The Role of Soft Information in Crowdfunding: Evidence from Crowdsourced Data

October 15, 2024

Abstract

We propose a new approach to capture the role of soft information in crowdfunding markets. We define soft information as not codeable, i.e. cannot be easily web-scraped by robots. Using a crowdsourced survey where participants rate live Kickstarter campaigns, we show that human raters improve the likelihood of correctly distinguishing between a funded and a failed campaign by 9%. Our survey design also reveals how investors evaluate information with respect to multiple qualitative factors in forming funding decisions. Our results suggest that human evaluations can detect soft information and thus still have value in a world dominated by big data.

keywords: Kickstarter, crowdfunding, crowdsourcing, online survey, soft information, project evaluation criteria

JEL codes: G24

1 Introduction

This paper studies the role of soft information in raising capital. Soft information is traditionally defined as information that is nonstandard or more difficult to quantify or verify. On the one hand, soft information can reduce information asymmetries by communicating important information otherwise unavailable to market participants (Uchida et al., 2012; Iyer et al., 2016). On the other hand, soft information can be used to exacerbate behavioral biases and even mislead investors, especially when they have limited attention and processing power (Hirshleifer and Teoh, 2003). Hence, it is important to examine whether the ability to convey information improves or hurts investment decisions. In this paper, we suggest that the recent rise in FinTech provides a new opportunity to answer this question. We differ from the traditional definition of hard information to include *codeable* data, i.e. attributes which can be easily web-scraped by robots to construct big data. Soft information, on the other hand, includes cues that cannot be fully captured by structured data. According to this definition, people see everything that machines do, but also more. That is, machines can observe only hard information, while humans can detect and process additional soft cues and can hence observe both hard and soft information. Accordingly, the residual value of soft information can be decomposed by comparing the predictive ability of computer algorithms to that of human raters.

To the best of our knowledge, this study pioneers the application of crowdsourcing to assess the overarching impact of soft information on capital acquisition. We find that human ratings of live Kickstarter campaigns, obtained using Amazon Mechanical Turk, outperform machine-only prediction for fund-raising success. The marked enhancement in predictive accuracy, stemming from the integration of annotations by human evaluators, is attributed to the incremental value of soft information, uniquely processed by human raters. This incremental value indicates that existing measures of soft information may be incomplete and have limited capability in capturing soft information in its entirety. Furthermore, our study explores the mechanisms through which soft information shapes funding success. Our survey approach allows for the identification of codeable campaign features that can, to some extent, unveil soft information. We also examine how various soft evaluation criteria are aggregated to reach the funding decision. Our innovative methodology for examining how potential funders assess crowdfunding campaigns offers valuable insights for entrepreneurs, investors, and intermediaries.

We use online crowdfunding, in which soft information is the primary mode in which information is presented, to run a horse-race between machines and humans. While crowdfunding entrepreneurs have almost full control over the information released to potential backers, other asset classes allow much less flexibility. Credit markets rely on standard variables and models informing a credit score, while public securities involve highly regulated disclosure (such as a prospectus, indenture, and so forth), due diligence by institutional investors, and information intermediaries such as security analysts. Overall, the crowdfunding market involves very little regulation, oversight and standardization, making it the perfect laboratory to examine the role of soft information.

Reaching the funding goal by the deadline is the main hurdle in crowdfunding, making funding success the most valuable outcome to predict.[1] Crowdfunding platforms typically employ a "provision point mechanism," otherwise known as an all-or-nothing fundraising scheme (Bagnoli and Lipman, 1989). This mechanism was put in place to ensure funders that their money will only be invested in a project if the creators raised all the funding they believe they need to achieve their stated goal. If the amount of money pledged by the deadline is still below the required funding goal, no money is collected from the pledgers.[2] For founders, failing to achieve the funding goal can have severe reputational consequences to future projects (Li and Martin, 2019). For funders, committing their funds to campaigns that fail to reach their funding goal comes at the expense of forgoing campaigns that get funded while missing on favorable projects (Kuppuswamy and Bayus, 2017).

Our first research question asks: Do human raters have incremental value in predicting funding success atop structured data (Q1)? It is ex ante unclear whether the ability of crowdfunding entrepreneurs to communicate information more freely—and in particular, more soft information—improves the precision of investment decisions.[3] We quantify the ability to predict whether a

[1]The crowdfunding literature focuses on funding success and not on delivery, as most funded campaigns deliver. Mollick (2014), for example, reports that non-delivery is well below 5%. That said, we acknowledge that funding success does not mean that the project team actually managed to deliver the promised goods, which is outside the scope of this study.

[2]Typically, if the funding goal is not reached by the deadline, the project will not start. We acknowledge that some Kickstarter projects that did not reach their funding goal managed to obtain funding from other sources and the projects still started.

[3]While rewards are typically not purely financial (as in interest or dividends), for brevity we use creators/initiators interchangeably with "entrepreneurs", as well as backers/pledgers/funders interchangeably with "investors" throughout the paper.

campaign would be successful in achieving its funding goal based on structured data, both independently and in conjunction with human ratings. This comparative analysis enables us to assess the residual value of soft information, representing the degree to which the prediction of funding success is enhanced with the incorporation of human ratings atop structured data.

We then attempt to uncover the underlying thought process of potential investors in forming the funding decision. Real-world pleaders likely go through several phases, including information gathering and processing, evaluation of alternatives, and ultimately an investment decision. We are particularly interested in how the information available on the campaign web page is translated into various soft facets. To model this task, we conjecture a simple two-phase hierarchical process, motivated by the Analytic Hierarchy Process commonly applied to model rational decision making in management, marketing, and psychology (Saaty et al., 2008), as well as in credit granting (Srinivasan and Kim, 1987). In the first phase, agents decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. In the context of crowdfunding, pledgers evaluate all available information with respect to multiple evaluation criteria. In the second phase, agents integrate all the elements of the hierarchy to form a decision. To identify the best decision choice, agents may either apply an intuitive rule of thumb, or more explicitly estimate a weighted score of the evaluation criteria. To implement this framework, we first conduct a preliminary survey which identifies five evaluation criteria specific to the crowdfunding domain: idea, presentation, value for money, entrepreneur, and likelihood of success. We then investigate two related research questions: In the decomposition phase, how do various campaign features map into the soft factors which potential investors consider when they evaluate a crowdfunding campaign (Q2)? And then, in the integration phase, how are those qualitative factors consolidated to reach the funding decision (Q3)? For example, what is the relative importance of each of the qualitative factors, and does soft information affect this relative importance?

To answer these questions, one cannot rely solely on the standard data commonly used in the existing crowdfunding literature. In particular, because crowdfunding markets enable entrepreneurs to communicate material information even when it is nonstandard or more difficult to quantify or verify, one cannot easily capture such soft information from the existing structured data. We thus employ an online survey to capture those campaign elements not immediately available to the researchers, allowing us to take a peek into the black box of the crowd's investment decision process. The survey participants were sourced through Amazon Mechanical Turk (www.mturk.com), a crowdsourcing marketplace that allows requesters to post tasks and workers to perform these tasks for an agreed fee. The survey asks participants to rate actual live Kickstarter campaigns—i.e., in real-time when the outcome is still unknown—to ensure that our survey results can serve as a valid indication for the behavior of crowdfunding participants in the real world.

We ask participants to rate whether the project *should* be funded (i.e. based on their subjective opinion, as opposed to whether it *would* be funded, mainly driven by speculation on others). We ask participants to rate not only the projects overall, but also to rate the perceived quality of various campaign factors, such as the perceived quality of the campaign's presentation, the qualities of the entrepreneur behind the project, the perceived value for money, and the perceived likelihood of the project delivering on its promise. Overall, we paid 1,206 participants to rate 936 live Kickstarter campaigns. Each campaign was rated by 5 different participants, which we then averaged to minimize estimation error.

We followed the campaigns that were evaluated in the survey and checked whether they were successful in achieving their funding goals. We then run a horse race of human raters sourced from Amazon's Mechanical Turk against the most common algorithms, to determine the incremental effect of soft information in predicting funding success (Q1). Specifically, we compare human judgement to binary classifiers, which have been found to perform well in the prediction of credit ratings (Jones et al., 2015). We use the "Area Under the ROC Curve" (AUC), a standard measure for aggregate classification performance used in the computer science and machine learning literature, to investigate how much one can rely on human annotations in predicting whether the campaign would succeed in attracting enough funds. As we explain in more detail in Section 2.4, the AUC statistic can be used to compare between different models based on their aggregate classification performance.[4] To capture the incremental benefit of human ratings, we control for an array of "hard" variables—i.e., those that could easily be web-scraped to construct big data. Following Mollick (2014), we incorporate both live progress metrics (the funding goal relative to the share of goal raised, days left, and the number of backers at scan) employed by funders primarily for a

[4] Importantly, we estimate the AUC using the full data set, i.e. training machines on the entire set in order to give machines their best shot.

rational estimation of funding success, as well as additional codeable variables that may capture soft information to some extent. For instance, as suggested by Mollick (2014), funders may try to estimate the founder's preparedness and effort by considering variables such as the number of images and the length of videos.

Overall, our analysis shows that information contained in human annotation of campaigns can complement predictive algorithms. That is, even though crowdsourced survey participants are likely to be less knowledgeable than—and have inferior information relative to—actual backers, we find that their annotations have substantial improvement in predicting funding success. A subsample analysis further reveals that the incremental contribution of human raters really shines in projects that are not about building a physical project. This result is consistent with the notion that soft information plays a more important role in projects that do not promise a more clearly quantifiable physical reward. We further split the sample by rater characteristics in order to verify that the added value of our crowdsourced ratings is not driven by self-selection into participating in our survey of crowdfunding experts prone to outperform machines. We find that the incremental contribution of human raters persists when we only include raters with no prior crowdfunding experience. We also perform a robustness test that shows that humans can contribute even without "live" information. That is, the improvement in predicting funding success persists even when crowdsourced survey participants cannot observe actual backers. Our results thus extend to entrepreneurs seeking to use crowdsourcing tools to obtain better predictions before launching their projects in the real world.

We next ask which campaign features are more important in carrying soft information. We do so by examining how various hard features of a campaign, which have been found to predict funding success, correlate with human ratings of various project features (Q2). In particular, we use natural language-processing methods to examine the effect of sentiment. While both machines and humans can ascertain the overall tone of the project description (hard variable), only humans can fully comprehend the project description (soft information). Hence, if the sentiment score is correlated with human ratings, the project description likely carries much more soft information than merely its tone. We employ the CoreNLP sentiment system which, unlike traditional systems that categorize individual words while ignoring their order, analyses whole sentences based on the sentence structure, and computes the sentiment based on how words compose the meaning of longer phrases. We find that while the rating given to the campaign is correlated with the sentiment in the project's description, sentiment is not significant in predicting funding success, and the incremental value of human rating remains after controlling for the widely used computer-generated sentiment score. That is, while the sentiment score captures only one type of linguistic style (e.g., how the text is written), more predictive power is contained within human analysis of the *content*. We further find the rating given to the campaign is correlated with various "hard" features in presentation, such as the number of images in the campaign's webpage, and whether the project has active Facebook and Twitter accounts. Ratings are also partially correlated with some of the entrepreneur's perceived demographic traits, such as gender, age, and ethnicity.

We further ask whether soft information has only a direct effect on individual evaluation criteria scores, or does it also affect how these scores are consolidated (Q3). It is possible, for example, that soft cues may affect not only a criterion-specific score, but also the criterion's weight, in turn indirectly affecting the overall rating assigned to the project and one's willingness to fund it. Our unique data allows us to disentangle the effect of soft information on criterion-specific scores from its effect on their aggregation. Recall that we ask participants to rate not only the projects overall, but also to rate the perceived quality of various campaign factors. This allows us to estimate the actual importance of individual evaluation criteria in the data, using their marginal effect on the overall rating given to a project. Specifically, we use partial correlations between the focused ratings and overall rating. We first observe that the actual importance is generally consistent with the self-reported one, suggesting that our survey participants do not tend to overweight some aspects while underweighting others. Yet while people expect their views to be influenced by their assessment of the project's chance of success, this investment criterion does not provide additional predictive power for the overall rating. Importantly, the discrepancy between the self-reported and actual importance of specific determinants of funding success is stronger in art and music projects, in which soft information is more influential. The latter result is inconsistent with a common bias such as social desirability bias, and consistent with soft information playing a role in the thought process of human raters as they consolidate their overall view. As such, we shed more light on the channels by which soft information can explain why human raters outperform machines in predicting funding success.

Our paper contributes to three streams of literature. First, our paper contributes to the literature on soft information in financial markets. Early papers in finance focus on lender-borrower relationships and rely on geographical distance to proxy for the ability to convey soft information through interpersonal communication (Berger et al., 2005; Agarwal and Hauswald, 2010; Uchida et al., 2012). With the recent rise in FinTech, more and more information can be easily web-scraped by robots to construct big data, and so soft information becomes information which is not codeable. Most recently, Barbopoulos et al. (2021) distinguish between machines and humans accessing company information (8-K filings). They find that, while machines are better at handling numerical information, humans can better handle sequential and soft information. We find that human ratings outperform machine-only prediction for funding success.[5] Our results suggest that a human crowd of non-experts is able to detect cues in soft information beyond raw campaign features. In other words, human evaluators can "read between the lines" and form a more accurate prediction compared to the common variables.

We further examine whether machines may be able to observe at least some of the soft information by employing the widely used computer-generated sentiment score (Tetlock, 2007). We find that the incremental value of human rating remains after controlling for the sentiment in the project's description. Our result suggest that the widely used computer-generated sentiment score is rather limited in the amount of soft information it can carry. Our results highlight the importance of the "human touch," suggesting that humans are able to process soft and abstract information better than fully automated systems. Most broadly, our results suggest that human evaluations that can detect and process such soft information still have value when the world has gone toward quantitative analysis based on big data.

Importantly, our survey design allows us to shed new light not only on whether but how humans process soft information, and how they use it to make investment decisions. The existing

[5] We obviously do not mean to argue that any human's prediction will always outperform any algorithm. VLMs (Vision / Language Models) can digest both text and images and identify items within them. LLMs (Large Language Models) such as ChatGPT are more capable at handling language, however we cannot incorporate any model trained on the internet without look-ahead bias, as there is no way to ensure that it was not trained on (or can directly access) the outcome. Further, relating to soft information in crowdfunding is not the typical micro task these tools were designed to automate. Even after digesting material related to crowdfunding campaigns over a long time, it is not clear whether such tools could mimic a human-like decision based on the trustworthiness of the creator, the excitement in the reward offered, or the emotional reaction to images, ranging from attraction to disgust. Human raters are constantly exposed to various initiatives, form their own individual opinions and observe the eventual success of those initiatives. Arguably, while the gap between human annotations and automated processes will decrease as better technologies will be trained on huge volumes of data, the incremental value in human annotations could still remain significant. One persistent advantage of using crowdsourcing over algorithms is algorithm manipulation. Cohn et al. (2024) suggest that big data algorithm performance may actually deteriorate over time as users learn how to manipulate it.

literature essentially takes a black-box approach, by considering only the information presented to investors and their investment decision—while ignoring anything in between. We employ the Analytic Hierarchy Process (Srinivasan and Kim, 1987; Saaty et al., 2008) to model the underlying thought process of potential investors in forming the funding decision. We suggest that humans first evaluate all available information with respect to multiple qualitative evaluation criteria, and then integrate all the qualitative evaluation criteria to form an investment decision. This framework allows us to disentangle the effect of soft information on evaluation criteria from its effect on their aggregation. Interestingly, we find that the weights people assign to the evaluation criteria may be affected by salient attributes (Bordalo et al., 2013). Our results suggest that factors unique to human raters play an important role in weighting and aggregating the individual criteria.

Second, our paper contributes to the crowdfunding literature. We inform creators whether and how to incorporate soft information in the campaign to facilitate funding success. Existing literature on soft information revolves around credit markets, including traditional and peer-to-peer lending, and explores particular soft signals for borrower trustworthiness. Herzenstein et al. (2011) show that narratives by borrowers that contain a trustworthy identity claim are associated with increased loan funding. Duarte et al. (2012) show that borrowers who appear more trustworthy based on their photograph have higher probabilities of having their loans funded. Lin et al. (2013) show that the identity of online friends of borrowers increase the probability of successful funding. Iyer et al. (2016) report that peer lenders demonstrated 87% of the predictive power of an econometrician who incorporates standard financial borrower information. Further, crowd lenders exhibited 45% greater accuracy than models that use the borrower's credit score in predicting an individual's likelihood of defaulting on a loan. Dorfleitner et al. (2016), on the other hand, show that soft factors that are derived from the description texts in peer-to-peer lending influence the funding probability but not the default probability.

Our paper focuses on the role of soft information in reward-based and equity crowdfunding. The information asymmetry that characterizes investment decisions in new ventures is usually high, making signaling by entrepreneurs crucial to their funding success. Furthermore, it is much more difficult for an entrepreneur to predict the response of potential pledgers than it is for a borrower to predict the response of lenders. While borrowers offer lenders interest payment schedules, crowdfunding project creators offer potential backers various types of rewards, ranging from early provision of the product to a return on investment. Most importantly, many of the projects are unique and/or novel, leaving both sides of the market in uncharted territory. This uncertainty is further exacerbated by the fact that project creators rely on pledgers from the general public, with almost no institutional, intermediary, and/or regulatory involvement. While the behavior of some potential backers is more predictable, e.g., friends or family of the entrepreneur, or a small group of enthusiasts and/or fans that will back the project almost automatically (Agrawal et al., 2015), predicting the behavior of the general public is the hardest task project creators face. We tackle this task by collecting human annotations using an online survey, uncontaminated by any offline social relationship between pledgers and creators.

Project creators are particularly struggling to predict how the crowd will respond to soft signals. Ahlers et al. (2015) show that crowdfunding entrepreneurs use signals that reduce uncertainty (e.g., providing information about risks) to increase funding success. Burtch et al. (2016) focus on soft information among campaign contributors, and it's effect on the likelihood of conversion, as well as on the average contributions conditional on conversion. Importantly, the paper focuses on soft cues only by peer pledgers, but not on signals by the creators. Consistent with the view that social interaction among investors helps reduce information asymmetry, Hervé et al. (2019) find that sociability in the investor's area of living is important in an equity crowdfunding context but does not affect participation in bond investments. Shafi (2021) find that technical information such as financial metrics disclosed in campaign descriptions does not predict funding success. Rather, funding success is related to founders' motivation and commitment. To date, existing literature has been scarce, and focuses on particular components of soft information.

We are the first to examine the *overall* effect of soft information in the crowdfunding domain. To decompose the residual effect of soft information, we control for a broad array of hard variables that are either publicly available, or that can be easily web-scraped by robots, to predict funding success. Following Mollick (2014), these include live progress, such as the share of the funding goal raised and the number of backers in the early days of a campaign (Burtch et al., 2013; Colombo et al., 2015; Kuppuswamy and Bayus, 2017; Vismara, 2018), as well as the entrepreneur's demographic traits, such as gender, age, and ethnicity (Dahlin et al., 2019; Gafni et al., 2021). We further include other codeable campaign characteristics, such as the number of images and length of videos (Xiao et al., 2014; Koch and Siering, 2015), sentiment and textual tone (Tetlock, 2007; Gafni et al., 2019), and social media and investor communication (Moritz et al., 2015).

Focusing on these hard attributes, funding success has been attracting a lot of attention within the computer science literature, which by now made machine prediction the standard measure and big data practices become the norm (Greenberg et al., 2013; Chen et al., 2014; Etter et al., 2013; Sawhney et al., 2015; Li et al., 2016; Jhaveri et al., 2019). Yet this approach may be missing soft information which machines are unable to process. We thus examine whether human annotations add value to machine prediction. We show that the machine-only prediction for funding success is outperformed by a human-augmented one. We further show that crowdfunding market participants evaluate campaigns based on soft cues embedded in five specific factors that may be predictive of a campaign's perceived quality: idea, presentation, value for money, entrepreneur, and likelihood of success. By considering soft dimensions in the context of funding success, our study offers an important incremental contribution to the crowdfunding literature. Our findings enable both entrepreneurs and crowdfunding platforms to better understand how campaigns are evaluated, and to adjust their plans accordingly.

Third, our paper contributes to the literature on the wisdom of the crowd in capital markets. Jame et al. (2016) find that crowdsourced forecasts are incrementally useful in forecasting earnings and in measuring the market's expectations of earnings, traditionally performed by professional analysts. Yet the question whether the wisdom of the crowd can add valuable information is particularly interesting in the crowdfunding domain. For crowdfounders, failing can have substantial reputational consequences and may compromise their ability to launch additional projects (Li and Martin, 2019). As such, the ability to evaluate a campaign in its very early days to try and predict funding success is a highly valuable tool. Entrepreneurs may seek to hire experts, such as venture capitalists and grant-making bodies, to evaluate their crowdfunding campaigns (Mollick and Nanda, 2015). We suggest that an easier and more affordable strategy is to employ an online crowdsourcing tool, readily available to any project creator (as opposed to proprietary or professionally designed survey) and at very low cost relative to the funding goal. Importantly, we apply classification performance measures from the machine learning literature to show that our results have external validity. As we explain in more detail in Section 2.4, the statistic we use to measure the predictive ability of crowdsourced ratings holds not only within our campaign sample, but rather also in campaigns it was not trained on. Our results thus suggest that crowdsourced ratings are informative even in highly skewed samples, in which one could easily predict funding success without using any information. We further show that crowdsourcing is valuable even without "live" information, i.e. even when survey participants are unaffected by actual backing behavior of users on the platform (Agrawal et al., 2015; Vismara, 2018). Our results thus suggest that crowdsourcing can be used before launching the project in the real world, markedly giving entrepreneurs the chance to abort a project that is unlikely to succeed and fend off the reputational consequences of a public failure. Most broadly, our results suggest that crowdsourcing the evaluation of investments can be utilized to gain insights into investment opportunities.

2 Data and Methods

2.1 Crowdsourced Ratings

We analyze the factors affecting how people rate a crowdfunding campaign using an online survey. The survey participants were sourced through Amazon Mechanical Turk (www.mturk.com), a crowdsourcing marketplace that lets requesters post tasks and allows workers to perform these tasks for an agreed fee.

In recent years, researchers have started using Amazon Mechanical Turk to recruit subjects of social science experiments and surveys. A number of benefits exist regarding the use of Amazon Mechanical Turk to study crowdfunding. Relying strictly on an online survey ensures that participants are all internet users, which corresponds more closely to potential crowdfunding backers. Importantly, given that our goal is to run a horse race between human ratings and computer algorithms, it is crucial to ensure that our survey participants do not have an offline social relationship with the founder, whether they are family, friends, or are otherwise members of the same social network as the entrepreneur (Mollick, 2014; Agrawal et al., 2015). The use of Amazon Mechanical Turk for our study minimizes any such concerns. Another important benefit to using Amazon Mechanical Turk as our survey tool is the heterogeneity in its subject pool. Classical surveys are commonly either concentrated geographically (i.e., all participants are from the same school, city, state or country) or homogeneous in other dimensions (age group, race, educational background and so forth). Goodman et al. (2013), for example, indicate that the commonly used university student subject pools have different cognitive capabilities compared to online communities in general, and thus crowdfunding surveys based on student participants would have much lower validity to actual Kickstarter participants than an online survey.

However, using Mechanical Turk to run experiments also presents a challenge. Given that Mechanical Turk participants are willing to complete tasks for little money, they may thus pay less attention than actual potential backers, and hence generate less informative assessments. To ensure data quality, we apply the standard filters in accepting participants (HIT Approval Rate of 95% on top of Mechanical Turk's automatic reputation score hurdle of 90%) (Peer et al., 2014). We further try to overcome this challenge by incorporating questions to verify that participants have actually examined the page, as well as open questions such as, "What are the main reasons for your rating?" to ensure that participants provide thoughtful ratings. To further minimize the noise in our ratings data, we assign the same campaign to several different raters, and then average all ratings across all raters assigned to the campaign. We can report that, despite substantial cross-sectional variation *between* projects, the degree of agreement among all raters *within* a specific campaign was surprisingly high, considering that the rating is very subjective: the inter-class correlation, $ICC(1, K)$ was 0.514.[6]

2.2 Live Crowdfunding Campaigns

Crowdfunding platforms, such as Kickstarter, Indiegogo, FundRazr and Crowdfunder have recently become prominent mechanisms for entrepreneurs to obtain funding for companies or projects. Kickstarter is the most popular reward-based crowdfunding website (Hoskins, 2014), and as of February 2022, it reports that over 215,000 projects have been launched, and more than 20 million backers have contributed approximately $6.4 billion (Kickstarter, 2022). Kickstarter allows users to ask for funding for various types of projects, including technical gadgets (such as the Pebble smartwatch (Chang, 2012)), movies (such as the Veronica Mars movie (Thomas, 2013)), art projects and games. Campaigns set up a scale of *rewards* for investment that range from simple acknowledgments, to providing a product to the backer, and onwards to even letting the backer play an active role in the project itself (Belleflamme et al., 2014). A creator posting their campaign on Kickstarter sets a *funding goal* that reflects the amount of funds required to start the project, and a *deadline*. Money can be pledged by backers until the project deadline expires. If the funding goal is reached, money

[6] See (Banerjee et al., 1999) for a discussion of measures of agreement between raters.

is collected from the pledgers and the creators begin the project; if the amount of money pledged by the deadline is still below the required funding goal, no money is collected from the pledgers. Hence, funders know that their money will only be invested in a project if the creators raised all the funding they believe they need to achieve their stated goal. This mechanism was put in place to remove a barrier to collecting funds from the general public, as in the past each funder was reluctant to donate money unless enough had been collected from others.

Our survey asks participants to rate actual live Kickstarter campaigns—i.e., in real-time when the outcome is still unknown. The participants were given links to Kickstarter project pages, which they were able to freely browse and examine. That is, different raters were looking at the same campaign over time (rather than given an identical snapshot at one point in time). At the time of each assignment—i.e., when a participant is assigned a campaign—we therefore recorded the share of the funding goal raised, the number of backers, and the number of days left until the deadline. We note that these real-time variables are easily available to both robots and humans as they scan through the project's page. We choose to provide direct links, rather than say a partial replica, to expose our survey respondents to the same information presented on the actual website. It is likely, for example, that Kickstarter participants are influenced by how much money was already raised and how close it is to the goal. We therefore need to ensure that our survey respondents are affected by any herding or information cascades that may be at play in real markets. Our goal is not to examine how people react to different information sets, but rather to minimize any such information gap between our survey respondents and actual Kickstarter participants. Given the same information set as actual Kickstarter participants, we can compare between human raters and the most common algorithms to determine which one prevails in predicting funding success.

2.3 Survey Design

We note that the Analytic Hierarchy Process framework (Srinivasan and Kim, 1987; Saaty et al., 2008) that we use only describes the classification process but does not dictate the particular qualitative evaluation criteria as these are domain-specific. Hence, to prepare the ground prior to conducting our main survey, we first need to explore what are the soft factors which potential investors consider when they evaluate a crowdfunding campaign. We identified these soft factors through a preliminary survey based on focus groups recruited (with no monetary incentive) from interns and staff at Microsoft research. Participants were first provided information on crowdfunding as well as a sample of Kickstarter campaigns. The preliminary survey included opened-ended/free-form questions relating to natural factors that may be predictive of a campaign's perceived quality. Five specific factors stood out: idea, presentation, value for money, entrepreneur, and likelihood of success. One can clearly see how each and every one of these five dimensions could be affected by soft cues that cannot be easily web-scraped by robots. We ran three independent groups to make sure that the factors as well as their definitions converge. Figure 1 provides the definition of each factor which we provided to our main survey participants.

Our main survey then consists of two parts. The first part is the background and criteria data set that contains the participants' responses to a questionnaire regarding their crowdfunding experience and the criteria they think should be used to rate a campaign's quality. The second part is the campaign rating data set that contains reports provided by participants assigned to rate randomly sampled crowdfunding campaigns.

To generate the background and criteria data set, our participants completed a questionnaire regarding experience with crowdfunding and knowledge of crowdfunding platforms. We asked participants if they were familiar with crowdfunding, if they had ever visited the Kickstarter webpage, and if they had ever invested in Kickstarter projects. In order to elicit the Analytic Hierarchy Process, we asked participants to compare between the five evaluation criteria which were identified in the preliminary survey by *allocating* 100 points among them according to how important they seem in determining whether campaigns, in general, are worthy of investment. The distribution of these values is plotted in Figure 2 with confidence intervals. By collecting this data we can confirm that all five dimensions identified in the preliminary survey indeed bear significant perceived importance among our main survey participants. In addition, we also use this self-reported perceived importance to examine how it compares to the actual importance of the various evaluation criteria in determining the overall rating given to a project.

To generate the campaign and rating data set, each of our survey participants examined several randomly selected Kickstarter crowdfunding campaigns and wrote a detailed report about them. We refer to the Kickstarter campaigns simply as the *campaigns* and to the participants examining them as the *raters*. Participants were paid a fee of $1 for each campaign. In order to ensure participants would not fill the questionnaire blindly, each participant reviewed at most 6 campaigns, and received \$6 if he/she completed all of them. To even the number of reports per campaign, we stopped allocating participants to review a campaign once it received more than 5 reviews. This allows us to use average campaign ratings across participants to minimize estimation error.

For each campaign, we record the funding goal and project category (whether physical/product or not, and whether artistic, game, film or musical project). We also record other features such as the number of images, length of videos, links to social media features (indicating whether the page includes a Facebook and/or Twitter link), as well as sentiment which we explain in more details below.

We asked participants to indicate the gender, age (in bins; e.g., 20–30, 30–40, 40+), and ethnicity (either White, Black, Hispanic, Chinese, Indian, or Other) of the entrepreneur, and whether the project is a charitable one. We note that the measurement of entrepreneur characteristics is likely highly uniform (if not unanimous) across raters rather than rater-specific.

Each rating report collected details about the campaign (enabling us to verify participants have actually examined the page) and an overall rating between 0 and 100 (0 being the worst and 100 being the best). Participants entered their rating scores using a drop-down menu, which did not indicate any default value, to avoid any bias elicitation in respondents' answers. We instructed participants to give a score of 20 for what they considered to be a very bad campaign, 40 for a below-average campaign, 50 for an average campaign, 60 for an above-average campaign and 80 for an excellent campaign. In addition to the overall rating given for each campaign, we asked raters to rate the campaign based only on one *specific* dimension of the evaluation criteria in Figure 1. For example, we asked raters to rate a campaign based only on the quality of idea (ignoring other criteria), or to rate the campaign based only on the quality of the presentation. We call the ratings that are based only on one aspect of the campaign *focused ratings*, as the raters focused only on one aspect of the campaign. The rating task data also allows us to estimate the relative importance of these evaluation criteria *in practice*. To estimate the *actual* importance of the evaluation criterion in determining an overall rating in the screening task, we used multiple linear regression. The regression target variable is the overall rating, and the predictor variables are the set of focused ratings (i.e., the set of ratings given to each of the evaluation criteria for that campaign). We compare the relative importance estimated from the rating data set with the self-reported one from the criteria data set. This allows us to quantitatively analyze which evaluation criteria people *expect* to affect their campaign ratings and which of them are *actually* important in determining the overall rating given to a campaign. In addition to ratings, we also asked raters what amount they would be willing to invest in the project, as well as their desire to invest in the project, in the project's category, and in Kickstarter in general.

After collecting the rating data set consisting of the participants' evaluation of the campaigns, we waited for the campaigns deadline to expire. Once the campaign's deadline passed, we could see whether it succeeded or failed in attaining the funding goal. We could then examine the relation between the participants' evaluation of a campaign and its success.

2.4 Classification Quality

Determining whether a campaign is likely to succeed or fail is a classification task: Given the ratings assigned by survey participants to a campaign, we try to assess whether the campaign would be successful in achieving its funding goal. The machine learning literature provides important insights regarding analyzing classification tasks (Bishop, 2006; Bradley, 1997; Fawcett, 2006). In the following, we discuss how such insights relate to our problem of evaluating people's ability to predict the success of a crowdfunding campaign.

Clearly, the higher a campaign is rated by survey participants, the more likely it is to succeed. If we are interested in predicting whether a campaign would succeed, we could take a small number of people sourced from a crowdsourcing market, ask them to evaluate the campaign, and examine how their ratings are distributed. For very high ratings, we would predict the campaign to succeed, whereas for very low ratings we would be quite confident the campaign would fail. In other words, the higher a campaign is rated, the more confident we are that it would succeed. However, what is the appropriate threshold rating above which we should predict the campaign would be successful?

We note that the choice of the rating threshold upon which we predict the campaign to be successful affects misclassification in two ways. One type of a misclassification is predicting a campaign to succeed when in fact it fails, which is referred to as a "false positive"; another type of misclassification is predicting a campaign to fail when in fact it succeeds, which is referred to as a "false negative." When we choose a high threshold rating to predict success, more campaigns are classified as likely to fail, increasing the number of false negatives, but decreasing the number of false positives. Conversely, if we choose a low threshold rating for predicting success, more campaigns are classified as likely to succeed, decreasing the number of false negatives, but increasing the number of false positives.

Therefore, rather than measuring the accuracy of one particular threshold, we need a summary tool that would enable us to obtain a general estimate of the extent to which one can rely on the wisdom of the crowd in predicting the success of campaigns. We use a statistical tool commonly used in machine learning called the "Area Under the ROC Curve" (AUC) as a measure of the quality of a classifier. It is based on a curve called the "Receiver Operating Characteristic" (ROC) curve. While one could easily convert the continuous predicted value of the logistic regression to a binary success/fail classification using the default threshold of 0.5, the ROC is a function plot depicting the performance of a binary classifier for different thresholds. The ROC is generated by plotting the proportion of "true positives" (instances correctly classified as positive) against the proportion of false positives (instances incorrectly classified as positive). The AUC is a summary statistic from the ROC plot. It is the total area between the x-axis and the ROC function, ranging between 0.5 (worthless, random classifier) and 1 (perfect classifier).[7]

The AUC is particularly informative when the classifier is likely to operate on a skewed distribution. Rather than merely reflecting the most prevalent outcome in the sample, the AUC measures quality as if campaign success was difficult to predict (as in close to 50-50).[8] The rating assigned by the classifier to an instance reflects whether the classifier believes the instance to be a positive or negative one. A possible "test" for a classifier is to take one positive instance and one negative instance (i.e., one campaign that is successful and one that is not), and see if the classifier can work out which is the positive instance and which is the negative instance. The AUC is the probability that the classifier would succeed in this goal, reflecting the model's accuracy on campaigns it was not trained on. In our case, it is the probability that given one randomly chosen successful campaign and one randomly chosen failed campaign, the sampled rating to the successful campaign would be higher than that of the failed campaign.

[7]A general scale for AUC values is: .90–1 = excellent, .80–.90 = good, .70–.80 = fair, .60–.70 = poor, .50–.60 = fail.

[8]Consider a crowdfunding platform in which very few campaigns actually succeed in reaching their funding goal, such as a platform in which only 10% of the campaigns are successful. A classifier that examines a sample of people's ratings of campaigns and correctly classifies whether a campaign is successful in 90% of cases sounds very impressive. However, because the success distribution is very skewed, achieving this classification quality is trivial: One degenerate classifier can simply ignore the sampled ratings and predict the campaign to fail. As 90% of the campaigns in this platform do actually fail, the classifier is likely to be correct in roughly 90% of the cases. Thus, using the overall proportion of correctly classified instances as a quality metric can be very misleading.

To estimate the AUC, we first run a (multivariate) Logistic regression to predict classification success. Given a threshold, one could easily convert the continuous predicted value to a binary success/fail classification. We plot the trade-off between false positives (*type 1* error) and false negatives (*type 2* error) for every possible threshold (ROC), which allows us to compute the area under the curve (AUC) (Bishop, 2006; Bradley, 1997; Fawcett, 2006).

2.5 Sentiment

In light of the wide use of computer-generated sentiment scores, it is of great interest to study the effectiveness of such scores in allowing machines to observe at least some of the soft information hidden in written communication. In order to examine the effect of sentiment, we apply a textual analysis on the project's description, and determine whether it is positive, negative, or neutral in tone. Sentiment analysis, also referred to as "opinion mining," is a natural language-processing technique that allows extracting a subjective speaker tone or other information in a natural language text. A sentiment-analysis system allows taking a piece of natural language text, such as a sentence, paragraph, or a longer piece of text and automatically determining the speaker's expressed tone or emotion. To build the sentiment analyzer, humans first annotate a large set of such utterances with the tone expressed in them; a machine learning algorithm then processes the text to determine which linguistic features are predictive of a specific expressed tone.

Generally speaking, a sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation (see appraisal theory), affective state (the emotional state of the author when writing), or the intended emotional communication (the emotional effect the author wishes to have on the reader). In many sentiment analysis applications, we are interested in mining the general attitude expressed in the text, and thus utterances are classified as either "positive," "negative," or a "neutral" sentiment. Such sentiment analysis has been widely applied to determine tone in an investment context (Tetlock, 2007).

We used the Stanford CoreNLP library.[9] This library uses various text analysis techniques to ascertain the tone of a body of text. Unlike the previous systems, which categorize individual words while ignoring their order, the CoreNLP sentiment system computes the sentiment based on how

[9] `http://stanfordnlp.github.io/CoreNLP/`

words compose the meaning of longer phrases. It sets up a neural network that enables building a representation of whole sentences based on the sentence structure. It trains this neural net on a large body of text, thus when we use it on the Kickstarter project descriptions in our sample, it can ascertain the overall tone of the text fairly well. The sentiment score ranges from 1 to 5, with "Very negative" = 1, "Negative" = 2, "Neutral" = 3, "Positive" = 4, and "Very positive" = 5.

3 Results

We gathered the data during February and March of 2014. In total, we had $n_p = 1,206$ survey participants, yielding $n_r = 6,098$ campaign reports on $n_c = 936$ Kickstarter campaigns.

3.1 Descriptive Statistics

Table 1 presents descriptive statistics of the background and criteria data set. Panel A presents the general questions we asked participants regarding their background and prior knowledge of Kickstarter. Our participants were 57.3% male and 42.7% female, with an average age of 31.09 (the median was 29). A strong majority of the raters, 81.7%, were from the U.S., with a small contingent of 15.4% from India. Importantly, our crowdsourced survey respondents largely overlap with actual crowdfunding users: 77.7% of our participants (82.5% of U.S. participants) were familiar with crowdfunding, and 58.8% (63.7% in the U.S.) had previously visited a crowdfunding platform website, while 18% (19.2% in the U.S.) indicated that they had previously invested in a crowdfunding project.

Panel B reports the perceived importance of campaign evaluation criteria. To assess perceived importance, we ask the survey participants to consider how important each of the evaluation criteria should be in affecting their overall rating of a crowdfunding campaign and their willingness to invest in it. They were then asked to allocate 100 points among the evaluation criteria accordingly. All five factors that were identified in the preliminary survey bear significant perceived importance. The idea seems to be the most important, with likelihood of success to follow.

In its basic form, the Analytic Hierarchy Process assumes that the criteria contribute independently to the overall decision. It is possible, however, that the five factors explored in this study may interact. In Panel C, we explore whether the interactions between criteria are significant.

Specifically, we examine cross-correlations in order to identify the level of substitution between the various factors. Recall that relative importance represents the weight of each factor in the aggregated score. As such, most correlations are negative, as by definition each weight comes at the expense of others. While the five factors may be to some extent confounded, it appears that no two factors have substantial negative correlation. The strongest substitution (negative correlation) is between value for money and the idea. That is, participants that allocate high relative weight to value for money do so mostly at the expense of the idea (and vice versa), suggesting that the two factors overlap to some extent. In Section 3.4, below, we complement this analysis by examining the partial correlation between the score given to each of the factors on the overall rating given to a project, controlling for the scores in all remaining factors. This allows us to estimate the actual weight of the various evaluation criteria scores on the overall rating given to a project, which we can compare to the self-reported importance. This exercise enables us to test whether soft information affects one's willingness to fund a campaign through attenuating or exacerbating the relative importance of individual evaluation criteria.

Table 2 describes our campaign and rating data set. We randomly selected live Kickstarter crowdfunding campaigns during February and March of 2014. For each campaign, we record the funding goal and project category (whether physical/product or not, and whether artistic, game, film, or musical project). We also record other "hard" features—i.e., variables available to both robots and humans as they scan through the project's page—such as the number of images, length of videos, links to social media features (indicating whether the page includes a Facebook or Twitter link), and a sentiment score between 1 and 5 based on textual analysis of the project description ("Very negative" = 1, "Negative" = 2, "Neutral" = 3, "Positive" = 4, and "Very positive" = 5).

We allocate campaigns at random to the participants. We direct participants to live links to Kickstarter project pages, so as to minimize any information gap between actual Kickstarter participants, survey respondents, and computer algorithms. At the time of each scan—i.e., when a participant is assigned a campaign—we record the share of the funding goal raised, the number of backers, and the number of days until the deadline. These variables are important controls to disentangle between the value of human ratings and any signaling available to them from live data. It is plausible, for example, that the likelihood of funding success correlates with the share of the funding goal raised at the time of the scan (Colombo et al., 2015). By controlling for the share of the funding goal raised at the time of the scan, we will later be able to test whether entrepreneurs can use crowdsourcing tools to obtain better predictions before launching their project in the real world. Each participant had to provide an overall rating, as well as a rating based only on one *specific* dimension of the evaluation criteria. For example, we asked raters to rate a campaign based only on the quality of idea (ignoring other criteria), or to rate the campaign based only on the quality of the presentation. Ratings range from 0 to 100. In Panel A, we pool together all campaigns used in this study. Starting with the hard variables, we see that the mean funding goal was around $18K. The project page typically included several images, a video, and a Facebook link, and the tone in the project's description was neutral on average. Moving on to our crowdsourced ratings, we first note that most campaigns were assigned in their early launch days, with only around 1% of the funding goal raised and around 17 days left (most campaigns have a 30 day deadline). While the average rating was 50.27, we note that subjects did not merely default to 50. The standard deviation was 25.96, which shows a high variance in the opinions regarding the different campaigns.

Once a campaign's deadline had passed, we record the amount raised both in absolute terms as well as percent of the funding goal. *Funding success* is a binary variable with a value of 1 if the amount raised is equal or higher than the funding goal, and 0 otherwise. With the mean funding success close to 50-50, predicting funding success is challenging, implying that achieving high classification quality would be nontrivial. Nonetheless, as an initial indication that soft information processed by human raters can significantly improve predicting funding success, we find that a classifier relying solely on the mean overall rating achieves an AUC of 0.719. In Section 3.2, we will attempt to decompose the overall effect of soft information by comparing the AUC for various combinations of hard and soft variables.

We next report the subsample analysis, which allows us to examine the ratings and funding success in more homogenous subsamples. Kickstarter allows entrepreneurs to classify their campaign into one of many categories. We aggregate the Kickstarter categories in two dimensions. The first is the project outcome, which can be either a product or non-product, while the second is the project topic, which can be art, game, film or music. Our category groups are not mutually exclusive and often overlap, and each campaign can fall into multiple category groups. For example, music is also an art. Some art projects (though not all) can be a product as well. For example, a craft work is both a product (a physical reward that backers are promised), and it is also an artwork. A musical or theatrical performance is also art, but is not a product.

Panel B splits the sample by outcome type while Panel C splits the sample by project topic. Panel B shows that while product-type projects have substantially lower success in reaching their funding goal, they typically achieve higher ratings across all categories than non-product projects. One explanation would be that soft information plays a more important role in projects that do not promise a physical reward. Panel C shows that project topics associated with higher funding goals typically achieve lower funding success. Game projects typically provide the highest number of images and achieve the highest presentation ratings, as opposed to music projects that provide the lowest number of images as well as sentiment scores, and achieve the lowest presentation ratings. Overall, the univariate inspection in Panels B and C yields no clear relation between the overall rating and actual funding success. We therefore test how people compare to machines in processing soft information using multivariate analysis.

3.2 The Added Value of Human Ratings in Predicting a Campaign's Success

To what extent can one rely on the wisdom of the crowd in classifying whether a campaign will succeed or fail? We note that the mean overall rating given by our participants to successful campaigns was 56.53. In contrast, the mean overall rating of unsuccessful campaigns was 44.81 (the differences between the average rating were more pronounced with physical products than for non-product projects, though it was still true for all). This difference was significant at the $p < 1 \times 10^{-55}$ level (using a Mann-Whitney U test). Figure 3 shows a boxplot for the ratings of successful vs. unsuccessful campaigns. One can see that all five numbers (the minimum, the maximum, the sample median, and the first and third quartiles) are higher for successful campaigns. The median rating for successful campaigns is higher than the third quartile for failed campaigns, whereas the median rating for failed campaigns is lower than the first quartile for successful campaigns.

Recall however that our main research question is not about the overall value of human ratings, but on the *incremental* role of soft information atop structured data in screening crowdfunding campaigns. Specifically, we incrementally add information that may predict funding success, and in each increment estimate the classification quality. Our main interest is the improvement in classification quality when we add human ratings on top of structured data. We use measures of classification quality that we borrow from the machine learning literature. Panel A of Table 3 shows the logistic regression results as well as the corresponding AUC for different input combinations. The dependent variable is binary, indicating whether the money raised at the deadline reached (or exceeded) the funding goal. To minimize estimation error, the unit of analysis is a campaign—that is, we average all variables (including binary ones) across all raters assigned to the campaign. Below each model, we present the corresponding AUC, which summarizes how the fitted values from the corresponding logistic regression would perform (in terms of the proportion of true positives against the proportion of false positives) under all possible thresholds. Importantly, we estimate the AUC using the full data set, i.e. training machines on the entire set in order to give machines their best shot.[10] The AUC is computed based on the ROC curves shown in Figure 4. As defined in Section 2.4, the ROC plots the proportion of true positives (instances correctly classified as positive) against the proportion of false positives (instances incorrectly classified as positive). The total area between the x-axis and the ROC function is the AUC.

Column 1 uses information available on Kickstarter, including real-time progress (the share of the funding goal raised, the number of backers, and the number of days until the deadline—all measured at the time of each scan) and project characteristics (funding goal, project type and various categories). Arguably, actual backers likely perceive live progress as confirmation, validation, or reassurance of their personal interest in the campaign that may reinforce their inclination to commit financial resources. Recall however that our survey participants examined randomly selected campaigns, and are hence expected to exhibit less sensitivity to live progress compared to actual backers. Interestingly, the number of days until the deadline is significant while both the share of goal raised and amount of backers at scan are insignificant. One explanation is that the number of days until the deadline outweighs the share of the funding goal raised and the number of backers, consistent with Burtch et al. (2013). Importantly, the AUC of 0.725 shows much improvement in predicting funding success relative to a random classifier (as suggested by a mean funding success of around 0.5 in Panel A of Table 2). The AUC in this specification reflects the classification quality of a basic computer algorithm emulating funders' rational estimation of funding success based on live progress. In Column 2, we add the sentiment score based on the textual analysis of the project description. Interestingly, sentiment does not improve the AUC, suggesting that this more

[10]Had we split the set, train machines on a training set and judge them on a different validation set, then by construction machines would be trained on a smaller set and humans would have been expected to add more value.

advanced machine-generated data does not capture much of the soft information. It is also possible that the effect of sentiment is non-linear: while positive tone may attract more irrational backers, overenthusiastic overstatements might act as a red flag for more rational backers. In Column 3, we add the entrepreneur characteristics (perceived gender, age and ethnicity), and whether the project is a charitable one. While these variables are collected from our human participants, these variables are relatively definitive and principally stem from hard information. In particular, gender is insignificant in predicting funding success, which seems at odds with its effect among actual backers (Gafni et al., 2021). Arguably, while actual backers self-select into campaigns based on gender bias (e.g. a tech project by a man or a fashion project by a woman), we reiterate that our survey participants were assigned campaigns at random.

The specifications thus far, relying exclusively on structured data, indicate that the AUC estimated within the training set is limited below 0.75. It is reasonable to anticipate that prediction accuracy would be even lower when estimated on a separate testing set. For comparison, computer algorithms have not been able to reach such prediction accuracy. Greenberg et al. (2013) uses various learning algorithms to predict whether a campaign would be successful by examining the project's Kickstarter page and reached 67% prediction accuracy, which is similar to the quality of a different model proposed by Chen et al. (2014), as well as to another similar technique by Etter et al. (2013). Sawhney et al. (2015) uses a model based on various features as well as language analysis, but only reached 71% prediction accuracy, while Li et al. (2016), using dynamic analysis of project success over days and real-time social media, approaches our results. Importantly, all these algorithms require an active Kickstarter page as well as actual real-world dynamic data from the first few days of the active campaign page, whereas our method can be utilized even before the project goes live.

In Column 4, we finally introduce human ratings by adding the (average) overall rating given to a campaign. As one would expect, human ratings exhibit a high correlation with funding success. Our focus, however, is not on the statistical significance of human ratings but rather on the notable spike in the AUC, approaching 9%. This improvement reflects the residual contribution of soft information atop structured data. The existence of such a substantial residual value in soft information, as uniquely processed by human raters, indicates that current measures of soft information may be incomplete and unlikely to capture soft information in its entirety. Column 5 adds all project-specific ratings for individual evaluation criteria, showing that the individual grades given to entrepreneur and likelihood of success are statistically significant. While the statistical significance of individual coefficients in this specification is downward biased, given the inherent multicollinearity of overall rating as the weighted average of the five factors, the AUC measure remains valid. The AUC spikes to an astounding 0.93, leaving the best computer algorithms behind.

One may argue that the high AUC that we report is over-fitted. It is possible, for example, that only crowdfunding enthusiasts self-select to complete our survey, hence resulting in an overestimated AUC which is not representative of the general population. To attend to this concern, we run the horse-race only using participants which have no prior experience (i.e., answered "No" in the background questionnaire when asked "Have you invested in a Kickstarter project before?"). The results in Panel B show the addition of overall rating in Column 4 increases the AUC by around 3%, while the inclusion of all data collected from human raters in Column 5 brings the AUC to 0.83. The persistent results in the inexperienced sub-sample alleviate the concern that the incremental contribution of our human raters is driven by a "haphazard" sample, consisting only of experienced crowdfunding enthusiasts who self-select into our survey mainly because they can easily outperform machines.

The results in Panels A–B of Table 3 suggest that the predictive ability of humans may outperform that of machines. More broadly, our results suggest that focus groups still have value even when the world of marketing has moved toward big data. We next perform a subsample analysis by splitting the sample according to project outcome. Given the overlap between project outcome and project topic as previously explained (e.g. all game projects are product rewards), we drop both project- and reward-type dummies ("Is project a physical object," "Is project artistic," "Is project a game," "Is project a film," and "Is project musical") from the subsample horse-race specification. Panel C shows that in physical rewards, classification quality starts off very high, even when it is based only on hard information. It is in the non-physical rewards where the incremental contribution of the human raters truly shines. This result is consistent with the notion that soft information plays a more important role in projects that do not promise a more clearly quantifiable physical reward.

In Panel D, we split the sample by the potential effect of actual live progress. We note that the variables capturing live progress at the time of each scan (e.g., the share of the funding goal raised) are statistically insignificant, suggesting that funding success can be predicted fairly well without live information. We consider this a very encouraging finding, since the ability to evaluate a campaign in its very early days and to predict whether it would succeed is a very powerful tool. Such an evaluation may enable both entrepreneurs and investors to better understand the strengths and weaknesses of a campaign and thus make decisions accordingly. One may still argue, however, that controlling for live progress is not the same as not having any live data at all. If so, crowdsourcing before launching a campaign may not be as effective. We next devise a robustness test that does not require live progress.[11] We do so by first observing that almost all campaigns are set up with a deadline of 30 days. This observation allows us to define a subsample consisting only of ratings collected in the first few days of a campaign, before any pattern or signaling has started to form. In particular, we first filter only ratings for which the number of days left at scan is above 20. By doing so, we include only ratings collected 20 days or earlier prior to the deadline. Since we assigned campaigns randomly, many projects were sampled only later throughout their lifecycle, and thus this filter leaves us with about one-third of the campaigns in Panel A. After applying the filter, we then go back to averaging all variables at the campaign level—however, this time based purely on ratings collected in the very early days.

The results in Panel D show that humans can contribute even without "live" information. The addition of overall rating in Column 4 increases the AUC by around 5%, while the inclusion of all data collected from human raters in Column 5 brings the AUC to over 0.85. Our results can thus be extended to entrepreneurs seeking to use crowdsourcing tools to obtain better predictions before launching their project in the real world.

As such, our study gives rise to a valuable practical implication. We show that crowdsourcing can be used as a powerful tool to evaluate a crowdfunding campaign in its very early days. Such an evaluation can enable both entrepreneurs and crowdfunding platforms to better understand the strengths and weaknesses of a campaign and to adjust their plans accordingly. Project initiators can use crowdsourcing to learn more about their chances, and thus improve their campaigns prior to listing them on a crowdfunding venue. The wisdom of the crowd can convey whether crowdfunding is the best funding path for the project. If a crowdfunding campaign is predicted to fail, then the

[11] We thank an anonymous referee for this suggestion.

project may potentially be better off with another form of financing such as angels, venture capital funds, or other types of early-stage investors. Most importantly, by employing a crowdsourcing tool, entrepreneurs are better equipped to decide whether to abolish a project altogether. Indeed, many campaigns fail to achieve their funding goal.[12] For crowdfounders, failing can have substantial reputational consequences and may compromise their ability to launch additional projects (Li and Martin, 2019). Gompers et al. (2010) suggest that the perception of performance persistence—the belief that successful entrepreneurs are more skilled than unsuccessful ones—can induce such persistence. Skirnevskiy et al. (2017) find that a strong track record encourages funding from loyal backers. The authors suggest that internal social capital develops between serial crowdfounders and their previous backers.

Given the costs associated with failing, entrepreneurs thus seek opportunities to learn about—and obtain early feedback on—the quality of their campaign. Howell (2021), for example, shows that new venture competitions can help entrepreneurs learn about their projects' quality. Winning has large positive effects on measures of subsequent venture success, including employment and financing. Receiving negative feedback, on the other hand, is shown to increase venture abandonment. Entrepreneurs may also seek to hire experts, such as venture capitalists and grant-making bodies, to evaluate their crowdfunding campaigns (Mollick and Nanda, 2015). We suggest that an easier and more affordable strategy is to employ an online crowdsourcing tool. In other words, our finding that crowdsourcing is informative regarding the decisions of actual potential backers gives rise to a valuable practical implication to crowdfunding market participants. That is, by posing questions to a crowd of non-expert raters regarding a crowdfunding campaign, one can gain insights into the campaign's potential to reach its funding goal, markedly giving entrepreneurs the chance to abort a project that is unlikely to succeed and fend off the reputational consequences of a public failure.

3.3 How are (Hard) Campaign Features Translated to (Soft) Human Evaluation Criteria?

According to the Analytic Hierarchy Process framework that we use, pledgers first translate various campaign features into qualitative evaluation criteria. We thus examine how various hard features

[12] Out of all the Kickstarter campaigns considered in our survey, only 47% were successful.

correlate with the perceived quality of the campaign. In particular, we test whether and how hard variables correlate with the human ratings. Using images to demonstrate, we test whether the number of images is correlated with overall quality score. One interpretation is that people simply like more pictures. This interpretation, however, is inconsistent with our previous results, as both machines and humans can count images, yet humans outperform predicting funding success. Rather, we argue that machines cannot "see" the images (soft information), but only count how many images are there (hard variable). As such, if more images are associated with a higher rating, the images then likely carry soft information that people can extract (Duarte et al., 2012; Gonzalez and Komarova Loureiro, 2014). In turn, entrepreneurs utilize images to convey soft information. A similar logic holds for the sentiment score. Machines cannot "read" the project description, but only ascertain its overall tone. If a more positive tone is associated with a higher human rating, the project description likely carries much more soft information than merely its tone.

Table 4 presents Pearson and Spearman cross-correlations between some of the hard features of the campaign and the different ratings provided by our human raters.[13] Recall that we asked participates to provide not only an overall rating for each campaign, but also a rating based only on one *specific* dimension of the evaluation criteria. We focus on the hard variables which are more elastic—i.e., more at the discretion of the entrepreneur. Some aspects of a crowdfunding campaign are core traits of the project, such as the idea and the outcome the project aims to achieve. Other aspects of a campaign are adjustable and can be altered while retaining the essential identity of the project. For example, the same idea may be pitched in a more appealing way, and a different entrepreneur or team may present or even carry out the same project. We therefore expect such elastic features—e.g. relating to presentation—to carry more soft information.

3.3.1 Visual Elements

While the number of images on the campaign's webpage and video length are significantly correlated with all campaign evaluation criteria, the correlation is strongest with presentation ratings. This correlation may indicate that having many images increases the perceived overall quality of the project, or that people who run successful campaigns (in terms of other factors, such as the idea or

[13]The Spearman correlation between two variables is equal to the Pearson correlation between the rank values of those two variables. While Pearson's correlation assesses linear relationships, Spearman's correlation assesses monotonic relationships (whether linear or not).

the value for money) tend to have a honed presentation that includes many images. We note that this correlation is particularly strong for physical products, which are more strongly linked to the number of pictures than non-physical products.

3.3.2 Social Media

The incorporation of social media may allow potential backers to readily extract soft information simply by viewing other people's opinions or by interacting directly with the project creators (Moritz et al., 2015). Projects associated with a Twitter account have a mean rating of 53.5, while those without an account have a mean rating of only 49.5 ($p < 1 \times 10^{-6}$). This effect is most pronounced for games, which have a difference in mean of almost 10 points between projects with a Twitter account and those without one. Surprisingly, the results for Facebook are the opposite.

3.3.3 Sentiment

As noted above, we apply a sentiment analysis to the textual description of the projects, which provides a sentiment score, reflecting the tone expressed in the text (see Section 2.5). In essence, a computer-generated sentiment score is an attempt to allow machines to observe at least some of the soft information.

The univariate analysis in Table 4 shows that the text chosen to present the project is closely related to the perceived quality of presentation, entrepreneur, and the project overall. Further analysis indicates that the mean overall score given to projects with a sentiment score above 3 was 52.13, while those below 3 had a mean rating of 49.45, a difference significant by 1×10^{-4}. Turning to realized success rates, we find that campaigns that reached their funding goal tend to have a more positive tone (mean sentiment score of 4.6, median 3) than campaigns that did not get funded (mean score of 3.75, median 2). This difference is significant at $p < 1 \times 10^{-6}$ (using a Mann-Whitney U test). Overall, the univariate results would indicate that computer-generated sentiment scores may be able to capture at least some of the soft information that is otherwise entirely unobservable to computer algorithms.

We next explore which channel may drive the association between sentiment and the overall rating. We study the correlation between sentiment and each individual evaluation criterion in isolation—i.e., holding all other evaluation criteria ratings fixed. Table 5 presents multivariate regressions of the sentiment score on the focused ratings provided to individual evaluation criteria. The dependent variable is sentiment, while the independent variables include the focused ratings provided to individual evaluation criteria. This specification does not imply that human ratings can "predict" sentiment scores, but merely a statistical exercise to examine the correlation between the two in a multivariate setting. The results show that sentiment is strongly correlated only with presentation ratings. The latter result holds in all campaigns, as well as in subsamples based on project type and topic. Interestingly, sentiment matters more in art projects, in which it also becomes correlated with the perceived value for money. Overall, while sentiment may be able to decrypt some of the soft information, our AUC analysis shows that sentiment cannot replace human ratings. Taken together, these results are again consistent with the notion that the widely used computer-generated sentiment score is rather limited in the amount of soft information it can carry.

Importantly, the sentiment analysis shows that the tone in the text, rather than just its length (e.g., number of words), captures at least some of the soft information. The latter result further validates our soft information interpretation of the association between hard variables and human ratings. All in all, the results in this section indicate that human raters are affected mostly by soft information conveyed by the project that cannot be readily extracted by computer algorithms, including those using techniques meant to capture soft information such as sentiment scores.

3.3.4 Entrepreneurial Traits

The horse race results in Table 3 show that entrepreneur traits are somewhat significant in some of the specifications. To complement our discussion of the extent to which elastic features can carry soft information, we verify that more definitive characteristics such as entrepreneur traits do not perform as well. Recall that while the perceived entrepreneur characteristics are collected from our human participants, the characteristics are relatively definitive and principally stem from hard information. We asked participants to determine—to the best of their ability—the gender, age (in bands), and ethnicity of the entrepreneur initiating the campaign.

Results are presented at the bottom of Table 4. There are no statistically significant differences in the participants' project ratings based on the entrepreneurs' gender. One possible explanation is that male participants sampled from Mechanical Turk favored campaigns with female entrepreneurs appearing in images and/or videos, which in turn obscured any gender effect apparent in actual funding success rates (Gafni et al., 2021).

While age is positively correlated with the overall rating, we note that the relation is not monotonic. When the participants thought the entrepreneur was younger than 30 (our bottom two bands), the mean rating was 50.06; when the participants thought the entrepreneur was between 30 and 40 (middle band), the mean rating was 52.94; and when the participants thought the entrepreneur was over 40 (our top three bands), the mean rating was 47.87. These differences were significant at the $p < 1 \times 10^{-3}$ level. The focused quality ratings of the entrepreneur show a similar pattern, with means of 48, 48.83, and 43.62 for the same respective age groups. The age group that tends to score the highest ratings are those between 30 and 40 (about 28% of the entrepreneurs were judged to be in this group). Entrepreneurs whose age was judged to be higher than 40 (accounting for 14% of the entrepreneurs) or lower than 30 (58% of the entrepreneurs) tended to score lower ratings. A possible interpretation is that older entrepreneurs are deemed less innovative.

We can further report that the mean rating is 50.71 when the participants thought the entrepreneur was White, whereas the mean rating is 43.6 for a Black entrepreneur. This difference was statistically significant at the $p < 1 \times 10^{-4}$ (using a Mann-Whitney U test). The mean rating for Chinese and Indian entrepreneurs was higher than for White entrepreneurs (51.24 and 62.65, respectively), though only the Indian ethnicity was statistically significant ($p < 1 \times 10^{-4}$ using a Mann-Whitney U test), but that significance disappeared when removing Indian participants. The ethnic correlation was most evident in art categories, possibly as it is a field where there is a less strict characterization of a "good" project, thus allowing people to display hidden bias.

3.4 Does Soft Information Affect How Individual Investment Criteria are Consolidated?

A funding decision is a result of an underlying thought process in which potential investors need to aggregate over multiple evaluation criteria. Recall that we model the thought process of potential investors using the Analytic Hierarchy Process (Saaty et al., 2008). our underlying assumption is that, in the integration phase, agents roughly estimate a weighted-average score. That is, agents consider the relative importance of each of the qualitative factors, and identify the best decision choice as the one with the highest weighted score.

It is possible, however, that human raters are affected by soft yet salient attributes that stand out in a campaign (e.g., explicit images, or a personal/emotional appeal by the entrepreneur, which are atypical in other campaigns). Such salient attributes may attenuate or exacerbate the relative importance of an individual evaluation criterion on that particular campaign, and in turn, indirectly affect the overall rating assigned to the project and one's willingness to fund it (Bordalo et al., 2013). That is, soft cues may affect not only a criterion-specific score, but also the criterion's weight. While people start off with some general weights (i.e., those that apply to the average project) in mind, they update these weights conditional on soft cues embedded in each individual project at hand. If so, it is difficult—if not impossible—to disentangle the effect of soft information on criterion-specific scores from its effect on their aggregation.

Unlike the existing literature, which can only infer the determinants of funding success from publicly available information, our survey enables us to also observe their self-reported importance. Recall that we previously reported the perceived importance of campaign evaluation criteria (Panel B of Table 1). In this section, we exploit these self-reported weights to study how soft information can play a role in the aggregation of the various evaluation criteria. Our unique data allows us to compare between the self-reported importance of various evaluation criteria to their actual effect on the overall rating given to a project. Any discrepancy between the two would be consistent with soft information playing a role in the thought process of human raters as they aggregate over multiple evaluation criteria. As soft attributes are unobservable to computer algorithms, machines cannot consolidate an overall score as efficiently as human raters, thus do not perform as well as human ratings in predicting funding success.

To estimate actual importance, we go back to the focused ratings in the rating task data (see Section 2.3), which reflect a rater's opinion regarding a specific campaign *based only on one specific aspect*. To generate these ratings, we asked raters to evaluate the campaign based only on a single evaluation criterion. Table 6 reports the correlations between the focused ratings of each evaluation criterion and the overall rating. The focused ratings of all the evaluation criteria were all strongly correlated with the overall rating in a statistically significant way (an F-test had a result of $p < 1 \times 10^{-200}$ for all evaluation criteria). This shows that knowing the focused rating of any single evaluation criterion *in isolation* is very informative regarding the overall rating given to the campaign. Given those factors shown to be significantly correlated with the overall rating, we quantify the relative importance of each factor. A standard method to obtain such relative importance estimates is to regress a target variable using a set of predictor variables and then apply a measure of relative weight, or "relative importance." Most such measures are based on multiple linear regression models, including partial correlations, normalized regression coefficients, or the change in the coefficient of multiple determination.

Panel A in Table 7 presents the results of the multiple linear regression model for predicting the overall rating given to a crowdfunding campaign from the focused evaluation criterion ratings. The predicted variable is the overall rating given to the crowdfunding campaign, and the predictor variables are the focused ratings given to the various evaluation criteria. In the pooled sample, the model has an overall fit of $R^2 = 0.575$ and is statistically significant ($F = 1390$, $p < 1 \times 10^{-200}$). The coefficients of all of the evaluation criterion focused ratings, with the exception of the likelihood of success, are significant at the $p < 0.0001$ level. The coefficient of the likelihood of success is not found to be statistically significant at the $p < 0.05$ level. This indicates that while the rating of the project's likelihood of success is correlated with the overall rating, this rating does not carry additional information regarding the overall rating beyond the information carried by all the other evaluation criteria. In other words, when predicting the overall rating of a project, once we know how the project is rated in terms of the other evaluation criteria, there is little to be gained by knowing the rating of the likelihood of success.

Panel B of Table 7 shows the significance levels and several measures of the relative importance of each of the factors: partial correlations (PC), normalized regression coefficients (β), and the change in the coefficient of multiple determination (CCMD).[14] The numbers in parentheses are the values normalized to 100%, which finally allows us to compare the actual importance with perceived importance. Recall that to assess perceived importance, we asked the survey participants to allocate 100 points among the evaluation criteria considering the importance of each of the evaluation criteria in affecting their overall rating.

Figure 5 summarizes our findings regarding the relative importance of factors related to the overall rating of a crowdfunding campaign. It plots the actual importance (the normalized partial

[14] By partial correlations we refer to the correlation between one predictor variable and the target variable, controlling for all remaining predictor variables. Partial correlation is computed by regressing one focused rating on the focused rating of remaining criteria, and overall rating on the remaining criteria, and then calculating the correlation between the residuals. By the change in the coefficient of multiple determination for a predictor variable x_i, we refer to the change in R^2 value between the regression containing all predictors (including x_i) and the regression containing all predictors except x_i.

correlations, PC, shown in Table 7) against perceived importance (values in Panel B of Table 1). Although the idea is the most important factor, in terms of both perceived and actual importance, the value for money as well as the presentation are also strong factors. The entrepreneur's evaluation is quite predictive of the overall rating, though to a smaller degree than the previous factors. This is consistent with Gafni et al. (2019), who find that the presentation of the entrepreneur plays an important role. Most interestingly, while our participants expected the project's likelihood of success to be a very influential factor on their opinions, our results regarding the actual importance of this factor indicate that it is not important in practice. Such a discrepancy is consistent with soft information playing a role in the thought process of human raters as they aggregate over multiple evaluation criteria. In particular, human raters may incorporate soft yet salient information from the project's presentation, leading to an overweighting of the importance of presentation at the expense of the likelihood of success. As such soft attributes are unobservable to computer algorithms, human ratings may outperform machines in predicting funding success.

One may argue that self-reported importance suffers from a common bias such as social desirability bias, by which survey respondents tend to answer questions in a manner that will be viewed favorably by others. If so, then we expect the discrepancy between the self-reported and actual importance of specific determinants of funding success to be consistent across different types of campaigns. We thus perform a cross-sectional analysis, drilling down into the various project categories of Kickstarter campaigns. We compute the normalized partial correlations (PC) within each of the subsamples based on project outcome and project topic. As previously mentioned, our category groups are based on the campaign's classification on Kickstarter, and as they are not mutually exclusive, each campaign can fall into more than one category group. Figure 6 shows that the relevance of the various evaluation criteria changes with the project category. Importantly, the discrepancy between the self-reported and actual importance of likelihood of success is driven by art and music projects, in which soft information is most influential. The latter result is inconsistent with a common bias such as social desirability bias, and consistent with soft information playing a role in the thought process of human raters as they consolidate their overall view. In many cases, the importance of the campaign's presentation moves in the opposite direction of the importance of the idea, and thus all non-product categories, including arts, film and music, exhibit a significant weight given to the presentation's quality. The outlier here is games, as ideas play an average role, while presentation plays a huge role, due to entrepreneurs carrying very little weight, as they also do in the film category.[15]

4 Conclusions

We suggest that the opinions of a human crowd can detect cues in soft information beyond raw campaign features. In particular, our analysis shows that information contained in human annotations of campaigns can complement predictive algorithms: By incorporating annotations made by human evaluators of campaigns, we are able to improve the accuracy of predicting whether a campaign would achieve its funding goal. We consider this a very encouraging finding, since the ability to evaluate a campaign in its very early days and to predict whether it would succeed is a very powerful tool. Such an evaluation can enable both entrepreneurs and investors to better understand the strengths and weakness of a campaign and thus make their decisions accordingly. More broadly, our results suggest that human evaluations still have value when the world has moved toward big data.

We examine the high-level criteria by which people evaluated crowdfunding campaigns. We show that while the most important factor for a campaign's success is the idea itself, many other issues are correlated with how people rate the campaign. In many projects, people certainly care about the value they will get in return for their money, indicating that Kickstarter is not used mostly for charity-like projects. We also examine several soft cues that influence the campaign's perceived quality. We show that the rating given to the campaign is correlated with the sentiment in the project's description. Nonetheless, sentiment cannot replace human ratings in predicting funding success. Interestingly, although people expected their ratings to be strongly influenced by the likelihood of the project succeeding, this factor does not seem to provide further information regarding the overall project rating beyond the information contained in the other factors. This makes Kickstarter somewhat of a platform of dreamers, in which even very difficult projects may be funded if they are perceived to offer a high value.

This study is not without limitations. Throughout this study, we consider funding success—that is, campaigns that managed to hit the minimal funding milestone to be launched. This does

[15] In all these fields, the difference in mean score between successful and unsuccessful projects is maintained.

not mean that the project team actually managed to deliver the promised goods. In particular, it is possible that human raters outperform machines in predicting funding success but fail in predicting the delivery outcome. While the value of soft information in predicting delivery outcomes would have vast implications on the efficiency of crowdfunding markets, it is outside the scope of this study.

Many questions still remain open for future research. First, can the perceived quality of a campaign be strategically manipulated? For example, would an otherwise-identical campaign be more likely to succeed when it has more images in its description, or is it more likely to succeed when proposed by a White individual between the ages of 30 and 40 than when proposed by a 50-year-old Black individual? Showing the existence of such biases requires a controlled experiment, which is outside the scope of this paper. Second, given many opinions of individuals about the success of a project, how can we best *aggregate* these different ratings so as to obtain the most accurate prediction regarding the campaign's success? In particular, what would happen if we let raters trade contracts contingent on funding success in a prediction market? Would the bets add even more to the value of human annotations that can outperform machines? While the common view is that prediction markets outperform prediction polls, Atanasov et al. (2017) show that polls, as long as they are carefully designed, remain an attractive alternative to prediction markets for distilling the wisdom of crowds.

References

Agarwal, S. and Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, 23(7):2757–2788.

Agrawal, A., Catalini, C., and Goldfarb, A. (2015). Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics and Management Strategy*, 24(2):253–274.

Ahlers, G. K., Cumming, D., Günther, C., and Schweizer, D. (2015). Signaling in equity crowdfunding. *Entrepreneurship theory and practice*, 39(4):955–980.

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., Ungar, L., and Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science*, 63(3):691–706.

Bagnoli, M. and Lipman, B. L. (1989). Provision of public goods: Fully implementing the core through private contributions. *The Review of Economic Studies*, 56(4):583–601.

Banerjee, M., Capozzoli, M., McSweeney, L., and Sinha, D. (1999). Beyond kappa: A review of interrater agreement measures. *Canadian journal of statistics*, 27(1):3–23.

Barbopoulos, L. G., Dai, R., Putniņš, T. J., and Saunders, A. (2021). Market efficiency in the age of machine learning. *NYU Stern School of Business Forthcoming.*

Belleflamme, P., Lambert, T., and Schwienbacher, A. (2014). Crowdfunding: Tapping the right crowd. *Journal of Business Venturing*, 29(5):585–609.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C. (2005). Does function follow organizational form? evidence from the lending practices of large and small banks. *Journal of Financial economics*, 76(2):237–269.

Bishop, C. (2006). *Pattern recognition and machine learning*, volume 4. springer New York.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159.

Burtch, G., Ghose, A., and Wattal, S. (2013). An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information System Research*, 24(3):499–519.

Burtch, G., Ghose, A., and Wattal, S. (2016). Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Information Systems Research*, 27(3):478–496.

Chang, A. (2012). Hands on with pebble smartwatch, the most successful kickstarter project ever. *Wired Magazine*.

Chen, K., Jones, B., Kim, I., and Schlamp, B. (2014). Kickpredict: Predicting kickstarter success. Technical report, CalTech.

Cohn, J., Rajan, U., and Strobl, G. (2024). Credit ratings: strategic issuer disclosure and optimal screening. *Review of Finance*.

Colombo, M. G., Franzoni, C., and Rossi–Lamastra, C. (2015). Internal social capital and the attraction of early contributions in crowdfunding. *Entrepreneurship Theory and Practice*, 39(1):75–100.

Dahlin, L., Rhue, L., and Clark, J. (2019). Crowdfunding community formation: Fundraiser race and gender homophily. Working paper.

Dorfleitner, G., Priberny, C., Schuster, S., Stoiber, J., Weber, M., de Castro, I., and Kammler, J. (2016). Description-text related soft information in peer-to-peer lending–evidence from two leading european platforms. *Journal of Banking & Finance*, 64:169–187.

Duarte, J., Siegel, S., and Young, L. (2012). Trust and credit: the role of appearance in peer-to-peer lending. *Review of Financial Studies*, 25(8):2455–2484.

Etter, V., Grossglauser, M., and Thiran, P. (2013). Launch hard or go home!: predicting the success of kickstarter campaigns. In *Proceedings of 1st ACM Conference on Online Social Networks (COSN)*, pages 177–182, Boston, Massachusetts.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Gafni, H., Marom, D., Robb, A., and Sade, O. (2021). Gender dynamics in crowdfunding (kickstarter): Evidence on entrepreneurs, backers, and taste-based discrimination. *Review of Finance*, 25(2):235–274.

Gafni, H., Marom, D., and Sade, O. (2019). Are the life and death of an early stage venture indeed in the power of the tongue? lessons from online crowdfunding pitches. *Strategic Entrepreneurship Journal*, 13(1):3–23.

Gompers, P., Kovner, A., Lerner, J., and Scharfstein, D. (2010). Performance persistence in entrepreneurship. *The Journal of Financial Economics*, 96(1):18–32.

Gonzalez, L. and Komarova Loureiro, Y. (2014). When can a photo increase credit?: The impact of lender and borrower profiles on online p2p loans. *Journal of Experimental and Behavioral Finance*, 2:44–58.

Goodman, J. K., Cryder, C. E., and Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of mechanical turk samples. *Journal of Behavioral Decision Making*, 26(3):213–224.

Greenberg, M. D., Pardo, B., Hariharan, K., and Gerber, E. (2013). Crowdfunding support tools: Predicting success & failure. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1815–1820, Paris, France.

Hervé, F., Manthé, E., Sannajust, A., and Schwienbacher, A. (2019). Determinants of individual investment decisions in investment-based crowdfunding. *Journal of Business Finance & Accounting*, 46(5-6):762–783.

Herzenstein, M., Sonenshein, S., and Dholakia, U. M. (2011). Tell me a good story and i may lend you money: The role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research*, 48(SPL):S138–S149.

Hirshleifer, D. and Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1):337–386.

Hoskins, R. (2014). Global 100 crowdfunding index. http://crowdfundingpr.wordpress.com/2013/06/12/crowdfund press-center-releases-the-first-global-100-crowdfunding-site-index/.

Howell, S. T. (2021). Learning from feedback: Evidence from new ventures. *Review of Finance*, 25(3):595–627.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Jame, R., Johnston, R., Markov, S., and Wolfe, M. C. (2016). The value of crowdsourced earnings forecasts. *Journal of Accounting Research*, 54(4):1077–1110.

Jhaveri, S., Khedkar, I., Kantharia, Y., and Jaswal, S. (2019). Success prediction using random forest, catboost, xgboost and adaboost for kickstarter campaigns. In *Proceedings of the 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, pages 1170–1173, Erode, India.

Jones, S., Johnstone, D., and Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance*, 56:72–85.

Kickstarter (2022). Stats. https://www.kickstarter.com/help/stats.

Koch, J.-A. and Siering, M. (2015). Crowdfunding success factors: The characteristics of successfully funded projects on crowdfunding platforms. In *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, Münster, Germany.

Kuppuswamy, V. and Bayus, B. L. (2017). Does my contribution to your crowdfunding project matter? *Journal of business venturing*, 32(1):72–89.

Li, E. and Martin, J. S. (2019). Capital formation and financial intermediation: The role of entrepreneur reputation formation. *Journal of Corporate Finance*, 59:185–201.

Li, Y., Rakesh, V., and Reddy, C. K. (2016). Project success prediction in crowdfunding environments. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining (WSDM)*, pages 247–256, San Francisco, California.

Lin, M., Prabhala, N. R., and Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management science*, 59(1):17–35.

Mollick, E. and Nanda, R. (2015). Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science*, 62(6):1533–1553.

Mollick, E. R. (2014). The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1):1–16.

Moritz, A., Block, J., and Lutz, E. (2015). Investor communication in equity-based crowdfunding: a qualitative-empirical study. *Qualitative Research in Financial Markets*, 7(3):309–342.

Peer, E., Vosgerau, J., and Acquisti, A. (2014). Reputation as a sufficient condition for data quality on amazon mechanical turk. *Behavior research methods*, 46:1023–1031.

Saaty, T. L. et al. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98.

Sawhney, K., Tran, C., and Tuason, R. (2015). Using language to predict kickstarter success. Technical report, Stanford University.

Shafi, K. (2021). Investors' evaluation criteria in equity crowdfunding. *Small Business Economics*, 56(1):3–37.

Skirnevskiy, V., Bendig, D., and Brettel, M. (2017). The influence of internal social capital on serial creators' success in crowdfunding. *Entrepreneurship Theory and Practice*, 41(2):209–236.

Srinivasan, V. and Kim, Y. H. (1987). Credit granting: A comparative analysis of classification procedures. *The Journal of Finance*, 42(3):665–681.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168.

Thomas, R. (2013). The veronica mars movie project. https://www.kickstarter.com/projects/559914737/the-veronica-mars-movie-project.

Uchida, H., Udell, G. F., and Yamori, N. (2012). Loan officers and relationship lending to smes. *Journal of Financial Intermediation*, 21(1):97–122.

Vismara, S. (2018). Information cascades among investors in equity crowdfunding. *Entrepreneurship theory and practice*, 42(3):467–497.

Xiao, S., Tan, X., Dong, M., and Qi, J. (2014). How to design your project in the online crowdfunding market? evidence from kickstarter. In *Proceedings of the 35th International Conference on Information Systems (ICIS)*, Auckland, New Zealand.

Table 1: General Survey Responses
The survey participants were sourced through Amazon Mechanical Turk during February and March of 2014. In total, we had $n_p = 1,206$ survey participants. Panel A includes the background information. We also asked participants whether they had previously heard of Kickstarter, whether they had ever visited the Kickstarter webpage, and whether they had invested in Kickstarter projects. Panel B includes the perceived importance of the five specific factors that may be predictive of a campaign's perceived quality. We asked participants to rate the different evaluation criteria by allocating 100 points among them according to how important they perceive them in determining whether campaigns, in general, are worthy of investment.

Panel A: General Questions

Gender	
Male	57.30%
Female	42.70%
Age	
25 >	25.09%
25-29	28.06%
30-34	20.37%
35-39	11.99%
40-49	9.35%
50 ≤	5.15%
Country of residence	
USA	81.86%
India	15.41%
Other countries	2.73%
Education	
Below high school	1.15%
High school	31.24%
College / Bachelor degree	49.93%
Master degree	13.66%
PhD.	1.56%
Average annual income (US $)	
20,000 >	34.65%
20,000-30,000	17.20%
30,000-40,000	15.40%
40,000-50,000	8.69%
50,000-60,000	7.81%
60,000-80,000	7.56%
80,000 <	6.15%
Prefer to not say	2.54%
Have you heard of Kickstarter before?	
Yes	77.74%
No	22.26%
Have you visited the Kickstarter website before?	
Yes	58.79%
No	41.21%
Have you invested in a Kickstarter project before?	
Yes	18.02%
No	81.98%

Panel B: Perceived Importance of Crowdfunding Campaign Evaluation Criteria

Evaluation Criterion	Average points	Median points	Std deviation
Idea	30.55	29	13.78
Presentation	15.94	15	8.83
Value for money	18.9	19	10.95
Entrepreneur	11.1	10	9.05
Likelihood of success	23.36	21	12.74

Panel C: Cross-correlation between Perceived Importance of Crowdfunding Campaign Evaluation Criteria

	Idea	Presentation	Value for money	Entrepreneur	Likelihood of success
Idea	1				
Presentation	-0.178	1			
Value for money	-0.411	-0.193	1		
Entrepreneur	-0.34	0.026	-0.072	1	
Likelihood of success	-0.337	-0.333	-0.21	-0.285	1

Table 2: Campaign Data
We randomly selected live Kickstarter crowdfunding campaigns during February and March of 2014. In total, we had $n_p = 1,206$ survey participants, yielding $n_r = 6,098$ campaign reports, regarding $n_c = 936$ Kickstarter campaigns. For each campaign, we record the funding goal and project categories (both project type and topic as detailed above). We also record other "hard" features—i.e., variables available to both robots and humans as they scan through a project's page—such as the number of images, length of videos, links to social media features (indicating whether the page includes a Facebook or Twitter link), and a sentiment score between 1 and 5 based on textual analysis of the project description ("Very negative" = 1, "Negative" = 2, "Neutral" = 3, "Positive" = 4, and "Very positive" = 5). At the time of each scan—i.e., when a participant is assigned a campaign—we record the share of the funding goal raised, the number of backers, and the number of days until the deadline. Each participant had to provide an overall rating as well as a rating based only on one *specific* dimension of the evaluation criteria, while ignoring all other criteria. Ratings range from 0 to 100. Once a campaign's deadline passed, we could see if it attained the funding goal. *Funding success* is a binary variable with a value of 1 if the amount raised is equal or higher than the funding goal, and 0 otherwise. We also record the amount raised in both absolute terms as well as the percent of the funding goal.

Panel A: All Campaigns (N=6,088)

	mean	median	std. deviation
Campaign hard variables:			
Funding goal	17,728	5,000	67,712
Image number	6.25	2	16.64
Video length	155 seconds	2-3 minutes	78.28
Facebook link	0.68	yes	0.47
Twitter link	0.2	No	0.4
Sentiment	2.46	2	0.89
Share of goal raised at scan	1.26	0.13	28.49
Amount of backers at scan	363.04	13	11178
Amount of days left at scan	17.46	16	32.25
Crowdsourced ratings:			
Overall	47.87	50	25.93
Idea	47.99	50	28.48
Presentation	46.82	50	28.67
Value for money	41.06	40	27.82
Entrepreneur	44.20	46	25.8
Likelihood of success	51.97	56	31.66
Funding outcome:			
Funding success	0.47	0	0.50
Project amount raised	8,399.27	1,410	40,785.02
Project percent raised	85.80	37.81	414.88

Panel B: Campaign Data by Project Outcome Type

Product rewards (N=2,961)

	mean	median	std. deviation
Funding goal	16,353	5,880	41,889
Image number	8.5	5	12.43
Video length	145 seconds	2-3 minutes	77.31
Facebook link	0.66	yes	0.47
Twitter link	0.21	No	0.41
Sentiment	2.43	2	0.86
Share of goal raised at scan	0.89	0.09	5.2423
Amount of backers at scan	483.62	13	13038
Amount of days left at scan	17.57	16	18.96
Overall	51.53	50	26.22
Idea	50.49	53	28.75
Presentation	49.97	50	28.83
Value for money	43.56	43	27.84
Entrepreneur	47.40	50	25.92
Likelihood of success	55.44	60	31.71
Funding success	0.40	0	0.49
Project amount raised	9,288.77	1,263	27,047.71
Project percent raised	137.11	24.5	570.85

Non-product rewards (N=3,137)

	mean	median	std. deviation
Funding goal	19,027	4,000	85,173
Image number	4.14	1	19.58
Video length	164 seconds	2-3 minutes	78.13
Facebook link	0.69	yes	0.46
Twitter link	0.18	No	0.39
Sentiment	2.48	2	0.92
Share of goal raised at scan	1.62	0.17	39.39
Amount of backers at scan	240.22	14	9078.5
Amount of days left at scan	17.36	16	41.02
Overall	49.16	50	25.60
Idea	48.51	50	28.18
Presentation	47.68	50	28.47
Value for money	40.31	39	27.72
Entrepreneur	45.48	50	25.68
Likelihood of success	55.87	60	31.60
Funding success	0.53	1	0.50
Project amount raised	7,556.99	1,768	50,441.36
Project percent raised	85.81	100.25	160.54

Panel C: Campaign data by Project Topic

Art projects (N=3,554)

	mean	median	std. deviation
Funding goal	17,842	4,000	80,298
Image number	4.71	1	18.84
Video length	162 seconds	2-3 minutes	78.57
Facebook link	0.69	yes	0.46
Twitter link	0.18	No	0.38
Sentiment	2.46	2	0.9
Share of goal raised at scan	1.55	0.18	37.03
Amount of backers at scan	228.72	14	8529.5
Amount of days left at scan	17.16	16	38.82
Overall	47.63	50	25.56
Idea	47.06	50	28.16
Presentation	46.65	50	28.47
Value for money	38.64	36	27.64
Entrepreneur	44.23	47	25.73
Likelihood of success	52.86	58	31.77
Funding success	0.53	1	0.50
Project amount raised	7,585.69	1,535	47,888.84
Project percent raised	73.63	55.79	217.66

Game projects (N=459)

	mean	median	std. deviation
Funding goal	31,108	7,500	83,195
Image number	11.97	6	19.25
Video length	150 seconds	2-3 minutes	84.9
Facebook link	0.64	yes	0.48
Twitter link	0.42	No	0.49
Sentiment	2.42	2	0.91
Share of goal raised at scan	0.78	0.13	1.56
Amount of backers at scan	218.36	23	664.53
Amount of days left at scan	18.11	17	26.17
Overall	53.51	56	25.10
Idea	53.10	58	27.67
Presentation	54.81	60	28.23
Value for money	46.29	50	27.38
Entrepreneur	47.93	50	25.17
Likelihood of success	56.35	60	31.12
Funding success	0.41	0	0.49
Project amount raised	12,004.91	2,053	28,327.97
Project percent raised	161.82	32.09	293.79

Panel C: Campaign Data by Project Topic (cont'd)

Film projects (N=1,159)

	mean	median	std. deviation
Funding goal	30,257	6,995	110,780
Image number	5.45	2	30.78
Video length	170 seconds	2-3 minutes	75.89
Facebook link	0.69	yes	0.46
Twitter link	0.19	No	0.39
Sentiment	2.52	2	0.81
Share of goal raised at scan	1.27	0.11	26.71
Amount of backers at scan	480.52	13	14686
Amount of days left at scan	16.53	16	16.63
Overall	47.67	50	25.90
Idea	45.77	48	28.24
Presentation	47.20	50	28.63
Value for money	37.48	34	27.72
Entrepreneur	43.66	48	25.46
Likelihood of success	50.19	54	31.18
Funding success	0.47	0	0.50
Project amount raised	6,783.46	1,603	16,196.28
Project percent raised	56.03	35.95	64.17

Music projects (N=972)

	mean	median	std. deviation
Funding goal	7,381.6	3,700	12,394
Image number	1.74	1	4.41
Video length	163 seconds	2-3 minutes	82.14
Facebook link	0.69	yes	0.46
Twitter link	0.21	No	0.41
Sentiment	2.38	2	0.98
Share of goal raised at scan	2.58	0.3	64.14
Amount of backers at scan	161.14	19	2954
Amount of days left at scan	19.67	16	69.47
Overall	49.57	50	25.31
Idea	48.06	50	27.53
Presentation	46.15	50	27.92
Value for money	42.95	43	27.68
Entrepreneur	46.12	50	25.75
Likelihood of success	58.42	62	31.18
Funding success	0.65	1	0.48
Project amount raised	5,652.72	2,560.5	11,117.54
Project percent raised	90.01	105.37	71.80

Table 3: Predicting Funding Success

We run a horse race of human raters against the most common algorithms to determine which one prevails in predicting funding success. The dependent variable is binary, indicating whether the money raised at the deadline reached (or exceeded) the funding goal. Column 1 uses information available on Kickstarter, including project characteristics (project type and various categories) and real-time progress (the share of the funding goal raised, the number of backers, and the number of days left until the deadline—all measured at the time of each scan). In Column 2, we add the sentiment score ranging from 1 ("Very negative") to 5 ("Very positive"). In Column 3, we add entrepreneur perceived characteristics (gender, age and ethnicity), and whether the project is a charitable one. In Column 4, we add the (average) overall rating given to a campaign. Column 5 uses all data collected from human raters, including project-specific ratings for individual evaluation criteria. To minimize estimation error, the unit of analysis is a campaign; that is, we average all variables (including binary ones) across all raters assigned to the campaign. For consistency, we include only projects with non-missing data in all variables throughout the table. Panel B includes only ratings collected from participants with no Kickstarter investment experience, while Panel D includes only ratings collected 20 days or earlier prior to the deadline. Standard errors are in parentheses, and *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

Below each model, we present the corresponding AUC, which summarizes how the fitted values from the corresponding logistic regression would perform (in terms of the proportion of true positives against the proportion of false positives) under all possible thresholds.

Panel A: All Campaigns

Evaluation Criterion	(1) Hard data	(2) +Sentiment	(3) + Definitive	(4) +People's overall grade	(5) All
Average share of goal raised at scan/100	0.065	0.064	0.049	0.069	−0.27
	(0.105)	(0.106)	(0.105)	(0.097)	(0.08)
Average amount of backers at scan/100000	0.015	0.014	0.222	-0.355	-0.13
	(0.55)	(0.55)	(0.56)	(0.52)	(0.43)
Average amount of days left at scan	-0.003**	-0.003**	-0.008**	-0.003***	-0.002*
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)
Funding goal (in 10K)	-0.008***	-0.008***	-0.009***	-0.006***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Image number/10	0.007***	0.007***	0.007***	0.003	0.0003
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)
Video length (in minutes)	0.029**	0.029**	0.031**	0.005	-0.001
	(0.011)	(0.011)	(0.011)	(0.01)	(0.01)
Facebook link	-0.11**	-0.11**	-0.09*	-0.06	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Twitter link	0.12**	0.12**	0.11**	0.07	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)
Is project a physical object	0.02	0.02	0.02	0.02	0.03
	(0.07)	(0.07)	(0.07)	(0.06)	(0.05)
Is project artistic	0.18**	0.17**	0.17**	0.21***	0.13**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
Is project a game	0.009	0.009	0.009	0.01	0.008
	(0.06)	(0.06)	(0.06)	(0.06)	(0.05)
Is project a film	-0.04	-0.04	-0.04	-0.04	0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)
Is project musical	0.13*	0.13*	0.13*	0.13**	0.1*
	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)
Sentiment		-0.006	-0.002	-0.01	-0.006
		(0.02)	(0.02)	(0.02)	(0.01)
Average entrepreneur gender estimate			0.04	0.05	0.05
			(0.02)	(0.02)	(0.02)
Average entrepreneur age estimate			-0.03**	-0.03***	-0.02*
			(0.01)	(0.01)	(0.009)
Is estimated ethnically white			0.08	0.07	0.05
			(0.05)	(0.05)	(0.04)
Is estimated ethnically black			-0.09	-0.06	-0.04
			(0.07)	(0.06)	(0.05)
Is estimated ethnically hispanic			-0.04	-0.05	-0.04
			(0.07)	(0.06)	(0.05)
Is estimated ethnically Chinese			0.14	0.14	0.112
			(0.11)	(0.1)	(0.08)
Is estimated ethnically Indian			0.07	0.01	0.106
			(0.34)	(0.31)	(0.26)
Is project charitable			0.007	0.1*	0.07
			(0.05)	(0.05)	(0.04)
Average people overall grade				0.01***	0.0006
				(0.001)	(0.002)
Average presentation grade					0.001
					(0.001)
Average idea grade					-0.001
					(0.002)
Average entrepreneur grade					-0.005**
					(0.002)
Average value for money grade					0.001
					(0.001)
Average likelihood of success estimate					0.017***
					(0.0009)
Intercept	34.058**	34.185**	26.085	-11.979	-11.138
	(12.96)	(12.97)	(22.12)	(20.61)	(16.966)
N-obs	936	936	936	936	936
Adjusted R-Squared	0.0979	0.097	0.11	0.244	0.49
AUC	0.7251	0.7251	0.7339	0.8205	0.926

Panel B: Predicting Funding Success with No Prior Kickstarter Investment Experience

Evaluation Criterion	(1) Hard data	(2) +Sentiment	(3) + Definitive	(4) +People's overall grade	(5) All
Average share of goal raised at scan/100	0.257	0.258	0.27	0.311*	0.355*
	(0.153)	(0.154)	(0.153)	(0.151)	(0.144)
Average amount of backers at scan/100000	-2.157	-2.158	-2.297	-2.666*	-2.839*
	(1.238)	(1.239)	(1.235)	(1.214)	(1.156)
Average amount of days left at scan	-0.002	-0.002	-0.002*	-0.002*	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Funding goal (in 10K)	-0.008**	-0.008**	-0.007**	-0.007**	-0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Image number/10	0.032	0.032	0.029	0.014	0.005
	(0.022)	(0.022)	(0.022)	(0.021)	(0.021)
Video length (in minutes)	0.036**	0.036**	0.042**	0.04**	0.042***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.012)
Facebook link	-0.068	-0.068	-0.075	-0.046	-0.041
	(0.046)	(0.046)	(0.047)	(0.046)	(0.044)
Twitter link	0.135**	0.135**	0.136**	0.122**	0.094*
	(0.047)	(0.047)	(0.047)	(0.046)	(0.044)
Is project a physical object	0.019	0.019	0.027	0.029	0.006
	(0.082)	(0.082)	(0.083)	(0.081)	(0.077)
Is project artistic	0.166*	0.167*	0.156*	0.18*	0.129
	(0.074)	(0.074)	(0.074)	(0.073)	(0.07)
Is project a game	-0.02	-0.02	-0.053	-0.059	-0.075
	(0.073)	(0.073)	(0.074)	(0.072)	(0.069)
Is project a film	-0.104	-0.104	-0.092	-0.102	-0.091
	(0.066)	(0.066)	(0.067)	(0.065)	(0.062)
Is project musical	0.072	0.073	0.075	0.044	0.041
	(0.067)	(0.067)	(0.068)	(0.067)	(0.063)
Sentiment		0.001	-0.003	-0.013	-0.008
		(0.021)	(0.021)	(0.021)	(0.02)
Average entrepreneur gender estimate			-0.019	-0.015	-0.017
			(0.036)	(0.035)	(0.034)
Average entrepreneur age estimate			-0.019	-0.019	-0.015
			(0.012)	(0.012)	(0.011)
Is estimated ethnically white			0.009	0.017	0.021
			(0.053)	(0.052)	(0.05)
Is estimated ethnically black			-0.207*	-0.173*	-0.152
			(0.088)	(0.087)	(0.083)
Is estimated ethnically hispanic			-0.01	-0.035	0.001
			(0.08)	(0.079)	(0.075)
Is estimated ethnically Chinese			-0.085	-0.065	-0.073
			(0.105)	(0.104)	(0.099)
Is estimated ethnically Indian			0.055	0.054	0.072
			(0.131)	(0.129)	(0.123)
Is project charitable			0.002	0.002	0.002
			(0.002)	(0.002)	(0.002)
Average people overall grade				0.004***	-0.0004
				(0.001)	(0.001)
Average presentation grade					-0.001
					(0.001)
Average idea grade					0.0003
					(0.001)
Average entrepreneur grade					-0.001
					(0.001)
Average value for money grade					-0.00004
					(0.001)
Average likelihood of success estimate					0.008***
					(0.001)
Intercept	21.991	21.998	34.853	24.066	21.806
	(14.755)	(14.767)	(18.152)	(17.94)	(17.206)
N-obs	655	655	655	655	655
Adjusted R-Squared	0.0741	0.0727	0.0823	0.116	0.203
AUC	0.7199	0.7199	0.7297	0.7603	0.8101

Panel C: Predicting Funding Success by Project Outcome

Product project

Evaluation Criterion	(1) Hard data	(2) +Sentiment	(3) + Definitive	(4) +People's overall grade	(5) All
Average share of goal raised at scan/100	1.438***	1.444***	1.501***	1.43***	0.41
	(0.377)	(0.377)	(0.378)	(0.347)	(0.293)
Average amount of backers at scan/100000	0.449	0.453	0.572	0.035	0.295
	(0.68)	(0.68)	(0.68)	(0.63)	(0.51)
Average amount of days left at scan	-0.003	-0.003	-0.002	-0.003*	-0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Funding goal (in 10K)	-0.015**	-0.015**	-0.015**	-0.014**	-0.004
	(0.005)	(0.006)	(0.006)	(0.005)	(0.004)
Image number/10	0.009***	0.008***	0.007**	0.003	-0.0005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Video length (in minutes)	0.013	0.012	0.016	-0.008	-0.008
	(0.015)	(0.015)	(0.016)	(0.015)	(0.013)
Facebook link	-0.115*	-0.115*	-0.118*	-0.068	-0.078
	(0.054)	(0.054)	(0.055)	(0.051)	(0.041)
Twitter link	0.155**	0.154**	0.146**	0.087	0.06
	(0.054)	(0.054)	(0.055)	(0.051)	(0.042)
Sentiment		0.018	0.02	0.007	0.009
		(0.025)	(0.025)	(0.023)	(0.019)
Average entrepreneur gender estimate			-0.025	0.009	0.018
			(0.043)	(0.04)	(0.033)
Average entrepreneur age estimate			-0.035*	-0.036*	-0.024*
			(0.015)	(0.014)	(0.012)
Is estimated ethnically white			0.082	0.074	0.059
			(0.065)	(0.06)	(0.049)
Is estimated ethnically black			-0.022	-0.005	-0.026
			(0.106)	(0.097)	(0.08)
Is estimated ethnically hispanic			0.026	0.034	0.016
			(0.089)	(0.082)	(0.067)
Is estimated ethnically Chinese			0.297*	0.277*	0.26**
			(0.132)	(0.122)	(0.1)
Is estimated ethnically Indian			-0.259	-0.209	-0.108
			(0.465)	(0.427)	(0.35)
Is project charitable			-0.016	0.082	0.05
			(0.098)	(0.09)	(0.074)
Average people overall grade				0.013***	0.002
				(0.001)	(0.002)
Average presentation grade					0.0002
					(0.002)
Average idea grade					-0.003
					(0.002)
Average entrepreneur grade					-0.002
					(0.002)
Average value for money grade					-0.001
					(0.002)
Average likelihood of success estimate					0.018***
					(0.001)
Intercept	37.008*	36.784*	59.755	6.7273	12.428
	(17.239)	(17.251)	(34.339)	(32.024)	(26.343)
N-obs	463	463	463	463	463
Adjusted R-Squared	0.105	0.104	0.112	0.253	0.499
AUC	0.9418	0.9418	0.9458	0.9416	0.9593

Panel C: Predicting Funding Success by Project Outcome (cont'd)

Non-product project

Evaluation Criterion	(1) Hard data	(2) +Sentiment	(3) + Definitive	(4) +People's overall grade	(5) All
Average share of goal raised at scan/100	0.01	0.003	-0.023	0.005	-0.02
	(0.12)	(0.12)	(0.12)	(0.11)	(0.09)
Average amount of backers at scan/100000	-0.902	-0.916	-0.563	-1.146	-0.811
	(0.96)	(0.96)	(0.95)	(0.897)	(0.743)
Average amount of days left at scan	-0.003*	-0.003*	-0.003*	-0.002	-0.0009
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Funding goal (in 10K)	-0.008**	-0.008**	-0.007**	-0.006*	-0.0001
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Image number/10	0.004	0.004	0.005*	0.001	0.0005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Video length (in minutes)	0.048**	0.048**	0.048**	0.022	0.01
	(0.016)	(0.016)	(0.016)	(0.015)	(0.013)
Facebook link	-0.089	-0.093	-0.078	-0.074	-0.027
	(0.061)	(0.061)	(0.061)	(0.057)	(0.047)
Twitter link	0.074	0.076	0.079	0.063	0.063
	(0.058)	(0.058)	(0.057)	(0.054)	(0.045)
Sentiment		-0.034	-0.029	-0.033	-0.022
		(0.025)	(0.025)	(0.023)	(0.019)
Average entrepreneur gender estimate			0.13**	0.121**	0.075*
			(0.046)	(0.043)	(0.036)
Average entrepreneur age estimate			-0.032	-0.034*	-0.012
			(0.017)	(0.016)	(0.013)
Is estimated ethnically white			0.05	0.035	0.017
			(0.076)	(0.072)	(0.06)
Is estimated ethnically black			-0.169	-0.134	-0.08
			(0.098)	(0.091)	(0.076)
Is estimated ethnically hispanic			-0.1	-0.123	-0.09
			(0.097)	(0.091)	(0.076)
Is estimated ethnically Chinese			-0.216	-0.204	-0.15
			(0.175)	(0.164)	(0.136)
Is estimated ethnically Indian			0.364	0.28	0.37
			(0.488)	(0.457)	(0.38)
Is project charitable			0.03	0.113	0.082
			(0.065)	(0.061)	(0.052)
Average people overall grade				0.012***	-0.0002
				(0.002)	(0.002)
Average presentation grade					0.003
					(0.002)
Average idea grade					-0.0006
					(0.002)
Average entrepreneur grade					-0.007***
					(0.002)
Average value for money grade					0.003
					(0.002)
Average likelihood of success estimate					0.017***
					(0.001)
Intercept	28.868	30.08	-11.497	-32.33	-32.041
	(19.478)	(19.478)	(30.254)	(28.464)	(23.601)
N-obs	473	473	473	473	473
Adjusted R-Squared	0.0577	0.0596	0.0913	0.202	0.454
AUC	0.7011	0.7042	0.7327	0.8054	0.9135

Panel D: Predicting Funding Success in Early Days

Evaluation Criterion	(1) Hard data	(2) +Sentiment	(3) + Definitive	(4) +People's overall grade	(5) All
Funding goal (in 10K)	-0.007**	-0.007**	-0.006**	-0.006*	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Image number/10	0.05*	0.05*	0.041	0.006	0.0002
	(0.02)	(0.002)	(0.025)	(0.025)	(0.024)
Video length (in minutes)	0.037*	0.038*	0.04**	0.029	0.018
	(0.015)	(0.015)	(0.015)	(0.015)	(0.014)
Facebook link	-0.105	-0.105	-0.111	-0.062	-0.057
	(0.06)	(0.06)	(0.056)	(0.053)	(0.057)
Twitter link	0.09	0.089	0.09	0.083	0.065
	(0.054)	(0.054)	(0.054)	(0.053)	(0.049)
Is project a physical object	0.108	0.113	0.139	0.162	0.099
	(0.094)	(0.094)	(0.095)	(0.092)	(0.087)
Is project artistic	0.05	0.182*	0.2*	0.246**	0.195*
	(0.084)	(0.084)	(0.085)	(0.082)	(0.077)
Is project a game	0.052	0.057	0.043	0.046	0.054
	(0.084)	(0.084)	(0.086)	(0.084)	(0.078)
Is project a film	0.018	0.017	0.007	0.012	0.032
	(0.073)	(0.074)	(0.075)	(0.072)	(0.068)
Is project musical	0.211**	0.216**	0.21**	0.205**	0.134
	(0.076)	(0.076)	(0.077)	(0.075)	(0.071)
Sentiment		0.02	0.018	0.019	0.022
		(0.024)	(0.024)	(0.024)	(0.022)
Average entrepreneur gender estimate			-0.005	-0.007	0.013
			(0.044)	(0.043)	(0.04)
Average entrepreneur age estimate			-0.033*	-0.031*	-0.03*
			(0.015)	(0.014)	(0.013)
Is estimated ethnically white			0.068	0.081	0.103
			(0.066)	(0.064)	(0.06)
Is estimated ethnically black			-0.017	-0.006	0.022
			(0.099)	(0.096)	(0.089)
Is estimated ethnically hispanic			-0.006	-0.017	0.036
			(0.085)	(0.082)	(0.076)
Is estimated ethnically Chinese			0.119	0.158	0.134
			(0.1118)	(0.114)	(0.106)
Is estimated ethnically Indian			-0.353	-0.316	-0.165
			(0.248)	(0.241)	(0.225)
Is project charitable			-0.04	0.03	0.004
			(0.073)	(0.072)	(0.067)
Average people overall grade				0.007***	0.0005
				(0.001)	(0.002)
Average presentation grade					0.001
					(0.002)
Average idea grade					-0.001
					(0.002)
Average entrepreneur grade					-0.005**
					(0.002)
Average value for money grade					0.002
					(0.002)
Average likelihood of success estimate					0.011***
					(0.001)
Intercept	33.535	33.6	57.449	22.149	21.917
	(19.71)	(19.715)	(30.759)	(30.512)	(28.649)
N-obs	513	513	513	513	513
Adjusted R-Squared	0.0779	0.0773	0.0806	0.134	0.259
AUC	0.7130	0.7138	0.7250	0.7767	0.8463

Table 4: Cross-correlation between Ratings and Hard Variables

We compute pairwise correlations between the ratings provided by our human raters and some of the hard features of the campaign. For each campaign, we record the funding goal and other "hard" features—i.e., variables available to both robots and humans as they scan through the project's page—such as the number of images, length of videos, links to social media features (indicating whether the page includes a Facebook or Twitter link), and a sentiment score between 1 and 5 based on textual analysis ("Very negative" = 1, "Negative" = 2, "Neutral" = 3, "Positive" = 4, and "Very positive" = 5). Each participant had to provide an overall rating, as well as a rating based only on one *specific* dimension of the evaluation criteria while ignoring all other criteria. Ratings range from 0 to 100. *, ** and *** denote significance at the 10%, 5% and 1% levels respectively. Statistical significance employs the Mann-Whitney U test.

Panel A: Pearson Correlation between Ratings and Hard Variables

	Presentation grade	Idea grade	Entrepreneur grade	Value for money grade	Likelihood of success grade	Overall grade
Image number	0.144***	0.088***	0.086***	0.073***	0.085***	0.102***
Video length	0.215***	0.098***	0.124***	0.075***	0.095***	0.104***
Facebook link	-0.088***	-0.081***	-0.082***	-0.081***	-0.057***	-0.1***
Twitter link	0.079***	0.037**	0.035**	0.053***	0.054***	0.061***
Sentiment	0.068***	0.033*	0.041**	0.016	0.024	0.033*
Average entrepreneur gender estimate	0.016	0.016	0.028*	-0.017	0.019	0.016
Average entrepreneur age estimate	0	0.005	0.056***	0.005	-0.002	0.011

Panel B: Spearman Correlation between Ratings and Hard Variables

	Presentation grade	Idea grade	Entrepreneur grade	Value for money grade	Likelihood of success grade	Overall grade
Image number	0.315***	0.158***	0.163***	0.152***	0.156***	0.184***
Video length	0.223***	0.104***	0.132***	0.081***	0.098***	0.11***
Facebook link	-0.087***	-0.08***	-0.08***	-0.082***	-0.054***	-0.102***
Twitter link	0.08***	0.04**	0.034**	0.055***	0.055***	0.062***
Sentiment	0.085***	0.04**	0.049***	0.024*	0.032*	0.043***
Average entrepreneur gender estimate	-0.008	-0.018	0.058***	-0.022	-0.004	-0.02
Average entrepreneur age estimate	0.012	0.023	0.061***	0.016	0.003	0.029*

Table 5: Multivariate Regressions of Sentiment on Focused Ratings

We applied a sentiment analysis to the textual description of the projects, reflecting the tone expressed in the text (“Very negative” = 1, “Negative” = 2, “Neutral” = 3, “Positive” = 4, and “Very positive” = 5). Sentiment is a computer-generated sentiment score using the Stanford CoreNLP library (see Section 2.5). The LHS variable is sentiment, and the RHS variables include the focused ratings provided to individual evaluation criteria. We use this specification to examine the cross-correlations between sentiment and individual evaluation criteria in a multivariate setting. Ratings range from 0 to 100. Each participant had to provide a rating based only on one *specific* dimension of the evaluation criteria while ignoring all other criteria. Standard errors are in parentheses, and *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

Evaluation Criterion	All	Product	Non-product	Art	Game	Film	Music
Idea	0.0002	0.0007	-0.0004	0.0003	$2.39 \cdot 10^{-6}$	0.0004	-0.0009
	(0.0006)	(0.0009)	(0.0009)	(0.0008)	(0.002)	(0.001)	(0.002)
Presentation	0.003***	0.003***	0.003***	0.003***	0.0005	0.004**	0.003
	(0.0006)	(0.0008)	(0.0009)	(0.02)	(0.002)	(0.001)	(0.002)
Value for money	-0.0014*	-0.0005	-0.002*	-0.003**	0.002	-0.002	-0.003
	(0.0006)	(0.0009)	(0.0009)	(0.0009)	(0.002)	(0.001)	(0.002)
Entrepreneur	0.0002	$-3.05 \cdot 10^{-5}$	0.0006	$-1.71 \cdot 10^{-5}$	0.0007	-0.001	0.004*
	(0.0007)	(0.0009)	(0.001)	(0.0009)	(0.002)	(0.001)	(0.002)
Likelihood of success	-0.0002	$7.01 \cdot 10^{-5}$	-0.0007	$-1.58 \cdot 10^{-5}$	0.0006	-0.002	-0.002
	(0.0005)	(0.0006)	(0.0006)	(0.0006)	(0.002)	(0.001)	(0.001)
Intercept	2.37***	2.28***	2.46***	2.38***	2.21***	2.55***	2.27***
	(0.03)	(0.037)	(0.04)	(0.04)	(0.11)	(0.05)	(0.08)
N-obs	6,098	2,961	3,137	3,554	459	1,159	972
Adjusted R-Squared	0.00491	0.00801	0.00395	0.00629	0.00127	0.00593	0.0107

Table 6: Correlations between Individual Evaluation Criteria Ratings and the Overall Rating
The table reports the correlations between the focused ratings of each evaluation criterion and the overall rating. In addition to an overall rating given for each campaign, we asked raters to give a rating based only on one *specific* dimension of the evaluation criteria. All ratings are between 0 and 100 (0 being the worst and 100 being the best).

Evaluation Criterion	Correlation (R^2)
Idea	0.7057
Presentation	0.6121
Value for money	0.6533
Entrepreneur	0.6019
Likelihood of success	0.4487

Table 7: Actual Importance of Evaluation Criteria in the Rating Data
To estimate the actual importance of the evaluation criteria in the data, Panel A uses a multiple linear regression model. The predicted variable is the overall rating given to the crowdfunding campaign, and the predictor variables are the focused ratings given to the individual evaluation criteria. In addition to an overall rating given for each campaign, we asked raters to give a rating based only on one *specific* dimension of the evaluation criteria. All ratings are between 0 and 100 (0 being the worst and 100 being the best). Standard errors are in parentheses, and *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.

In Panel B we report several factor importance measures based on the multiple regression in Column 1 in Panel A: partial correlations (PC), normalized regression coefficients (β), and the change in the coefficient of multiple determination (CCMD). The numbers in parentheses are the values normalized to 100%. By partial correlations we refer to the correlation between one predictor variable and the target variable, controlling for all remaining predictor variables. By the change in the coefficient of multiple determination for a predictor variable x_i, we refer to the change in R^2 value between the regression containing all predictors (including x_i) and the regression containing all predictors except x_i.

Panel A: Linear Regressions

Evaluation Criterion	All	Product	Non-product	Art	Game	Film	Music
Idea	0.35***	0.39***	0.31***	0.32***	0.33***	0.35***	0.24***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
Presentation	0.16***	0.13***	0.19***	0.18***	0.23***	0.14***	0.26***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
Value for money	0.19***	0.19***	0.18***	0.19***	0.18***	0.20***	0.17***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)
Entrepreneur	0.10***	0.09***	0.11***	0.10***	0.03	0.06**	0.10***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.05)	(0.03)	(0.04)
Likelihood of success	0.02**	0.03**	0.01	-0.00	0.03	0.03	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)
Intercept	11.96***	11.38***	12.57***	12.77***	11.23***	13.57***	12.34***
	(0.50)	(0.72)	(0.70)	(0.66)	(1.94)	(1.14)	(1.32)
N-obs	6,098	2,961	3,137	3,554	459	1,159	972
Adjusted R-Squared	0.58	0.60	0.56	0.57	0.57	0.56	0.54

Panel B: Several Relative Importance Measures based on the Multivariate Regression (All Campaigns)

Evaluation Criterion	Significance	PC	β	CCMD
Idea	$p < 0.0001$	0.3534 (41.74%)	43.27%	0.0596 (63.49%)
Presentation	$p < 0.0001$	0.1763 (20.82%)	19.50%	0.0134 (14.27%)
Value for money	$p < 0.0001$	0.1965 (23.08%)	23.23%	0.0168 (17.86%)
Entrepreneur	$p < 0.0001$	0.0958 (11.31%)	11.95%	0.0039 (4.12%)
Likelihood of success	$p < 0.1$	0.0247 (2.92%)	2.05%	0.0003 (0.27%)

Evaluation Criterion	Comments
Idea	The quality of the proposed idea.
Presentation	The quality of the campaign's presentation.
Value for money	Whether the reward given to backers upon project completion provides a high value relative to the investment.
Entrepreneur	The qualities of the entrepreneur behind the project.
Likelihood of success	How likely it is that the campaign would succeed in delivering up on its promise

Figure 1: Evaluation Criteria Examined in this Study

To help us decide on the evaluation criteria examined further on, we conducted a preliminary survey which included free-form questions relating to natural factors that may be predictive of a campaign's perceived quality.

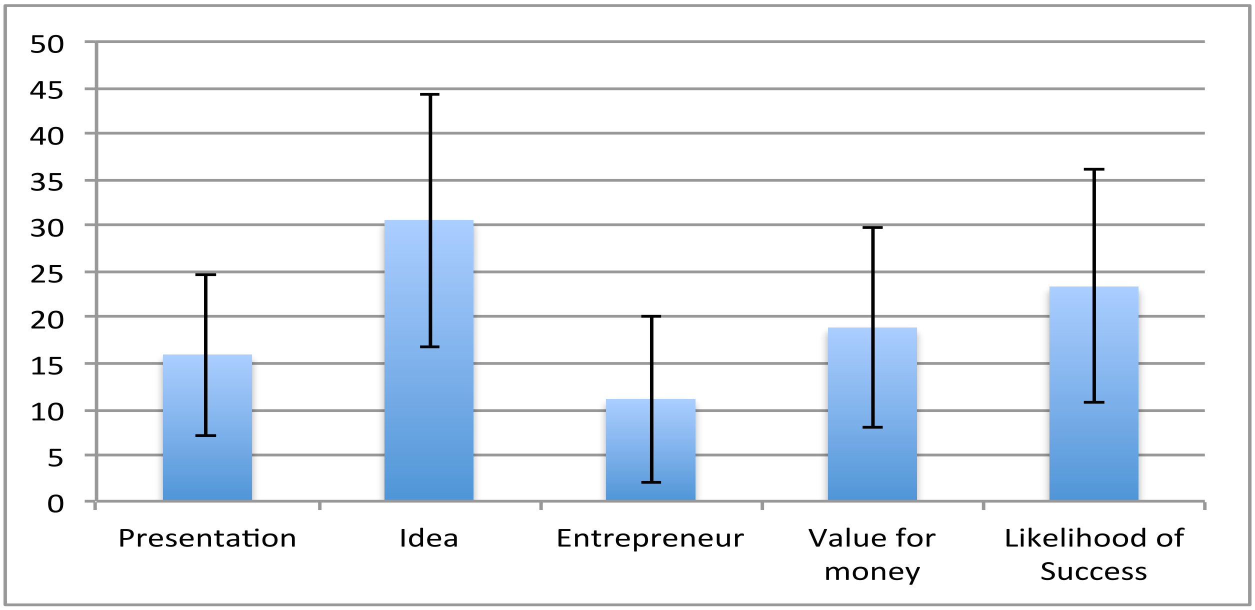

Figure 2: Perceived Importance of Crowdfunding Campaign Evaluation Criteria
We measure the perceived importance of the five specific factors that may be predictive of a campaign's perceived quality. We asked participants to allocate 100 points among them according to how important they perceive them in determining whether campaigns, in general, are worthy of investment. The distribution of these weights is plotted with confidence intervals.

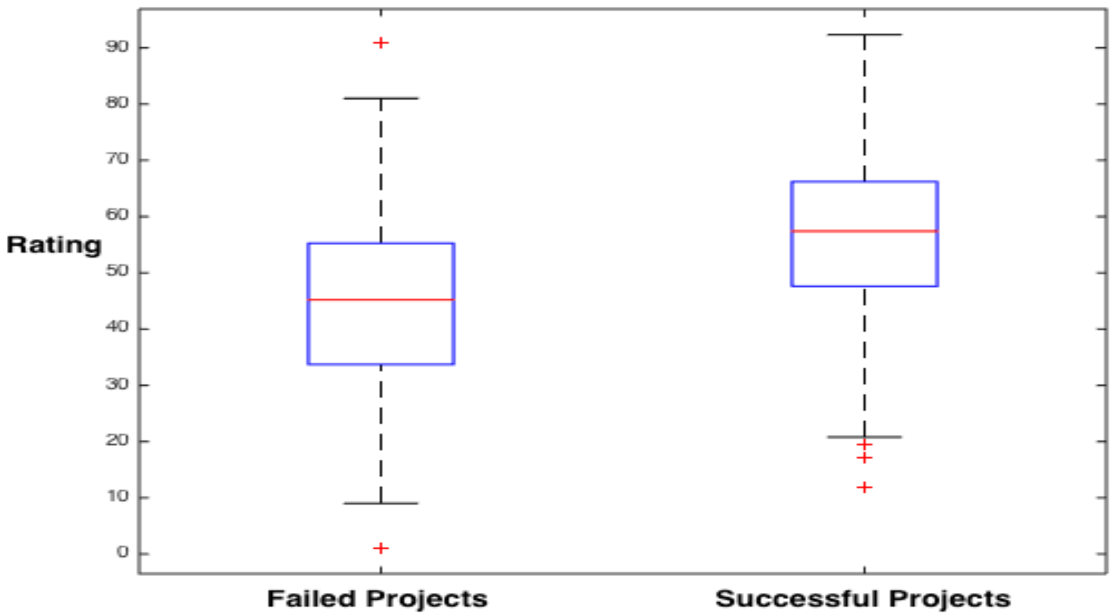

Figure 3: Crowdsourced Rating of Successfully and Unsuccessfully Funded Projects
The survey asks participants to rate actual Kickstarter campaigns while they are live. After collecting the ratings, we waited until the deadline for each campaign had expired and record whether it succeeded in attaining the funding goal. The boxplot shows the minimum, maximum, median, and first and third quartiles of the overall rating given to successfully and unsuccessfully funded campaigns.

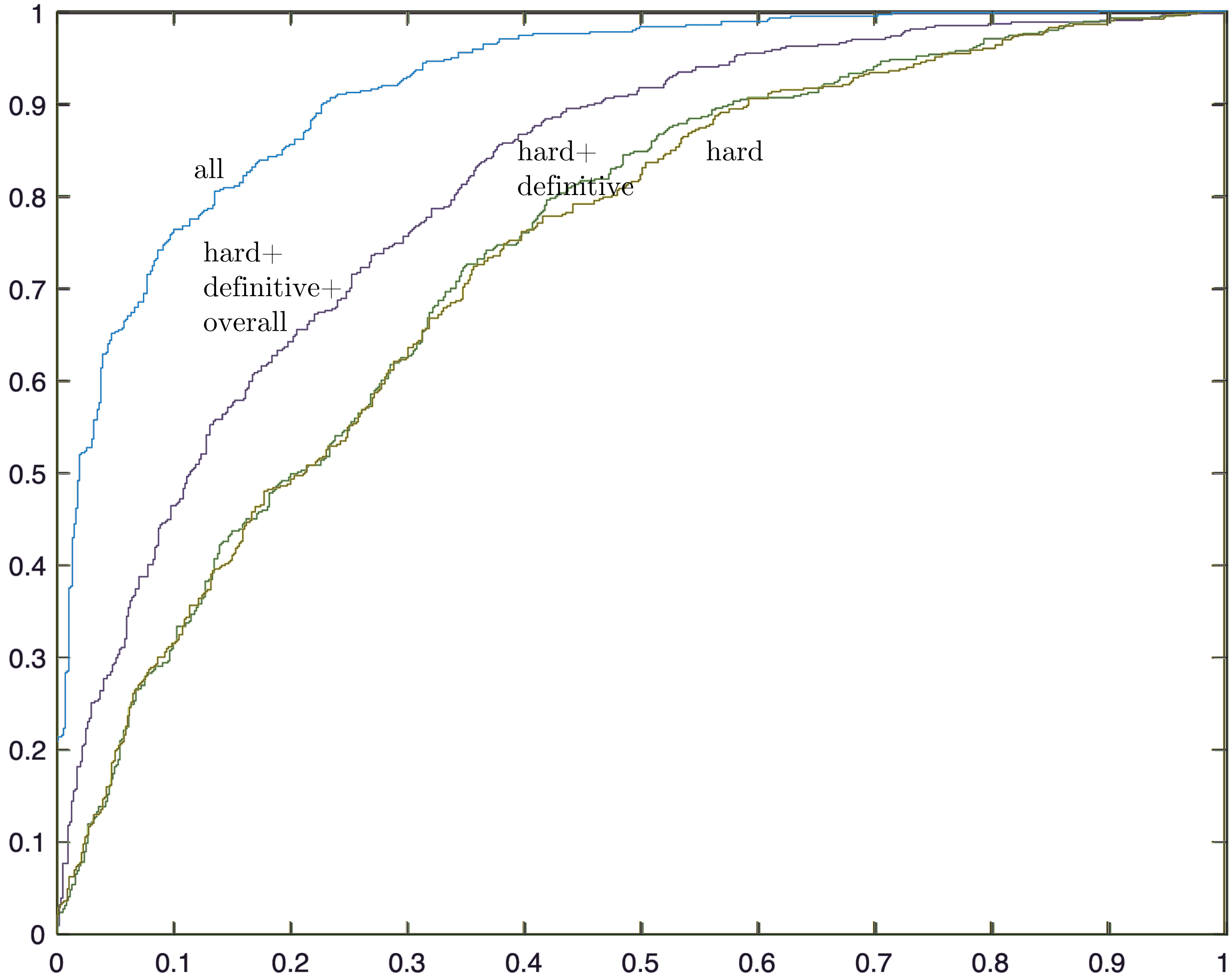

Figure 4: ROC Curves for Predicting Campaign Funding Success
A receiver operating characteristic (ROC) curve plots the true positive rate (the probability that an actual successful campaign will be classified as successful) against the false positive rate (the probability that an actual unsuccessful campaign will be classified as successful) as the discrimination threshold parameter (used to classify the estimated probability from the logistic regression of funding success) is varied between 0 and 1. As such, the more upper-left the curve is, the better. The different ROC curves correspond to different information sets used (only hard data, hard and definitive data, hard and definitive data as well as people's overall score, and all data, i.e., with "soft" evaluations).

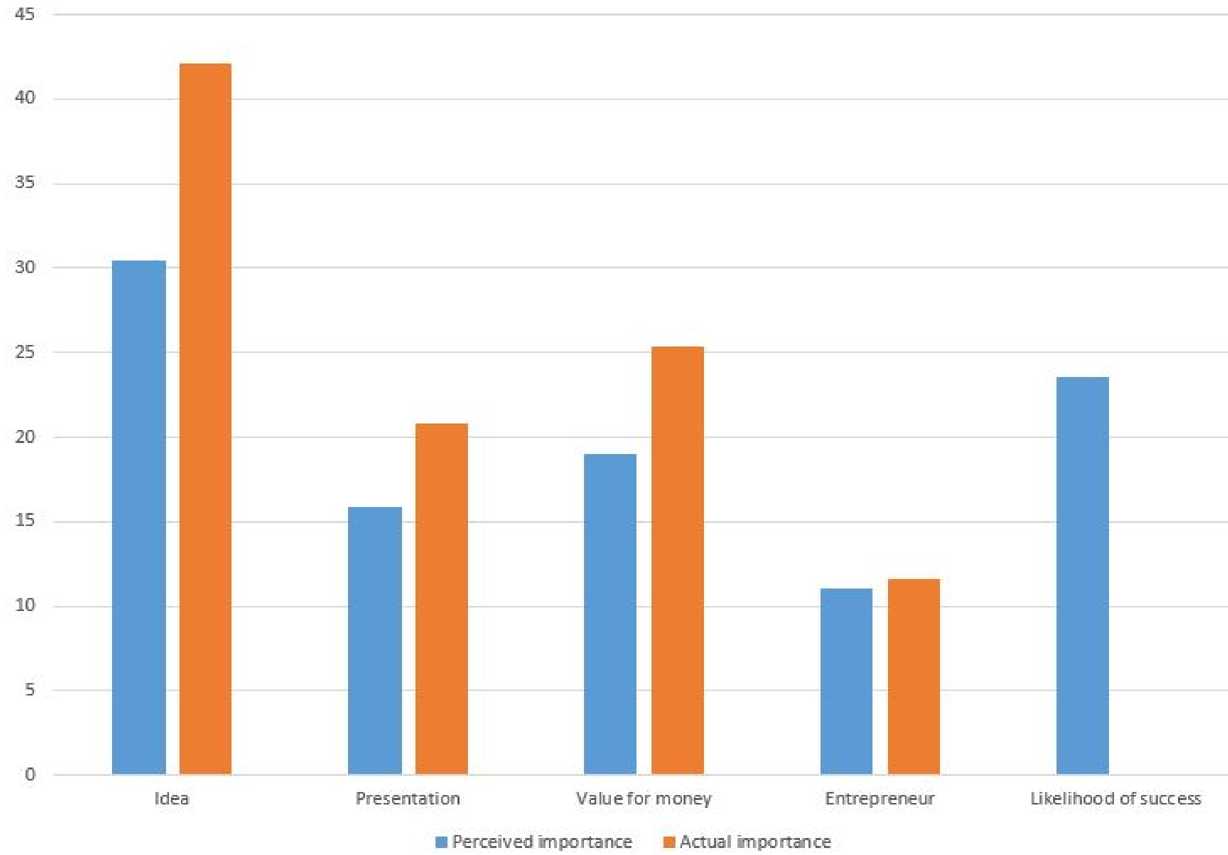

Figure 5: Perceived and Actual Importance of Evaluation Criteria
Our survey allows us test whether factors that generally seem more important to human raters—i.e., the relative influence people *believe* a factor would have on the overall rating provided to the campaign—also carry more weight in the data. The perceived importance is from Panel B of Table 1. We ask participants to allocate 100 points among the evaluation criteria considering the importance of each evaluation criterion in affecting their overall rating. The actual importance is the normalized partial correlations (PC) shown in Panel B of Table 7. By partial correlations we refer to the correlation of the focused ratings given to each individual evaluation criterion and the overall rating controlling for all the remaining evaluation criteria, renormalized to 100%.

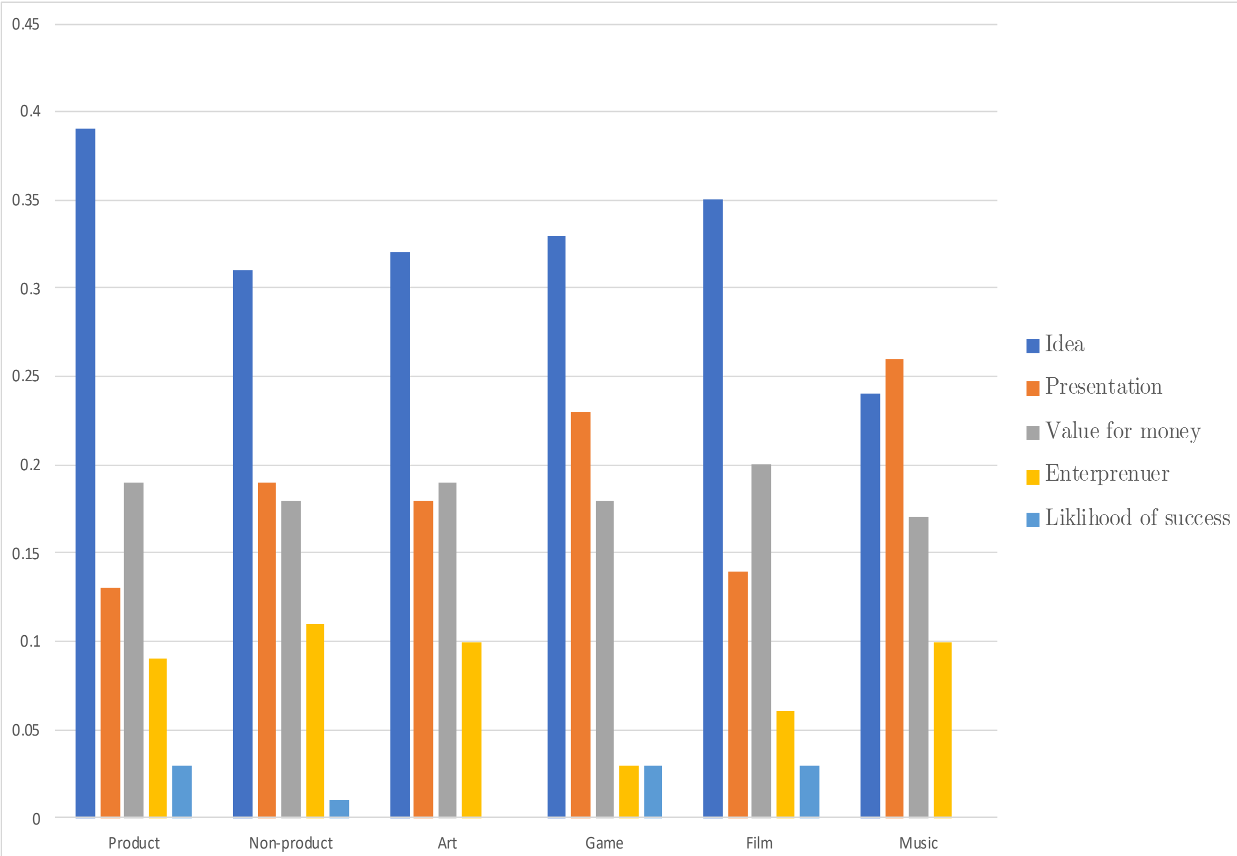

Figure 6: Actual Importance of Evaluation Criteria for Various Project Categories
To estimate actual importance, we regress the overall rating given for each campaign on the focused ratings based on each evaluation criterion. We plot normalized partial correlations (PC) based on the subsamples in Panel A of Table 7. By partial correlations we refer to the correlation between the focused ratings given to each individual evaluation criterion and the overall rating, controlling for all remaining evaluation criteria, and normalized to 100%.