

Social Entrepreneurship, Language, and Funding: Evidence from Tech Startups in Sub-Saharan Africa

June 28, 2016

Abstract

Social ventures, characterized by the “double bottom line” of profitability and social impact, have become an increasingly recognized model of entrepreneurship. Particularly in developing economies, in which economic growth in itself is often characterized as a social good, the line between social entrepreneurship and more traditional commercial entrepreneurship can be unclear. I investigate this tension by employing computational methods of text analysis on a sample of over 800 startups in sub-Saharan Africa. Using both supervised and unsupervised methods, I create measures of the degree to which each firm is oriented towards social impact based on their marketing language. I then examine the relationship between this orientation and funding outcomes.

Keywords: Social Entrepreneurship, Africa, Venture Capital, Natural Language Processing, Latent Dirichlet Allocation

1 Introduction

Social entrepreneurship has emerged as a nascent field of study in recent years, as researchers have recognized it as a distinct and separate phenomenon from commercial entrepreneurship. While precise definitions continue to be a matter of debate, social ventures are generally characterized as having different missions, goals, skill sets, and performance measures from more traditional entrepreneurial ventures (Dees 1998, Peredo and McLean 2006, Martin and Osberg 2007). To borrow some of the buzzier language characteristic of its enthusiasts: social entrepreneurship harnesses the transformative power of entrepreneurial behavior to tackle persistent social problems, thereby providing a “millennialist vision of harmony between private sector initiatives and public sector values” (Cho 2006). Social entrepreneurs, then, rather than being merely profits-driven, pursue the “double bottom line” of financial success and social impact.

One of the key features emphasized by scholars of social entrepreneurship is that it exists to correct instances of market failure – to step in when traditional market forces fail to meet a social need (Austin, Stevenson et al. 2006). But what precisely constitutes market failure when considering a developing country context, particularly one in which a majority of economic activity takes place in informal markets to begin with? In countries facing

entrenched problems related to poverty, infrastructure and public health, nearly any entrepreneurial venture that generates profits and employment could plausibly be construed as serving a social need. In this type of setting, then, does the line between commercial and social entrepreneurship become fuzzier?

These questions are particularly relevant to sub-Saharan Africa, where a surge in mobile and internet penetration has led to unique opportunities for innovation. In the trailblazing paradigm of M-PESA, the Kenyan mobile payment system, entrepreneurs have sought to develop lightweight, mobile- or web-based solutions that are unique to local needs (Hersman 2013). These innovations are frequently touted as engendering a “leapfrog” effect, bypassing the need for outdated infrastructure (Bornman 2012). While technology sectors from Silicon Valley to Singapore tend to be characterized, at the most fundamental level, as existing to solve society’s problems, the ICT sector in SSA is more likely to be described in terms of social impact, particularly by multinational players in the sector (Marchant 2015). Meanwhile, the “hubs” frequently utilized by tech entrepreneurs – typically co-working/incubator hybrids, a regional answer to the expense of workable office space and reliable internet access – have varying levels of emphasis on social impact. Likewise, some, but not all, venture capital firms with a focus on Africa explicitly mention social impact in their mission statements, typically alongside profitability and sustainability (Norton 2015).

Receiving venture funding from external sources is a crucial outcome for early-stage entrepreneurs, particularly in credit-constrained regions like sub-Saharan Africa. A survey of entrepreneurs conducted by the Tony Elumelu Foundation across 53 African countries found that 86.9% of respondents said that access to financing was “not easy at all” (2015). Virtually all the respondents in the survey had initially financed their company with their own savings, and by borrowing from family and friends. Ben White, founder of the VC4Africa network, has noted: “Innovative early stage ventures that have the potential to yield high social and environmental impact but require less than \$1 million in capital are the most difficult segment of the SME pipeline to reach Often times the ventures have a minimal track record and lack the collateral needed to secure, e.g., debt capital from a local bank” (2015). Access to a significant source of external funding is likely then to be a highly relevant performance metric for small African technology firms.

If it is possible for ventures to position themselves deliberately along the continuum from social impact to profits-oriented, by means of marketing cues, are there funding-related advantages to presenting in a particular way? What are the language cues that most strongly indicate this orientation, and are there relationships between these cues and the amount, type and source of funding?

I seek to answer these questions through a unique application of text analysis tools, employing supervised and unsupervised methods to understand linguistic features associated with social entrepreneurship, and then examining the relationship between these features and funding outcomes. This analysis was inspired in part by interviews with entrepreneurs in Lagos, Nigeria, the largest city in Africa and an emerging center for technology entrepreneurship.

nant aid discourse that permeates much ICT for development work in Africa, as well as the more recent ways in which multinational tech companies view their purpose in the country—such large economic actors might have a difficult time justifying their participation in the ecosystem without referring in some way to social impact” (2015).

This strategy can be seen reflected in the language used by venture capital firms active in the region. While the primary focus of these funds is profitable returns, positive social impact is often mentioned in the same breath. As an example, Netherlands-based eVA Fund notes that their purpose is “investing capital and knowledge to strengthen small and medium sized internet related enterprises thus securing and creating jobs and income for large African communities and with that, generating attractive financial returns for investors” (2016). Novastar Ventures stresses that they are “in search of businesses where positive social impact for lower-income households is a result of commercial success, not an end in itself” (2016).

Ventures that mirror this type of language might then be attractive to these potential funding sources – to a degree. As the director of an incubator in Lagos told this author, regarding venture capitalists and social impact: “Sure, they care, but they don’t really care.” It seems, then, that paying lip service to social impact may be helpful in attracting funding, but is by no means a substitute for a strong profit-generating business model. This would suggest that the best funding outcomes are achieved by startups that employ both types of language in their company marketing.



Figure 2: Word Cloud Generated from Nigerian Startups

The idea that firms that span the boundary between commercial and social goals may be most successful in attracting funding runs counter to main thesis of categorization theory, which suggests that firms or products that span multiple categories are disadvantaged (Hsu, Hannan et al. 2009). A possible explanation for this is that boundary spanning in mission (profit versus social goals) is fundamentally different from boundary spanning across industries or product categories. Indeed, the idea of a dual mission is part of the fundamental appeal of social entrepreneurship: essentially, the notion that you can have your cake and eat it too.

It is also possible that this particular effect is specific to the developing country context and may not be present in rich economies.

It is also important to note that while acquiring funding is crucial to these small firms' success, this is only one metric that is correlated with, but not entirely representative of, a firm's performance. Particularly because most of the firms in this sample have only one round of seed funding, this metric represents more of a "gut level" indicator of success that may or may not correspond with viable growth. It is likely that this commercial-social orientation may matter less for acquiring subsequent rounds of funding, as actual financial performance will become more important for securing further support. This is a question that I would like to investigate further, as the technology sector in sub-Saharan Africa continues to grow and more data becomes available.

3 Methodology and Hypotheses

To understand and track how entrepreneurs present themselves, the data for the text analysis is taken from Crunchbase, a global platform with detailed data on startups. Using this platform, I created a sample of 844 startups in sub-Saharan Africa (across Kenya, Nigeria, Uganda and Ghana). Each company provides a description of roughly one or two paragraphs to describe their activities, and the primary purpose of the platform is to generate exposure to potential funding sources. Figures 1 and 2 show word clouds created from the descriptions of startups based in Kenya and Nigeria.

The first approach taken to understand the relationship between language and social entrepreneurship is a supervised one, leveraging the Mechanical Turk (mTurk) platform to recruit subjects to label the firms as either social entrepreneurship or not based on their descriptions. This type of crowdsourcing approach to annotating data is well established in sentiment analysis and other natural language tasks (Snow, O'Connor et al. 2008). Each rater was asked to read the descriptions of 30 firms and give a rating of zero ("not social entrepreneurship") or one ("social entrepreneurship"). By averaging across all raters, I generate a score for each firm between that ranges between zero and one, from most to least social.

The second approach is unsupervised and provides a complimentary method of assessing the social orientation of the firms. I employed Latent Dirichlet allocation (LDA), a well-known generative topic model, to discover topic distributions across the startups' descriptions (Blei, Ng et al. 2003). By finding the most stable topic model and labeling the resulting topics, I computed the relative proportions of the social impact topics and the commercial, profit-related topics within each firm's description. These proportions serve as a measure of the social orientation of the firm that can be validated against the mTurk ratings.

The initial conjecture discussed above lends itself to the following two related hypotheses:

Hypothesis 1: Firms with greater mTurk rater disagreement on whether or not they constitute social entrepreneurship will have better funding outcomes.

Hypothesis 2: Firms with high proportions of both highly social and highly commercial topics in their descriptions, based on the LDA model, will have better funding outcomes.

These measures of social orientation should be closely related, and similar results across both methods of measuring the independent variable should add strength to the argument that startups that emphasize both their profitability and their potential for social impact will achieve the greater success in attracting funding.

4 Results: mTurk Scores and Funding

To acquire external measures of the firms' social orientations, social entrepreneurship was defined for the raters on mTurk as "innovative, social value creating activity that can occur within or across the nonprofit, business or government sectors" (Austin, Stevenson et al. 2006). Each rater was given 30 companies to rate, and quality checks were inserted to ensure that they were reading carefully. Ultimately, each firm was rated by an average of between 15 and 16 mTurk users. The resulting "mTurk Score" for a particular firm is the mean of all these users' ratings, ranging from zero (if all users agreed that firm did not constitute social entrepreneurship) to one (if all users agreed that it did). Across the firms, the mean mTurk Score was 0.44, with a standard deviation of 0.21.

As a check on whether the users understood the task, Table 1 shows the words selected by a Lasso regression of the mTurk Score on the words in the companies' descriptions.¹ The words are ordered by the size of the coefficient, with the most negative coefficients being predictive of a lower, less social score, and the largest positive coefficients predicting a higher, more social score. The resulting model is highly intuitive, with words like "poverty", "charity", "communities", "youth" and "schools" having the largest coefficients, and the most negative coefficients belonging to words like "immediately", "leading", "provider", "develops", and "branches". Appendix Tables 1 and 2 display results from a similar process run only on the lower half and the upper half of the scores, respectively. This provides some insight into what type of language distinguishes low from middle scores, and middle from high scores.

Hypothesis 1 predicted that firms with more user disagreement over whether or not they constituted social entrepreneurship – that is, scores in the middle of the range – would have better funding outcomes than those on either extreme. Figure 3 displays the proportion of firms funded at each range of the mTurk Score. An inverted-U pattern is clearly evident, with the peak appearing to be slightly toward the less-social part of the spectrum. Table 2 shows evidence for this same effect, with a logistic regression of the funding indicator run on a quadratic model of the mTurk Score. This effect persists even when controlling for country

¹This model treats each description as a sparse document-term matrix, in which the covariates are each word in the vocabulary and the values are the count of times those words appear in the description. Regularization via the Lasso adds a penalization term (proportional to the sum of the absolute values of the coefficients) to the optimization process (Tibshirani 1996). A shrinkage parameter λ controls the extent of the regularization (when λ is equal to zero, the results are identical to OLS). Because of the nature of the absolute value term, Lasso produces sparse coefficient vectors, which is highly useful in this case due to the number of terms in the vocabulary. The optimal level of the shrinkage parameter was selected via cross-validation with ten folds, choosing the model that minimized mean squared error.

(Model 1), for the four most common industries (Model 2), and for all industry fixed effects (Model 3).

Table 1: Words Selected by LASSO model predicting mTurk score

Feature	Coefficient	Feature	Coefficient
immediately	-0.03	solving	0.01
leading	-0.02	start-up	0.01
provider	-0.02	entrepreneurs	0.01
develops	-0.02	hub	0.01
branches	-0.01	children	0.01
nairobi	-0.01	directions	0.01
email	-0.01	sustainability	0.02
firm	-0.01	accessibility	0.02
company	-0.01	programs	0.02
securities	-0.01	patients	0.02
delivery	-0.01	together	0.02
management	-0.01	entrepreneurship	0.02
agency	-0.01	healthcare	0.02
strong	-0.01	create	0.02
exchange	-0.01	improve	0.02
limited	-0.01	clean	0.02
telecommunication	-0.00	affordable	0.02
eastern	-0.00	solve	0.02
bank	-0.00	access	0.02
solution	-0.00	women	0.02
web	-0.00	farmers	0.03
online	-0.00	urban	0.03
subsidiary	-0.00	raise	0.03
water	0.00	health	0.03
enough	0.00	teach	0.03
give	0.00	non-profit	0.03
mentoring	0.00	enhancing	0.03
job	0.00	awareness	0.03
solar	0.00	social	0.03
share	0.00	scale	0.03
foundation	0.00	lives	0.03
pool	0.01	people	0.04
materials	0.01	community	0.04
study	0.01	educational	0.04
better	0.01	civil	0.04
interact	0.01	sustainable	0.05
opportunity	0.01	africans	0.05
discover	0.01	households	0.05
incubator	0.01	students	0.06
education	0.01	schools	0.06
leveraging	0.01	youth	0.07
care	0.01	communities	0.08
supporting	0.01	charity	0.10
rural	0.01	poverty	0.10
forum	0.01	(Intercept)	0.40

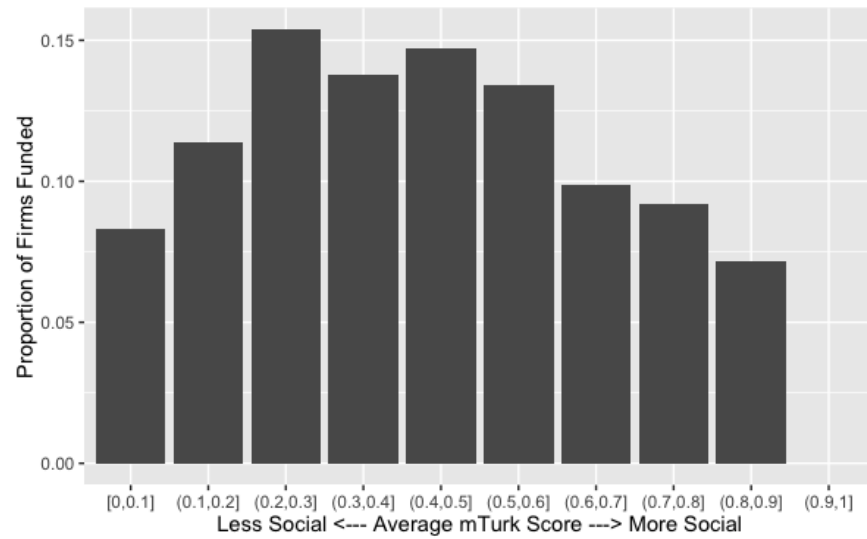


Figure 3: Proportion of Firms Funded by mTurk Score

Table 2: Logistic Regression of Funding Dummy on mTurk Score

	Funding indicator		
	Model 1 (1)	Model 2 (2)	Model 3 (3)
mTurk Score	4.37* (2.55)	4.82* (2.65)	7.46** (3.74)
mTurk Score*2	-5.17* (2.69)	-5.27* (2.78)	-7.75* (3.98)
Kenya	-0.03 (0.37)	0.14 (0.38)	0.49 (0.57)
Nigeria	0.14 (0.36)	0.24 (0.37)	0.81 (0.56)
Uganda	0.45 (0.44)	0.53 (0.45)	0.72 (0.69)
Software		1.01*** (0.34)	2.03*** (0.50)
Mobile		0.82** (0.33)	0.30 (0.65)
E-Commerce		0.73* (0.38)	1.28** (0.56)
Education		0.59 (0.44)	1.85*** (0.62)
Constant	-2.84*** (0.64)	-3.40*** (0.69)	-5.66*** (1.05)
All 200+ Industries	No	No	Yes
Observations	844	844	844

Note:

*p<0.1; **p<0.05; ***p<0.01

Because this sample consists of primarily early-stage African startups, most of which have no external funding, the funding indicator is the most useful outcome variable for this analysis. However, it is useful to see whether any particular type of funding drives the observed effect. Table 3 displays logistic regression models for various different funding indicators. The quadratic effect discussed above can be observed for firms acquiring their first funding round, but does not appear to predict subsequent funding rounds (although there may not be enough data to demonstrate this effect, as only a minority of firms have more than one funding round). Similarly, there appears to be a borderline significant quadratic effect for acquiring seed funding, but not for venture or private equity funding. Ideally, follow-up research with access to a larger sample of funded firms would plumb these contradictions further. However, it may be that spanning the commercial-social boundary is most crucial for acquiring early stage funding, and this orientation matters less for subsequent funding acquisition.

Table 3: Specific Types of Funding Outcomes

	Funding Indicators			
	First Funding Round	Subsequent Funding Rounds	Seed Funding	Venture/PE Funding
	(1)	(2)	(3)	(4)
mTurk Score	8.58** (3.98)	-10.48 (14.49)	8.40* (4.40)	0.18 (8.18)
mTurk Score*2	-8.67** (4.20)	6.72 (18.75)	-7.30 (4.48)	-0.38 (9.24)
Kenya	0.68 (0.64)	-3.19** (1.58)	0.01 (0.56)	1.93 (1.46)
Nigeria	1.14* (0.61)	-4.72** (2.17)	-0.49 (0.58)	1.01 (1.46)
Uganda	0.74 (0.77)	-0.51 (2.19)	-0.11 (0.70)	0.69 (1.96)
Constant	-6.33*** (1.14)	-2.57 (2.90)	-5.30*** (1.20)	-5.35** (2.17)
All 200+ Industries	Yes	Yes	Yes	Yes
Observations	824	844	844	844

Note:

*p<0.1; **p<0.05; ***p<0.01

5 Results: Topics

The topic modeling portion of the analysis employed a LDA model to discover a distribution of topics across the company descriptions. Figure 4 displays the top ten most likely words from each topic for a model with 22 topics, a number that was chosen with a combination of human semantic judgment and the maximization of the harmonic mean (Griffiths and Steyvers 2004).² Many of these topics are industry specific – for example, Topic 2 appears to

²There are various statistical methods to validate the choice of the optimal number of topics, although some research has found that measures of perplexity or log likelihood on a held-out corpus often do not match well with human judgments of semantic meaning (Chang et al. 2009). I determined, based on semantic judgment, that the most meaningful topics emerged somewhere between 20 and 25 topics. Appendix Figure 1 displays the harmonic mean as the number of topics varies from two to 50, reaching the maximal point at

be related to health, Topic 6 is energy, and Topic 19 is real estate. Others appear to indicate more universal language, such as Topic 1, a collection of words related to product or service quality. The most obviously social topic is Topic 14 – “development”, “people”, “economic”, “sustainable” – while still others, like Topic 10, appear to relate to more general business concerns. Figure 5 displays these same topics, sorted by average estimated prevalence across the documents.

[1] service world one market quality years services customers value products	[2] health medical insurance access care healthcare patient million services child	[3] bank company financial services group banking business one limited exchange	[4] games network social create new end creation currently lagbook include	[5] mobile platform service money payments payment via services phone provides	[6] energy fuel solar water waste local distribution power clean cost	[7] data information use management easy organizations communication accounting manage solution	[8] company media technology limited newspaper graphic nairobi ict printing east	[9] news social content music business online entertainment users website one
[10] solutions business marketing services development technology web company software design	[11] business entrepreneurs companies technology capital market businesses investment community provide	[12] students university education school learning research institutions educational institute higher	[13] farmers small market smallholder scale seed markets food quality farm	[14] development people economic sustainable foundation organization non youth life social	[15] travel location events city services information hotel local bus app	[16] first waste company name years united now since year time	[17] platform people help service make online us world way free	[18] delivery online ng range latest food fashion order get shop
[19] real estate job project jobs property projects properties professionals employers	[20] sms mobile tech premium content voice messaging data bulk technological	[21] online buyers products sellers commerce shop cars consumer marketplace offer	[22] tours national safaris safaris wildlife gorilla rwanda packages east ment					

Figure 4: Top Ten Most Probable Words from Each Topic

As a check on the validity of determining where the topics fall on the social-commercial orientation, Figure 6 shows the coefficients from a regression of the average mTurk Score of each firm on its estimated topic proportions. These coefficients are sorted by magnitude, from most predictive of a high score (Topic 14: Development/Non-Profit) to least (Topic 3: Banking).³ While the topic proportions are correlated with the mTurk Scores, there is potential for variance. Table 4 explores this variance by regressing the funding indicator on the topic proportions and on each of the topic proportions interacted with the mTurk Score. The results of Model 2 indicate that firms with a high proportion of two of the less social topics (Topic 10: Business Solutions, and Topic 18: Delivery) are significantly more successful in achieving funding when they have a higher overall mTurk Score – that is,

22 topics.

³These topic labels are my own, for ease of reading the tables.

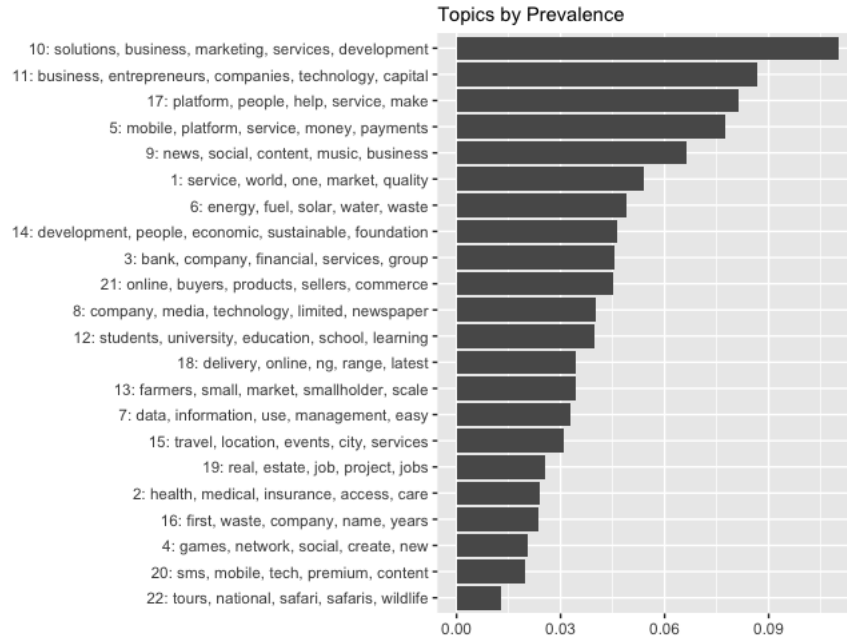


Figure 5: Topics Sorted by Estimated Prevalence in Company Descriptions

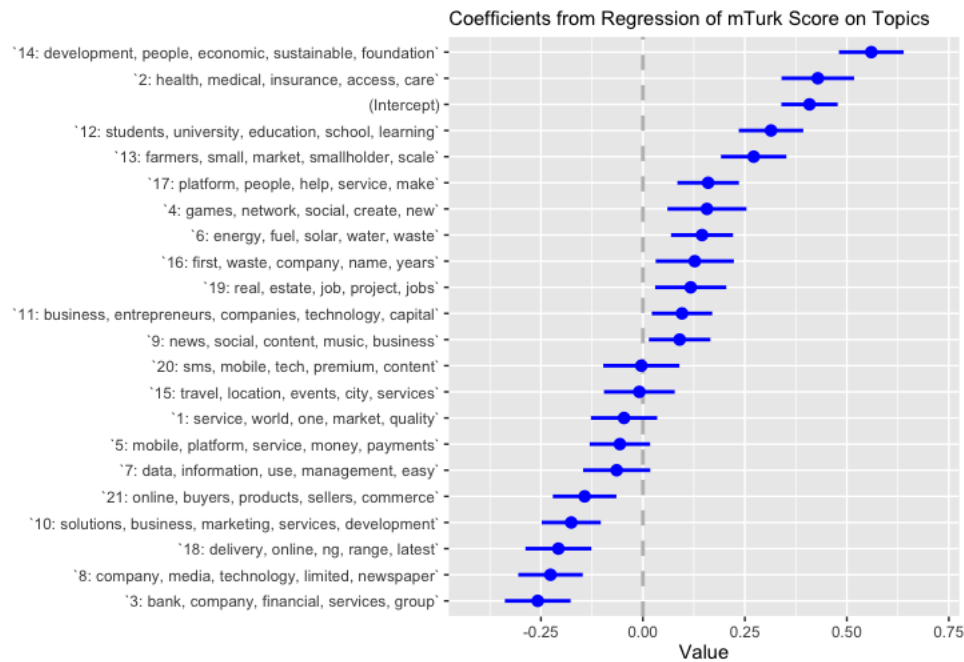


Figure 6: Topics Sorted by Relationship with mTurk Score (most to least social)

perhaps when these topics are balanced out by some more social content. Similarly, two of the more social topics (Topic 12: Education and Topic 17: Service) have negative interactions with the mTurk Score, indicating that firms with high proportions of these topics are more successful when deemed by the mTurk users as less social overall.

Table 4: Logistic Regression of Funding Dummy on Topic-mTurk Score Interaction

	Funding Indicator	
	Model 1 (1)	Model 2 (2)
T1: quality	-2.83 (1.88)	-4.87 (5.15)
T2: health	-2.33 (2.17)	-7.72 (8.41)
T3: banking	-4.95** (2.50)	-14.55** (7.26)
T4: gaming/networks	-0.10 (1.84)	-5.43 (5.67)
T5: mobile platforms	-1.34 (1.51)	-6.52 (4.64)
T6: energy	-0.51 (1.50)	-4.56 (4.68)
T7: data	-1.12 (1.70)	-8.75* (5.29)
T8: media	0.42 (1.49)	-4.31 (4.36)
T9: content	-0.05 (1.46)	-5.88 (4.62)
T10: business solutions	0.25 (1.40)	-7.13* (4.33)
T11: entrepreneurship	-0.59 (1.46)	-3.61 (4.41)
T12: education	-0.09 (1.52)	-1.17 (4.81)
T13: farming	-1.02 (1.67)	-8.24 (5.96)
T14: development/non-profit	-0.14 (1.56)	-8.90 (6.24)
T15: events	0.22 (1.64)	-7.84 (5.06)
T16: time	-3.38 (2.47)	-5.63 (6.20)
T17: service	-0.24 (1.46)	-1.96 (4.60)
T18: delivery	-0.79 (1.64)	-10.53** (5.15)
T19: real estate	0.21 (1.65)	-5.15 (4.91)
T20: SMS	-0.38 (1.80)	-2.50 (4.79)
T21: e-commerce	-1.51 (1.66)	-6.01 (4.76)
Score x T1: quality		-7.99 (7.63)
Score x T2: health		0.22 (10.80)
Score x T3: banking		12.45 (11.55)
Score x T4: gaming/networks		0.15 (7.23)
Score x T5: mobile platforms		0.46 (4.72)
Score x T6: energy		-2.08 (3.72)
Score x T7: data		6.52 (6.95)
Score x T8: media		-1.41 (4.10)
Score x T9: content		1.96 (3.50)
Score x T10: business solutions		7.82*** (2.81)
Score x T11: entrepreneurship		-4.46 (2.94)
Score x T12: education		- 7.59* (4.36)
Score x T13: farming		3.47 (6.47)
Score x T14: development/non-profit		5.19 (6.14)
Score x T15: events		6.15 (5.67)
Score x T16: time		-6.27 (10.10)
Score x T17: service		- 7.02* (3.91)
Score x T18: delivery		13.90** (7.08)
Score x T19: real estate		0.53 (5.39)
Score x T20: SMS		-7.69 (6.21)
Score x T21: e-commerce		-1.85 (6.08)
Score x T22: safaris		-13.62 (12.44)
Kenya	-0.11 (0.38)	-0.12 (0.40)
Nigeria	0.03 (0.37)	0.04 (0.39)
Uganda	0.18 (0.45)	0.24 (0.48)
Constant	-1.37 (1.37)	3.63 (4.18)
Observations	844	844

Note:

*p<0.1; **p<0.05; ***p<0.01

Hypothesis 2 suggested that firms with high proportions of both highly social and highly commercial topics would be most successful in attracting funding. As a way of testing this, I interacted the least and most social topics, separately, with each of the other topics. Table 5 displays a logistic regression of the funding dummy on the topic proportions and on each of the topics interacted with Topic 3: Banking (for brevity's sake, only the Topic 3 main effect and the significant interactions are displayed). Three of the most significantly positive interactions occur with highly social topics (energy, education, and farming). However, there are also positive interactions with low-social topics (media). Table 6 displays a similar analysis on the most social topic, Topic 14: Development/Non-Profit. Again, the two most significantly positive interactions are with more commercial topics, media and content, though there is also a borderline significant interaction with a more social topic, energy. As many of these topics appear to be closer to industry indicators, some subtlety may be lost in interacting them. However, the evidence tentatively indicates that a high prevalence of both social and commercial topics can be a successful strategy for attracting funding.

Table 5: Interactions with Banking Topic

	Funding Indicator	
	(1)	(2)
T3: banking	-4.95** (2.50)	-137.99* (73.29)
T3 x T1: quality		156.66* (93.64)
T3 x T5: mobile platforms		165.38* (91.92)
T3 x T6: energy		209.25** (97.88)
T3 x T8: media		224.72** (111.20)
T3 x T9: content		174.20* (92.83)
T3 x T11: entrepreneurship		162.26* (86.09)
T3 x T12: education		243.36** (112.34)
T3 x T13: farming		236.77** (113.92)
T3 x T14: development/non-profit		130.53* (74.01)
T3 x T19: real estate		231.65* (126.10)
Kenya	-0.11 (0.38)	-0.02 (0.40)
Nigeria	0.03 (0.37)	0.15 (0.39)
Uganda	0.18 (0.45)	0.35 (0.47)
Constant	-1.37 (1.37)	-0.49 (1.45)
Observations	844	844

*p<0.1; **p<0.05; ***p<0.01

All individual topics and interactions with T3 are included.

Only significant interactions are displayed.

Table 6: Interactions with Development/Non-Profit Topic

	Funding Indicator	
	(1)	(2)
T14: development/non-profit	−0.14 (1.56)	−3.88 (3.17)
T14 x T6: energy		18.90* (9.65)
T14 x T8: media		34.85** (17.08)
T14 x T9: content		47.06** (20.54)
Kenya	−0.11 (0.38)	−0.13 (0.40)
Nigeria	0.03 (0.37)	0.09 (0.38)
Uganda	0.18 (0.45)	0.34 (0.47)
Constant	−1.37 (1.37)	−1.37 (1.45)
Observations	844	844

*p<0.1; **p<0.05; ***p<0.01

All individual topics and interactions with T14 are included.

Only significant interactions are displayed.

6 Discussion and Conclusion

This analysis is limited both by the nascent state of the African startup industries and by the availability of good data on these companies. However, the results point to an interesting pattern, in which blurring the lines between social and commercial entrepreneurship appears to be an advantageous strategy for sub-Saharan African ventures seeking early-stage funding. The mechanisms behind, and the limitations of, this effect are questions that are ripe for further exploration. Additional data, as well as industry growth, should help to answer in particular these questions: Does this effect only persist for first-round funding? Does the origin of the funding matter? Does spanning the commercial-social boundary correlate positively with actual financial performance? And, perhaps most interestingly: is this effect limited to developing economies?

In future investigations, I hope to probe the limitations of the theory laid out in this paper and the extent of its validity in other geographical contexts. While the context and motivation of investors in rich countries may be different, there may be parallels found in the much-parodied, vaguely inspirational language of Silicon Valley. A future analysis will examine a similar sample of U.S.-based companies, to determine whether social entrepreneurship takes on a different meaning in a rich-world context, and how funding acquisition in this context is impacted by firms' self-presentation.

7 References

- (2015). Unleashing Africa's Entrepreneurs: Improving the Enabling Environment for Start-ups, The Tony Elumelu Foundation Africapitalism Institute.
- (2016). "eVA Fund." from <http://www.eva-fund.com/about>.
- (2016). "Novastar Ventures: About Us." from <http://www.novastarventures.com/-about>.
- Austin, J., et al. (2006). "Social and Commercial Entrepreneurship: Same, Different, or Both?" *Entrepreneurship Theory and Practice* 30(1): 1-22.
- Blei, D. M., et al. (2003). "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3.
- Bolino, M. C., et al. (2008). "A Multi-Level Review of Impression Management Motives and Behaviors." *Journal of Management* 34(6): 1080-1109.
- Bornman, E. (2012). "The Mobile Phone in Africa: Has it Become a Highway to the Information Society or Not?" *International Conference on Communication, Media, Technology and Design*.
- Carstens, M. (2013). Dear African VC, angel: international investors are stealing your lunch, Ventureburn.
- Chang, J., et al. (2009). "Reading Tea Leaves: How Humans Interpret Topic Models." *Neural Information Processing Systems*.
- Cho, A. H. (2006). "Politics, values and social entrepreneurship: A critical appraisal." *Social entrepreneurship* 280.
- Dees, J. (1998). "The Meaning of Social Entrepreneurship."
- Griffiths, T. L. and M. Steyvers (2004). "Finding scientific topics." *PNAS* 101.
- Hersman, E. (2013). "Mobilizing Tech Entrepreneurs in Africa." *innovations* 7(4).
- Highhouse, S., et al. (2009). "An Organizational Impression Management Perspective on the Formation of Corporate Reputations." *Journal of Management* 35(6).
- Hsu, G., et al. (2009). "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." *American Sociological Review* 74(1): 150-169.
- Lounsbury, M. and M. A. Glynn (2001). "Cultural Entrepreneurship: Stories, Legitimacy, and the Acquisition of Resources." *Strategic Management Journal* 22.
- Marchant, E. (2015). "Who is ICT Innovation For? Challenges to Existing Theories of Innovation, a Kenyan Case Study." *CGCS Occasional Paper Series on ICTs, Statebuilding and Peacebuilding in Africa*.
- Martin, R. L. and S. Osberg (2007). "Social Entrepreneurship: The Case for Definition." *Stanford Social Innovation review*.
- Norton, R. (2015). "Venture Capital Investors." 2016, from <http://www.siliconcape.com/page/venture-capital-investors>.
- Peredo, A. M. and M. McLean (2006). "Social entrepreneurship: A critical review of the concept." *Journal of World Business* 41(1): 56-65.

Snow, R., et al. (2008). "Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks." Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Journal of the Royal Statistical Society 58(1).

White, B. (2015). Angel Investors Invest \$27 Million In African Startups Listed On VC4Africa. M. Nsehe, Forbes.

Yang, T. and H. E. Aldrich (2011). "All Organizations Were Once New: Revisiting Stinchcombe's Liability of Newness Hypothesis." Frontiers of Entrepreneurship Research 31(4).

Zott, C. and Q. N. Huy (2007). "How Entrepreneurs Use Symbolic Management to Acquire Resources." Administrative Science Quarterly 52.

8 Appendix

Table A1: Words that distinguish low from middle mTurk scores

Feature	Coefficient	Feature	Coefficient
firm	-0.64	language	0.15
price	-0.58	users	0.16
(Intercept)	-0.57	africa	0.17
ltd	-0.42	startups	0.17
solution	-0.41	employers	0.18
booking	-0.40	send	0.19
analytics	-0.38	helps	0.20
africa's	-0.36	seed	0.20
record	-0.36	whole	0.20
brand	-0.28	promote	0.22
turn	-0.27	ideal	0.23
name	-0.23	major	0.23
nigeria's	-0.23	university	0.23
within	-0.21	initiative	0.24
post	-0.19	energy	0.24
organisation	-0.17	small	0.25
management	-0.15	model	0.26
states	-0.14	public	0.27
headquartered	-0.14	free	0.28
bank	-0.11	financing	0.32
banking	-0.10	unlimited	0.33
mobility	-0.09	african	0.33
sale	-0.08	cook	0.34
kingdom	-0.07	open	0.34
goods	-0.06	founders	0.39
provider	-0.04	take	0.39
always	-0.04	realized	0.39
using	0.02	june	0.42
school	0.03	members	0.45
potential	0.03	used	0.45
unlike	0.04	text	0.45
family	0.04	activities	0.48
government	0.04	hot	0.50
enhancing	0.05	favorite	0.51
sustainable	0.05	platform	0.51
investment	0.05	social	0.53
enable	0.05	places	0.55
source	0.06	annually	0.57
long	0.07	export	0.58
videos	0.09	news	0.59
lifestyle	0.09	countries	0.63
income	0.09	b2b	0.63
material	0.11	improvement	0.64
manage	0.11	people	0.68
portal	0.12	event	0.71
provide	0.12	established	0.78
formed	0.13	spend	0.79
extra	0.13	around	0.80
daily	0.13	relevant	0.81
local	0.14	territory	0.82
europe	0.15	channels	1.37
classes	0.15		

Table A2: Words that distinguish middle from high mTurk scores

Feature	Coefficient	Feature	Coefficient
engineering	-0.67	connect	0.03
whole	-0.57	another	0.04
app	-0.45	ideas	0.04
user	-0.45	women	0.04
investment	-0.40	solving	0.07
software	-0.39	post	0.08
ideal	-0.39	improved	0.08
leading	-0.39	life	0.09
enable	-0.38	africans	0.10
seamless	-0.36	people	0.10
customers	-0.36	organisation	0.10
(Intercept)	-0.35	booking	0.11
sell	-0.35	waste	0.11
entertainment	-0.31	awareness	0.12
places	-0.31	economic	0.13
manage	-0.31	start-up	0.13
unlimited	-0.31	towards	0.13
word	-0.30	urban	0.13
fashion	-0.30	united	0.16
capital	-0.27	farmers	0.16
general	-0.26	feed	0.18
product	-0.25	education	0.18
google	-0.24	scale	0.18
clients	-0.24	recruitment	0.20
analysis	-0.24	households	0.22
media	-0.21	mission	0.23
magazine	-0.20	lives	0.23
sellers	-0.18	price	0.24
company	-0.17	social	0.24
companies	-0.17	healthcare	0.24
news	-0.17	pool	0.25
investor	-0.16	patients	0.26
generate	-0.16	health	0.26
government	-0.15	study	0.27
started	-0.15	share	0.28
owned	-0.15	access	0.29
regulatory	-0.14	sells	0.30
territory	-0.12	improve	0.31
transactions	-0.10	entrepreneurship	0.31
reviews	-0.09	networking	0.32
nairobi	-0.09	supported	0.33
money	-0.09	raise	0.34
strategy	-0.09	interests	0.35
wallet	-0.07	foundation	0.40
base	-0.07	problems	0.46
required	-0.06	enhance	0.46
african	-0.04	affordable	0.51
shopping	-0.04	students	0.52
hot	-0.02	strive	0.53
internet	-0.02	charity	0.64
advertising	-0.01	educational	0.66
stories	-0.00	youth	0.74
service	-0.00	informal	0.79
non-profit	0.01	teach	0.85
child	0.01	poverty	0.90
answers	0.02	communities	0.96
environment	0.03		

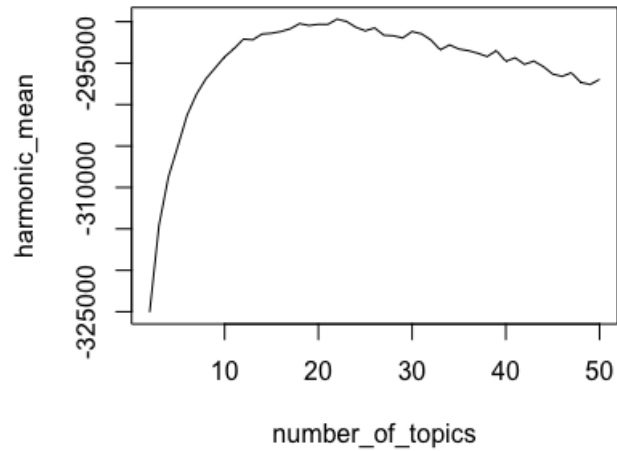


Figure A1: Maximization of Harmonic Mean for Topic Number Selection

Table A3: Regressions of Total Funding on mTurk Score

	Funding Amount: Less than...					
	\$10m	\$1m	\$400k	\$200k	\$50k	\$25k
	(1)	(2)	(3)	(4)	(5)	(6)
mTurk Score	188,883 (340,299)	21,205 (44,280)	43,249** (20,577)	24,880* (13,101)	2,031 (3,169)	2,048 (1,908)
mTurk Score*2	-150,817 (354,938)	-16,363 (46,167)	-39,834* (21,436)	-26,372* (13,642)	-1,459 (3,298)	-1,676 (1,984)
Kenya	61,072 (53,155)	-5,430 (6,979)	-4,817 (3,259)	-1,573 (2,074)	-72 (507)	310 (306)
Nigeria	50,273 (52,802)	-11,617* (6,920)	-2,791 (3,228)	-706 (2,056)	34 (501)	268 (304)
Uganda	32,511 (69,215)	-12,483 (9,075)	-6,135 (4,236)	-282 (2,697)	174 (658)	957** (397)
Constant	-46,310 (88,906)	4,573 (11,558)	-6,770 (5,393)	-3,648 (3,429)	-434 (831)	-883* (501)
All 200+ Industries	Yes	Yes	Yes	Yes	Yes	Yes
Observations	835	816	808	800	782	770
R ²	0	0	0	0	1	0

Note:

*p<0.1; **p<0.05; ***p<0.01