

**The Narrative Advantage:
Gender and the Language of Crowdfunding**

Andreea Gorbatai

Haas School of Business

UC Berkeley

Laura Nelson

Kellogg School of Management

Northwestern University

August 2015

Abstract

In this study, we set out to examine the role of language in the success of online fundraising—a new form of entrepreneurial project financing. In particular, we evaluate the influence of linguistic content on fundraising outcomes, above and beyond type of product or service offered. Online fundraising settings pose an interesting empirical puzzle: women are systematically more successful than men, an outcome contrary to offline gender inequality. We propose that this outcome is partially explained by linguistic differences between men and women in terms of language they use, and we test this mechanism using data from the online crowdfunding platform Indiegogo. The results support our theory, suggesting a link between micro-level linguistic choices and macro level outcomes: the institution of crowdfunding may reduce gender inequalities in the fundraising arena by benefitting the communication style of women.

INTRODUCTION

Economic and social arrangements in markets and organizations have been shown to systematically disadvantage women across a wide range of outcomes ranging from hiring, performance evaluations, rewards, and promotion in the labor market to financial support in the capital market. Research has identified several mechanisms through which this inequality is perpetuated, ranging from homophily, tokenism and structural constraints to negative stereotyping and women's own beliefs about their skill level and worth in the labor market.

In particular, research on financing, small business, and entrepreneurship has shown that women are at a marked disadvantage compared against men with similar skill and experience levels. This difference has been largely attributed to choice homophily among predominantly male funders, and to the type of businesses that women start. In the venture capital industry, male venture capitalists acts as gatekeepers; this results in less funding and mentorship for female entrepreneurs. In other entrepreneurial ventures, women gravitate towards small business ventures where they are often the sole employee instead of choosing scalable business projects.

In all these contexts, the long term outcomes – financing, and the terms of the financing deals – are the results of many difficult to quantify factors resulting from the interaction between the funder and the entrepreneur. It is thus difficult to isolate the effects that non-verbal behavior, paralinguistic cues, contextual factors, and interactions between the entrepreneur and the funder have on the final decision regarding funding. Some of these factors have been studied experimentally in the laboratory (Brownlow and Zebrowitz 1990; Carney, Cuddy and Yap 2010; Kramer 1977). The majority of these studies rely on evaluating the effect of these factors on the audience, controlling for language content. But the language we use is intimately connected to

how we think, and how others evaluate what we are saying. Moreover, language is connected with socio-demographic characteristics of the speaker or writer, such as gender, age or occupation.

In this study we aim to examine the effect of language on the success of crowdfunding campaigns, and the relationship between linguistic content and gender. Online, text-based campaigns are ideal for examining the effect of language content apart from characteristics of the delivery medium, message sender, and audio-visual information because the message is delivered to the potential donors via an information-poor, asynchronous text interface. In turning our attention to the language used in crowdfunding campaigns, we examine four different dimensions of language content in campaign descriptions: positive (sentiment) language, vividness, inclusive language, and business language. We then suggest that three of these types of linguistic content (positive sentiment, vividness and inclusive language) are both more likely to be rewarded in crowdfunding campaigns, and more likely to be used by women, while the use of language related to money is more likely to be penalized in the crowdfunding context and more likely to be used by men. We then test and confirm our theory that language mediates the relationship between gender and fundraising outcomes using data from the online crowdfunding site Indiegogo.

Our findings indicate that gender-specific language partially mediates the success of women in fundraising money through crowdfunding. This study identifies an economic institution (crowdfunding) where female-specific linguistic patterns are preferred over male-specific patterns, leading to a reversal in gender inequality with respect to funding. Additionally, this study contributes to economic sociology research on gender by quantifying the impact of linguistic choices on fundraising outcomes. Lastly, this research contributes to research in

computational sociology by employing topic models to refine the product classification of crowdfunding campaigns and quantify crowdfunding campaign text along several content dimensions using the Linguistic Inquiry Word Count dictionary (Tausczik and Pennebaker 2010).

THEORY AND HYPOTHESES

MONEY AND LANGUAGE

Financial investments represent a particular type of economic exchange driven by expectations of future returns. Research interested in the antecedents of successful access to financial resources have found that bank loans, one major category of funding, more often go to larger businesses, more established businesses, and businesses with collateral, while small businesses that do not own the collateral necessary to secure a bank loan often seek equity investors (Berger and Udell 1998). Research on equity investors suggests that previous ties with funders increases the chance of being funded (Shane and Cable 2002), and that equity investment is geographically concentrated—it is difficult to get funded far away from an industry's center (Stuart and Sorenson 2003). In whole, small businesses, businesses that are not in high-growth sectors such as technology, and businesses away from major city centers are at a disadvantage when seeking funding. As a result, the majority of new small businesses rely solely on personal funds and personal debt, at times supplemented by funds from friends and family, rather than outside funds from banks or investors (Coleman and Robb 2009).

More recently, many people and small organizations have turned to funding their entrepreneurial ideas through donation-based crowdfunding sites. The global crowdfunding market reached 1.1 million campaigns worldwide in 2012, and raised approximately 2.7 billion

dollars, of which 1.6 billion dollars in the United States (Massolution. 2013). To get funded using traditional methods requires taking a financial risk via loans, or entering the often unfriendly world of investors. Donation-based crowdfunding sites offer a completely different institutional setting in which to obtain funding for new small businesses or products. Taking appeals for funds directly to the public, over the impersonal medium of the internet, bypasses many risks and difficulties inherent in traditional methods of fundraising.

Crowdfunding sites differ from traditional channels for business financing—such as bank loans, venture capital, and microcredit groups (Anthony 2005), along three main dimensions: the interaction medium, the types of investors, and the motivations for investing. These three aspects combined suggest that entrepreneurs seeking funding through crowdfunding sites may be evaluated differently than when seeking funding through a traditional channel, but little is known about what leads to success in this setting.

The interaction medium in a crowdfunding setting is different from other investment settings in that crowdfunding is computer-mediated and a-synchronous, such that potential contributors make decisions in absence of face-to-face interaction or other visual cues. In traditional investing settings entrepreneurs will make a pitch directly in front of potential investors, and the investors have a chance to ask questions about the entrepreneur and her product or business. Crowdfunding takes place online, absent of face-to-face interaction or other visual cues, and absent of the need to be in the same location as your investor. Second, the types of investors are different. Traditional investors are typically specialized in business and finance, and many traditional forms of investment rely on one or a small number of individuals or institutions providing large amounts of money. Projects pitched on crowdfunding sites, alternatively, target the general population, not specialized individuals or institutions, and they

seek small amounts of money from a large number of people. Many investors on crowdfunding sites provide only \$5 or \$10 each, and few provide more than a few thousand dollars.

Third, the institutional difference between traditional financing and crowdfunding-based financing alters the reasons why an individual might provide money to finance a project.

Traditional investors seeking to invest in a new product or business, for example, will typically invest because they believe the project or business will succeed, that is, it will become profitable and the investor will see a return on their investment. The investor does not need to like the product or business or find it valuable—they must believe that there is a significant group of people who will like it or find it valuable, and the business will thus bring in a profit.

On crowdfunding sites, alternatively, people invest in a product or business for reasons other than financial returns. Individuals give money in order to support projects that they value for social or value-related reasons—receiving symbolic gifts and messages in exchange for their contribution to the writing of a book, recording of a music album, or launch of an educational initiative —, and in order to receive products or services that will be available, if the proposed project is successful. Similar to other forms of investment, contributors are not guaranteed the success of the projects they invest in, but, in contrast to many other types of investment, contributors are not eligible to receive their money back if the project fails to produce results. Individuals contributing on crowdfunding sites thus must either personally like the project or business, or they want to be part of a project that they think is valuable in some way.

The process of convincing investors to invest in a particular product, project, or business on a crowdfunding site is thus much different than in a traditional setting. The lack of face-to-face interaction alters how entrepreneurs build relationships with investors. In traditional

investment settings visual and verbal cues are important, such as making eye contact, having an erect posture, dressing appropriately, and having a strong voice. Instead, crowdfunding campaigns must catch the attention of potential investors, depending much more on written language rather than visual and verbal cues.

Prior research on language and health-related communication indicates that the way a message is framed is consequential for decision making, holding constant the message itself (Rothman and Salovey 1997). As crowdfunding campaigns rely largely on written words to persuade people to fund a business campaign, and the literature shows that the way something is written affects the messages' efficacy, we expect the language used to frame the business idea behind a campaign to affect the ability to attract online donations.

Second, the type of language that will succeed in attracting money is different. While specialized financial language is necessary to attract traditional investors, this type of language can sound formulaic and dry to the general population and may be off-putting or confusing for those donating money on crowdfunding sites. Furthermore, individuals contributing to donation-based crowdfunding sites are not doing so for financial gain, so financial language is unlikely to convince them to donate. For these reasons we expect that proposals detailing monetary or financial aspects of the project are less likely to succeed than campaigns that choose to focus on other aspects of the business plan:

Hypothesis 1a. Crowdfunding proposal success decreases with the use of **money-related language**.

Entrepreneurs seeking funding must get potential investors excited about their projects; this is difficult to do in a traditional financing setting, but even more so in crowdfunding, where

financial gain is not factored into the decision. One way to do this is to tell an exciting story about the product or business, building excitement through narrative. Research focused on business pitches in a traditional fund-raising environment suggests that entrepreneurs are successful when they adequately address both the “what”—storytelling aspect of the business idea—and the “how”—narrative aspect of it (Pollack, Rutherford and Nagy 2012: 918). The interest in a vivid story is high among evaluators of movie pitches as well. For example, one of the informants in Elsbach and Kramer’s (2003) study stated: “I think that magic is perhaps the most important part of the pitch [...] it’s a seduction, a promise of what lies ahead” (p. 296). This would suggest that crowdfunding proposals are more successful when potential investors are able to visualize the idea, and when the narrative is compelling. Rather than the dry language of finance, crowdfunding pitches require colorful, vivid language:

Hypothesis 1b. Crowdfunding proposal success increase with the use of **vivid language**.

Aside from being personally excited by a campaign, individuals on crowdfunding sites may contribute to a project because the idea resonates with them and they feel emotionally invested in the cause. One way to get people to feel good about giving to a business idea is by using positive language and appealing to people’s emotional side. Elsbach and Kramer’s (2003) qualitative study of Hollywood pitches examined 28 videos of television and movie pitches and found that ideas with great potential are those that generate positive emotions. Linguistic studies have similarly shown that people who use more positive emotion terms are more likely to be optimistic and confident, and are perceived as such (Tausczik and Pennebaker 2010).

Hypothesis 1c. Crowdfunding proposals success increases with **positive emotion**.

Relatedly, individuals decide to donate to a project because to feel like they are helping a great idea succeed—and may incidentally benefit from it too. For example, one evaluator in Elsbach and Kramer’s study stated: “the best approach is sort of relational in that I have to feel good and feel as if I’ve ‘won’ something as well” (Elsbach and Kramer 2003: 295) from investing into an idea. Therefore language that will make potential investors feel connected to the cause or the founders, and included as part of the venture will increase likelihood of success.

Hypothesis 1d. Crowdfunding proposal success increases with the use of **inclusive** (relational) **language**.

GENDER AND LANGUAGE

Language use, however, is correlated with socio-demographics and geography. People with different backgrounds and from different demographic groups have different communication styles, and this often is exhibited through written language. In particular, the link between gender and language has been extensively studied in linguistics. Research shows readers can often guess the gender of an author through their language use, and machines can fairly easily correctly assign a text a gender (Thomson and Murachver 2001), meaning gender is correlated with language use in predictable ways. For example, several studies have shown that women and men have substantively different writing styles, with women exhibiting more interpersonal sensitivity (McMillan et al 1977), referencing emotional states more often (Mulac and Lundell 1994: 307), and showing a higher level of social engagement compared to men. Thus, if language is important in online funding platforms, gender will be important too, whether or not the gender of the writer is made visible.

In particular, research on gender and negotiation has found that – women are more relational, express more emotion in their language (Kray and Thompson 2005), and they approach social interactions as a “negotiation for closeness and connection” (Tannen 1990, cited in Kray and Thompson 2005). Additionally, because women view morality through a care-based perspective (Gilligan 1982), they are more likely to consider both parties’ interests in negotiations and exchanges. Findings from these prior studies would suggest that:

Hypothesis 2a. Women use more **inclusive** language than men do.

Hypothesis 2b. Women use **positive emotional** language more than men do.

Prior research has found that men are more likely to use language related to money (Newman et al. 2008). Additionally, as suggested above, women are generally more empathetic to others’ needs, being able to position their idea as desirable from the perspective of its ideational characteristics. Coupling this with the fact that women tend to use more perceptual words (Newman et al. 2008)—such as “feel,” “taste,” or “delicious”—, we expect women are better at telling a story that resonates with potential crowdfunding investors, and are less likely to use money- or finance-related language than men are:

Hypothesis 2c. Women use more **vivid** language than men do.

Hypothesis 2d. Women use less language related to **money** than men do.

LANGUAGE, GENDER AND CROWDFUNDING SUCCESS

Prior research has found that gender socialization affects the type of jobs and projects men and women choose (Barbulescu and Bidwell 2013)—for example, women identify less with “masculine” jobs and expect to be less successful in them. Women-owned small businesses are

also heavily concentrated in the service and retail sectors and are poorly represented in the technology sector. These structural differences are often cited as the reason for the lack of external funding for women-owned businesses (Anna, Chandler, Jansen, and Mero 1999). However we propose that, even holding the type of project constant, women and men make different linguistic choices in terms of framing these projects, and these linguistic choices mediate the relationship between gender and crowdfunding success:

Hypothesis 3. Language mediates the relationship between gender and fund-raising success.

DATA & METHODS

DATA

The data for this study was collected from the online crowdfunding platform Indiegogo. Indiegogo is a successful peer-to-peer crowdfunding platform founded in 2008 with the initial purpose of raising funds for film-making, and, soon after, for other artistic endeavors. As of 2014, Indiegogo hosts a wide variety of campaigns, from music and film to small business, technology, and non-profit causes, allowing campaign founders to solicit money for an idea, charity or start-up business. Indiegogo has raised a total of approximately 271 million dollars since its founding, and, as of December 2014, over nine million people visit the site on a monthly basis.

For the purposes of this analysis, we have elected to examine the relationship between language used in the request for money, gender of founder-entrepreneur and of donor, and

campaign success in the **Technology** and **Small Business** categories. Our choice was motivated by the desire to constrain the type of fundraising market we are looking at to commercial products and services. By way of comparison, other crowdfunding categories such as Community, Education, and Health include donation requests for non-profit and humanitarian causes; and the remaining categories pertain to sponsorship of artistic projects such as Music, Film, Photography or Theatre.

Among these campaigns, we have selected campaigns in the English language, run by solo entrepreneurs (teams of one), with campaign text longer than 50 words. This allows us to compare and contrast the effect of language on campaign success. In order to assess the role of gender in crowdfunding success, we have created a list including all participant names (founder-entrepreneurs and donors-contributors) and used the Social Security database to probabilistically assign gender based on recorded correlations between newborn gender and name choice.

The final dataset includes 9,943 campaigns from business and technology created between February 2, 2010 and December 25, 2013. Of these, 66 percent were campaigns related to small business projects, and 34 percent technology project campaigns. Women were solo entrepreneurs on 35 percent of the small business campaigns, and on approximately 10 percent of the technology campaigns.

Topic Models

In order to allay concerns that linguistic patterns identified are correlated with the nature of the business projects proposed we have used a topic modeling algorithm to categorize campaigns (Blei and Lafferty 2009). Topic modeling looks for words that co-occur in documents in rates higher than expected if the words were distributed randomly. Words that have a relatively high

co-occurrence are grouped together into topics, with each topic represented as a weighted distribution over all unique words in the corpus. The content of each topic is related to the top weighted words per topic. Each document is then represented as a weighted distribution over every topic. The top weighted topics for each document indicates what that document is about. Topic modeling has been used in Political Science to determine what issues are salient in political texts, and is often used to analyze how focal issues change over time or how focal issues differ between politicians.

The majority of campaigns on Indiegogo include a written description of the proposed project or product. Each campaign in our dataset is one document, and is represented as a weighted distribution over 10 topics. For examples of the highest weighted terms for each topic, which suggests what that topic is about, see **Table 1**.

Insert Table 1 here

We are interested in analyzing how language affects how much money a campaign raises regardless of the type of project being proposed. We include the topic weights in our models to hold constant the nature of the proposed campaign, isolating word use as key variables.

Dependent Variables

Money Raised. In our analysis, we focus on the amount of money raised by a campaign as a dependent variable for testing Hypotheses 1 and 3. We have chosen money raised as a measure of success because many Indiegogo campaigns have “flexible” terms – entrepreneurs receive all the money regardless whether they have met their fundraising target or not.

Content. In order to evaluate Hypotheses 2 a-d, we have sought to operationalize inclusive-, emotional-, financial- and vivid language and use them as dependent variables. To operationalize these concepts, we have employed the Linguistic Inquiry Word Count (LIWC) dictionary (Tausczik and Pennebaker 2010). We compared this dictionary against campaign texts, and scored the relative frequencies of words in each category as a ratio to total words in that category in a campaign text to the total length of the campaign text. The linguistic content categories chosen were: **positive emotion, perception** (to measure the vividness of the campaign description), **money** (to measure the profit-orientation), and **inclusiveness** (to measure relational orientation). Examples of words from each category can be seen in **Table 2**.

Insert Table 2 here

Independent Variables

Composition. In order to address concerns that we may be capturing holistic qualities of the text that are correlated with the topics of interest, we control for three compositional characteristics: lexical diversity, readability, and concreteness. *Lexical diversity* is widely accepted as a measure of writing and speech proficiency which may affect investor decision and be correlated with our content categories. The lexical diversity index is measured as a type-token ratio by comparing the number of unique, different words (types) to the total words used (tokens). *Readability* estimates the years of education needed to understand a piece of writing. To calculate readability we used McLaughlin's SMOG formula (McLaughlin 1969), which is a function of the average sentence length and the number of polysyllable words (words with three or more syllables) in each sentence. In the context of crowdfunding we may expect simpler language (more

readability, lower reading grade) is correlated with broader appeal from a wide range of investors, but also with certain content—for example, business vocabulary may be lower in readability than language related to emotion.

Concreteness measures the extent to which the language in a text is specific as opposed to general. Both storytelling and business explanations may include both concrete and abstract language, but we would expect business language to be more concrete. More abstract language may allow more potential investors to connect with an idea, or it may be too vague to convince. The concreteness score is calculated using the lexical resource WordNet (Princeton University 2010). Among other things, WordNet organizes English nouns and verbs hierarchically through hypernyms and hyponyms. Each sense of each noun has set paths of hypernyms to reach 25 root (general) words, and each verb the equivalent to 9 root verbs. For example, the word "furniture" has this path to its root word, "entity": furniture -> furnishing -> instrumentality -> artifact -> whole -> object ->physical_entity -> entity. Compare this to chair: chair -> seat -> furniture-> furnishing -> instrumentality -> artifact -> whole -> object ->physical entity -> entity.

The length of the path to the root word is equivalent to a word's number of hypernyms—the more hypernyms a word has the more specific it is. “Furniture” has 7 hypernyms while “chair” has 9, making “chair” more specific than “furniture.” We calculated the number of hypernyms for each noun and verb in the text. The specificity score is an average over all of the nouns and verbs in a text. Since the maximum depth (concreteness) encountered in the text was slightly over 9.6, we have transformed this into a fraction between zero and one, where one is the most concrete campaign text in the dataset.

Controls – Campaign Attributes. As additional controls few included the campaign start date; campaign duration (log transformed, in days); the number of campaign images, log- transformed ($\ln(\text{image}+1)$); campaign video (binary measure, yes or no); the length of the campaign text (log-transformed); campaign goal (sum of money requested, log-transformed); and currency type (pound or dollar). Descriptive characteristics of our dataset are presented in **Table 3**.

Insert Table 3 here

ANALYTICAL STRATEGIES

Our analytical approach consists in three types of models. We first test Hypotheses H1 a-d and H3 using an ordinary least squares estimator, with log-transformed money raised as a dependent variable, fixed effects for campaign type, and robust standard errors. Second, we test Hypotheses H2 a-d using the proportion of words in a campaign text from each type of linguistic category as dependent variable, and a fractional logit estimator, implemented via STATA as a generalized linear model with logistic link and robust standard errors.

Lastly, in order to allay concerns that men- and women-ran campaigns differ significantly not only in campaign gains, but also in terms of other campaign characteristics as well, we construct a matched sample of campaigns using coarsened exact matching (CEM), with entrepreneur gender (female) as treatment and male-run campaigns as control (Iacus, King and Porro 2008). We apply the CEM algorithm using maximal information, and generate strata with different numbers of treated and control units; to compensate for the differential strata sizes, CEM returns weights to be used in subsequent analyses. We then rerun the first two estimators

explained above with CEM weights on the subsample of matched data consisting in 4,180 campaigns out of the total 8,793 campaigns in the sample.

FINDINGS

Our first hypotheses propose a relationship between linguistic content and crowdfunding campaign success. **Table 4**, model 1 represents a baseline model assessing the relationship between campaign characteristics and money raised. Model 2 includes the linguistic terms we are interested in. The results confirm our Hypotheses 1a, c, and d indicating that business language is negatively correlated with money raised, while positive emotion and inclusive language are positively correlated with fundraising success. We fail to find support for Hypothesis 1 b, linking vivid language to campaign success.

Insert Tables 4, 5 and 6 here

Table 5 presents the relationship between gender of campaign founders and language used, as estimated using fractional logit estimators. These results confirm our Hypotheses 2 a-d, indicating that women are more likely to use language related to positive emotion, vividness, and inclusion, and less likely to use money/ business related terms. **Table 6** models 1-4 tests these relationships using the CEM algorithm, with gender (female) as treatment. These models suggest effects consistent with those presented in Table 5.

Lastly, we turn to Hypothesis 3 to examine the relationship between the entrepreneur's gender, language and money raised, and the mediating role of language, in **Table 4** models 3 and

4, and **Table 6** models 5 and 6. The former results suggest that the effect of entrepreneur's gender (female) on money raised by the crowdfunding campaign is slightly diminished when considering linguistic attributes; a multiple mediation analysis reveals this effect to be around 10.4 percent. However, the mediation effect of language is stronger in the matched sample results presented in **Table 4**, accounting for 15-20 percent of the effect of gender on money raised.

ROBUSTNESS CHECKS

In order to evaluate the robustness of our findings we have examined several alternative specifications. We have considered controlling for the pre-defined campaign types (small business and technology) instead of the ten types resulting from the topic model classification. We have also examined only the subset of campaigns that raised at least one dollar (under the assumption that campaigns that failed to raise any money may have not been seen). Lastly, we have examined only campaigns that did not feature a video, in order to eliminate paralinguistic cues, body language, and visual and auditory gender cues, and found results consistent with our theory. When matching this latter subset of campaigns on all other characteristics with gender as the treatment variable, we find that linguistic content (positive emotion, inclusiveness, vividness, and business language) fully mediates the relationship between gender and money raised by the crowdfunding campaign.

Prior research has suggested that women are less likely than men to benefit from investments due to the fact that many gatekeepers in the venture capital industry are male, and they are prone to in-group bias. More recent research on crowdfunding campaign outcomes indicates that women perform better than men in crowdfunding campaigns due to the female

investors (Greenberg and Mollick 2014). Using our data, we have examined patterns of male and female donations (ratio of male and female donations and total money donated to total donations and total money donated) and have found that the linguistic characteristics proposed have similar impacts on male and female donors. Thus, although women are more likely than men to donate to other women's campaigns, language partially accounts the positive relationship between female entrepreneurs and female donors. Conversely, the negative relationship between female entrepreneurs and donations by men becomes stronger when we control for linguistic characteristics, suggesting that both male and female donors respond with the same valence to the linguistic content our theory focuses on.

Another study examining the relationship between gender and crowdfunding has proposed that men and women donors are motivated by different benefits from crowdfunding, such that women are more likely to donate for altruistic reasons, while men sometimes donate in order to receive attractive products and perquisites (Marom, Robb, and Sade 2014). We are currently in the process of manually classifying the types of benefits awarded to investors for every level of participation in the campaigns we are studying, such that we will be able to ascertain whether participants requested a significant benefit in exchange for their investment or not.

DISCUSSION

Our results are robust, but they are not without limitations. One of these limitations the lack of a quality variable in our dataset, such that we cannot account for the fact that some campaigns may receive more funding than others because they offer a more interesting product, or a more

credible business idea. We have sought to overcome this problem in two main ways. First, we have controlled for composition-level characteristics of campaign text such as level of abstractness, lexical diversity and readability. Second, we have used coarsened exact matching on campaign characteristics with gender as a treatment variable in order to restrict the analysis to a region of common support in the data.

In terms of external validity, crowdfunding campaigns represent an investment context different from traditional settings. First, investors are not experts, but rather individuals from the general population who are making investment decisions without an expectation of financial returns. In this sense, crowdfunding is more similar to investment in non-profit causes, where individuals contribute for non-material reasons, but may nevertheless evaluate ideas in order to estimate their soundness or odds of success. Second, traditional investment agents (such as angel investors or bank loan officers) make investments with the purpose of diversifying the risk: this consideration is less important in crowdfunding, where interest in the idea itself or in the promised products and services may drive investment decisions. However, prior research on business and movie pitches suggests that the linguistic dimensions considered (positive emotion, vividness, inclusiveness, and business language) are important in traditional settings as well. While in these settings entrepreneurs' business plans and financial forecasts may be subject to close scrutiny, the pitch itself often serves to estimate the entrepreneur's vision and ability to inspire—both dimensions that may be correlated with linguistic content of the type proposed in this study. In future research we plan to examine linguistic content characteristics of successful and failed social entrepreneurship business ideas, in order to evaluate the role that these choices play in an offline, centralized investment decision.

CONCLUSION

In this study we identify an institutional setting (crowdfunding platforms) that benefits female communication patterns and contributes to the reduction of gender inequality in the economic sphere. Prior research on organizations and negotiations (for a review, see Kray and Thompson 2005) is replete with social mechanisms accounting for why women have worse outcomes in the economic sphere than men, from receiving funding and support as independent entrepreneurs (Wu and Chua 2012), to being hired and receiving raises and promotions within organizations. Many of these studies key in on differences in self-confidence (Hargittai and Schafer 2006), social capital, risk behaviors, and gender discrimination or stereotyping; and insignificant differences in skill or actual performance (Hargittai and Schafer 2006; Kalleberg and Leicht 1991). This research suggests that women are less like to ask for favor, and more likely to be modest in appraising their skills (Hargittai and Schafer 2006).

However, most these field studies are predicated on interactional gender categorization, which is inherent in face-to-face interactions such as interviewing, negotiating, presenting a pitch, or working together. Our study suggests that online environments may mitigate interactional gender categorization because the mediated (virtual) environment allows participants independence from their physical persona and can downplay the prominence of gender as a salient social category.

Our theory also suggests scope conditions on gender categorization, such that women and men may be evaluated differently eliciting free-form descriptions because of systematic differences in the linguistic choices of men and women. This contributes to a new research stream that is bringing computational methods, sociolinguistics, and sentiment analysis into

sociology and organization theory to inform our understanding of social, cultural, and organizational practices.

Table 1. Topic models for campaign categories (Latent Dirichlet Allocation, 10 topics)

Energy	Campaigns	Family	Small Business	Education
energy	folk	children	busi	book
electr	explain	dream	company	student
water	trust	busi	servic	product
product	campaign	love	loan	busi
system	contributor	would	family	publish
power	describe	family	invest	social
engin	cant	donat	provid	market
manufactur	remind	kid	job	com
technolog	nois	life	money	women
light	ask	community	local	communiti
design	introduce	school	pay	compani
develop	break	mother	month	global
patent	doesn't	home	market	educ
vehicle	tell	live	small	service
research	contribut	know	income	young
test	entir	abl	property	event
car	real	give	owner	entrepreneur

Art	Perks	Phone apps	Devices	Food & Beauty
game	shirt	app	3D	food
artist	websit	user	print	anim
music	print	devic	machin	product
site	donat	softwar	batteri	bar, natur
video	cloth	applic	product	restaur
youtub	item	develop	prototype	room
art	shop	mobil	design	bodi
com	store	data	charg	line
advertis	com	android	color	paint
websit	thank	system	devic	love
record	card	phone	manufactur	open
platform	onlin	server	price	organ
studio	design	file	materi	made
creativ	plus	iphon	case	beauty
busi	page	control	power	local
fan	busi	access	ship	car
entertain	inventori	featur	part	man

Table 2. Categories of language content, and sample words (LIWC dictionaries).

Money	Inclusive	Vividness	Positive emotion
bank	and	aroma	accept
bargain	along	audible	admire
bought	both	bright	charm
dollar	include	delicious	enthusiastic
donate	inside	delicate	laugh
fee	open	harmony	love
sale	with	rotten	splendid
own	plus	heavy	glamorous
invest	we	picture	adore

Table 3. Descriptive statistics.

Variable	Mean	Std. Dev.	Min	Max
Money (ln)	1.548	2.610	0.000	14.223
Vividness	0.011	0.010	0.000	0.145
Pos emo	0.030	0.013	0.000	0.111
Business	0.025	0.017	0.000	0.146
Inclusive	0.030	0.025	0.000	0.154
Goal (ln)	9.336	1.699	6.215	21.416
Start date	19308	246	18306	19680
Duration (ln)	3.778	0.617	0.000	6.346
Images (ln)	0.959	0.528	0.000	3.829
Video	0.032	0.176	0.000	1.000
Lexical div	0.557	0.096	0.226	0.923
Concrete	0.746	0.034	0.627	0.961
Readability	11.221	2.262	3.000	31.983
Words (ln)	6.008	0.718	3.850	8.808

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Money (ln)													
2 Vividness	0.08												
3 Pos emo	0.08	0.08											
4 Business	-0.16	-0.17	-0.03										
5 Inclusive	0.04	-0.04	0.13	0.04									
6 Goal (ln)	-0.04	-0.05	-0.07	-0.02	0.06								
7 Start date Duration	0.08	0.03	0.02	-0.03	-0.04	0.07							
8 (ln)	-0.07	0.00	-0.01	-0.01	0.01	0.21	-0.24						
9 Images (ln)	0.18	0.10	0.01	-0.11	-0.02	0.02	0.09	-0.02					
10 Video	0.17	0.03	0.02	-0.06	0.02	0.04	0.05	-0.02	0.06				
11 Lexical div	-0.17	-0.02	0.02	0.13	0.01	-0.17	-0.06	-0.05	-0.21	-0.05			
12 Concrete	0.06	0.10	-0.05	0.03	-0.02	-0.04	0.03	-0.01	0.05	-0.02	0.06		
13 Readability	0.03	-0.07	-0.09	-0.01	0.02	0.15	0.01	0.04	0.03	0.01	-0.08	-0.14	
14 Words (ln)	0.26	0.06	-0.02	-0.19	-0.01	0.21	0.07	0.05	0.28	0.08	-0.88	-0.02	0.16

Table 4. Language and money raised by campaign (Hypotheses 1 a-d).

	Money raised Model 1	Money raised Model 2	Money raised Model 3	Money raised Model4
Female			0.500*** (0.064)	0.461*** (0.064)
Business		-11.229*** (1.498)		-11.015*** (1.493)
Vividness		1.839 (2.645)		1.261 (2.636)
Inclusive		2.916* (1.225)		2.634* (1.219)
Positive emotion		11.544*** (1.906)		10.464*** (1.902)
Goal (ln)	-0.125*** (0.015)	-0.124*** (0.015)	-0.110*** (0.015)	-0.111*** (0.015)
Start date	0.000*** 0.000	0.000*** 0.000	0.000*** 0.000	0.000*** 0.000
Duration (ln)	-0.234*** (0.044)	-0.236*** (0.044)	-0.229*** (0.044)	-0.231*** (0.044)
Images (ln)	0.438*** (0.062)	0.419*** (0.062)	0.427*** (0.062)	0.411*** (0.061)
Video	2.071*** (0.197)	2.021*** (0.195)	2.074*** (0.196)	2.028*** (0.195)
Pounds (baseline \$)	-0.230** (0.071)	-0.230** (0.071)	-0.211** (0.070)	-0.212** (0.070)
Text length (ln)	1.625*** (0.077)	1.547*** (0.078)	1.583*** (0.077)	1.512*** (0.078)
Readability	-0.013 (0.011)	-0.008 (0.011)	-0.014 (0.011)	-0.009 (0.011)
Lexical diversity	6.317*** (0.550)	6.006*** (0.553)	6.163*** (0.548)	5.881*** (0.550)
Concreteness	2.711*** (0.779)	3.323*** (0.782)	2.686*** (0.777)	3.264*** (0.780)
R squared	15.2	16.1	15.9	16.6
Deg freedom	19	23	20	24

Note: Constant term and control for product type omitted from the analysis. N=8,793. ~ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 5. Gender and language (fractional logits).

	Vividness	Positive emotion	Business	Inclusive
	Model 1	Model 2	Model 3	Model 4
Female	0.069*** (0.020)	0.078*** (0.011)	-0.030* (0.015)	0.029*** (0.008)
Goal (ln)	-0.013* (0.006)	-0.004 (0.003)	0.001 (0.004)	0.017*** (0.002)
Start date	0.000 (0.000)	0.000* (0.000)	0.000 (0.000)	0.000 (0.000)
Duration (ln)	0.021 (0.016)	0.004 (0.009)	-0.004 (0.011)	-0.006 (0.006)
Images (ln)	0.089*** (0.016)	0.021* (0.009)	-0.050*** (0.014)	-0.005 (0.007)
Video	0.071~ (0.043)	0.057* (0.025)	-0.077~ (0.040)	0.069*** (0.019)
Pounds (baseline \$)	-0.014 (0.024)	0.027~ (0.014)	0.029 (0.018)	0.007 (0.010)
Text length (ln)	0.166*** (0.034)	0.038* (0.017)	-0.235*** (0.022)	-0.039** (0.012)
Readability	-0.022*** (0.006)	-0.014*** (0.003)	0.007* (0.003)	0.009*** (0.002)
Lexical diversity	0.966*** (0.246)	0.326** (0.118)	-0.749*** (0.157)	-0.230** (0.088)
Concreteness	1.309*** (0.330)	-1.171*** (0.165)	0.707*** (0.210)	-0.265* (0.126)
Chi-squared	875.8	666.7	2502.4	280.8
Deg freedom	20	20	20	20

Note: Constant term and control for product type omitted from the analysis. N=8,793. ~ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

Table 6. Coarse Exact Matching – Gender, Language and Success.

	Vividness	Pos. emotion	Business	Inclusive	Money received	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Female	0.054~ (0.030)	0.100*** (0.014)	-0.044* (0.021)	0.025* (0.011)	0.461*** (0.074)	0.539*** (0.074)
Vividness					10.230** (3.710)	
Positive emotion					14.299*** (2.808)	
Business					-19.461*** (2.303)	
Inclusiveness					4.694** (1.810)	
Lexical diversity	0.383 (0.301)	0.03 (0.097)	0.689*** (0.151)	-0.046 (0.080)	-3.560*** (0.438)	-3.852*** (0.441)
Readability	-0.009 (0.012)	-0.018*** (0.004)	0.002 (0.005)	0.015*** (0.003)	0.095*** (0.017)	0.090*** (0.017)
Concreteness	2.505*** (0.593)	-1.106*** (0.242)	0.922* (0.372)	-0.158 (0.181)	4.849*** (1.100)	4.168*** (1.104)
Chi-squared	30.88	84.35	33.79	34.39		
R-squared					0.0595	0.0326
Deg freedom	4	4	4	4	8	4

Note: N=4,810. ~ p<0.1, * p<0.05, ** p<0.01, *** p<0.001

REFERENCES

- Anna, A. L., G. N. Chandler, E. Jansen and N. P. Mero. 1999. "Women Business Owners in Traditional and Non-Traditional Industries." *Journal of Business Venturing* 15(3): 279-303.
- Anthony, Denise. 2005. "Cooperation in microcredit borrowing groups: identity, sanctions, and reciprocity in the production of collective goods." *American Sociological Review* 70(3):496-515.
- Barbulescu, Roxana, and Matthew Bidwell. 2013. "Do women choose different jobs from men? Mechanisms of application segregation in the market for managerial workers." *Organization Science* 24(3):737-56.
- Berger, Allen N. and Gregory Udell. "The Economics of Small Business Finance: The Roles of Private Equity and Debt Markets in the Financial Growth Cycle." *Journal of Banking and Finance* 22(6-8):613-673.
- Blei, David M. and John D. Lafferty. 2009. "Topic Models." Pp. 71-94 in Srivastava, Ashok N. and Mehran Sahami (eds), *Text Mining: Classification, Clustering, and Applications*. Boca Ration, FL: Chapman & Hall/CRC.
- Brownlow, Sheila, and Leslie A Zebrowitz. 1990. "Facial appearance, gender, and credibility in television commercials." *Journal of Nonverbal Behavior* 14(1):51-60.
- Carney, Dana R, Amy JC Cuddy, and Andy J Yap. 2010. "Power posing brief nonverbal displays affect neuroendocrine levels and risk tolerance." *Psychological Science* 21(10):1363-68.
- Coleman, Susan and Alicia Robb. 2009. "A Comparison of New Firm Financing by Gender: Evidence from the Kaufmann Firm Survey Data." *Small Business Economics* 33(4):387-411.

- Elsbach, Kimberly D. and Roderick M. Kramer. 2003. "Assessing Creativity in Hollywood Pitch Meetings: Evidence for a Dual-Process Model of Creativity Judgments." *The Academy of Management Journal* 46(3):283-301.
- Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Boston: Harvard University Press.
- Greenberg, Jason, and Ethan R Mollick. 2014. "Leaning In or Leaning On? Gender, Homophily, and Activism in Crowdfunding." Retrieved on December 31, 2014 from http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2462254
- Hargittai, Eszter and Steven Shafer. 2006. "Differences in Actual and Perceived Online Skills: The Role of Gender." *Social Science Quarterly* 87(2): 432- 448.
- Iacus, Stefano M., Gary King and Giuseppe Porro. 2008. "Matching for Causal Inference Without Balance Checking." http://gking.harvard.edu/_les/abs/cem-abs.shtml.
- Kalleberg, Arne L. & Kevin T. Leicht. 1991. "Gender and Organizational Performance: Determinants of Small Business Survival and Success." *Academy of Management Journal* 34(1): 136-161.
- Kramer, Cherris. 1977. "Perceptions of female and male speech." *Language and Speech* 20(2):151-61.
- Kray, Laura J. and Leigh Thompson. 2005. "Gender Stereotypes and Negotiation Performance: An Examination of Theory and Research. In Staw, Barry and Roderick M. Kramers (eds) *Research in Organizational Behavior Series*, 26:103-182. Greenwich, CT: JAI Press.
- Marom, Dan, Alicia Robb, and Orly Sade. 2014. "Gender Dynamics in Crowdfunding (Kickstarter): Evidence on Entrepreneurs, Investors, Deals and Taste Based Discrimination." *NBER working papers*.

- Massolution. 2013. 2013CF- Crowdfunding Industry Report. Retrieved from <http://www.crowdsourcing.org/editorial/2013cf-the-crowdfunding-industry-report/> on December 29, 2014
- McMillan, Julie R., A. Kay Clifton, Diane McGrath, and Wanda S. Gale. 1977. "Women's Language: Uncertainty or Interpersonal Sensitivity and Emotionality." *Sex Roles* 3(6): 545-599.
- McLaughlin, G. Harry. 1969. "SMOG Grading – A New Readability Formula." *Journal of Reading* 12(8):639-646.
- Mulac, Anthony and Torborg Louisa Lundell. 1994. "Effects of Gender-Linked Language Differences in Adults' Written Discourse: Multivariate Tests of Language Effects." *Language and Communication* 14(3):299-309.
- Newman, Matthew L., Carla J. Groom, Lori D. Handleman, and Janes W. Pennebaker. 2008. "Gender Differences in Language Use: An Analysis of 14,000 Text Samples." *Discourse Processes* 45:211-236.
- Pollack, Jeffrey M., Matthew W. Rutherford, and Brian G. Nagy. 2012. "Preparedness and Cognitive Legitimacy as Antecedents of New Venture Funding in Televised Business Pitches." *Entrepreneurship Theory and Practice* 36(5):915-939.
- Princeton University. 2010. "About WordNet." Word Net. Princeton University. <http://wordnet.princeton.edu>.
- Rothman, Alexander J. and Peter Salovey. 1997. "Shaping Perceptions to Motivate Healthy Behavior: The Role of Message Framing." *Psychological Bulletin* 121(1):3-19.
- Shane, Scott and Daniel Cable. 2002. "Network Ties, Reputation, and the Financing of New Ventures." *Management Science* 48(3):364-381.

- Stuart, Toby and Olav Sorenson. 2003. "The Geography of Opportunity: Spatial Heterogeneity in Founding Rates and the Performance of Biotechnology Firms." *Research Policy* 32:229-253.
- Tannen, Deborah. 1990. *You Just Don't Understand: Women and Men in Conversation*. New York: William Morrow & Company.
- Tausczik, Yla R., and James W Pennebaker. 2010. "The psychological meaning of words: LIWC and computerized text analysis methods." *Journal of language and social psychology* 29(1):24-54.
- Thomson, Rob and Tamar Murachver. 2001. "Predicting Gender from Electronic Discourse." *British Journal of Social Psychology* 40(2):193-208.
- Wu, Zhenyu and Jess H. Chua. 2012. "Second-Order Gender Effects: The Case of U.S. Small Business Borrowing Cost." *Entrepreneurship Theory and Practice* 36(3):443-463.