The Language that Gets People to Give: Phrases that Predict Success on Kickstarter

Tanushree Mitra
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, USA
tmitra3@cc.gatech.edu

Eric Gilbert
School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA, USA
gilbert@cc.gatech.edu

ABSTRACT
Crowdfunding sites like Kickstarter—where entrepreneurs and artists look to the internet for funding—have quickly risen to prominence. However, we know very little about the factors driving the "crowd" to take projects to their funding goal. In this paper we explore the factors which lead to successfully funding a crowdfunding project. We study a corpus of 45K crowdfunded projects, analyzing 9M phrases and 59 other variables commonly present on crowdfunding sites. The language used in the project has surprising predictive power—accounting for 58.56% of the variance around successful funding. A closer look at the phrases shows they exhibit general persuasion principles. For example, *also receive two* reflects the principle of *Reciprocity* and is one of the top predictors of successful funding. We conclude this paper by announcing the release of the predictive phrases along with the control variables as a public dataset, hoping that our work can enable new features on crowdfunding sites—tools to help both backers and project creators make the best use of their time and money.

Author Keywords
Crowdfunding; CMC; natural language processing (NLP)

ACM Classification Keywords
H.5.3. Group and Organization Interfaces; Asynchronous interaction; Web-based interaction

INTRODUCTION
Kickstarter is a crowdfunding website, a site where artists and entrepreneurs alike look to the internet for capital. At the time of this writing, a small startup called Pebble[1] is Kickstarter's most-funded project. An e-paper watch, Pebble only asked for $100K, but more than 18,000 people flocked to the idea and pledged a staggering $2.6M in just three days [31]. Yet, for every success story like Pebble, there is a failure like Ninja Baseball[2]. A PC game, Ninja Baseball was appealing enough to generate press attention [29], yet only attracted one-third of its requested $10K. As per Kickstarter's "all-or-nothing" funding rules, the Ninja Baseball team received no money.

Successful projects like Pebble have already raised over $300M on Kickstarter [30]. Furthermore, crowdfunding sites seem poised to rise to greater prominence soon. With President Obama's signing of the JOBS Act and its subsection the CROWDFUND ACT [7], industry observers expect several new crowdfunding sites to emerge very soon [13]. They will join prominent ones already on the internet, like StartSomeGood[3] and Flattr[4], in addition to crowdfunding platforms like Kickstarter, IndiGoGo and RocketHub.

While research in crowdfunding has recently gained attention in the HCI and CSCW communities [17, 21, 27, 35], we still know very little, for example, about what drives the "crowd" to take projects to their funding goals. What makes some projects like Pebble succeed while others, like Ninja Baseball, fail? In this paper, we look to answer this question. Results from our statistical model show that several project attributes (e.g., duration, presence of a video, etc.) have substantial predictive power, yielding a model with fairly high accuracy: 17.03% error under 10-fold cross-validation. We find, however, that the error-rate drops to 2.4% when we include the phrases used in the project's pitch—a non-random improvement over a strong null, controls-only model. This suggests that the language used by creators to pitch their project plays a major role in driving the project's success, accounting for 58.56% of the variance around success.

Adopting a corpus of 45K Kickstarter projects and natural language methods, we closely study 20K phrases, filtered from a corpus of 9M phrases. The aim is to see how they affect success on Kickstarter. Simultaneously, we control for 59 other variables commonly present on crowdfunding sites, like a project's goal amount, its duration and whether it has an associated video. Applying penalized logistic regression, we demonstrate the predictive power of phrases relative to these controls. Next, by comparing our phrase set with the Google 1T Corpus [5], we surface a more generalizable subset of phrases (i.e., phrases not specific to Kickstarter).

Taking a closer look at the words and phrases used in project pitches, we see general persuasion principles at work [10].

[1] `http://kck.st/OuioZA`
[2] `http://kck.st/Rb8ToC`

[3] `http://startsomegood.com`
[4] `http://flattr.com/`

CSCW'14, February 15–19, 2014, Baltimore, Maryland, USA.

ACM 978-1-4503-2540-0/14/02...$15.00.
http://dx.doi.org/10.1145/2531602.2531656

For example, the phrase *also receive two* (predicting a funded project), offers a favor in return for donating. In other words, this reflects *reciprocity* from the persuasion literature [10]. Reciprocity says that people tend to return a favor (donate money) after receiving one (gifts and offers). Another example is the phrase *given the chance*, promising something *scarce* to the funders. We are releasing these 20K phrases and their associated β weights as a public dataset.

We believe that the phrases dataset may enable new features on crowd-funding sites, such as tools to help backers and project creators make the best use of their time and money. For instance, a crowd-funding site might analyze the content of a project pitch while a project creator types it and notifies her whenever the words and phrases shift towards negative predictors. We see this work as particularly timely and important because crowdfunding provides small businesses a new way to access capital. Moreover, project creators can craft their language any way they see fit, as opposed to other indicators over which they have relatively little control, like social connections to influential people.

RELATED WORK

Here we provide an overview of related work in crowd-funding. Next we discuss work that adopts text analysis approaches similar to ours. Finally, we conclude this section by laying out work from the persuasion literature which we use to interpret our findings.

Dynamics of crowdfunding

Recent enthusiasm around crowdfunding startups has sparked some research to understand their properties. A qualitative study of the motivations behind crowdfunding shows that project creators use crowdfunding to spread awareness of their work and to receive online validation of their creativity [17]. This is in accordance with theories of self-efficacy which states that people's perception of their ability increases on receiving public recognition [2]. Other factors associated with crowd-funding initiatives are incentives to establish long-term relationships with funders, enjoyment in being part of a community [17] and an urge to keep control over one's projects [17, 24]. Similar results have also been reported by Ordanini et al.'s study of crowd-funding participants [36]. They found that for some participants the driving factor is social participation, or extending financial help and encouragement, while for others the motivation comes from the expectation of a monetary payoff [36] (see also [3]). These findings echo work on distinguishing intrinsic (e.g., enjoyment, involvement) and extrinsic (monetary rewards) motivations for funding [52].

Research on crowd-funding has also been extended to the enterprise level. Recent work on enterprise crowd-funding found that crowd-funding leads to enhanced inter-departmental collaboration and supports collective concerns over individual self-interest [35]. Other parallel studies using quantitative approach have found that higher funding goals and longer project durations lead to lower chances of successful funding on crowd-funding platforms like Kickstarter [34, 35]. Inclusion of a video in a project pitch increases the likelihood of full funding [34]. Additionally, one's social network has strong association with a project's success [34]. While these studies examined the determinants of successful funding and the motivating factors behind crowdfunding, they did not take into account the actual content in the project's description. The present work specifically examines the predictive power of content, and more precisely the words and phrases project creators use to pitch their projects.

A recent study used machine learning classifiers to predict the chances of successfully funding a Kickstarter project [21]. Similar to earlier work [34], the researchers used several attributes of a crowd funded project and surprisingly found that their accuracy hit an upper bound of 67%. The authors acknowledge that they were possibly ignoring additional factors, like the content of the project pitch. We build on their prior work and extend it by taking into account the language of the project pitch along with additional project attributes.

Analyzing Text for Social Information

Social scientists have studied the links between language and social behavior for decades. By analyzing the words and language people use in everyday life, they have developed tools like LIWC [38], a dictionary for inferring cognitive styles and social behavior from unstructured text. Using machine learning techniques, researchers have built SentiWordNet: a lexical resource to determine if text is opinionated by attaching valence scores to terms [37]. Positive and negative valence dictionaries have also been used to detect emotions during text-based CMC [23]. Emotions expressed via language even act as good predictors of future stock market prices, where emotions were estimated from blog posts [19]. In fact, research on tapping emotions and sentiment from language has grown enormously in the past few years [37]. Studies have shown, for example, that a Twitter user's personality (influential or not) is linked to the type of words he uses while tweeting [40], and that MySpace users express grief via certain words and linguistic styles [8].

More recently, researchers have demonstrated that certain words and phrases predict whether a corporate email will be sent up or down the organizational hierarchy [18]. Using statistical techniques, they built and announced the release of a reusable dictionary of power and hierarchy: one that can predict the direction of message flow from the raw text of the email. Following in their footsteps, our work looks to build a dictionary of phrases which signal successful crowdfunding. We base our study on statistical techniques, drawing inferences from analysis and visualizations of language usage. The interpretive framework we use is borrowed from the persuasion literature [10]. Next we provide a brief overview of this framework.

Theories of Persuasion

Theories of persuasion have played an important role in several spheres of scholarly research— advertising, marketing, consumer behavior research (see [44] for an overview) and more recently persuasive design [14]. Extensive work in this area has determined the basic principles that govern getting compliance from people [9, 10]. One of these principles is the rule of reciprocity, a widely studied idea among social psychologists [6, 20, 39, 41, 56]. Reciprocity is the sense of obligation to return a favor after receiving one [10]. Thus often requestors offer a gift in return for their requests [10].

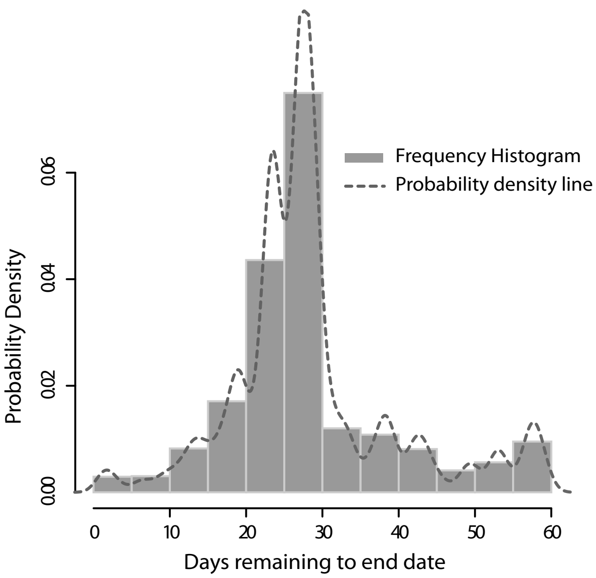

Figure 1. Histogram showing the days remaining for projects to reach their end date. The peak at around 30 days depicts that most projects were scheduled to finish collecting funds in the next 30 days.

They also exploit the power of scarcity principle—the tendency to attach more value to an item as soon as it becomes rare. Another factor which has been found to influence people's decision is whether others are also making that same choice, called social proof [1, 11]. Along similar lines, social identity theory posits that people identify themselves with a group when they perceive they have attributes in common [26]. These perceptions can then influence later decisions and choices [49]. Additionally, people often defer to expert opinions, an authority [10]. Moreover, liking for the requester can increase chances of compliance towards his request [10].

Using this interpretive framework of persuasion we delve into the qualitative explanation of our findings, looking for evidence of these theories in our predictive variables.

METHOD

We work from a list of 45,815 Kickstarter project URLs originally collected by Appsblogger [28] that includes all projects launched as of June 2, 2012. We scraped these project URLs on August 3, 2012 (when we performed our analysis) restricting our dataset to only those projects which had reached their last date of fund collection. We discuss our collection process in more detail in our next section. For reference and overview, Figure 2 illustrates the steps of our method.

On Kickstarter, you can fund creative projects in various categories. Table 1 lists the complete set of categories (13 in total) and sub-categories. A project creator pitches her project idea by providing a project description. She often uploads a video to substantiate the description. Every project on Kickstarter has a funding goal and an end date by which the goal needs to be reached. Kickstarter works on the "all-or-nothing" funding principle, whereby a project receives pledged money only if it meets or exceeds its funding goal by its end date.

Unit of Analysis: End-Dated Projects

The units of analysis in this paper are all projects which have reached their last date for collecting money. This ensures that our dataset contains only those projects which have reached their end date and thus have a clear outcome: *funded* or *not-funded*. This process eliminated 5 projects, and we were left with a dataset of 45,810 projects. Such a small number of eliminations might look strange at first, but recall that there is a lag between collection of project URLs (June 2, 2012) and scraping the project's page (August 3, 2012). When we plot the histogram of the days remaining for projects to reach their end date as of June 2nd, we obtain a unimodal distribution with a clear hump at 30 days (Figure 1). In other words, when we scraped data from the project's page, most projects had already reached their end date. In our final dataset of 45,810 projects, 51.53% (23,604) were successfully funded while 48.47% (22,206) not funded. Thus we had a fairly balanced dataset to do our statistical analysis.

Response Variable (dependent measure)

Our dependent variable is a binary response variable representing whether the project is *funded* or *not funded*.

Predictive Variables: Phrases

For each project, we scrape its textual content from its Kickstarter homepage. This includes scraping both the project's textual description and the promised rewards published by the project creator. We used Beautiful Soup[5] for scraping. It is a widely used Python library for scraping web content and allowed us to only select the text on the page related to pitches. Next, we convert all text to lowercase and tokenize text to every possible unigram, bigram and trigram, following the conventional bag of words model. Thereafter, we remove all phrases solely comprised of stop words. Finally, we end up with a corpus of 9,071,569 unique phrases.

However, including all phrases in our model will likely subtract information and produce poor results. This is because our collection of 9,071,569 phrases includes several obscure ones: those which would not matter in other domains or on other sites. We did not want our model to take into account these edge-case phrases and cloud our results. Rather, we wanted our model to reflect more general phrases. Hence we decided to build our model around relatively common English phrases. We throw away any phrase which occur less than 50 times in the corpus.

Moreover, some phrases might be specific to certain project categories. For example, the phrase *game credits* has a considerable presence in the entire corpus (140 occurrences), but unsurprisingly appears most often in projects belonging to the Games category. The phrase *our menu* has a large presence in the 'Food' category, but rarely appears in other categories. Because of this, we need a way to guard against phrases that uniquely identify categories and threaten generalizability. So, we keep only those phrases which are present in all thirteen categories. To summarize, by restricting phrases which occur at least 50 times and in all project categories, we ensure that we only build our model around relatively common English

[5] http://www.crummy.com/software/BeautifulSoup

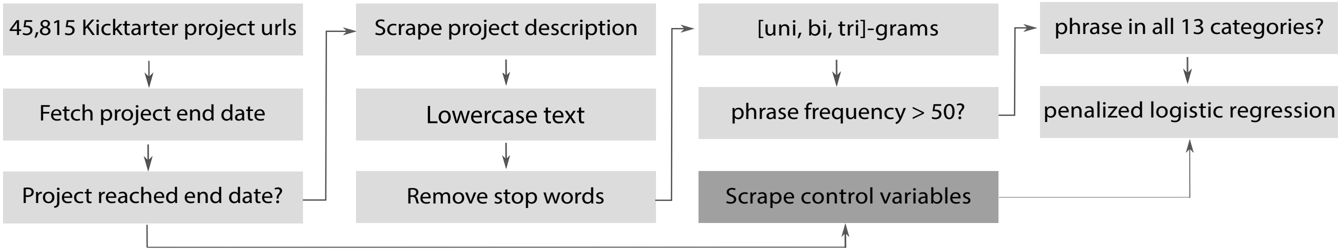

Figure 2. A flowchart of steps taken to extract predictive variables for modeling. We employ a series of filters to build our set of words and phrases. All phrases and control variables are fed into the penalized logistic regression model.

phrases and mitigate concerns associated with bias towards Kickstarter specific language. After these steps, we are left with 20,391 phrases, all of which are used as predictive features in our model.

Predictive Variables: Controls

In addition to these phrases, we include a set of 59 other Kickstarter variables as potential predictors of *funding*. This step is shown as a dark shaded box ("Scrape control variables") in Figure 2. Certain Kickstarter features—funding goal, project duration, project category, presence of a video in the project pitch, being featured on Kickstarter, and stating Facebook connections—can predict whether a project will be funded or not [34]. These variables are not just specific to Kickstarter. Most crowdfunding platforms (e.g., IndiGoGo, RocketHub) allow the project creators to mention their monetary goals, time line by which they plan to achieve their goal, category of the project, etc [17]. We control for these variables by adding them into the model's feature space (Table 1).

Project Goal: The amount of money (in USD) the project creator wants to raise for a project.

Project Duration: The time limit set by the project creator, up to which the project can accept funds.

Number of Pledge Levels: Project creators publish different levels of pledge amounts to seek donations from potential backers. They can be as small as $1, or as large as $10,000. For each pledge tier, they also promise rewards to backers in exchange for pledging.

Minimum pledge amount: The lowest tier amount.

Featured in Kickstarter: A project is said to be "Featured" in Kickstarter if it appears on any of these Kickstarter pages: *Staff Picks, Popular, Recently Launched, Ending Soon, Small Projects, Most Funded, Curated Pages*. Being featured is a strong indicator of successful funding [34].

Video Present: Creators often include a video explaining their project. Kickstarter claims that projects with a video report higher success rates than those without [6].

Video Duration: Since videos are often claimed to make projects compelling on Kickstarter, we wanted to test if the duration of videos actually has an effect. However, this is a coarse measure that does not reflect a video's quality.

[6] http://kck.st/UhMwge

Control Variables

Project Goal	Featured	No. of Updates
Project Duration	Video Present	No. of Comments
No. Pledge Levels	Video Duration	FB* Connected
Min. Pledge	Categories	

List of **Categories** & Sub-Categories

Art	**Music**	**Publishing**
Conceptual Art	Classical Music	Art Book
Crafts	Country & Folk	Children's Book
Digital Art	Electronic Music	Fiction
Illustration	Hip-Hop	Journalism
Painting	Indie Rock	Nonfiction
Performance Art	Jazz	Periodical
Mixed Media	Pop	Poetry
Public Art	Rock	**Technology**
Sculpture	World Music	Open Hardware
Design	**Film & Video**	Open Software
Graphic Design	Animation	**Dance**
Product Design	Documentary	**Theater**
Games	Narrative Fim	**Photography**
Board & Card Games	Short Film	**Food**
Video Games	Webseries	
Fashion	**Comics**	

Table 1. List of 59 control variables in our statistical model. Of these 59 controls, 49 are category and sub-category names. The 13 bolded items are the category names with their corresponding sub-categories listed below them. We include dummy variables for each of these in our model. FB stands for Facebook.

Number of updates: Project creators have the option to post updates stating the progress of the project to current backers and also to inspire new backers to donate.

Number of comments: Current and potential backers can use comments to ask questions to the project creator, which are then publicly visible.

Facebook connected: An entrepreneur's social ties influence the decisions of potential investors [43, 53]. We wanted to check if integrating Facebook connections to a Kickstarter project (an affordance the site offers) affects the likelihood of funding. We thus include a binary predictive variable to indicate whether the project creator has linked their Facebook account to the project page.

Project Categories: Kickstarter provides a list of categories and sub-categories for a new project. A project can be assigned either, but not both. Table 1 shows the full list of 49 categories and sub-categories. We include all as predictive variables.

Model Limitation

One potential confound with this approach is that the project's page might be updated during the lifetime of the project. For example, the project creator may add the names of the backers in the 'Backers' page, post updates in the 'Updates' page or even change the project description in the project 'Home' page. We made sure to guard against this by only crawling the content of the 'Home' page, ignoring any updates that have been made in the other pages. While this does not give us the exact mirror image of the project as it was first launched, we see it as a reasonable first-order estimate of the original project pitch. Follow-up work may find traction studying how dynamics affect funding success.

Statistical Technique

We employ penalized logistic regression [16] to predict the dependent variable *funded*. Penalized logistic regression is well-suited for our purposes because it guards against collinearity and sparsity, both of which are prevalent in our phrase dataset. Collinearity is a well-known property of English phrases. Consider an example from our phrase dataset: the phrase *week* will follow *every* more often then *fashion*. The regression technique handles this by moving the coefficient's weight to the most predictive feature. A parameter α determines how the weights will be moved among the predictive variables. $\alpha = 0$ includes all correlated terms but shrinks the weights towards each other (known as Ridge regression), while $\alpha = 1$ (Lasso regression) does both shrinkage and variable selection. In other words, Lasso includes only one variable per correlated cluster of variables by shrinking the coefficients of other variables to zero. We opted for a parsimonious model, letting our $\alpha = 1$.

A common problem with a large number of predictive variables is overfitting – a phenomenon where the statistical model fits the data under study too closely, explaining minor fluctuations but sacrificing overall predictive power. Cross-validation is a common technique to guard against overfitting. Thus we use an R implementation of penalized logistic regression with ten fold cross-validation, `cv.glmnet`[7] to handle phrase collinearity and guard against overfitting. `glmnet` also takes care of sparsity, another feature associated with texts, where a single project description spans only a small percentage of all possible phrases. Taking a predictive feature vector as input, `glmnet` predicts a binary response variable (*funded* or *not funded*), and upon fitting the data it associates β coefficients with the predictive variables. These β coefficients determine the relative power of the variables (both controls and phrases) in predicting whether the project will be *funded* or not. As our first step in building the predictive model we included only the control variables to measure model performance and the explanatory power of these variables. Next,

[7] `http://cran.r-project.org/web/packages/glmnet`

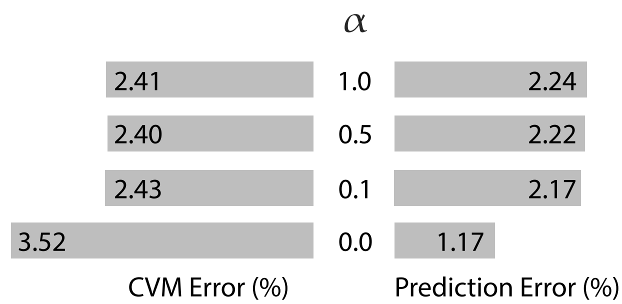

Figure 3. Percentage cross validation error and prediction error for different values of α. α=0 denotes ridge regression, while $\alpha = 1$ is lasso regression. Prediction error is measured on an 80/20 train/test split.

Model	Dev	df	χ^2	p	Error
Null	63.4K	0			
Controls-only	37.5K	58	25.9K	$< 10^{-15}$	17%
Phrases + Controls	363.4	5.4K	37.2K	$< 10^{-15}$	2.4%

Table 2. Summary of different model fits. *Null* is the intercept-only model. *Dev* denotes deviance which measures the goodness of fit. Error is the cross validation error rate reported by the model.

we expand our feature list to include our set of phrases to see if there is any substantial gain in explanatory power..

RESULTS

Using a Controls-only model as our baseline model (instead of an intercept-only model), we compare the deviance reported by our Phrases + Controls model. The deviance is related to the log-likelihood of the model and is a useful measure of goodness of fit. Standalone comparisons with the null model will fail to expose the relative power of phrases in contrast to the other control variables. While the null model has a deviance of 63,463.47, the addition of control variables reduced the deviance to 37,525.52, suggesting that they have considerable explanatory power. However, a dramatic reduction in deviance happens upon addition of phrases to the model, going from 37,525.52 to 363.42. Table 2 summarizes these results.

The difference between the deviances of two models approximately follows a χ^2 distribution, with degrees of freedom equal to the number of additional predictors in a model. By comparing the deviances (Table 2), we see that the information provided by the control variables has significant predictive power: $\chi^2(58, N{=}45{,}810) = 63{,}463.47 - 37{,}525.52 = 25{,}937.95$, $p < 10^{-15}$. The cross validation error reported by the control only model is 17.03%. Adding phrases results in another massive reduction of deviance (58.56% drop), suggesting the substantial predictive power of the phrases: $\chi^2(5{,}427 - 58, N{=}45{,}810) = 37{,}525.52 - 363.42 = 37{,}162.1$, $p < 10^{-15}$. There is also a dramatic reduction in error rate, with cross-validation error being only 2.4%. The low cross validation errors ensure that our models are guarded from excessive overfitting.

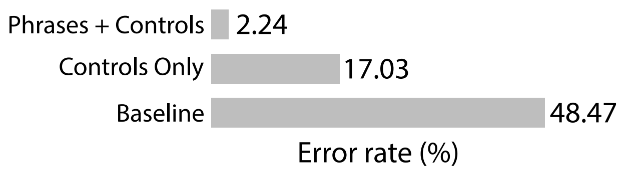

Figure 4. Comparison of error rates between the baseline, controls-only and phrases+controls model. The control variables alone result in a fairly high accuracy model. Addition of phrases gives an additional boost of about 15%.

The low error rate at first might imply that phrases are solely dominating the prediction results. However, recall that the control variables alone result in a fairly high accuracy model (cross validation error rate = 17.03%). Adding phrases gives an additional boost of about 15%. Figure 4 shows a comparison between the error rates of the baseline, Controls-only and Phrases + Controls model. The accuracy of the baseline model corresponds to picking the most likely class (*not funded*) every single time. To further gain confidence in our results we repeated our regression technique varying the parameter α and comparing two heuristics: 1) cross-validation error rate and 2) prediction accuracy on an 80/20 train/test split. Figure 3 shows the comparison. Lasso regression ($\alpha = 1$) performs better than ridge ($\alpha = 0$), and there is negligible difference among the other settings of α.

Looking at the β weights generated by our Phrases + Controls model we find that the top 100 positive and negative β predictors are solely comprised of phrases. Table 3 lists the 100 phrases with the most positive β weights. These are the ones which predict that a project will be funded *(F)*. Conversely, Table 4 presents 100 phrases with the most negative β weights, strongly predicting that a project will not be funded *(NF)*.

Can these phrases be grouped under meaningful categories? We turn to the LIWC (Linguistic Inquiry and Word Count) program [38] to find an answer to our question. LIWC is a text analysis technique based on a hand-built dictionary of words and word stems, assembled into categories. It matches the words from input text to words and stems of the dictionary, generating the percentage of words that match these categories. We tested against all 82 LIWC categories. Since LIWC performs simultaneous tests, we apply a Bonferroni correction, letting $\alpha = 0.05/82 = 0.0006$. We find that there is significant difference in 74 categories. Figure 5 shows the membership in seven of these 74 language dimensions (p-values $< 10^{-15}$). We interpret these results further in our Discussion section.

To get a glimpse of the structure of the predictive phrases, we present Word Tree visualizations [55] in Figure 6. We used the online site Many Eyes [54] to build these visualizations. Here we present the searches for the phrases *pledgers will* and *even a dollar* in the *funded* and *not funded* parts of the corpus, respectively. We see that *pledgers will* is often used to convey the gifts one would receive after funding the project. On the other hand, *even a dollar* perhaps reads as groveling for money and is less appealing.

How do these phrases map to more general text? In other words, are there phrases which occur fifty times and in all

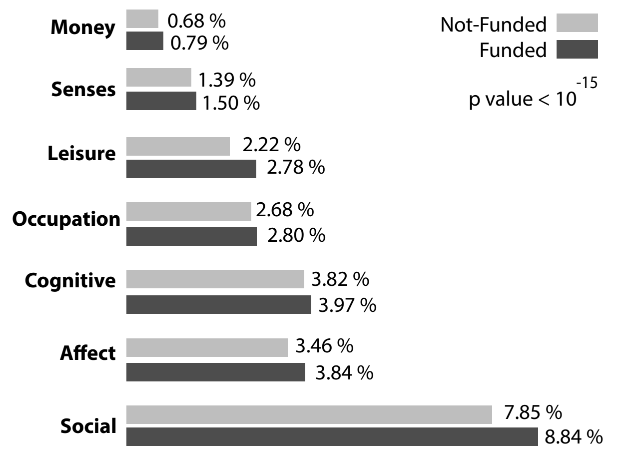

Figure 5. Testing for groupings of predictive phrases. We use LIWC to find membership of predictive phrases in 82 language dimensions. We have shown seven categories here. Phrases which predict that the project will be funded exhibit more social, cognitive, affective or emotional, and sensory or perceptual processes. They also have higher presence in 'Person Concern' categories (e.g., occupation, leisure, money and financial issues).

categories, but still only matter to Kickstarter alone? A cursory glance through the non-zero β weights pointed us to these phrases: *kickstarter goal*, *[via, of their, our, on, using] kickstarter*. All of them are specific to Kickstarter's culture and yet were not discarded by our filtering steps. To present a more generalizable set of phrases, we turn to the Google 1T corpus [5], a vast database of the frequency of n-grams Google has seen on the web. We made this choice because both Kickstarter and Google 1T corpus correspond to web text. Scanning the Google corpus, we look for all non-zero β phrases that our model generated from Kickstarter. We note their frequencies in the Google corpus. Next, we perform χ^2 tests of independence between the frequencies of phrases in Google and Kickstarter corpus. Since we performed simultaneous tests against the sets of Google unigrams, bigrams and trigrams, we apply Bonferroni correction, letting α = 0.05/5.4K. Finally, we search for phrases with significant statistical differences and whose membership is higher in Google's 1T corpus. This discarded 4,440 phrases from our set of 5.4K phrases with non-zero β weights, yielding a rather small phrase set of 494 positive and 453 negative predictors. We mark (in gold) the occurrence of these phrases in our list of top 100 β predictors (Tables 3 and 4).

DISCUSSION

We find that among 59 control variables and 20,391 phrases, the top 100 predictors of *funded* and *not funded* are solely comprised of phrases. 29 of the 59 control variables had non-zero, non-random predictive power: presence of a video, number of updates, Facebook connected, number of pledge levels, and eleven categories and sub-categories (Graphic Design, Theater, Food, Games, Documentary, Art, Board & Card Games, Webseries, Fashion, Periodical, Animation) had positive β weights (Figure 5). Project duration was associated with *not funded* (β = -0.0129). This matches reports from recent research that longer project durations lower the chances of funding [34]. Additionally, thirteen categories and subcategories had negative β weights (Figure 6).

Figure 6. Word Tree visualizations showing membership of phrases in two sets of our dataset. The top visualization shows the occurrence of *pledgers will* (positive β weight) in the *funded* dataset. It occurs 70 times in the corpus. The visualization shows phrases that fork off from *pledgers will* in this part of the corpus. The font size of a phrase is proportional to its frequency of occurrence. We see that the most commonly occurring trigram is *pledgers will receive*. The visualization at the bottom depicts the search for the phrase *even a dollar* (negative β weight) in the *not funded* dataset. It occurs 56 times in this part of the corpus. The visualization shows that the most likely phrases are *even a dollar short*, *even a dollar will*, *even a dollar can*, perhaps reading as groveling for money.

(F) phrases	β	(F) phrases	β
project will be	18.48	difference for	5.60
has pledged	5.42	pledged will	4.01
pledged and	3.98	december of	3.21
we can afford	2.94	trip in	2.83
used in a	2.82	par	2.79
around new	2.78	trash	2.75
their creative	2.71	given the chance	2.69
mention your	2.69	your continued	2.65
to build this	2.65	cats	2.64
option is	2.59	inspired me	2.57
workshop and	2.56	project will allow	2.56
the coming	2.55	dollar pledged	2.54
we have chosen	2.53	accessible to the	2.52
and an invite	2.51	christina	2.51
all supporters	2.48	from the past	2.44
pledgers will	2.44	finding out	2.43
lane	2.39	plus recognition	2.37
want them to	2.31	farm	2.31
got you	2.31	atlantic	2.30
and encouragement	2.28	some help with	2.26
that exists	2.25	as people	2.25
in this new	2.22	projects will	2.21
would greatly	2.20	we are fully	2.20
dates and	2.15	a national	2.14
conception	2.14	problem of	2.12
and added	2.11	kind to	2.08
unveiling	2.07	good karma and	2.04
commemorating the	2.04	shows that	2.02
girl and	2.00	il	1.99
two friends	1.96	secure the	1.95
future is	1.94	testament	1.93
that i feel	1.91	the meaning	1.91
fundraising goal	1.89	their thoughts	1.89
nv	1.88	support at	1.87
a personal tour	1.86	are raising money	1.85
the brooklyn	1.85	good as	1.84
administration	1.83	and develop	1.83
also receive two	1.83	the inside of	1.81
upfront	1.81	to play the	1.79
looking for your	1.77	as a small	1.77
for two years	1.76	changed my	1.76
gain a	1.76	our social	1.76
answering	1.74	design elements	1.74
funding will help	1.73	guarantee a	1.73
company for	1.72	all previous rewards	1.72
thanks a	1.72	a detailed	1.71
sharing with	1.71	the correct	1.71
be called	1.70	and share it	1.70
of hot	1.70	a lot about	1.70
message and	1.70	poster of your	1.69

Table 3. The top 100 phrases signaling that the project will be funded. The phrases obtained after comparison with the Google 1T corpus are marked in gold. All phrases are significant at the 0.001 level.

(NF) phrases	β	(NF) phrases	β
pledged	-7.12	dressed up	-4.64
not been able	-4.02	trusting	-3.91
all the good	-3.89	based in the	-3.87
models of	-3.84	school that	-3.75
information at	-3.65	kids of all	-3.55
of the leading	-3.53	on a larger	-3.44
new form of	-3.43	that uses	-3.42
we have lots	-3.24	to enjoy a	-3.20
way for us	-3.18	room on	-3.18
an honorable mention	-3.17	panel of	-3.17
is time for	-3.14	even a dollar	-3.10
nm	-3.08	be followed	-3.02
easy and	-2.97	later i	-2.96
and to provide	-2.91	will surely	-2.90
word out about	-2.87	picture in	-2.87
logo on it	-2.84	also work	-2.83
location of the	-2.80	people into	-2.78
you message from	-2.76	blanket	-2.76
provide us	-2.76	every time you	-2.73
need one	-2.69	help support our	-2.68
the culture of	-2.68	us from the	-2.67
unseen	-2.67	in school	-2.65
a door	-2.59	a masters	-2.59
a blank	-2.57	discretion	-2.57
volunteers to	-2.56	we raise will	-2.55
to the cost	-2.54	reusable	-2.53
the profits	-2.52	hand made by	-2.52
educate and	-2.51	get to pick	-2.48
based upon the	-2.47	will soon	-2.47
unified	-2.46	illustration	-2.46
to identify	-2.45	the production costs	-2.45
product will be	-2.43	refined	-2.43
space at	-2.41	continue with	-2.41
hope to get	-2.39	no extra	-2.39
present in	-2.37	definitely a	-2.35
occur in	-2.34	you start	-2.34
the needed	-2.34	addition to being	-2.33
decide what	-2.32	tuning	-2.32
deeper into	-2.30	help to bring	-2.29
known and	-2.28	underway	-2.27
campaign will help	-2.25	for decades	-2.23
goes in	-2.23	notoriety	-2.22
get to vote	-2.22	make you an	-2.21
air and	-2.20	an alternative	-2.19
be creative	-2.19	shows the	-2.19
post card with	-2.19	website for more	-2.19
signed postcard	-2.18	varies	-2.18
on different	-2.16	left my	-2.16
of their choice	-2.16	who like	-2.15
name or logo	-2.14	piggy	-2.14

Table 4. The top 100 phrases signaling that the project will not be funded. The phrases obtained after comparison with the Google 1T corpus are marked in gold. All phrases are significant at the 0.001 level.

Control Variables	β	Control Variables	β
Graphic Design	1.35	Video Present	0.60
Theater	0.57	Food	0.48
Games	0.33	Documentary	0.32
Updates Count	0.25	Art	0.13
FB* Connected	0.13	Board & Card Games	0.12
Webseries	0.11	Fashion	0.05
Pledge levels	0.05	Periodical	0.04
Animation	0.01		

Table 5. The 15 control variables which have non-zero predictive power to signal that the project will be *funded*. All control variables are significant at the 0.001 level.

Taking a closer look at Tables 3 and 4, we find an intriguing view of words and phrases providing cues of *funded* and *not funded*. As is perhaps to be expected, phrases which exude negativism (*not been able*), or lack assurance (*later i*, *hope to get*) are predictors of *not funded*.

> i have **not been able** (β = -4.07) to finish the film because none of my editors will see the project through to the end.

> i can't take size orders and possibly **hope to get** (β = -2.47) them all made in time for christmas.

Perhaps unsurprisingly, phrases which signal lucrative offers to potential backers (*also receive two*, *mention your*) are positive predictors of successful funding.

> i'll thank you by name in each and every one of season one's episodes, and **mention your** (β = 2.69) own project or message at the top of one of them.

> add $40 and you will **also receive two** (β = 1.83) vip tickets to the premiere screening.

In more formal terms, these positive predictors reflect the principle of *reciprocity* [10] from the persuasion literature. We next illustrate how the usage of certain phrases exhibit subtle hints of persuasion, possibly motivating people to donate [10, 12, 15]. Finally, we conclude with the design and theoretical implications of our findings.

Reciprocity

Reciprocity is the tendency to return a favor after receiving one [10]. Social psychologists have studied the norms of reciprocity for many years [6, 20, 39, 41, 56]. People often use persuasive appeals when reflecting norms of reciprocity (i.e., "If you grant the request, I will reward you") [12]. We see similar phenomena in our phrase dataset: *mention your*, *also receive two*, *we can afford*, *pledged will* are among the top 100 positive predictors. Taking a closer look, we see that these phrases are often used to offer a reward or a gift in return for donation funds: *mention your name* in the [film, program, introduction, thank you section, acknowledgment section, credits of the film], *pledgers will* have [their pick, a special credit], *pledgers will* [get, also receive, have], *also receive two* [free, full passes, tickets, copies of].

> everyone who **has pledged** (β = 5.42) $xx [or more] will get ...

> if i make or exceed my goal then you will be charged what you **pledged and** (β = 3.98) you will get your fantastic rewards

Control Variables	β	Control Variables	β
Illustration	-2.55	Journalism	-1.12
Pop	-0.79	Rock	-0.5
Performance Art	-0.46	Film & Video	-0.44
Children's Book	-0.40	Mixed Media	-0.35
Country & Folk	-0.2	Music	-0.11
Public Art	-0.09	Electronic Music	-0.08
Short Film	-0.04	Project Duration	-0.01

Table 6. The 14 control variables which have non-zero predictive power to signal that the project will be *not funded*. All control variables are significant at the 0.001 level.

> we'll **mention your** (β = 2.69) name in the sleeve of our full length album (which you'll get the download code for) and you get our summer darling ep with additional never before released bonus track

A particularly interesting top predictor for *funded* projects was *good karma and* (β = 2.04).

> i will thank you on my website, send you **good karma and** (β = 2.04) give you a free digital download of the big spoon ep.

> you'll get an mp3 of a previously unreleased song delivered to you via email before anyone else hears it. i also trust that you'll accrue some **good karma and** i'll be truly grateful to have you on the team.

However, not all offers are alluring enough to attract backers. Low offers are often rejected [22, 42]. For example, the phrase *dressed up* is one of the top negative predictors in our dataset. On searching the corpus of *not funded* projects, the phrase usage reveals offers which are perhaps too low to attract backer attention.

> you get a physical copy of the ep, a rigoletto shirt, a rigoletto sticker, and a skype date with us **dressed up** (β = -4.52) as celebrities.

> **extremely limited** you will be our vip victim! we'll get you all **dressed up** in period clothing and have you be a victim in one of the paranormal crime scene set ups! (you are responsible for transportation and accommodations in utah. don't worry, i'll get you a deal with a hotel!) now that's killer!

Another dimension of reciprocity is 'personalization', a tactic often used by recommendation systems [33] and web personalization engines [50], whereby the person trying to persuade offers products and services that appeal to people.

> but we are going to ask that you send us a photo of yourself to be **used in a** (β = 2.82) collage featured in a piece of selfless season artwork.

> your vocal will be **used in a** similar way as a line or two of lead vocal - not just a faded background vocal.

Scarcity

People attach more value to products and opportunities which are rare, distinct, limited in supply or are available for a limited time [10]. For example, the following excerpt from a successful project pitch emphasizes limited time availability:

> for anyone who comes by and was thinking of pledging, the **option is** (β = 2.59) still there until 5:04 pm on 17 october. if you want the lower calendar price and no shipping charges, or if any of the rewards tickle your fancy, they're still available through kickstarter until monday afternoon.

Additionally, exclusivity is often harnessed while making offers, leading to higher chances of acceptance [10, 11].

> this custom color **option is** ($\beta = 2.59$) exclusive to kickstarter and will not be available at product launch!

> donating to this project is an opportunity contribute to something really special.first and foremost you are being **given the chance** (β = 2.69) to become part of something at its earliest stages.

> also, you will be **given the chance** to purchase our small batch pieces before the public domain.

> blast **from the past** ($\beta = 2.44$) – pledge and receive one of only ten remaining copies of the johnny starlings "aiming too high", my first ever cd release, long since out of print.

Social Proof

Social proof is the idea that people depend on others for cues on how to act [10]. Persuasive tactics use this principle by making people aware of what others are doing to increase their likelihood to follow along [11]. We see traces of social proof in the language of *funded* projects, often signaling the attention the project has already received.

> [name] **has pledged** ($\beta = 5.42$) some gas money.yay! thank you! so, you can see that i already have people willing to support my art.

> we've hit the 50% mark! huge thanks to everyone who **has pledged** so far. it is truly inspiring!

> dennis mckenna eethnopharmacologist): "now this is a worthy kickstarter project to support! i have **pledged and** ($\beta = 3.98$) suggest you do too! this is a possibility to change policy with respect to the use of ayahuasca in addiction treatment. support it if you can and tell your friends.

Social Identity

Social identity is an individual's knowledge that he belongs to a social group, in which individuals have common attributes and identify themselves in similar ways [25, 46]. They adopt the group member's attitudes as their own and their decisions are influenced by the group's majority opinion [47, 48, 57].

> collect is dedicated to making the arts **accessible to the** ($\beta = 2.52$) community, creating work and events that enhance and celebrate culture within the civilization of new york city

> a large portion of our community has come together **to build this** (β = 2.65) project.

Perhaps there are deeper underlying questions of whether funding efforts are driven by the group's interest. Our current work does not answer these; we hope future researchers can explore these questions.

Liking

People are more likely to comply with a person or product if they like them. Positive remarks about another person's attitudes and performance increases liking [4].

> with your help, this **project will be** ($\beta = 18.48$) a success, and you'll be able to enjoy our movie at a festival near you! thanks very much.

> thank you for your support **and encouragement** ($\beta = 2.28$)

Additionally, the two most important factors to increase liking are similarity and praise [10]. People use similarities to create bonds, which are later leveraged to garner support [10]. For instance, the following example shows how the project pitch leverages on the fan community and also extends appreciation—perhaps an attempt to draw other fans to fund the project.

> the support from the disney fan community and everyone who **has pledged** ($\beta = 5.42$) so far has been amazing. thank you all so much.

Authority

People often resort to expert opinions for making efficient and quick decisions [10]. Many of the top predictive phrases were used to portray the expertise of project creators and developers:

> that means having the best-quality exhibition master **we can afford** ($\beta = 2.94$), attending the film festivals in person to meet with potential buyers, and even hiring a professional publicist and graphic designer to help promote the film.

> the **project will be** ($\beta = 18.48$) sag certified and meet all union standards ensuring the highest level of professionalism.

> the **project will be** produced by dove award winning producer rusty varenkamp

> music development for the **project will be** coordinated by platinum-selling recording artist morten kier and supervised by gmcla artistic director e. jason armstrong.

However, research shows that people rely on experts more when they have to decide matters of low personal relevance, but engage in more thoughtful personal consideration when the matter is highly personal to them [57]. Unsurprisingly, higher thought process is associated with successfully funded projects, which we will discuss shortly in our LIWC results. Our study does not examine how the personal relevance of a project effects its success. We believe this is an interesting direction for future research.

LIWC and Sentiment

Our LIWC results in Figure 5 reveal some interesting findings: successfully funded projects demonstrate more active thinking (*Cognitive Process*), a higher degree of *Social Process*, higher perception rates (*Senses*), higher levels of emotions (*Affect*) and exhibit *Personal concerns* via references to Money, Occupation, Leisure and Home. We also examined the sentiment of the phrases, by performing sentiment analysis using the Natural Language Text Processing API provided by `text-processing.com`. Results reveal that *funded* projects have higher positive and negative sentiment compared to *not funded* projects, though the difference is not statistically significant. As is perhaps expected, LIWC results show that the membership in *Death* category (not shown in Figure 5) is significantly different in *funded* and *not funded* parts of the corpus, with higher presence in *not funded*. However, all of these differences are relatively minor and likely only show statistical significance because of our sample size.

Other phrases

While many of the top phrases relate to persuasion, others fell outside this framework, yet had high predictive power. As is perhaps to be expected, some phrases were used in the context of incentivizing donations.

> if you can't make a larger contribution, you can still contribute by telling all your friends about this. if ten people **mention your** (β = 2.69) name when they make a donation, you will get a free download of the film, a thanks in the credits"

A particularly interesting phrase is: *december of* ($\beta = 3.21$). *december of* mostly talks about projects and products scheduled to be released by December of a particular year. Despite our attempts to control for specificity, some phrases (e.g., *december of*) might be specific to Kickstarter's culture. That said, perhaps these projects subtly signal the tax exemptions people can take in the U.S. by donating by December 31 – a possible way to incentivize donations.

Another perplexing finding was the occurrence of phrases like *christina* (β = 2.51) and *cats* (β = 2.64) in our top predictors. While *christina* (β = 2.33) mostly referred to famous celebrity (i.e., Christina Aguilera), we had no clear explanation for the occurrence of *cats*—except for the commonly accepted wisdom that the internet loves them.

General Phrases

To map the phrases captured in the Kickstarter platform to more general text, we used Google's 1T corpus [5]. Their presence in the top 100 phrase predictors is marked in gold in Tables 3 and 4. Phrases like *and encouragement* ('Liking'), *given the chance* ('Scarcity'), *as people* ('Social Proof') make a reappearance. Additionally, pitching a completely *new form of* expression is associated with *not funded* projects, while drawing inspiration from something *that exists* seems to work better. We note the following examples to put this in context.

> it was to play educational games allowing students to experience a *new form of* (β = -3.33) discovery, curiosity, and mystery.

> we have one remaining trip to japan, where we hope to cover the creation of home video technology, as well as the extremely unique direct-to-video market **that exists** (β = 2.25) there.

Other intuitively positive predictors include *secure the* (β = 1.95), *gain a* (β = 1.76), *guarantee a* (β = 1.73).

> in exchange we offer the chance to **secure the** extremely limited special edition of the album and other items

> so pledging $10 is the only way to **guarantee a** copy!

Again, explicit groveling for money is less appealing as depicted by the phrases *provide us* (β = -2.87), *need one* (β = -2.69).

> we **need one** (β = -2.80) thing we don't have is money

Not surprisingly, more forward-looking phrases (*next step is*, *in the upcoming*,*will be published*, *to announce*), phrases indicative of ongoing project work (*and published*, *are preparing*, *working on a*, *worked in*, *teamed up with*), phrases exhibiting 'Authority' (*with a professional*) and phrases with positive sentiment (*wow*, *good for*) are predictors of *funded*. We have released the set of all non-zero positive and negative beta weights with membership in Google 1T corpus [8,9].

Phrase + Control Variables Dataset

Additionally we have also released all the predictive variables that form the core of our penalized logistic regression[10]. It consists of all possible phrases available to the regression model, in addition to the control variables. Each entry in the file lists predictive variables followed by their associated β weights, sorted in decreasing order. We also include entries with zero β weights, so that researchers can examine phrases which do not affect funding.

[8] http://cc.gatech.edu/~tmitra3/data/Gg.KS.pos
[9] http://cc.gatech.edu/~tmitra3/data/Gg.KS.neg
[10] http://cc.gatech.edu/~tmitra3/data/KS.predicts

Theoretical Implications

Despite advances in multi-media, computer mediated communication (CMC) systems mostly generate text [32]. Content analysis of CMC text can lead to meaningful inferences about human behavior in different social contexts, something we believe is demonstrated in this research. To our surprise, language has a strong effect on funding success, and many of these results seem to fit squarely within existing persuasion literature. We hope that this work stimulates new research in the CSCW and HCI community on crowdfunding. By releasing the model's predictive phrases and control variables, we inform two bodies of research: work on computerized text analysis to draw inferences of real-world behaviors [18, 51], and an emerging class of crowdfunding sites.

Though these phrases are associated with *funded* and *not funded* projects in very distinct ways, we do not claim that using these phrases guarantees success. Neither do we claim that using these phrases is the only means of successful funding. One should also note the importance of other project attributes, like certain project categories, the number of project updates and the number of pledge levels, all of which have substantial effect on the project's success. There could well be several other intervening variables which ultimately control a project's funding—perhaps they only manifest as language. We see this as a rich area for further study.

Design Implications

We foresee that our work may enable new features on crowdfunding sites—tools which can help both backers and project creators make best use of their time and money. Commentators have attached importance to observing both successful and failed projects before starting a new one [45]. With this as a backdrop, imagine the design of a "Help Center" or "FAQ" page on a crowdfunding site. The page lists words, phrases and language-style guides for project pitches: those which are associated with successful funding and those which are not. Moreover, perhaps while a project creator types her project pitch on the funding website, the site alerts her whenever the words and phrases shift toward negative predictors. It could perhaps also provide her with alternative language to increase the chances of success.

Study Limitations

Like many large-scale analysis of observational data, we cannot make any causal claims, such as whether the project's pitch was able to persuade people to donate money. Our work shows the factors which relate to successful funding and surprisingly found that along with some project attributes, the language used in the project pitch has a strong effect. Secondly, our study is a quantitative analysis followed by observations from our phrase dataset. While our approach is meaningful in describing what happens, without an accompanying qualitative analysis we can only speculate on the why. We hope our findings will motivate future researchers to address these limitations.

FUTURE WORK & CONCLUSIONS

The successful startup Pebble from the beginning of the paper has faced repeated problems delivering its product to market. There is certainly more work to do looking for traits like this—what some might call "over-promising and under-delivering."

Also, there is almost certainly additional predictive information outside the text of project pitches and our list of control variables. Future work could build on our results by exploring other attributes. For example, we saw that 'Facebook Connected' is a positive predictor of *funded*, but we did not explore the rich space of social network predictors.

Overall, our findings indicate a fundamental force which drives the "crowd" to fund crowd-projects: language. In this paper, we uncover the phrases which signal whether a project will be crowdfunded and announce the public release of a set of persuasive phrases. These phrases highlight the interesting ways in which raw CMC text can expose real world social behaviors. We hope our work motivates future researchers to delve deeper into the dynamics of crowdfunding, viewing it both through the lens of linguistic style and other traits of crowdfunding. Our results may also lead to a new class of applications, such as systems which can help project creators craft pitches with the best chances of success.

ACKNOWLEDGEMENTS

We would like to thank our reviewers, the comp.social lab at GeorgiaTech and Michael Muller for their valuable comments that improved this work. We also thank DARPA for supporting this research.

REFERENCES

1. Aronson, E., Wilson, T. D., and Akert, R. M. Social psychology. 5th, 2005.
2. Bandura, A. The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology 4*, 3 (1986), 359–373.
3. Belleflamme, P., Lambert, T., and Schwienbacher, A. Crowdfunding: Tapping the right crowd. In *International Conference of the French Finance Association (AFFI)* (2011), 11–13.
4. Berscheid, E., and Hatfield, E. Interpersonal attraction (2nd ed.). MA: Addison-Wesley (1978).
5. Brants, T., and Franz, A. Web 1T 5-gram Version 1. *Linguistic Data Consortium, Philadelphia* (2006).
6. Brehm, J. W., and Cole, A. H. Effect of a favor which reduces freedom. *Journal of Personality and Social Psychology*, 3 (1966), pp. 420–426.
7. Brown, Merkley, and Bennet. Crowdfund Act (S. 2190). Tech. rep., 2012.
8. Brubaker, J. R., Kivran-Swaine, F., Taber, L., and Hayes, G. R. Grief-stricken in a crowd: The language of bereavement and distress in social media. In *Proc. ICWSM'12* (2012).
9. Cialdini, R. B. *Influence (rev): The Psychology of Persuasion*. HarperCollins, 1993.
10. Cialdini, R. B. In *Influence: Science and practice*, Boston: Allyn & Bacon (2001).
11. Cialdini, R. B., and Goldstein, N. J. Social influence: Conformity and compliance. *Annual Review of Psychology. 55* (2004), 591–621.
12. Danet, B. The language of persuasion in bureaucracy: "modern" and "traditional" appeals to the israel customs authorities. *American Sociological Review 36*, 5 (1971), pp. 847–859.
13. Empson, R. Ready, set, crowdfund: President obama to sign jobs act tomorrow. http://tcrn.ch/hmhj67. retrieved sep. 19, 2012. (2012).
14. Fogg, B. J. Persuasive technology: using computers to change what we think and do. *Ubiquity 2002*, December (2002), 5.
15. French Jr, J. R., and Raven, B. The bases of social power. In *Group dynamics*, D. Cartwright and A. Zander, Eds., New York: Harper and Row (1960), 607–23.
16. Friedman, J., Hastie, T., and Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* (2010).
17. Gerber, E., Hui, J., and Kuo, P. Crowdfunding: Why people are motivated to post and fund projects on crowdfunding platforms. In *CSCW Workshop* (2012).
18. Gilbert, E. Phrases that signal workplace hierarchy. In *CSCW* (2012), 1037–1046.
19. Gilbert, E., and Karahalios, K. Widespread Worry and the Stock Market. In *Proc. ICWSM* (2010).
20. Goranson, R. E., and Berkowitz, L. Reciprocity and responsibility reactions to prior help. *Journal of Personality and Social Psychology*, 3 (1966), 227–232.
21. Greenberg, M. D., Pardo, B., Hariharan, K., and Gerber, E. Crowdfunding support tools: predicting success & failure. CHI EA '13 (2013).
22. Güth, W., Schmittberger, R., and Schwarze, B. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior &; Organization 3*, 4 (1982), 367 – 388.
23. Hancock, J., Landrigan, C., and Silver, C. Expressing emotion in text-based communication. In *Proc. CHI* (2007), 929–932.
24. Harms, M. What drives motivation to participate financially in a crowdfunding community?
25. Hogg, M. A., Abrams, D., Otten, S., and Hinkle, S. The social identity perspective. *Small Group Research 35*, 3 (2004), 246–276.
26. Hogg, M. A., and Turner, J. C. Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology 26*, 4 (1987), 325–340.
27. Hui, J., Greenberg, M., and Gerber, E. Understanding crowdfunding work: implications for support tools. CHI EA '13 (2013).
28. Jeanne, P. Kickstarter failures revealed! What can you learn from Kickstarter failures? http://www.appsblogger.com/kickstarter-infographic. Retrieved Sep. 19, 2012. (2012).
29. Jensen, K. The best kickstarter projects. http://www.ugo.com/web-culture/best-kickstarter-projects-ninja-baseball. retrieved sep. 19, 2012. (2012).

30. Kickstarter. http://kickstarter.com/help/stats. retrieved sep. 19, 2012. (2012).
31. Kosner, A. Pebble watch for iphone and android, the most successful kickstarter project ever. http://onforb.es/pp6dpw. retrieved sep. 19, 2012. (2012).
32. Lee Rainie, Kristen Purcell, A. S. The Social Side of the Internet. Tech. rep., Pew Internet & American Life Project, 2011.
33. Li, T., and Tsekouras, D. Reciprocity in effort to personalize: examining perceived effort as a signal for quality. Proc. ICEC '12, 1–8.
34. Mollick, E. The dynamics of crowdfunding: Determinants of success and failure. SSRN scholarly paper, Social Science Research Network, Rochester, NY, July 2012.
35. Muller, M., Geyer, W., Soule, T., Daniels, S., and Cheng, L.-T. Crowdfunding inside the enterprise: employee-initiatives for innovation and collaboration. Proc. CSCW'13, ACM (2013), 503–512.
36. Ordanini, A., Miceli, L., Pizzetti, M., and Parasuraman, A. Crowd-funding: transforming customers into investors through innovative service platforms. *Journal of Service Management 22*, 4 (2011), 443–470.
37. Pang, B., and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval 2*, 1-2 (2008), 1–135.
38. Pennebaker, J. W., and Francis, M. E. *Linguistic Inquiry and Word Count*. Lawrence Erlbaum, August 1999.
39. Pruitt, D. G., and Drews, J. L. The effect of time pressure, time elapsed, and the opponent's concession rate on behavior in negotiation. *Journal of Experimental Social Psychology*, 5 (1969), 43–60.
40. Quercia, D., Ellis, J., Capra, L., and Crowcroft, J. In the mood for being influential on twitter. In *Privacy, security, risk and trust (passat), 2011 IEEE SocialCom)* (2011), 307 –314.
41. Regan, D. T. Effects of a favor and liking on compliance. *Journal of Experimental Social Psychology*, 1 (1971), 621–639.
42. Roth, A. Bargaining experiments. In *Handbook of Experimental Economics.*, J. Kagel and A. Roth, Eds., vol. 3, Princeton University Press (1997), 253–348.
43. Shane, S., and Cable, D. Network ties, reputation, and the financing of new ventures. *Management Science* (2002).
44. Shrum, L., Liu, M., Nespoli, M., and Lowrey, T. M. Persuasion in the marketplace.
45. Steinberg, S., and DeMaria, R. In *The Crowdfunding Bible: How to Raise Money for Any Startup, Video Game or Project*, J. Kimmich, Ed., read.me (2012), 14–47.
46. Tajfel, H. Social categorization. In *Introduction à la psychologie sociale*, S. Moscovici, Ed., vol. 1, Paris: Larousse. (1972), 272–302.
47. Tajfel, H. In *Human Groups and Social Categories: Studies in Social Psychology.*, Cambridge, UK: Cambridge Univ. Press (1981).
48. Tajfel, H. Social psychology of intergroup relations. *Annual Review of Psychology 33*, 1 (1982).
49. Tajfel, H., and Turner, J. C. An integrative theory of intergroup conflict. *The social psychology of intergroup relations 33* (1979), 47.
50. Tam, K. Y., and Ho, S. Y. Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Info. Sys. Research 16*, 3 (2005), 271–291.
51. Tausczik, Y. R., and Pennebaker, J. W. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology 29*, 1 (2010), 24–54.
52. Van Wingerden, R., and Ryan, J. Fighting for funds: An exploratory study into the field of crowdfunding. Tech. rep., Lund University School of Economics and Management, 2011.
53. Venkataraman, S. The distinctive domain of entrepreneurship research: An editor's perspective. In *Advances in entrepreneurship, firm emergence, and growth*, J. Katz and R. Brockhaus, Eds., vol. 3, Greenwich, CT: JAI Press (1997), 119–138.
54. Viegas, F., Wattenberg, M., Van Ham, F., Kriss, J., and McKeon, M. Many Eyes: a site for visualization at internet scale. In *InfoVis*, Published by the IEEE Computer Society (2007), 1121–1128.
55. Wattenberg, M., and Viégas, F. The Word Tree, an interactive visual concordance. In *InfoVis* (2008), 1221–1228.
56. Wilke, H., and Lanzetta, J. T. The obligation to help: The effects of amount of prior help on subsequent helping behavior. *Journal of Experimental Social Psychology*, 6 (1970), 488–493.
57. Wood, W. Attitude change: Persuasion and social influence. *Annual Review of Psychology 51*, 1 (2000), 539–570.