswerEmpathy_{i-1} \times Difficulty_i$ is also positive and highly significant (0.0001, $p < 0.01$).

In the third model, we extend the model further by also interacting empathy with cognitive processing. We find here that while the interaction of empathy and cognitive processing in answers is positive, the interaction of empathy and cognitive processing in questions is negative.

In estimating all models, we control for several variables, and we briefly discuss the effects of the control variables here for completeness. Longer questions get more answers (0.003, $p < 0.01$) and interestingly, students with private profiles get more answers (-5.456, $p < 0.01$). Questions on longer lectures spur more conversation (0.009, $p < 0.01$), and those on advanced topics spur less conversation (-2.09, $p < 0.01$).

Finally, we see that the level of conversation exhibits substantial state dependence. The lag of conversation is influential in determining current conversation level (0.904, $p < 0.01$). Accounting for state dependence in our model ensures that the estimated effects of emotions on conversation are truly salient.

In the next section, we discuss our main results and provide interpretations. We then share implications for MOOC students, instructors and providers.

3.6. Discussion and Conclusion

Using daily student interaction from Khan Academy, we conduct one of the first Information Systems studies on student behavior on MOOCs. We adopt the lens of digital empathy and study whether empathy in online classrooms affect student conversation levels. We show that empathetic expression plays a crucial role in increasing student engagement.

Perhaps even more interestingly, we find that the effect of empathy becomes more pronounced as the difficulty of the content increases.

Our results have useful implications for MOOC providers, instructors and students. Firstly, our study paves way for a deeper understanding of online learning environments. The substantial role of empathy suggests that MOOCs are not just websites with video content and mechanical activity by students, but indeed platforms characterized by social interaction, wherein the human touch not only matters, but also facilitates engagement.

Secondly, given this new understanding, MOOC providers should design their platforms in ways that enable and encourage expressiveness and discourage shying away or the masking of emotions. A related motivational example would be the recent introduction of reactions by Facebook that allows for more ways to easily and quickly express how something we sees makes us feel in an online environment.

Our study reaffirms that MOOCs are more than about moving the physical classroom online. They have the potential to fundamentally transform the way education is imparted. This innovation has a unique benefit from the perspective of researchers and practitioners; data on student-behavior is now available like never before. This, combined with the lexical, machine learning and statistical tools now at our disposal, allows us to investigate the nuances of online learning at a scale that was not possible before. Findings from such studies will feed back into the innovative process around MOOCs and can lead to even better models of engaged learning.

Our work has several limitations. First, in this study, we limit the scope of the outcome of interest to the level of conversation and do not delve into the nature of conversation. Second, we work with the premise that engagement, is overall, desirable on online communities, especially,

in the context of education. But, this may not always be the case, and we defer making any comments about the relationship between student conversation and learning outcomes. Finally, Khan Academy is an asynchronous MOOC for K-12 students. While our findings are likely to be applicable to large scale learning platforms in general, we are unable to claim that these results are necessarily replicable on data from MOOCs of other breeds, such as synchronous MOOCs or MOOCs catered for university-level courses. These limitations provide avenues for further research in this emerging area and can help shed light on how information systems can play a vital role in transforming the way education is imparted in an increasingly interconnected world.

Chapter 4: Conclusion

In this dissertation, I explored two types of online collaborative communities. In the first two essays, I studied collaborative fundraising on a crowdfunding platform. In the third essay, I study collaborative learning.

The first study on crowdfunding has significant contributions to both, the crowdfunding literature and practitioners. To the best of our knowledge, this is one of the early studies to explore factors and strategies outside the fund-raising window that affect fund-raising outcomes. We complement the research stream that highlights the importance of early traction by studying the means to achieve this early traction. For crowdfunding practitioners, both entrepreneurs and platforms, we provide a set of actionable insights on the effects of prefunding and on how to employ it effectively to gain an edge in the competitive crowdfunding market.

In the second study, using daily student interaction from Khan Academy, we conduct one of the first studies in Information Systems on student behavior on MOOCs. We adopt the lens of digital empathy and examine whether empathy in online classrooms affect student conversation levels. We show that empathetic expression plays a crucial role in increasing student engagement. Perhaps even more interestingly, we find that the effect of empathy becomes more pronounced as the difficulty of the content increases. These results have useful implications for MOOC providers, instructors and students. Our study paves way for a deeper understanding of online learning environments. Findings from such studies will feed back into the innovative process around MOOCs and can lead to even better models of engaged learning.

BIBLIOGRAPHY

References for Chapter 1: Introduction

Beinhocker, Eric D. The origin of wealth: Evolution, complexity, and the radical remaking of economics. *Harvard Business Press*, 2006.

References for Chapter 2:

Launch on a High Note: How Prefunding Affects Crowdfunding Outcomes

- Aakvik A (2001) Bounding a matching estimator: the case of a Norwegian training program. *Oxford Bulletin of Economics and Statistics* 63(1): 115-143.
- Ahlers GKC, Cumming D, Gunther C, Schweizer D (2015) Signaling in equity crowdfunding. *Entrepreneurship Theory and Practice* 39(4): 955-980.
- Alois JD (2014) Chinese e-commerce giant, Jingdong launches crowdfunding platform. *Crowdfund Insider*. Available: <https://www.crowdfundinsider.com/2014/07/43214-chinese-e-commerce-giant-jingdong-launches-crowdfunding-platform/>
- Austin PC (2009) Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulations. *Biometrical Journal* 51(1): 171-184.
- Benveniste LM, Spindt PA (1989) How investment bankers determine the offer price and allocation of new issues. *Journal of financial Economics* 24(2): 343-361.
- Burtch G, Ghose A, Wattal S (1989) An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research* 24(3): 499-519.
- Burtch G, Anindya G, Wattal S (2015) The hidden cost of accommodating crowdfunder privacy preferences: a randomized field experiment. *Management Science* 61(5): 949-962.
- Bhattacharya N (2001) Investors' trade size and trading responses around earnings announcements: An empirical investigation. *The Accounting Review* 76(2): 221-244.
- Calic, G, Elaine, M (2016) Kicking off social entrepreneurship: how a sustainability orientation influences crowdfunding success. *Journal of Management Studies* 53(5): 738-767.
- Colombo MG, Franzoni C, Rossi-Lamastra C (2015) Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice* 39(1): 75-100.
- Cumming DJ, Leboeuf G, Schwienbacher A (2014) Crowdfunding models: Keep-it-all vs. all-or-nothing. *Paris December 2014 Finance Meeting EUROFIDAI-AFFI paper*. Vol. 10.
- Derrien F, Womack KL (2003) Auctions vs. bookbuilding and the control of underpricing in hot IPO markets. *Review of Financial Studies* 16(1): 31-61.
- Fischer, E, Reuber, R (2007) The good, the bad, and the unfamiliar: The challenges of reputation formation facing new firms. *Entrepreneurship Theory and Practice* 31(1): 53-75.

- Frydrych D, Block AJ, Kinder T, Koeck B (2014) Exploring entrepreneurial legitimacy in reward-based crowdfunding. *Venture Capital* 16(3): 247-269.
- Greenberg J, Mollick E (2017) Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly* 62(2): 341-374.
- Griffin PA (2003) Got information? Investor response to form 10-K and form 10-Q EDGAR filings. *Review of Accounting Studies* 8(4): 433-460.
- Hirshleifer DA, Myers JN, Myers LA, Teoh SH (2008) Do individual investors cause post-earnings announcement drift? Direct evidence from personal trades. *The Accounting Review* 83(6): 1521-1550.
- Kuppuswamy V, Bayus BL (2015) Crowdfunding creative ideas: The dynamics of project backers in kickstarter. *Social Science Research Network Scholarly Paper* ID 2234765.
- Leibenstein H (1950) Bandwagon, snob, and Veblen effects in the theory of consumers' demand. *The Quarterly Journal of Economics* 64(2): 183-207.
- Lehner OM, Nicholls A (2014) Social finance and crowdfunding for social enterprises: A public-private case study providing legitimacy and leverage. *Venture Capital* 16(3): 271-286.
- Li J, Chen X, Kotha S, Fisher G (2017) Catching fire and spreading it: A glimpse into displayed entrepreneurial passion in crowdfunding campaigns. *Journal of Applied Psychology*. Forthcoming.
- Lin M, and Viswanathan S (2015) Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* 62(5): 1393-1414.
- Lin M, Sias R and Wei Z (2015) "Smart Money": Institutional investors in online crowdfunding. *Working Paper*.
- Lomas N (2013) Prefundia, a platform for crowdfunding projects to gain backers ahead of launch, exits beta. *TechCrunch*. Available: <https://techcrunch.com/2013/11/09/prefundia/>
- Loughran T, McDonald B (2014) Measuring readability in financial disclosures. *The Journal of Finance* 69(4): 1643-1671.
- MacKinnon DP (2008) Introduction to statistical mediation analysis. New York: Erlbaum.
- McLaughlin B (2016) How Asia is adopting crowdfunding from the West. *Forbes*. Available: <https://www.forbes.com/sites/baymclaughlin/2016/12/01/how-asia-is-adopting-crowdfunding-from-the-west>
- Mollick E (2014) The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing* 29(1): 1-16.

- Moritz A, Block J, Lutz E (2015) Investor communication in equity-based crowdfunding: a qualitative-empirical study. *Qualitative Research in Financial Markets* 7(3): 309-342.
- Muniz AM, and O'guinn TC (2001) Brand community. *Journal of Consumer Research* 27(4): 412-432.
- Ordanini A, Miceli L, Pizzetti M, and Parasuraman A (2011) Crowd-funding: Transforming customers into investors through innovative service platforms. *Journal of Service Management* 22(4): 443-470.
- O'Hara M (2000) *Market Microstructure Theory*. Malden, MA: Blackwell.
- Richardson V (2014) This site lets you get backers for your crowdfunding project before it launches. *Entrepreneur*. Available: <https://www.entrepreneur.com/article/231400#>
- Rosenbaum, PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41-55.
- Rosenbaum, PR (2002) *Observational Studies*. New York: Springer.
- Scharfstein DS, Stein JC (1990) Herd behavior and investment. *American Economic Review* 80(3): 465-479.
- Skirnevskiy V, Bendig D, Brettel M (2017) The influence of internal social capital on serial creators' success in crowdfunding. *Entrepreneurship Theory and Practice* 41(2): 209-236.
- Schnitzlein CR (1996) Call and Continuous trading mechanisms under asymmetric information: An experiment investigation *The Journal of Finance* 51(2): 613-636.
- Venkataraman S (1997) The distinctive domain of entrepreneurship research. In *Advances in Entrepreneurship, Firm Emergence and Growth*. J Katz, R Brockhaus, eds. Greenwich, CT: JAI Press.
- Viotto J (2015) Competition and regulation of crowdfunding platforms: a two-sided market approach. *Communications & Strategies* 99(3): 33-50.
- Zhang J (2010) The sound of silence: Observational learning in the US kidney market. *Marketing Science* 29(2): 315-335.

References for Chapter 3:

Human Touch in the Online World: The Role of Empathy in Student Conversation on MOOCs

- Adamopoulos, Panagiotis. "What makes a great MOOC? An interdisciplinary analysis of student retention in online courses." (2013).
- Bateman, Patrick J., Peter H. Gray, and Brian S. Butler. "Research note—the impact of community commitment on participation in online communities." *Information Systems Research* 22.4 (2011): 841-854.
- Berger, J. and Milkman, K. L. 2012. What Makes Online Content Viral?, *Journal of Marketing Research*, 49(2), pp. 192-205.
- Byrne, J. A. 2015. Whats Behind Whartons Massive Bet on Online Learning, *Fortune*, 2015.
- Carini, Robert M., George D. Kuh, and Stephen P. Klein. "Student engagement and student learning: Testing the linkages." *Research in higher education* 47.1 (2006): 1-32.
- Chiu, Chao-Min, Meng-Hsiang Hsu, and Eric TG Wang. "Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories." *Decision support systems* 42.3 (2006): 1872-1888.
- Comfort, Alex. *Reality and empathy. Physics, mind, and science in the 21st century.* SUNY Press, 1984.
- Lauren Owen, 2015. "Empathy in the classroom: Why should I care?" *Edutopia*. November 2015
- Friedman, T. 2013. *Revolution Hits the Universities.* *The New York Times*, Jan, 26, 2013.
- Garrison, D. Randy, Terry Anderson, and Walter Archer. "Critical thinking, cognitive presence, and computer conferencing in distance education." *American Journal of distance education* 15.1 (2001): 7-23.
- Harden, N. 2012. *The end of the university as we know it,* *The American Interest*, 8(3). Heller, N. 2013. *Is college moving online?* *The Newyorker*, 2013.
- Leber, J. 2013. *In the Developing World, MOOCs Start to Get Real,* *MIT Technology Review.*
- Ma, M. and Agarwal, R. 2007. *Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities,* *Information Systems Research*, 18(1), pp. 42-67.
- MarketsAndMarkets. 2015. *Massive Open Online Course Market by Platform,* Report Code: TC 3945.

- Pennebaker, J. W. and Francis, M. E. 1996. Cognitive, Emotional, and Language Processes in Disclosure, *Cognition and Emotion*, 10(6), pp. 601-626.
- Pennebaker, J. W. and Stone, L. D. 2003. Words of Wisdom: Language Use Over the Life Span, *Journal of Personality and Social Psychology*, 85(2), pp. 291-301.
- Rabe-Hesketh, S. and Skrondal, A. 2012. *Multilevel and Longitudinal Modeling Using Stata*, 1.
- Ramesh, Arti, et al. "Modeling learner engagement in MOOCs using probabilistic soft logic." *NIPS Workshop on Data Driven Education*. Vol. 21. 2013.
- Ray, S., Kim, S. S., and Morris, J. G. 2014. The Central Role of Engagement in Online Communities, *Information Systems Research*, 25(3), pp. 528-546.
- Russell, J. A. 1980. A Circumplex Model of Affect, *Journal of Personality and Social Psychology* 39(6), pp. 1161-1178.
- Terry, Christopher, and Jeff Cain. "The emerging issue of digital empathy." *American Journal of Pharmaceutical Education* 80.4 (2016): 58.
- Tumbat, Gülnur, and Russell W. Belk. "Marketplace tensions in extraordinary experiences." *Journal of Consumer Research* 38.1 (2010): 42-61.
- Turkle, Sherry. *Reclaiming conversation: The power of talk in a digital age*. Penguin, 2016.
- Wasko, Molly McLure, and Samer Faraj. "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice." *MIS quarterly* (2005): 35-57.
- Xiao, Bo, et al. "Computational analysis and simulation of empathic behaviors: A survey of empathy modeling with behavioral signal processing framework." *Current psychiatry reports* 18.5 (2016): 49.
- Yang, Diyi, et al. 2013 "Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses." *Proceedings of the 2013 NIPS Data-driven education workshop*. Vol. 11. 2013.
- Yoganarasimhan, Hema. "Impact of social network structure on content propagation: A study using YouTube data." *Quantitative Marketing and Economics* 10.1 (2012): 111-150.

TABLES

Table 2.1 Variable Descriptions

Variable	Description
Prefunding	Binary variable that indicates whether a campaign underwent a prefunding phase. It is coded as 1 if the project went through prefunding, and 0 otherwise.
Success	Binary variable that indicates whether a campaign was successful in raising its target amount. It is coded as 1 if the target is met, and 0 otherwise.
Backers	Total number of backers at the end of the first day of fund-raising. We have a panel data set of number of backers at the end of every day. In our analyses, we use the number of backers on the first day.
Amount	Average contribution amount (in CNY) raised on the first day of fund-raising for a campaign. We also refer to it as contribution size.
Goal	Target funding amount for the campaign in CNY
Duration	Number of days of fund-raising period
Category	Binary variable that indicates whether a project is a technology project. It is coded as 1 if the project is a technology project (JD Categories: Appliances, Technology, Design) and 0 otherwise (JD Categories: Journalism projects, Charity, Entertainment, Health, Others).
Level	Number of backing options for this project, also known as reward levels or perk levels. Typically, each contribution amount is associated with different number of units of the product.
Link	Binary variable that indicates whether the campaign page provides links to other promotion channels.
Pictures	Number of pictures used in the project pitch
Experience	Number of campaigns previously launched by the entrepreneur who launched this campaign
PreDuration	Duration of prefunding period in days
PreUpdates	Number of updates by the entrepreneur in the prefunding period
PreDiscussions	Number of discussions between potential backers and the entrepreneur in the prefunding period

Table 2.2. Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Prefunding	0.67	0.47	0	1
Success	0.77	0.42	0	1
Backers	220.201	777.86	0	25,479
Amount	204.44	568.27	1	10417.71
Goal	122,984.30	197,889.90	1,000	1,000,000
Raised	55,430.86	210,676.60	0	5,047,543
Duration	35.89	11.38	7	60
Category	.725	.446	0	1
Level	8.31	3.81	2	96
Link	0.45	0.50	0	1
Pictures	10.83	8.96	0	72
Experience	0.41	1.64	0	20
PreDuration	2.39	2.89	0	53
PreUpdates	0.1412	0.544	0	6
PreDiscussions	9.68	34.36	0	1,068

Table 2.3. Logit Regression on Prefunding Decision

Variable	Coefficient
<i>Goal</i>	0.298*** (0.032)
<i>Tech</i>	0.968*** (0.089)
<i>Duration</i>	-0.040 (0.126)
<i>Level</i>	0.374** (0.173)
<i>Pictures</i>	0.391*** (0.068)
<i>Link</i>	0.168** (0.083)
<i>Experience</i>	0.136 (0.096)
<i>Constant</i>	-4.803*** (0.657)

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 2.4. Average Treatment Effect of Prefunding on Project Success

Group	Observation	Mean	Std. Dev.
<i>0 (Control)</i>	891	0.672278	0.469646
<i>1 (Treatment)</i>	891	0.829405	0.376366
<i>Diff</i>		0.15713	

Table 2.5. Effects of Prefunding Activities on Initial Performance

Dependent Variables	Initial Backers	Initial Amount
ln(PreUpdates)	2.045 (0.253)	7.899*** (0.026)
ln(PreDiscussions)	1.491*** (0.000)	1.574*** (0.046)
ln(PreDuration)	-1.462*** (0.002)	-1.723*** (0.006)
ln(PreFollowers)	-0.0002* (0.074)	0.0005* (0.068)
ln(Goal)	0.092** (0.037)	0.372 (0.000)
Inverse Mills Ratio	0.823 (0.334)	3.199 (0.509)
Cons	2.456*** (0.001)	-2.267*** (0.089)

*Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$*

Table 3.1. Variable Descriptions

Variable	Description
ConversationLevel	Number of answers attracted by a question
QuestionEmpathy	Level of empathetic expression in a question
QuestionCognitive	Level of cognitive processing reflected in a question
AnswerEmpathy	Level of empathetic expression in the answers up to the previous period
AnswerCognitive	Level of cognitive processing reflected in the answers up to the previous period
QuestionerProfile	Type of profile of questioner: 0 for Private, 1 for Public
QuestionLength	Length (characters) of question
LectureLength	Length (seconds) of video
LectureDifficulty	Difficulty level of topic of lecture (scale of 0 (Arithmetic) -8 (Calculus))

Table 3.2. Summary Statistics

Variable	Mean	SD	Min	Max
ConversationLevel	0.05	0.24	0	30
QuestionEmpathy	1.31	4.84	-33.34	40
QuestionCognitive	14.56	4.84	0	66.67
AnswerEmpathy	0.09	1.25	-50	50
AnswerCognitive	0.49	3.11	0	83.33
QuestionerProfile	0.29	0.46	0	1
QuestionLength	111.18	143.39	1	19020
LectureLength	440.82	248.16	44	1739
LectureDifficulty	3.11	2.15	0	8

Table 3.3: Effects of Empathetic Expression on student conversation on MOOCs

Independent Variable	Model 1	Model 2	Model 3
Lag (Conversation)	0.241*** (0.003)	0.206*** (0.003)	0.374*** (0.003)
QuestionEmpathy	0.052*** (0.000)	-0.016*** (0.002)	0.267*** (0.002)
QuesCognitive	0.089*** (0.000)	0.085*** (0.000)	0.122*** (0.000)
AnswerEmpathy	0.0002*** (0.000)	0.002*** (0.000)	-0.002*** (0.000)
AnswerCognitive	0.004*** (0.000)	0.005*** (0.000)	-0.001*** (0.000)
Difficulty	-0.655*** (0.002)	-0.677*** (0.002)	-0.596*** (0.002)
QuesEmpathy X Difficulty		0.033*** (0.000)	
AnsEmpathy X Difficulty		0.001*** (0.000)	
QuestionEmpathy X QuesCognitive			-0.015*** (0.000)
AnswerEmpathy X AnswerCognitive			0.001*** (0.000)
LectureLength	0.002*** (0.000)	0.002*** (0.000)	0.001*** (0.000)
QuestionLength	0.001*** (0.000)	0.001*** (0.000)	0.001*** (0.000)
QuestionerProfile	-1.735*** (0.017)	-1.853*** (0.017)	-1.674*** (0.017)
Cons	0.708*** (0.014)	0.736*** (0.014)	0.366*** (0.014)

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

FIGURES

Figure 2.1 Sample Project in Prefunding Phase

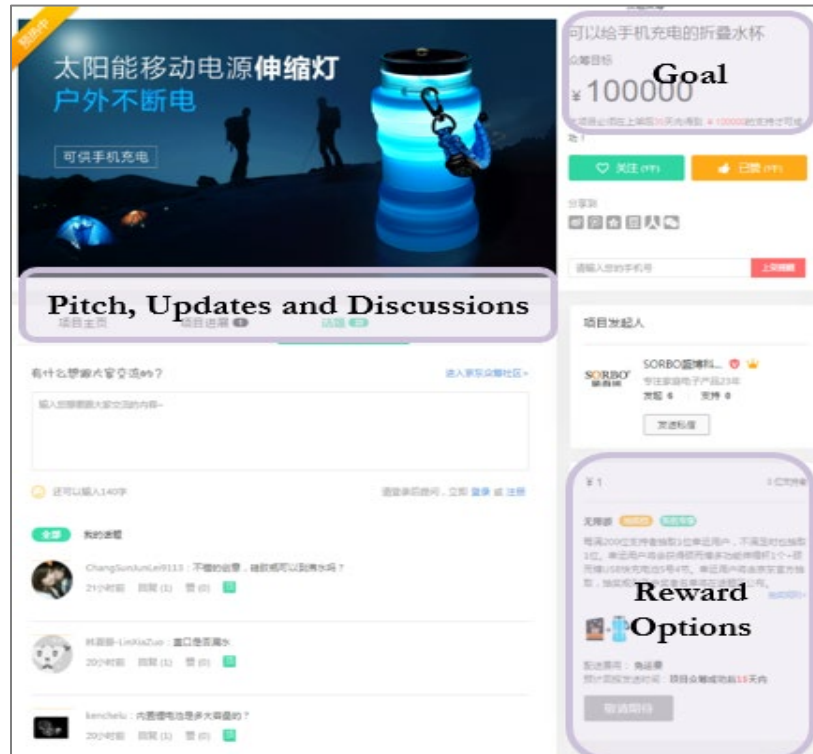


Figure 2.2. Theoretical Model

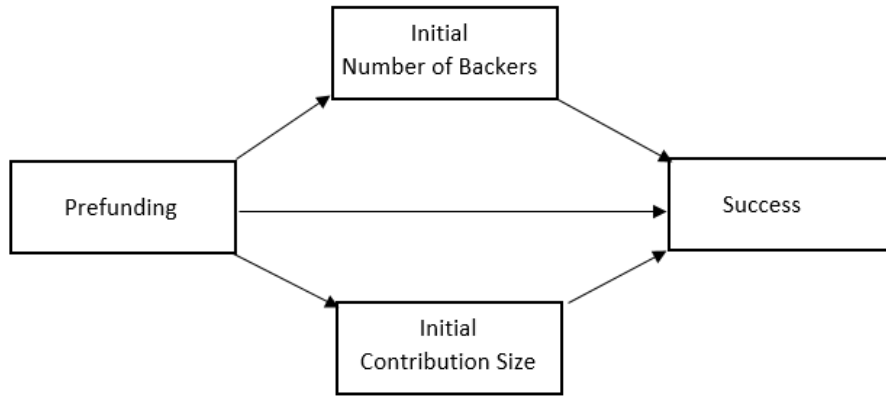


Figure 2.3. Propensity Score Distributions for Non-Prefunding and Prefunding Projects

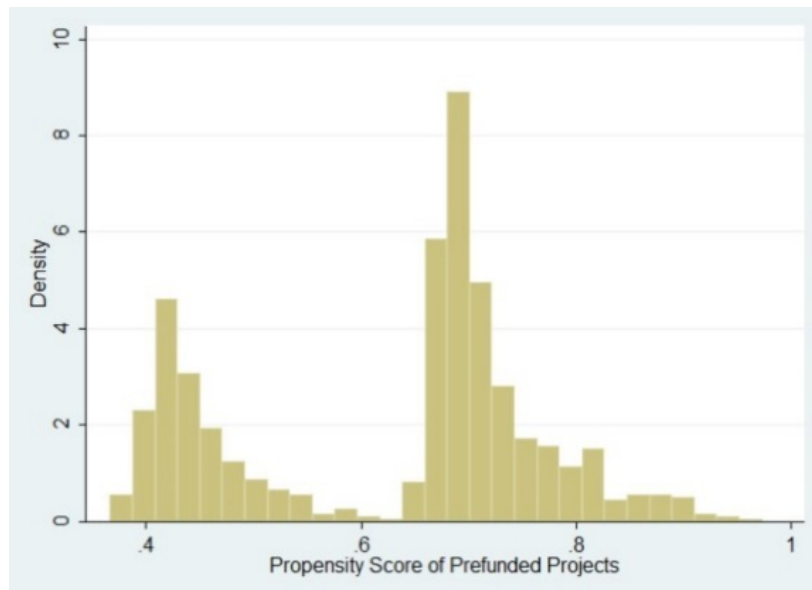
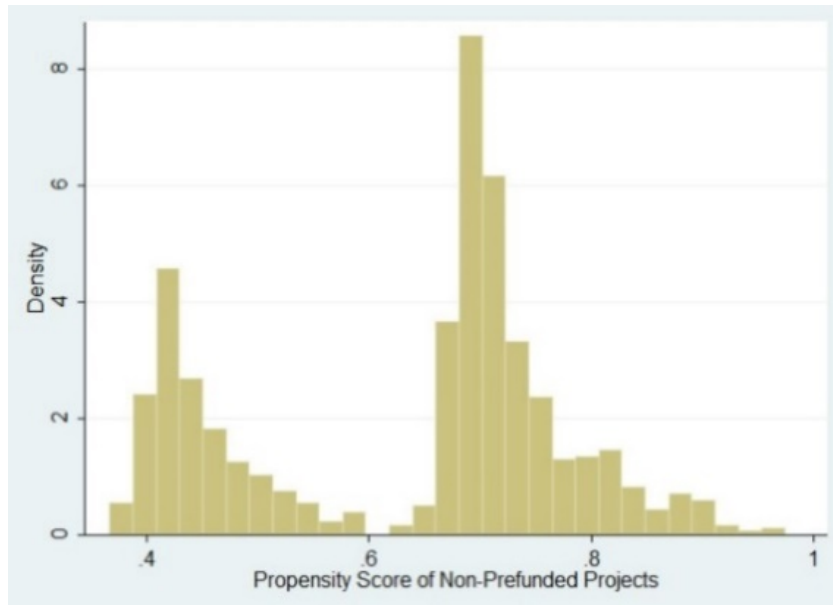


Figure 2.4. Number of Backers of Prefunding and Non-Prefunding Projects

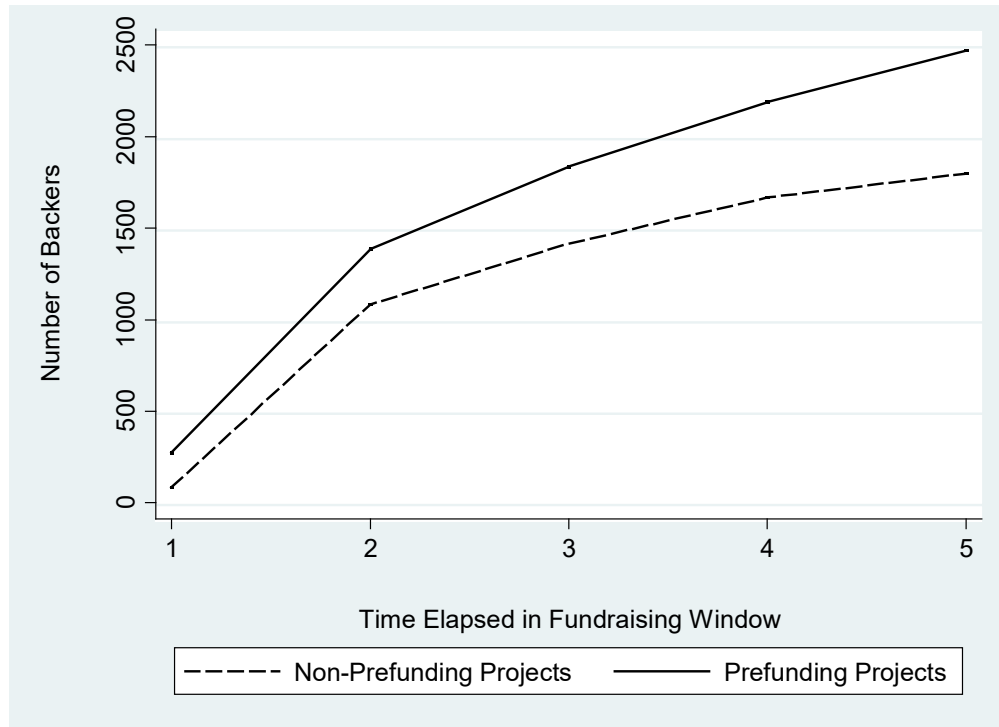


Figure 2.5. Contribution Size of Prefunding and Non-Prefunding Projects

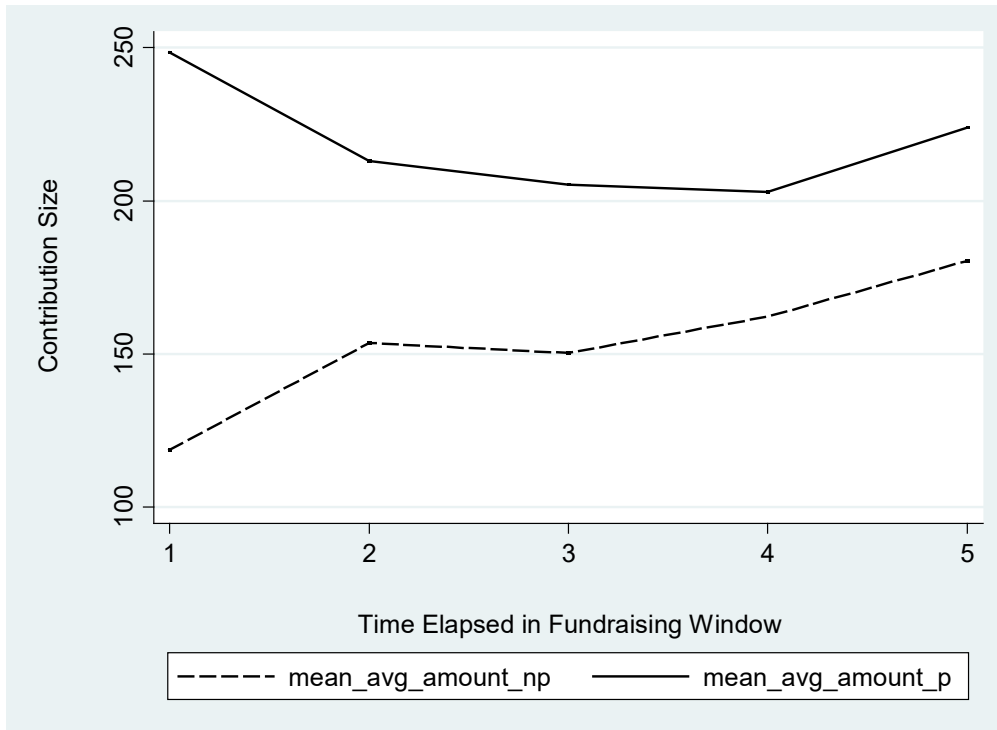


Figure 2.6. Distribution of Average Contribution Size on First Day of Fund-Raising

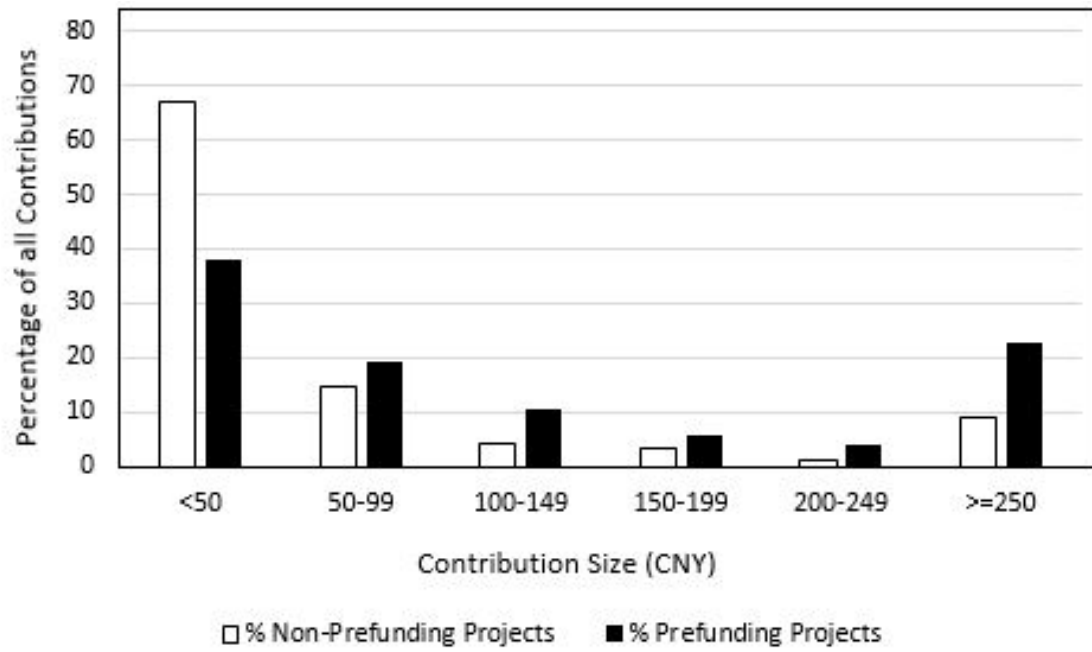


Figure 2.7. Distribution of Average Contribution Size on Last Day of Fund-Raising

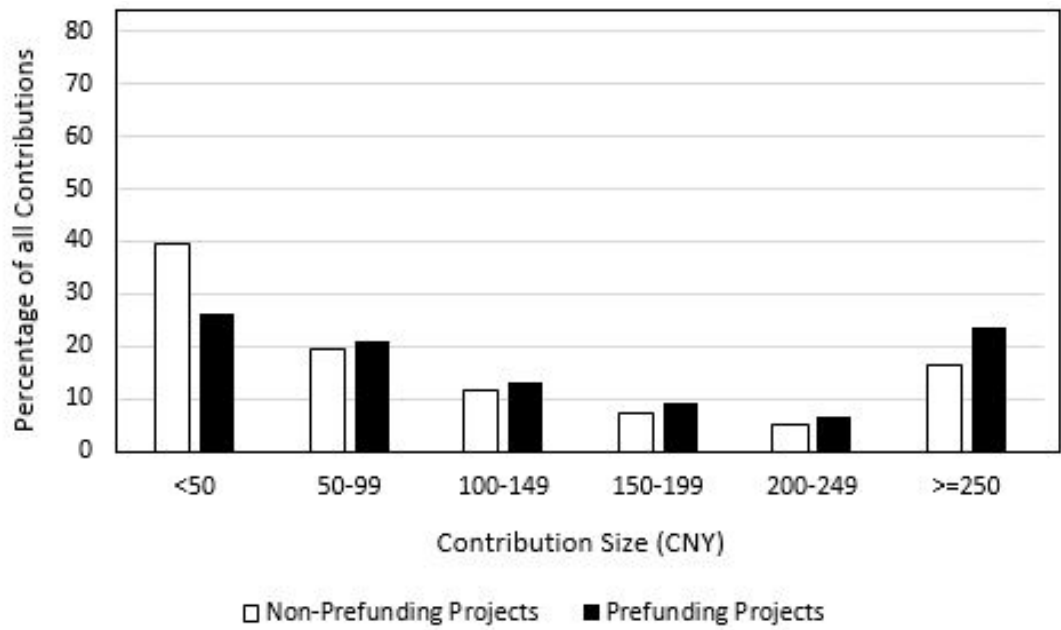
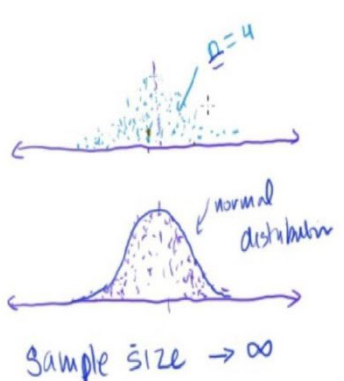


Figure 3.1. Example of a micro-lecture and student dialogues on Khan Academy

central limit theorem

Watch the entire video



Sample size $\rightarrow \infty$

What is the fundamental difference between the Law of large numbers & the Central limit theorem ?

59 votes · Comment · Flag · 5 years ago by Venky.vigi

Watch this video for the Law of Large Numbers:
<http://khanexercisesspot.com/video/law-of-large-numbers?playlist=Statistics>

The Law of Large Numbers basically tells us that if we take a sample (n) observation: our random variable & avg the observation(mean)– it will approach the expected val $E(x)$ of the random variable.

The Central Limit Theorem, tells us that if we take the mean of the samples (n) and p the frequencies of their mean, we get a normal distribution! And as the... [\(more\)](#)

87 votes · 1 comment · Flag · 5 years ago by Amruta.t

Show all 6 answers · Answer this question

If the sample-size approaches infinity (or the size of the population), wouldn't the form of distribution of the sample means go towards a distribution with a deviance of 0, practically becoming useless? I hope this makes sense.

15 votes · Comment · Flag · 4 years ago by redt

You are correct, the deviation go to 0 as the sample size increases, because you wo get the same result each time (because you are sampling the entire population). However, the deviation of the sampled means is not an indicator of the deviation of t entire population (as opposed to the mean of the sampled means, which IS an indic of the mean of the entire population).

Figure 3.2. Attenuation of conversation on Khan Academy: Most conversations die down in a span of four weeks.

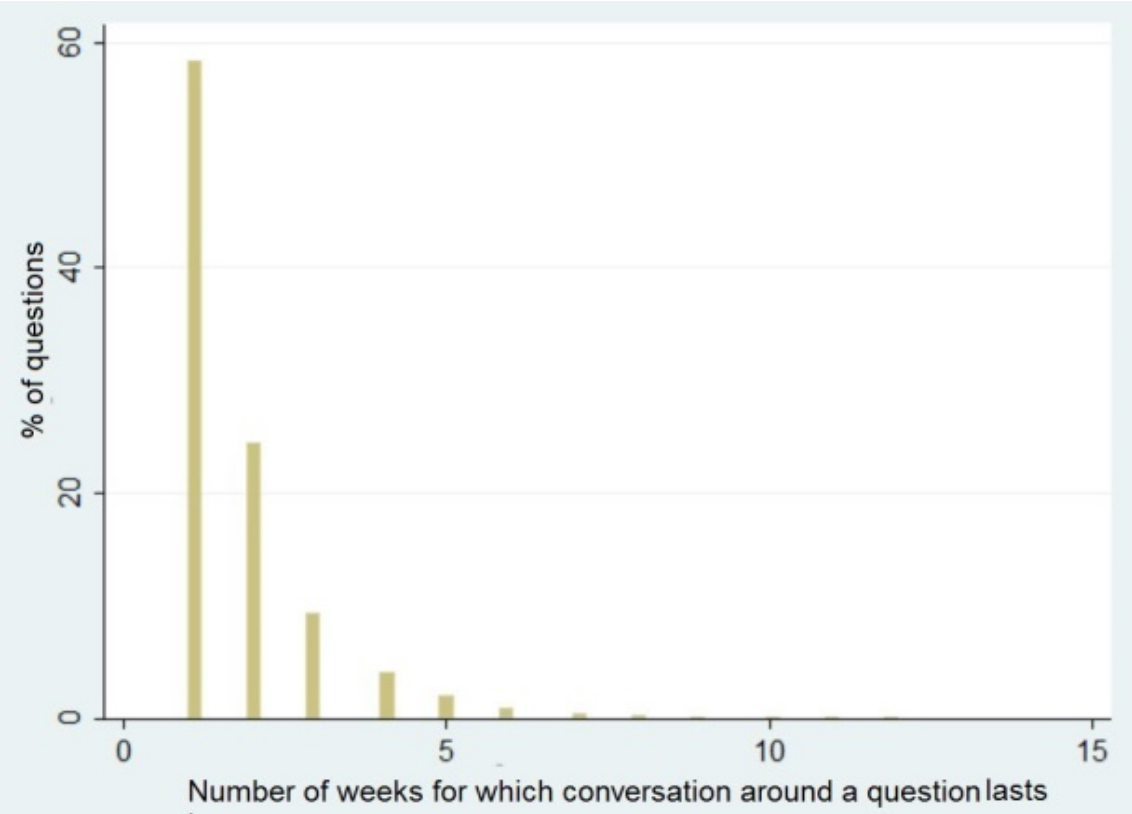


Figure 3.3: CDF of conversation level per question

