NBER WORKING PAPER SERIES

PERSUADING INVESTORS:
A VIDEO-BASED STUDY

Allen Hu
Song Ma

Working Paper 29048
http://www.nber.org/papers/w29048

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

The paper was previously circulated under the title "Human Interactions and Financial Investment: A Video-Based Approach." We thank seminar participants at ASSA, Boston College, Chicago, Colorado, EFA, Finance in the Cloud, Georgia State, IESEG, Indiana, MFA, NBER Behavioral Finance Meeting, NBER Summer Institute (Corporate Finance; Entrepreneurship), RFS/GSU FinTech Conference, SFS Cavalcade, Washington St. Louis, WFA, Yale, and numerous colleagues for helpful comments. Ran You provided excellent research assistance. The project received IRB approval at Yale University. All errors are our own. For code examples and instructions to implement the method, please email the authors. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Persuading Investors: A Video-Based Study
Allen Hu and Song Ma
NBER Working Paper No. 29048
July 2021
JEL No. C55,D91,G24,G4,G41

ABSTRACT

Persuasive communication functions not only through content but also delivery, e.g., facial expression, tone of voice, and diction. This paper examines the persuasiveness of delivery in start-up pitches. Using machine learning (ML) algorithms to process full pitch videos, we quantify persuasion in visual, vocal, and verbal dimensions. Positive (i.e., passionate, warm) pitches increase funding probability. Yet conditional on funding, high-positivity startups underperform. Women are more heavily judged on delivery when evaluating single-gender teams, but they are neglected when co-pitching with men in mixed-gender teams. Using an experiment, we show persuasion delivery works mainly through leading investors to form inaccurate beliefs.

Allen Hu
Yale School of Management
165 Whitney Ave
New Haven, CT 06511
allen.hu@yale.edu

Song Ma
Yale School of Management
165 Whitney Avenue
New Haven, CT 06511
and NBER
song.ma@yale.edu

Many economic decisions are made after interpersonal persuasive communications, e.g., pitches to investors, sales presentations, and fundraiser events (McCloskey and Klamer, 1995). These interactions are often formalized by persuasion models (DellaVigna and Gentzkow, 2010). These models, and the empirical explorations guided by them, mostly focus on the content in persuasive interactions. The content may be informational, like the NPV of a project and the key function of a new product (Stigler, 1961; Milgrom and Roberts, 1986; Dewatripont and Tirole, 1999; Kamenica and Gentzkow, 2011). Conversely, the content may be noninformational, like the "framing" (Mullainathan, Schwartzstein, and Shleifer, 2008), the appealing peripheral content catering to people's intuition and attracting attention (Bertrand et al., 2010), or the "models" that lead receivers to interpret data and facts in a certain way (Schwartzstein and Sunderam, 2020).

Beyond content, however, it is widely believed that the delivery in persuasive communications matters for the final outcome—features like facial expression, tone of voice, or diction of speech—can be impactful. These persuasion delivery features go beyond static traits of persuaders like how they look. Instead, these features are dynamic and multi-dimensional. As William Carlos Williams wrote, "It is not what you say that matters but the manner in which you say it..." Despite its importance, empirical research on the non-content features in delivery, especially outside the laboratory,[1] remains very scarce. This scarcity can partially be attributed to the empirical challenges. A viable real-world setting is necessary to capture, represent, and quantify numerous features in the persuasion delivery. In addition, one needs to observe the economic decisions subsequent to the persuasion in order to quantify the impact and understand the economics mechanism.

This paper overcomes these challenges through two new innovations. First, we use a non-lab, real-world setting of persuading investors in which entrepreneurs pitch to experienced venture investors for early-stage funding. We collect full pitch videos as data inputs, allowing us to capture complete persuasion deliveries; and we observe investor decisions and future startup performance. Second, to quantify persuasion delivery, we exploit machine learning (ML) algorithms to quantify features along visual, vocal, and verbal dimensions. This method

[1]For example, social psychologists use laboratory studies to examine how "charisma" affects the effectiveness of leadership and outcomes in everyday life; see Awamleh and Gardner (1999), Antonakis et al. (2011), and Tskhay et al. (2018).

is efficient, reproducible, and easy to be adapted for other settings.

To capture features in persuasion delivery, our video-based method first projects pitch videos onto what we call the "three-V" spaces: the visual (e.g., facial expressions), vocal (e.g., tone of voice), and verbal (e.g., the word choices in the script). Given the purpose of capturing the whole persuasion process, we process the full video of each pitch. The visual dimension is represented as a series of images at a frequency of ten frames per second (i.e., six hundred images for a one-minute pitch). We extract the full audio stream from the video. The audio file encompasses both the vocal component and the verbal script generated through a speech-to-text algorithm.

Next, we construct persuasion delivery features from these three-V channels using ML algorithms. We use easily accessible computation services, such as Face++, Microsoft Azure, and Google Cloud, to perform the computation and construct measures. The key algorithms are trained, tested, and cross-checked by reputable providers using millions of human-rated training observations. The approach also allows for high replicability and transparency, and it lowers the computation burdens for researchers.

The algorithms generate detailed pitch features including visual emotions (e.g., positive, negative), vocal emotions (e.g., positive, negative, valence, arousal), textual sentiment (e.g., positive, negative), and psychological features (e.g., warmth). Based on those detailed features, we create an overall measure, the Pitch Factor, that summarizes "how well" the startup team delivers the pitch using a factor analysis, similar to that in Tetlock (2007) and common in unstructured data analysis. This process collapses detailed features into a single factor that captures the maximum variance in the set of pitch features. Empirically, the Pitch Factor loads positively on positivity dimensions and negatively on negativity dimensions, so intuitively the measure can be interpreted as the overall level of positivity—e.g., happiness, passion, warmth, enthusiasm—in a pitch.

Using the Pitch Factor to capture the delivery corresponds well with the entrepreneurial pitch setting and relevant social psychology models. In a large-scale venture capitalist survey conducted in Gompers, Gornall, Kaplan, and Strebulaev (2020), passion, broadly defined, is consistently ranked a top-three factor when selecting portfolio companies. Indeed, under the general impression that entrepreneurs are positive, energetic, and optimistic (Åstebro, Herz, Nanda, and Weber, 2014), the positivity feature revealed in the pitch may have the power to swing the opinions of agents thinking categorically and coarsely (Fryer and Jackson, 2008; Mullainathan et al., 2008). Additionally, positivity demonstrated in pitches may be contagious and may be particularly salient in affecting investors' emotional states (DellaVigna, 2009), which in turn influence both beliefs of future prospects and assessments of risk.[2]

Earlier studies, for example in behavioral sciences and political science, have shown the usefulness of videos as data. These studies generally use the "subject rating, thin sliced data, static perception" norm. That is, researchers recruit subjects to view thin-sliced data, most often still images (thus no movements or audio information), and sometimes also very short video clips. The subjects are then asked to rate static features of the speaker such as attractiveness, trustworthiness, and competence; these features in turn are correlated with outcomes. For illustration, a classic set of research asks subjects to review images or short video clips from political campaign activities to rate candidates, and it finds the first impressions rated by subjects correlate with election outcome.[3]

By introducing ML techniques, our method improves upon two dimensions. First, our method captures dynamic and more complete information and across multiple information channels. Instead of focusing on static perceptions, we quantify the complete persuasion process across all three-V information dimensions through the whole pitch video. Roughly speaking, given a set of static personal features (e.g., one with an above-average appearance), our method focuses on how the individual looks, sounds, and talks in a dynamic persuasive communication. In fact, we show dynamic delivery features work independently of static traits. Second, our method has high scalability and replicability. Even though the underlying algorithms are trained and cross-checked using millions of subject-rated data points, our method does not involve subject recruiting. The algorithm can be viewed as a speedy, tireless, and well-trained rater following a consistent standard; thus the method is replicable and

[2]For emotional contagion, see Hatfield, Cacioppo, and Rapson (1993) and Barsade (2002). For emotions, beliefs, and their impact in economics, see Johnson and Tversky (1983), Arkes, Herren, and Isen (1988), Clore et al. (1994), Loewenstein et al. (2001), Hirshleifer and Shumway (2003), Dahl and DellaVigna (2009), and Kuhnen and Knutson (2011).

[3]See for example Rosenberg et al. (1986), Ambady and Rosenthal (1992), Ambady and Rosenthal (1993), Schubert et al. (1998), Todorov et al. (2005), and Benjamin and Shapiro (2009). For some recent economic papers adopting the same norm, see Berggren, Jordahl, and Poutvaara (2010), Brooks et al. (2014), Blankespoor, Hendricks, and Miller (2017), and Huang et al. (2018).

ready to scale up computationally.

Armed with the measures for pitch delivery features, our analysis uncovers four main findings. First, startup teams that score a higher Pitch Factor, i.e., those showing more positivity, passion, and enthusiasm in their pitches, are more likely to obtain funding. This pattern is consistent across individual measures from the vocal, visual, and verbal dimensions. A one-standard-deviation change in the more passionate direction is associated with a three percentage point increase in the probability of receiving funding, or a 35.2 percent increase from the baseline in the probability of receiving funding. Following DellaVigna and Kaplan (2007), we calculate that an inter-quintile move of Pitch Factor has a persuasion rate of roughly six percentage points. The analysis also demonstrates the advantage of using the video data at full length and simultaneously considering the three-V dimensions: when running horse-race models, we find that Pitch Factor generated using the full video dwarfs that constructed using thin slices of videos; similarly, measures constructed from individual V-channels are less robust compared to the Pitch Factor. Moreover, the dynamic persuasion delivery works independently of static personal traits like beauty.

The impact of pitch delivery features does not seem to be simply explained by the omission of quality controls in the baseline analysis. Indeed, if the style of pitches correlates with entrepreneur-level quality traits (say, better founders communicate more passionately), the pitch feature may simply be picking up omitted quality metrics. We adopt a test following Altonji, Elder, and Taber (2005) and Oster (2019). We include an extensive set of controls used in the literature, such as founders' education, employment background, and startup experience, in the investment regression (Bernstein, Korteweg, and Laws, 2017). When we do so, the estimated impact of pitch features remains stable in economic magnitude and statistical significance. The statistical tests show that the impact of pitch features is robust to a wide set of reasonable parameters regarding omitted startup quality.

Second, we examine whether the "better pitch, more funding" finding is driven by investors incorporating pitch delivery features to improve their investment decisions. We make use of the long-term performance of startups. If the above finding simply reflects that a person who makes a better pitch runs a better startup, the invested companies with higher levels of pitch positivity would likely perform better than those with poorer pitch features (Fisman, Paravisini, and Vig, 2017; Ewens and Townsend, 2020).

The evidence does not support this view. We track startup performance using employment, the probability of failure (i.e., website no longer operational), and the probability of attracting follow-up financing and the long-term funding amount. None of the positive pitch features link to better long-term startup performance conditional on obtaining the accelerator funding. In fact, many positive pitch characteristics are associated with poorer performance. This analysis certainly does not intend to establish any causal interpretation between pitch features and performance. Instead, our preferred interpretation is that investors lower the bar (or equivalently, assign a high investment probability) for startups that demonstrate more positivity and passion in their pitch delivery, which lowers the true average project quality of the portfolio.

Third, we report evidence that the impact of delivery features on funding differs when judging entrepreneurs of different genders and does so in a direction consistent with gender biases. Previous research shows that women are more often judged based on appearance and not on substance (Fredrickson and Roberts, 1997). In addition, women and men are expected to follow different gender stereotypes (Bordalo et al., 2019)—for example, women are expected to portray warmth, empathy, and altruism, more than men are (Kite, Deaux, and Haines, 2008; Ellemers, 2018). Women also receive less recognition in group work (Sarsons et al., 2020). Biasi and Sarsons (2021) and Cullen and Perez-Truglia (2019) show that social interactions with managers often put female employees in a disadvantageous position compared to men.

In single-gender teams, women and men are evaluated along similar dimensions and in the same directions, but with different intensity. The penalty for not showing positivity and warmth is nine times bigger for women than men. This result does not seem to be explained away by different speaking styles, industry compositions of startups, or algorithmic accuracy across gender. Next, in mixed-gender teams, pitch features of men remain relevant, but pitch features of women become statistically and economically irrelevant. This suggests that women are essentially overlooked when co-presenting with their male teammates.

Finally, we investigate the economic mechanisms through which persuasion delivery features influence investors' decisions. We follow DellaVigna and Gentzkow (2010) and broadly categorize the potential mechanisms to (inaccurate) belief-based and preference/taste-based. We try to separate them using a venture investing experiment designed following the structure of Bohren, Haggag, Imas, and Pope (2019). In the experiment, we ask subjects to watch ten pitch videos and make investment decisions to maximize their payoff. Importantly, we also directly elicit their beliefs on the pitching startups. In accordance with the previous analyses, subjects are more likely to invest in startups with more positive pitch features.

Consistent with the inaccurate beliefs channel, we find that investors mistakenly think that startups with more positive pitch features are more likely to succeed, even though the realized performance of those companies is not higher, as discussed above—hence the inaccurate beliefs. After controlling for the elicited belief upon which the investment decision is based, pitch features remain influential as a standalone determinant, consistent with the preference-based channel. Our quantitative decomposition shows that the inaccurate beliefs and preference-based channels contribute 82 percent and 18 percent, respectively, to the relation between the non-content persuasion features and investment decision.

The key conceptual contribution of this paper is the focus on the delivery features in interpersonal persuasions. These features are widely regarded as important, but they are not well-studied in an economic setting. Persuasion models often treat persuasive features, such as the content and framing, as signals transmitted to the receiver. The delivery feature can be treated as an additional dimension of such signals and fit into those models. Our evidence suggests that the closest theories that could explain our findings are those that allow for persuasive features to change investor beliefs, instead of those in which persuasion features enter the utility independent of beliefs. More specifically, our evidence favors models that allow beliefs to be inaccurately formed. We want to acknowledge that we do not directly study the fundamental mechanism behind how incorrect beliefs are formed. As briefly discussed earlier, those beliefs could be incorrectly swayed by coarse and categorical thinking, double-counting repeated information, or appeals to emotions.

Our paper provides a ML-based method to systematically explore unstructured video data in economic research. An emerging line of work uses unstructured data and ML techniques in economics. Gentzkow, Kelly, and Taddy (2019) thoroughly review the progress in using textual data in economics and finance in the past decade.[4] Hobson, Mayew, and Venkatachalam (2012) and Mayew and Venkatachalam (2012) use voice indicators in analyst conference calls to study CEO private information. Gorodnichenko, Pham, and Talavera (2021) study the influence of vocal emotions of Fed chairs in FOMC meetings on financial markets. A nascent literature uses ML-based algorithms to code static images (Boxell, 2018; Joo and Steinert-Threlkeld, 2018; Peng, 2018; Abrams, 2019; Peng et al., 2020). Our method improves upon this by allowing dynamic feature aggregation of all channels using complete video data, presenting opportunities to answer new questions in economics, such as in education (e.g., classroom recordings), labor (e.g., job interviews), and innovation (e.g., academic seminars).

Even though persuasion is prevalent and consequential in the functioning of financial markets, most empirical literature of persuasion focuses on marketing and advertising, i.e., persuading consumers, and on political economy, i.e., persuading voters and donors.[5] The literature on persuading investors has focused on how firms and analysts persuade investors through earnings announcements or stock recommendations. Our paper presents a unique setting in which interpersonal persuasion is particularly important for investment decisions. Given the economic significance of entrepreneurship and venture capital, the setting itself is economically important (Kortum and Lerner, 2000; Haltiwanger et al., 2013).

1. Data and Setting

Our empirical analysis investigates venture investment decisions after startup pitches. The data set consists of two main parts—entrepreneurs' pitch videos for accelerator applications and startups' company- and team-level information. The two parts are manually merged using company names. The sample spans from 2010 to 2019.

[4]For textual analysis in specific research settings, see Antweiler and Frank (2004), Tetlock (2007), Tetlock et al. (2008), Hoberg and Phillips (2010), Loughran and McDonald (2011), Hoberg and Phillips (2016), Loughran and McDonald (2016), among others.

[5]Instead of citing specific papers, we encourage readers to refer to DellaVigna and Gentzkow (2010) as a comprehensive survey on the empirical literature on this topic.

A. Video Data and the Pitch Setting

We use pitch videos when startups apply to five large and highly ranked accelerators in the US: Y Combinator, MassChallenge, 500 Startups, Techstars, and AngelPad. Accelerator investments are important for entrepreneurship and innovation (Hochberg, 2016; Lerner, Schoar, Sokolinski, and Wilson, 2018). As of July 2019, these accelerators have accelerated more than six thousand startups, which in total have raised over $48 billion of total funding. Many leading entrepreneurial companies were funded by accelerators, such as Dropbox (2007), Airbnb (2009), and DoorDash (2013). Accelerator investment typically grants a standard contract with a fixed amount of investment (ranging between $20,000 and $150,000, fixed within accelerator-year). This allows our study to focus on one clean "yes or no" investment decision and ignore other investment parameters like amount or term sheet negotiation.

When startups apply to accelerator programs, they are required (or highly recommended) to record and submit a self-introductory pitch video of standard length as part of the application process. These videos are typically one- to three-minute long, and they present the founder(s) introducing the team and describing the business idea. These videos, rather than being submitted to the accelerators directly, are uploaded to a public platform such as YouTube and links to these videos are provided in application forms. This procedure provides researchers an opportunity to access those videos. We develop an automatic search script for two public video-sharing websites, YouTube and Vimeo. The web crawler returns a list of videos using a set of predefined search keywords, such as "startup accelerator application video", and "accelerator application videos", among others. Appendix A provides more details on this process.

[Insert Table 1 Here.]

This process yields 1,139 videos used in our analysis. In Table 1 Panel A we report basic information at the level of video pitches. The median length of a pitch video is 68 seconds, and the mean is 83 seconds. These videos are not viewed often, and most do not attract any likes or dislikes. This is consistent with the fact that these videos are generic pitch videos for application purposes rather than for any marketing campaign or product promotion. Regarding team composition, we find that 46 percent of the startups have only one member pitching, and 54 percent have multiple members. The average number of members per video is 1.74. Forty-nine percent of the teams have only male founders, 27 percent have only female founders, and 24 percent have mixed genders.

We want to note that the videos in our analysis are an incomplete sample of all pitch videos ever made by accelerator applicants. Many startups may have chosen to unlist or privatize their videos to make them unavailable to us, as researchers, to search and view. We formally discuss the sample selection issue and its implications for our analysis in Section 3.A.1, and we show that this sample selection does not affect our findings or interpretations.

In our setting, investors did not view these pitches in person—these pitches are recorded and uploaded for investors to review during decision-making. As a result, when interpreting our findings, one needs to be mindful of how the same features (e.g., smile, passionate voice, word choices) that affect decision-making when watching the video can translate to in-person interactions. Reassuringly, Dana, Dawes, and Peterson (2013) show that experiencing in-person interviews and watching video interviews lead to similar biases. Using videos also facilitates the connection to empirical studies in other fields on persuasive materials, like advertisement, media materials, etc.

B. Startup Information and Team Background

We also collect startup information on both the companies and the founders. In venture investment, investors value human assets like education and work experience (Bernstein, Korteweg, and Laws, 2017; Howell, 2019). They may also be, often mistakenly, influenced by discriminative factors, most noticeably gender (Gompers and Wang, 2017; Ewens and Townsend, 2020; Gornall and Strebulaev, 2019; Hebert, 2020). We incorporate those investment determinants in our analysis.

Startup information is collected from two widely used entrepreneurship databases, Crunchbase and PitchBook. We start by searching for companies in these two databases according to the names identified in application videos using video title, subtitle, and uploader ID. For startups with duplicate names or name changes, we identify companies by names of founders, business descriptions, and time of founding. Startup-level variables include the year of founding, location, operating status, total funding round and amount, number of investors, and number of employees.

Beyond company-level information, we also collect information on the founding teams. We compile a list of founders for each startup company using Crunchbase, PitchBook, and video content. For each startup team member, we use LinkedIn to extract the five most recent education experiences and the ten most recent work experiences. This information is used to construct variables that indicate each team member's education (university, degree), job seniority, entrepreneurship experience, etc. We also standardize companies' industry classifications using the Global Industry Classification Standard (GICS). We categorize all companies into one of 24 GICS industry groups, which then form 11 GICS sectors, using the industry information in Crunchbase and PitchBook along with the video scripts.

Among these 1,139 applications, 462 are included in Crunchbase, and 217 can be found in PitchBook. The summary statistics of these startups are reported in Panel B of Table 1. On average, each startup, conditional on receiving funding from at least one investor, raises \$12,292,000 in total, and the median is \$365,000. Regarding the startup founder teams' backgrounds, we find that 30 percent of the founders have startup experience at the time of application. Nineteen percent and 3 percent of them hold Master's degrees and PhD degrees, respectively.

2. Method: Processing Video Data with ML Algorithms

Our video processing method proceeds in three steps. First, we decompose the information embedded in the videos into three-V dimensions—visual, vocal, and verbal. Second, for each dimension, we adopt ML algorithms to create visual, vocal, and verbal features from the raw data. Third, we aggregate these measures within and across these dimensions to characterize features representing each pitch video. Below we first give an overview of the key properties of video data relevant to economic research, and then we describe the three-step method in depth. Appendix B provides additional technical details.

A. Video as Data: Key Properties

Videos are a pervasive form of data. More than 80 percent of the world's internet traffic consists of transmissions of video, and more than 60 percent of the total digital data stored worldwide are video. However, videos are underexploited in economics research largely due to the complexity and computational burden. We begin by discussing some basic properties of videos and illustrate how these properties relate to our video processing and measure construction process.

First, video data are information intensive. To better understand the richness of video data, one can make a size comparison between video files and other data files. The csv-format startup panel in this study is around 1 MB in size. It includes 150 company-level characteristics for 1,139 startups. In contrast, a one-minute high-resolution pitch video in mp4 format can be as large as 200 MB in our sample. To put this into perspective, one second of a high-definition video, in terms of size, is equivalent to over 2,000 pages of text (Manyika et al., 2011; Gandomi and Haider, 2015).

Second, video data are unstructured and high-dimensional, making them more complicated to process relative to other data formats such as panel data or even unstructured textual data. Consider a one-minute video with a resolution of 1280×720 (720p) and two 48 kHz audio channels. In this case, there are 1280×720=921,600 pixels in each image frame. If we use a sampling rate of ten frames per second, the video can be represented as a series of 600 images. These 600 images need 552 million pixels (dimensions) to be represented. Further, to represent the acoustic information, we need 48,000×2 dimensions per second and thus, around 5 million dimensions for one minute. In total, to represent such a video, we need around 560 million dimensions.

Third, video data have a low signal-to-noise ratio (SNR) and low economic interpretability. Regarding the SNR, most videos have a large amount of background noises that are irrelevant to the primary economic question. In our setting, these noises include background noise, furniture in the video, among other things. Moreover, the information units of video data (e.g., the pixels and the sound waves) are not directly interpretable as accounting variables or textual words. Thus, when processing video data, we need to impose structures and algorithms to guide the extraction of information that is useful and meaningful for economic research.

B. Step 1: Information Structure and Representation of Video Data

The first step in video data processing is to decompose videos into the three-V structure and to represent them in a data format. Three-V means visual, vocal, and verbal; this structure is intuitive and widely accepted in the social psychology and communication literature (Mehrabian, 1972; Strahan and Zytowski, 1976; Krauss et al., 1981; Knapp et al., 2013).

We first extract these three dimensions from the two streams of digital information in raw video records—the image stream and the sound stream. For the visual component, we represent the video using images sampled at ten frames per second and employ face-detection ML algorithms to identify human faces in each video frame.[6] The unit for visual analysis is thus at the level of each video frame. For the vocal component, we extract the audio files from the video. For the verbal component, a speech-to-text ML algorithm is used to extract speech from the sound.

We now consider each of these three dimensions and discuss their data representations one by one.[7] For visual, human faces exist in the format of digital images, which can be numerically represented as two-dimensional matrices. Moreover, since we have a stream of such digital images, the information in the facial dimension is coded as a time series of two-dimensional matrices. The vocal signal, essentially sound waves, can also be digitized as a time series of amplitudes given a specific sampling rate (the number of times per second that the amplitude of the signal is observed) and resolution (the number of discrete levels to approximate the continuous signal). If the audio is multi-channel, it can be represented as a matrix with channels as columns/rows. Finally, we transform human speech into a term-frequency matrix, which tracks the use of words in the verbal content.

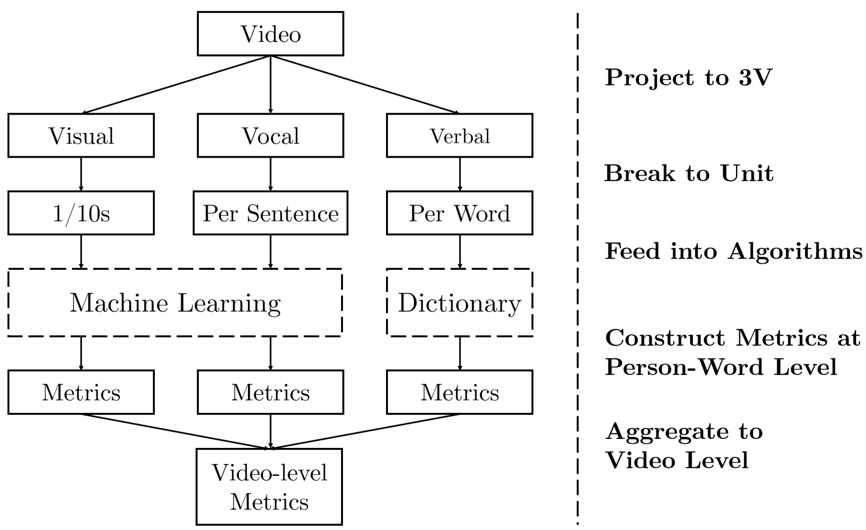
[6] 0.1 seconds is a very short time interval for our study and for the goal of capturing facial expressions. Even fast facial movements, like blinking, take on average around 0.25 seconds, and will be captured by our algorithm at the 1/10th-second intervals.

[7] This step builds on and extends the `pliers` package available at http://github.com/tyarkoni/pliers and McNamara, De La Vega, and Yarkoni (2017).

C. Step 2: Constructing Measures with Machine Learning

In the second step, we construct economically interpretable measures from the represented video data using ML algorithms. One way to think about our ML algorithms is as a robust and objective super robot that can rate and record the video data at a high frequency along the three-V dimensions. The ML algorithms are trained using millions of observations rated by subjects and are now automatic and easy to scale up. To do so, we leverage many recent groundbreaking advances in computer vision, speech recognition, and textual analysis.

Illustration of the Data Processing Framework



C.1. Visual. We identify human faces embedded in each video image using face detection algorithms. For replicability and transparency, we directly adopt the established implementation of Face++.[8] The Face++ platform provides APIs through which we feed our raw images into the cloud computing system and receive a host of face-related measures constructed by Face++'s ML algorithms. We also check the robustness using an alternative computation platform, Microsoft Azure Cognitive Services.[9]

[Insert Figure 1 Here.]

[8]Face++ can be accessed at: https://www.faceplusplus.com/emotion-recognition.

[9]Microsoft Azure Cognitive Services can be accessed at: https://azure.microsoft.com/en-us/services/cognitive-services/face.

The process is as follows. First, the algorithm detects locations of facial landmarks (e.g., nose, eyes) from the raw images using face detection technology. These coordinates allow us to detect facial movements, such as smiles, eye blinks, or the lowering of eyebrows. They also detect mouth movements that can help identify the speakers continuously. Second, this information enters emotion recognition algorithms that categorize facial emotions into one of the following six dimensions—happiness, sadness, anger, fear, disgust, and neutral. In the empirical analysis, we combine the measures of sadness, anger, fear, and disgust into a composite facial negative emotion measure. We exclude neutral facial emotion in our analysis because the sum of facial positiveness, negativeness, and neutrality is one, which induces collinearity. Third, we obtain a face beauty measure and some demographic characteristics of individuals, including gender and predicted age.

These variables are different from the static features of beauty and impressions (attractiveness, competence, etc.) that are used in prior economic studies. Intuitively, we take out the static features (e.g., people can be more or less attractive) that can be viewed as baseline features. We then track the movement of the face to detect facial emotions (e.g., passionate—which can be a feature of speakers with both more and less attractive appearances).[10]

Example. In Figure 1 we present sample frames of high-positivity and low-positivity facial expressions. As discussed earlier, the calculations of positivity, negativity, and other facial features are done at a frequency of every 1/10 of a second.

C.2. Verbal (Text). Next, for verbal, we extract human speech from audio data using the speech-to-text conversion API provided by Google Cloud.[11] This ML-based API converts audio into a text transcription. These transcripts include a list of words, time stamps (onsets, offsets, and durations) of these words, and punctuation. We then merge the speech corpus with two dictionaries. The first dictionary is the Loughran-McDonald Master Dictionary (LM

[10]This image processing technology is related to some recent work that explores images as data (Joo and Steinert-Threlkeld, 2018; Peng, Teoh, Wang, and Yan, 2020), particularly in communication and political science (Peng, 2018; Boxell, 2018). These papers typically focus on static images. Our images are continuous and extracted from videos, and the processing is conducted for the purpose of capturing dynamic interaction features from the video.

[11]Google Cloud Speech-to-Text API can be accessed at: `https://cloud.google.com/speech-to-text`.

hereafter), which is commonly used in financial text analysis and provides text categories such as negative and positive, among others (Loughran and McDonald, 2011). The second dictionary is developed by social psychologists using Wordnet and word embeddings (Nicolas, Bai, and Fiske, 2019). This dictionary (NBF hereafter) includes word categories along the dimensions of social psychological traits such as ability and warmth, which helps us to measure verbal characteristics from the angle of social perception (Fiske et al., 2007).

[Insert Figure 2 Here.]

Example. Given that researchers in economics and finance are already familiar with the negativity and positivity categorizations in textual analysis, we here provide an example of the "ability" and "warmth" dimensions in verbal expressions. In Figure 2, we show a high-ability script in Panel (a) and a high-warmth script in Panel (b). High-ability pitches focus on the ability of the entrepreneur and the operational efficiency of the business idea. In contrast, high-warmth pitches focus on communion, pleasing personalities, and a description of a bright future working with VCs and customers.

C.3. Vocal (Voice). Finally, we analyze information embedded in the vocal channel. In this part, we regard voice as digital signals and focus on the physical information not captured by textual transcription of human speech. Different from images that can provide rich information independently, audio data, essentially sound waves, code information in the audio's dynamics and auto-dependence structure. In other words, if we split the audio into fixed high-frequency segments like a series of images, we may lose the information embedded in the auto-dependence structure. We address this technical problem by focusing on word or sentence units by leveraging the outputs of speech-to-text algorithms. Specifically, we split each audio stream into segments by words and sentences. These units naturally reflect the auto-dependent information structure in the video and are also good approximations of human cognitive processes.

We employ `pyAudioAnalysis` (Giannakopoulos, 2015), which is a Python package for audio feature extraction, classification, segmentation, and application. We extract 34 low-level features in total. These features include but are not limited to spectrogram, chromagram, and energy. These features capture the physical characteristics of the vocal channel.[12]

We then construct high-level cognitive measures by feeding these low-level audio features into ML algorithms performing vocal emotion analysis. We adopt two conceptual frameworks for vocal emotion. The first framework models vocal emotion along two dimensions—arousal and valence (Bestelmeyer et al., 2017). Valence measures the positive affectivity of the vocal feature, while arousal captures the strength of such an attitude (exciting versus calm)—in some sense, it captures the concept of "being passionate" often referenced in entrepreneurship studies (Gompers et al., 2020). We use `pyAudioAnalysis`'s established implementation and pre-trained ML models to obtain vocal arousal and valence. Another framework models vocal emotion along three emotional dimensions—happy, sad, and neutral. To implement this conceptual model, we use `speechemotionrecognition`, which includes deep-learning-based speech emotion recognition algorithms. We adopt the pre-trained ML models in this package and feed our audio segments into these models.[13]

[Insert Figure 3 Here.]

Example. In Figure 3 we provide the waveform amplitude plot for two pitches, one with high arousal (i.e., high excitement) and one with low arousal.[14] The general patterns of the sound waves differ significantly and can be coded by our ML algorithm. Even though it is more difficult to visualize other features, the logic applies—analyzing the sound waves allows the categorization of vocal features.

D. Step 3: Measurement Aggregation

After constructing measures for each channel separately, we merge these measures and aggregate them to the video level. An appendix table of measures constructed in this paper along with their definitions and construction algorithms/procedures can be found on page 39.

[12] For a complete list of these audio features, please check `https://github.com/tyiannak/pyAudioAnalysis/wiki/3.-Feature-Extraction`.

[13] The deep-learning models in this package are trained on the Berlin Database of Emotional Speech (Burkhardt et al., 2005). Vocal-Neutral is dropped from our analysis due to collinearity.

[14] The high-arousal pitch can be downloaded from `https://www.dropbox.com/s/ipluo2w9tsszu2m/High%20Arousal%20Example.wav?dl=0`, and the low-arousal pitch can be downloaded from `https://www.dropbox.com/s/7igoqkjla72usdc/Low%20Arousal%20Example.wav?dl=0`.

D.1. Simple Aggregation. We first aggregate all measures at the video level by taking the mean of measures across everyone in the pitch and across the whole video. We create *Visual-Positive*, *Visual-Negative*, *Vocal-Positive*, *Vocal-Negative*, *Vocal-Valence* and *Vocal-Arousal* to capture facial and vocal emotions. For each of these variables, we compute the average proportion of time in the pitch that a team member shows certain facial or vocal emotion. We create *Visual-Positive* as the score of visual happiness and create *Visual-Negative* as the combined score of visual anger, sadness, fear, and disgust. This allows us to mitigate potential measurement errors introduced by the ML algorithm in classifying subtly different negative visual emotions. For vocal measures, we label vocal happiness as *Vocal-Positive* and vocal sadness as *Vocal-Negative*. For verbal content, we focus on *Verbal-Positive* and *Verbal-Negative* in the LM financial dictionary and *Verbal-Warmth* and *Verbal-Ability* in the NBF social psychology dictionary. *Verbal-Positive* and *Verbal-Negative* are calculated as the word counts (in each category) scaled by the total number of words in each pitch. *Verbal-Warmth* and *Verbal-Ability* are calculated as the directed word counts (+1, if in the positive direction of the category; −1, if in the negative direction; 0, if unrelated to the category), which are also scaled by the total number of words. Figure 4 illustrates the data structure and shows how we transform unstructured video data into our final structured panel data set.

[Insert Figure 4 Here.]

Table 2 presents the summary statistics for pitch videos. We show in Panel A the mean, median, and 25th and 75th percentiles. There are substantial cross-sectional variations in startup pitches along the three-V features. Take *Visual-Positive* as an example. The mean, 0.17, can be roughly seen as indicating that in a pitch video, on average, the speakers show clear happy visual features about 17 percent of the time. But this number varies quite dramatically. At the 25th percentile, one speaker could show happy facial features only 5 percent of the time, while the 75th percentile most positive team could score 25 percent in this measure. Some features, especially those capturing negativity, have low means. For instance, the vocal and verbal negativity measures both score 1 percent at the mean—this is not surprising given that entrepreneurs would likely try to hide negativity during pitches.

This does not mean, however, that those negativity measures are less meaningful—on the contrary, as will be shown below, the negativity features are important in our analytical results.

[Insert Table 2 Here.]

In Table 2 Panel B we show the correlation between metrics from different channels. We find that within a given dimension (the same V), features are highly correlated—for example, videos showing more positive attitudes naturally will show fewer negative attitudes. Meanwhile, across dimensions, similar metrics correlate in very interesting ways. Positivity in vocal features is positively correlated with positive visual expressions, confirming the validity of our metrics that are actually generated using completely different information and algorithms. But the correlation is mild, suggesting that vocal and visual expressions are correlated yet separate signals. Verbal information is uncorrelated with vocal and visual features. This could be because textual scripts can be more easily prepared and recited, and thus they can be disjointed from the vocal and visual delivery.

D.2. Generating a "Pitch Factor". Beyond the set of detailed features, one may wonder—can we create one variable to summarize "how well" entrepreneurs deliver their pitches? We achieve this goal using a factor analysis to extract the most important common component from the variance–covariance matrix of the features in pitch videos. This process allows us to eliminate the redundant features in the complex pitch structure.

Operationally, we estimate the factor analysis using the principal component method similar to Tetlock (2007). The method chooses the vector in the pitch feature space with the greatest variance, where each feature is given equal weight in the total variance calculation. We explore other factor analysis estimation methods, such as principal factor analysis and maximum-likelihood factor analysis. The qualitative empirical conclusions are not sensitive to the method chosen, and the quantitative conclusions change only minimally. Effectively, principal components factor analysis performs a singular value decomposition of the correlation matrix of pitch features. The single factor selected in this study is the eigenvector in the decomposition with the highest eigenvalue—we label it as the "Pitch Factor."

The Pitch Factor not only summarizes the pitch delivery but bears a clear and intuitive interpretation. The last two columns of Table 2 Panel A report each variable's loading on the Pitch Factor and its "uniqueness". The loadings are positive for all measures that have positive and affirmative economic meanings. For example, the factor loads positively on *Visual-Positive* (+0.08), *Vocal-Positive* (+0.39), *Vocal-Arousal* (+0.91), *Vocal-Valence* (+0.88), *Verbal-Warmth* (+0.06), and *Verbal-Ability* (+0.06). Meanwhile, the factor loadings are negative for all measures in the negative direction, such as *Visual-Negative* (−0.14), *Vocal-Negative* (−0.30), and *Verbal-Negative* (−0.14). Together, the Pitch Factor can be viewed as a composite index that integrates the information from three-V channels and represents the overall level of positivity, passion, and warmth reflected in the pitch video. The uniqueness is the percentage of variable variance that cannot be explained by the factor. The uniqueness is low on average, so the factor is well-behaved and powerful.

D.3. Cross-Validation with Human Raters. As a way of cross-validation, we compare our Pitch Factor with the ratings solicited from respondents at Amazon Mechanical Turk (the MTurkers). Appendix C provides a detailed description of the survey designs, and here we overview the exercise only briefly in order to avoid distractions from the main method.

We use two surveys to document the high correlation between our Pitch Factor and the ratings provided by MTurkers. In the first design, we directly elicit ratings from participants to compare with Pitch Factor. The MTurk survey follows previous research to validate ML-based measures using human raters (Peng et al., 2020).[15] We recruited 115 respondents to watch pitch videos and provide ratings on the overall level of positivity in the pitches on a scale of 1-9, where positivity is defined as enthusiasm and passion for the respondents. We show in Figure A.3 and Table A.3 that a strong correlation exists between the Pitch Factor generated from our ML-based method and the human-rated positivity score, even after controlling for rater FE.

In our second design, we ask MTurker respondents to compare pitch positivity for pairs of randomly-drawn videos. For each of these random pairs, we evaluate the consistency between our ML-based ranking and the human ranking. In other words, does the algorithm pick

[15]MTurk is increasingly used for other purposes in economic research (DellaVigna and Pope, 2018; Lian, Ma, and Wang, 2019).

the same winners as the raters? We find that the same winner is picked with nearly 89.5% consistency. Interestingly, we find strong disagreement among MTurker raters themselves when the two videos in the same pair have close algorithm-generated Pitch Factors. In other words, our method seems to be able to provide a more decisive ranking when there are high levels of noise.

3. Empirical Analysis

A. Baseline Result: Positivity Pitch Features and Venture Investment

Our first analysis examines whether delivery features in startup i's pitch relate to its likelihood of obtaining funding when applying to accelerator j during application year t. This is a cross-sectional data set with 1,139 pitches, since each investment evaluation happens only once in the sample. The analysis is performed using the following specification:

$$I(Invested)_{ijt} = \alpha + \beta \cdot X_{it} + \delta_j + \varepsilon_{ijt}. \tag{1}$$

The key outcome variable is $I(Invested)$, which takes a value of 1 if the startup was chosen by the accelerator and 0 otherwise. On the right-hand side, all pitch features are standardized into a zero-mean variable with a standard deviation of one, making the economic magnitudes easier to interpret and compare across variables.

We control for accelerator fixed effects to account for the possibility that certain accelerators might attract specific types of startup founders or that they have different investment criteria or preferences that could correlate with pitch features. Standard errors are clustered at the accelerator-year level. This takes into account the fact that an investment decision regarding one startup automatically correlates with the accelerator's decisions about others applying in the same year given the investment quota constraint accelerators face.[16]

[Insert Table 3 Here.]

[16]In Appendix Table A.4, we show that the results are qualitatively identical and quantitatively similar when using visual measures constructed using Microsoft Azure instead of the Face++ API. In Table A.5 we show that the results are robust to alternative fixed effects combinations such as industry FE, and estimation method such as OLS.

Table 3 presents the results of Eq. (1) estimated using a logit model. For each feature, we show the marginal effect calculated at the sample mean, the standard error, and the pseudo R^2 in each row. We first focus on the left panel of the table, in which the models do not add startup/team controls. The Pitch Factor, which captures the overall level of positivity, passion, and warmth of the pitch, positively and strongly correlates with the probability of receiving funding from the accelerator. The 0.030 coefficient means that a one-standard-deviation increase of the factor is associated with a change of three percentage points in funding probability, which is equivalent to a 35.2 percent increase from the baseline funding rate of 8.52 percent.

We also interpret this economic magnitude through persuasion rate (DellaVigna and Kaplan, 2007). Consider the following thought experiment: judges view either a pitch with a high pitch factor (top twenty percent) or a low pitch factor (bottom twenty percent). The former is labeled as treated, by the passionate pitch, and the latter as untreated. In our sample, the probabilities of funding in the treated and control group are 12.33 percent and 6.14 percent, respectively. The persuasion rate is calculated as[17]

$$f = \frac{I_{HighPitchFactor} - I_{LowPitchFactor}}{1-0} \cdot \frac{1}{1 - I_{Baseline}} = \frac{12.33\% - 6.14\%}{1-0} \cdot \frac{1}{1-6.14\%} = 6.60\% \tag{2}$$

Essentially, the approach scales the change of investment probability after being exposing to the treatment (the first fraction) by the effective room for persuasion after excluding the baseline investment rate, which is unobserved and conventionally approximated using control group behavior (i.e., $I_{Baseline} = I_{LowPitchFactor}$). This is a large economic magnitude—for comparison, Engelberg and Parsons (2011) find that local news coverage has a persuasion rate of 0.01 percent on trading behaviors. This suggests that, not surprisingly, persuasion in our venture pitch setting is more influential.

We next focus on individual visual, vocal, and verbal measures, which presents a consistent message. A one-standard-deviation increase in happiness reflected in the visual dimension is associated with a 1.5 percentage point increase in investment likelihood, a 17.6 percent increase from the unconditional funding rate. Startup teams that show more negative facial

[17] For a more general discussion on persuasion rates, we recommend DellaVigna and Gentzkow (2010).

expressions, captured as Visual-Negative, are less likely to receive funding. The absolute economic magnitude is 80 percent larger than it is for Visual-Positive (0.027 versus 0.015), suggesting that the negative component is even more relevant in driving investment decisions. This is in the same spirit as findings that the negative spectrum is more relevant in research on "beauty premium" (Hamermesh and Biddle, 1994) and textual "market sentiment" (Tetlock, 2007; Loughran and McDonald, 2011).

Note again that the visual delivery measures are captured dynamically, and are independent of static facial traits such as beauty (and similarly attractiveness, competence, etc.). To get a sense of this independence and a benchmark for understanding the above economic magnitude, we lean on the well-established concept of the "beauty premium", which is shown to be important in the labor market and other economic decisions (Hamermesh and Biddle, 1994; Mobius and Rosenblat, 2006; Graham et al., 2016). We confirm that more beautiful entrepreneurs are more likely to receive investment. The economic magnitude of the beauty effect is roughly the same as it is for Visual-Positive, and smaller than Visual-Negative.

In the vocal dimension, we rely on two vocal emotion categorizations. In the first positive-negative categorization, the pattern is quite similar to the findings on visual features. More positive (negative) tones in pitches are associated with a higher (lower) probability of receiving accelerator financing. In the second categorization, the audio channel is projected to a two-dimensional space of arousal and valence. We find that high-valence and high-arousal pitches are more likely to attract investment; a one-standard-deviation change in either measure is associated with an increase of 2.3 percentage points (27.1 percent from baseline) in the probability of receiving investment.

Regarding the verbal dimension, the use of positive versus negative words in pitches matters for the investment decision. Consistent with prior research using this categorization in economics, negative words are more relevant for economic outcomes (Loughran and McDonald, 2011). Verbal-Warmth and Verbal-Ability dimensions are based on the social psychology dictionaries and capture how each word influences the listener's perceptions. Warmer pitches (i.e., friendlier and happier) attract more investment. In contrast, teams that choose to talk more about their ability and competitiveness more often drive investors away, which is somewhat surprising given that entrepreneurial investment is a professional decision involving identifying more capable entrepreneurs.

Overall, the baseline result means that non-content delivery features in visual, vocal, and verbal channels all matter for financial investment decision-making in a persuasive communication. We do not want to underemphasize the seemingly obvious message. The alternative could be entirely possible—pitch features would not matter if investors were diverse in their favored styles, or if, at the more extreme, pitches are all about the content rather than the delivery.

Below we would like to further our discussion on this main specification.

A.1. Sample Selection of Videos. Pitch videos that can be accessed and collected from our internet search are a subsample of all such videos in accelerator applications. This is because some video files are made private, unlisted, or removed from the hosting websites after the application. The empirical regularity that governs the video selection process would naturally affect the validity and accuracy of the findings thus far. For instance, if pitch videos available to researchers are selected based on having more positive features and simultaneously are more likely to be kept available by invested startups, our main results could be driven by the selection mechanism.

We address this issue by directly tracking sample selection. Our goal is to explore the primary concern—whether the selection is related to the pitch features and to whether the startup gets invested in. All the videos used in this paper were available (searchable and viewable) in July 2019. We focus on all 527 videos uploaded within the 18-month window prior to July 2019, that is, the 2018 and 2019 cohorts, since attrition is more active immediately after the uploading and program selection. By the end of March 2020, 126 videos, or 23.9 percent, were selected out (unlisted, privatized, or removed) from the hosting platforms.

[Insert Table 4 Here.]

Table 4 shows the determinants of the selection. The selection, or equivalently the selection out, is unrelated to the Pitch Factor, the investment decision of the accelerator, or broadly any future VC financing. All coefficients are statistically weak and economically minimal across different specifications of the selection model. For instance, the 0.006 coefficient in column (1) means that a one standard deviation change of Pitch Factor shifts the selection probability by only 0.6 percent points, or 2.5 percent from the baseline, with a wide standard error. Thus, the sample selection can be considered as quasi-random for the purpose of our study.[18]

A.2. The Value of the Video-Based Method. Our video processing method has two distinct features from the prior literature that are particularly valuable in studying persuasion delivery. First, we use the complete pitch video when constructing the measure, differing from the literature exploiting thin slices of data (like the first few seconds) to capture perceptions. Second, we jointly use information from the three-V dimensions, raising the curiosity regarding the value of our method over those earlier works that examined some dimensions individually.

[Insert Table 5 Here.]

How valuable are those two features? The general design of the tests is to horse-race the Pitch Factor with measures constructed using thin-sliced video clips and with those using single-V dimensions. We report the analysis in Table 5. We first compare our measure with those using thin slices of data. First, we construct two new Pitch Factor measures, one using only a clip of when the first word is spoken (roughly one second), and the second using a random second of video clip in the pitch. In columns (1) and (3), we find that thin-sliced measures do provide useful information in explaining the investment choice. However, in columns (2) and (4), once we incorporate the full-video Pitch Factor to horse-race with the thin-sliced versions, the full-video Pitch Factor dwarfs the thin-sliced versions. This suggests that the marginal informational gain from using the full video is significant.

In columns (5) and (6), we investigate the value of using Pitch Factor over the information from individual V-dimensions. In our simplistic approach, we apply the factor analysis approach on each of the three-V dimensions and construct vocal-, visual-, and verbal- factors. We find that even though those individual dimensions matter in the decision-making, they are later dominated by the Pitch Factor which aggregates all various channels. In other words,

[18]One caveat is that the selection analysis conditions on a video once being publicly available at some time. This leaves the possibility that some videos were never made public. However, the selection trade-off at the initial uploading should be fairly comparable to the selection model estimated in Table 4.

jointly using the three-V channels indeed provides a more comprehensive reflection of the pitch delivery.

The factor analysis method used in Pitch Factor construction, though simple by design, accounts for not only the three-V channels, but importantly, the covariance matrix of information in those channels. A natural implication is that our finding is less prone to be confounded by the omitted variable bias than those using a single dimension alone. For example, if positive facial expressions are positively correlated with other features of delivery (e.g., passionate voices), it is hard to correctly estimate the effect of facial expressions when only analyzing the facial information alone without accounting for other channels.

B. Is Omitted Startup Quality Driving The Results?

The ability to deliver pitches is not randomly allocated—it may be affected by education and experience, among other factors. Is the documented impact of persuasion delivery just due to the omitted quality proxies for the startup that are not explicitly controlled for in the baseline analysis?

Building on the statistical discrimination literature, we can add control variables that are good proxies of startup and team quality and track the stability of coefficients associated with pitch features. Oster (2019) suggests a test for omitted variable bias that uses the information contained in the change in coefficient and the change in R^2 when moving from uncontrolled to controlled regression. The basic intuition is that if the coefficients are stable as we add (good but imperfect) quality controls, then the estimated effect is probably not due to an omitted quality variable and should be interpreted as arising through other independent channels. Formally, Altonji, Elder, and Taber (2005) and Oster (2019) show that if selection on the observed controls is proportional to the selection on the unobserved controls, then we can compute an identified set and test whether the identified set for the treatment effect includes zero.

We repeat the analysis in Table 3, adding control variables for founders' education (whether they have a master's degree or PhD degree, and whether they attended elite universities, as defined by U.S. News & World Report's Top 10), and founders' entrepreneurship experience, prior work experience and gender. These variables cover a large set of the information that the investor sees in addition to the pitch, and these variables have been shown to correlate with entrepreneurship quality, success, and the probability of obtaining venture financing (Bernstein et al., 2017; Ewens and Townsend, 2020).

The results are reported in the right panel of Table 3. All that we learned from the no-additional-control regression remains statistically robust. But perhaps more importantly, the coefficients remain very stable after introducing controls that are likely correlated with team quality. The formal test incorporates the change in R^2 induced by adding controls, and argues that the size of the R^2 change is informative in judging whether the stability of the estimated coefficient of interest is sufficient to argue away the omitted variable problem. Given that the Oster-test is designed for a single key explanatory variable, we focus on the exercises involving Pitch Factor, which we will call the uncontrolled (u) and controlled (c) regressions, respectively. We denote their estimates and R^2 as (β_u, R_u^2) and (β_c, R_c^2). Moreover, since the test is designed for linear models, we switch the estimation from Logit to OLS for this analysis.

To obtain an identified set of coefficients, the test strategy relies on assumptions on two more parameters—δ and R_{max}^2. δ (often referred to as the proportionality parameter) captures the level of selection on unobservables relative to selection on observable controls; a higher δ means that the omitted variable problem is more severe. R_{max}^2 is the hypothetical overall R^2 of the model with observables and unobservables. This measure indicates how much of the variation in the outcome variable can be explained by controlling for everything. The bias-adjusted coefficient, denoted as β_{adj} and determined by parameters δ and R_{max}^2, is closely approximated by the equation below (Section 3.2 in Oster (2019)):

$$\beta_{adj} \approx \beta_c - \delta \frac{(\beta_u - \beta_c)(R_{max}^2 - R_c^2)}{R_c^2 - R_u^2}. \tag{3}$$

With this adjusted coefficient β_{adj}, the recommended identified set is the interval between β_{adj} and β_c. We test whether the set safely excludes zero for reasonable parameterizations of δ and R_{max}^2.

[Insert Table 6 Here.]

In Table 6 we report the test results for different combinations of parameters. Table A.5

in the Appendix shows the raw OLS estimation results used in the Oster-test. Following the application of the test in Mian and Sufi (2014), our baseline test takes the values $R^2_{max} = \min(2.2R^2_c, 1)$ and $\delta = 1$. We show that the adjusted β is close to the estimated value and that we can easily reject the null that $\beta = 0$. In fact, the unobservable quality controls appear to be quite unimportant for our estimation. When pushing the δ to take a value of 2 and thus implementing the unrestricted estimator in Oster (2019), the identified set is still quite tight at $[0.019, 0.026]$. Only when we push the parameterization to almost unreasonably high values—$R^2_{max} = 1$ and $\delta = 2$—can we not reject the null.

This means, in a large set of scenarios, the effect of omitted quality controls is fairly minimal, and that the relation between the pitch features and investment decisions remains robust. In other words, pitch features do not seem to be correlated with funding decisions only because they are proxies for omitted startup quality. We want to acknowledge that it remains as an important assumption that unobservables are not more than twice as important as the vast set of observables ($\delta = 2$). Altonji et al. (2005) and Oster (2019) suggest this is appropriate, and the reasoning is that researchers often first focus on the most important set of controls (Angrist and Pischke, 2010). Even though $1 - R^2_c$ means sizable variations are unexplained by the model, this is a shared feature in related literature (Bernstein et al., 2017; Ewens and Townsend, 2020) pointing to the nature of venture investment.

C. Performance of Startups

The evidence so far suggests that the pitch delivery features exert independent, robust, and sizable impact on investors' decisions. Do they help investors reach better investment decisions? This interpretation could very well be true. For example, entrepreneurs may behave more positively and energetically if their startups are of higher quality. Another possibility is that communication and interpersonal skills reflected in pitch delivery may be productive for ventures. In addition, being positive and warm might be desirable and success-enhancing personality traits, given that the path of entrepreneurship is accompanied by challenges and difficulties. Under this line of thinking, the result can be interpreted as "better pitch is a valuable signal for better startups, so these startups receive more funding."

We test this stream of interpretation by making use of the long-term performance of startups. This test is motivated by Fisman, Paravisini, and Vig (2017) and Ewens and Townsend (2020). If the impact of persuasion features on investment decisions is indeed due to their correlation with startup potential, we would expect those features to be associated with better future performance. In contrast, if startups with preferred features underperform other companies conditional on obtaining funding, it serves as a sign that investors are subject to biases induced by those persuasion features. Regardless of the findings, the relation should bear limited causal interpretations. Instead, they are useful correlations that indicate if the decision based on the persuasion features is associated with better performance.

This exercise focuses on entrepreneurial ventures that were initially funded, allowing us to track the growth and performance of the startups. The analysis is performed using the following model:

$$Performance_i = \alpha + \gamma \cdot X_i + \delta \cdot Controls_i + \delta_{FE} + \varepsilon_i. \tag{4}$$

We measure performance in three ways. First, we examine the total employment of the company, which is a standard real-outcome performance measure used in entrepreneurship research (Puri and Zarutskie, 2012; Adelino, Ma, and Robinson, 2017). Second, we examine whether a startup has raised a follow-on round from a VC, typically a Series A round, and the total amount of capital raised from VCs. This measure serves as an interim measure of startup success (Ewens and Townsend, 2020). Finally, we examine whether a startup remains alive based on whether its website is still active. For total employment and amount of VC financing, we use an inverse hyperbolic sine transformation to transform the variables for better empirical properties, and the results remain almost identical to alternative transformation methods such as the log transformation.

Observations are at the startup level, and we focus on startups that received investment from an early-stage investor in or prior to 2017, so that those firms have at least two years to develop. All performance measures are as of July 2019. The key explanatory variable is the Pitch Factor. All regressions include startup and team controls used in Table 3. We also include controls for firm age and the squared term of firm age at the time of measurement in order to control for the growth stage. Accelerator fixed effects are included to account for variations in investor nurturing and value-adding.

[Insert Table 7 Here.]

Table 7 presents the results. We estimate a negative and significant coefficient across all the performance measures. This means that startups that show more positive pitch features grow more slowly in employment, are less likely to obtain more VC financing, and are less likely to survive. This is inconsistent with the explanation that features in pitches allow investors to form more accurate beliefs for investment decisions.

Overall, startups with a high Pitch Factor underperform in the long run. We do not interpret this as that the ability to deliver a passionate pitch is counter-productive. Instead, our preferred interpretation is that investors are too reluctant to invest in startups with a less positive pitch, and therefore only do so for the most promising companies, which in turn leads to better performance. In other words, a passionate pitch could lead investors to fund startups which may not merit the funding, suggesting a potential bias.

In Appendix D we provide a detailed conceptual framework to present those potential biases. In this framework, investors fund startups according to a simple cutoff rule, offering funding to all startups above a certain quality threshold. When investors are biased, either due to a taste-based channel or inaccurate beliefs about quality, it is possible that high-positivity startups may underperform. In the taste-based bias case, the investor continues to derive utility from startup performance. But she or he now also derives disutility from investing in startups with low positivity pitches—consequently, the investor sets a higher cutoff for them. As a result, expected performance, conditional on funding, will be higher for these low-positivity investments. In the inaccurate beliefs case, there is a gap between the investor's perceived performance distribution for low-positivity (or high-positivity) startups and the true performance distribution. Inaccurate beliefs can also lead investors to fund high-positivity founders with greater probability while having lower (true) expected performance for those investments.

We want to acknowledge that the test is not directly addressing a more nuanced version of startup performance—the right-tail "home-runs". Since our sample is concentrated in only the past five years, few startups in our sample have gone through an IPO or acquisition events, as startups usually stay private for longer during this period (Ewens and Farre-Mensa, 2020). As a result, we are not able to use the standard home-run measure, an indicator for IPO and acquisition event. In unreported explorations, we find a similar negative or noisy relation between Pitch Factor and in-sample right-tail measures like top-decile funding or employment.

D. Heterogeneous Effects Across Gender

The evidence thus far does not seem to support the explanation that fully rational agents learn from non-content delivery features in persuasion to improve investment decisions. This leaves room for other explanations such as those based on inaccurate beliefs, and/or investor taste and preference. Under these mechanisms, the relation between pitch features and investment choices would vary among subsamples in which those mechanisms may work differently.

One subsample is defined by gender. Women are often judged differently and treated differently in social occasions and economic settings such as hiring or promoting (Bordalo et al., 2019). For instance, Fredrickson and Roberts (1997) show that women are often more heavily judged on appearance and non-substantive features. Women and men are also expected to follow different gender stereotypes—there are generally higher expectations of men in general ability and task performance domains, while women are expected to be high in terms of warmth, empathy, and altruism (Kite, Deaux, and Haines, 2008; Fiske, 2010; Ellemers, 2018). Meanwhile, competent but less-warm women are biased against, particularly in leading roles (Rudman and Glick, 2001; Eagly and Karau, 2002), such as entrepreneurs. These gender biases govern a wide range of economic activities and outcomes (Goldin and Rouse, 2000; Bagues and Esteve-Volart, 2010; Brooks et al., 2014; Bohren et al., 2019). In recent literature, Sarsons et al. (2020) show that women receive less credit for group work. Cullen and Perez-Truglia (2019) show that employees' social interactions with their managers often favor men, contributing to the gender pay gap. Biasi and Sarsons (2021) show that women are less willing to engage in negotiations over pay, using a setting of public school teachers.

We first separately study startups with male-only or female-only teams (including one-person startups, which for convenience are also called teams). For each team, the metrics are naturally calculated for only people of the same gender and standardized as above but within gender. We apply the same empirical specification as in Eq. (1). By way of comparing the estimation results for the female and male subsamples, we are able to explore whether pitch delivery is more or less relevant for different gendered entrepreneurs.

[Insert Table 8 Here.]

Table 8 columns (1) and (2) present the result. We separately report the regression results for men and women entrepreneurs. Investment decisions on woman-only startups are significantly more sensitive to the performance in the pitch, with coefficients of 0.018 versus 0.170. Column (3) confirms that the magnitudes are statistically different when we perform the analysis using both types of teams and test the coefficients associated with men and women. This result is consistent with the literature on gender stereotyping, which shows that women and men are evaluated differently in social interactions and economic decisions. Not only are women judged more based on appearance and non-substantive features in pitches, but also the sensitivity is in the same direction as the gender stereotype. Investors reward women who fit their stereotypes—that is, those whom they see as warmer and more positive—and aggressively avoid investing in women entrepreneurs who do not fit this profile.

The result cannot be explained by several alternative explanations often discussed in the gender-finance literature. For example, the difference is not explained by different industry compositions of startups founded by male and female entrepreneurs. In our sample, male-only and female-only teams have similar industry distributions over GICS sectors (see Table A.6). In fact, in this analysis, we control for fixed effects at the industry level. Moreover, this is not due to different probabilities of obtaining funding or different distributional characteristics (e.g., mean, variance) of pitch features among women and men. Additionally, as will be discussed below, this also does not seem to be driven by a simple algorithm-bias explanation in which there are different levels of errors when measuring men and women.

What if a team has both male and female entrepreneurs? In Table 8 column (4), we focus on the subsample of startups that have both male and female team members pitching. For each team, we separately calculate the three-V dimensions and the Pitch Factor for women on the team and men on the team. We put those measures jointly in Eq. (1) so we can examine whether the features from women or men carry more weight for the probability that the team will receive funding.

At first glance, men drive the majority of the relations between pitch features and investment decisions in mixed-gender teams. One (very depressing) way to interpret this finding is that women are ignored in the pitches when they co-present with a man—thus, the features of their pitches matter less. Note that this is even though men and women actually speak for similar amounts of time in pitches on average. We acknowledge that the statistical significance of this result is weak, likely due to the small sample size.

Methodological-wise, the divergence of women-men comparisons in single-gender and mix-gender teams also provides some assurance of the algorithm measurement error problem introduced above. For example, one may worry that it might be more accurate for the algorithm to capture positive emotions from women (Hess et al., 2009; Sun et al., 2019). But in that case, the systematic measurement error would attenuate the results on men in both single-gender and mixed-gender teams, which is not the case. Conversely, if measurement errors are larger for female subjects, the systematic errors would attenuate the results on men in both analyses.

Overall, the evidence suggests that persuasion delivery affects male and female founders differently, in a direction consistent with gender stereotyping and inequality. One further test of this mechanism is to examine the role of reviewers' gender. This is unfortunately not feasible in our setting since investment decisions are often made by groups rather than one individual. We also do not have reliable information on the gender of the lead investor for each startup. It is possible that gender bias is more severe in cross-gender evaluations, while some research shows that gender-related social psychology forces are often salient among both male and female reviewers (Hentschel, Heilman, and Peus, 2019). Similarly, the sample's small number of startups founded by minority racial groups limits any analysis of race.

4. Experiment: (Inaccurate) Beliefs vs. Taste?

In this last section, we explore the economic mechanisms through which pitch delivery affects persuasion effectiveness. We follow DellaVigna and Gentzkow (2010) and explore the mechanisms in two broad categories—taste-based models and miscalibrated/inaccurate beliefs. We want to understand: which of these two mechanisms better explains how the non-content delivery features affect investor decision; and if both exist, how much do they contribute to the investment bias?

A. A Simple Conceptual Framework

We model venture investors' investment problems by incorporating the role of persuasion features through two possible channels: a *beliefs channel* and a *taste-based channel*. The beliefs channel works through investors' expectations: if true, investors would use startup pitch features to form their beliefs about startups' chances to succeed, accurately or inaccurately. The taste-based channel operates through a standalone component favoring certain pitch features in investors' preferences: if true, even with the same perceived quality of the startup, investors would still be more likely to invest in certain features.

Formally, an investor j makes the investment decision on startup i based on pitch delivery features θ_i; the investor's beliefs about the success probability of the venture, μ_{ij}; and the precision or confidence level of the belief σ_{ij}. The investment is based on a simple threshold investment rule:

$$I_{ij} = \mathbb{1}_{\{U_{ij} \geq \bar{U}\}}, \quad \text{where } U(\mu_{ij}, \sigma_{ij}, \theta_i) \equiv \gamma_\mu \mu_{ij} + \gamma_\sigma \sigma_{ij} + \kappa \theta_i. \tag{5}$$

In a wide class of decision models, $\gamma_\mu > 0$—investors are more willing to make an investment in startups that they believe to have a higher success probability. σ captures the second moment of the belief about the success probability—the larger is σ, the lower is precision and confidence. We should expect $\gamma_\sigma < 0$ for a risk-averse agent.

The beliefs channel is modeled by allowing μ and σ to be determined by hard information about the venture, Q_i, and the features in the pitch delivery θ_i:

$$\mu_{ij} = \lambda_\mu Q_i + \psi_\mu \theta_i, \tag{6a}$$

$$\sigma_{ij} = \lambda_\sigma Q_i + \psi_\sigma \theta_i. \tag{6b}$$

Under this framework, θ_i enters the investment decision in two ways—through the impact on beliefs μ (the size of the impact is $\psi_\mu \gamma_\mu \theta_i$) and σ (the size of the impact is $\psi_\sigma \gamma_\sigma \theta_i$); and/or through the direct utility gain through the term $\kappa\theta_i$. This means the β coefficient in Eq. (1), under our framework, is the combined effect of $\kappa + \psi_\mu \gamma_\mu + \psi_\sigma \gamma_\sigma$.

The experiment can help determine whether those channels exist and the relative importance of the channels in driving the main effect. We can expect three potential scenarios:

Scenario	$\psi_{\mu,\sigma}$	κ	Beliefs Channel	Taste Channel	Decompose β in Eq. (1)
1	$\neq 0$	$= 0$	✓	✕	$\beta = \psi_\mu \gamma_\mu + \psi_\sigma \gamma_\sigma$
2	$= 0$	$\neq 0$	✕	✓	$\beta = \kappa$
3	$\neq 0$	$\neq 0$	✓	✓	$\beta = \kappa + \psi_\mu \gamma_\mu + \psi_\sigma \gamma_\sigma$

In these different scenarios, the ψ captures whether the beliefs channel exists and its strength. The existence and strength of the preference channel hinge on whether $\kappa > 0$ when we explicitly control for the beliefs of the investor μ and σ.

B. Experiment Design

Our experiment constructs a setting to allow participants to act as venture investors. The experiment randomly allocates 10 pitch videos to each subject to review, with random ordering. The video pool consists of 62 videos that are highly standardized; they are from the same incubator program and have comparable lengths and resolutions. After viewing each video i, the subject is asked to answer questions around three main themes: (1) whether she/he would invest in company i, denoted as I_{ij}; (2) her/his expectation of the company's success probability, μ_{ij}, measured between 0 and 100%; and (3) her/his confidence level on her/his decision and expectation, σ_{ij}, measured on a scale of 1 to 5. When eliciting beliefs, the experiment asks both conditional and unconditional expectations, and the experiment also uses different definitions of success (staying alive, becoming a unicorn, etc.). We obtain the pitch features, θ_{ij}, using the same method as in the earlier part of the paper.

The subject pool consists of Master's students from the Yale School of Management (Yale SOM). All subjects have basic training in core business skills and a basic understanding of entrepreneurial finance. They all completed the Entrepreneurial Finance (MGT 897) class at Yale SOM, with the same instructor, in which they were exposed to startup evaluation in qualitative and quantitative dimensions and VC investments, among other topics. For each subject, we collect basic characteristics (including age, gender, academic background, work experience, ethnicity, etc.). In addition, we elicit her/his unconditional expectations of startup success probability and confidence level before the experiment. The experiment was a bonus assignment in the Yale SOM Entrepreneurial Finance (MGT 897) class. The response rate is 63.75 percent, and 102 subjects (in a class of 160) participated in and successfully finished the experiment, which on average takes 30 minutes.[19]

The subject pool is incentivized to participate in the experiment first with a flat bonus grade for the course on a scale of 50, equivalent to 20 percent of the total participation grade. Participants also receive a performance-based pay that is calculated based on the performance of startups they choose: 10 points for each startup that scores in the top 10 percent of performance in either funding amount or total employment among its cohort, an additional 5 points for each startup that stays alive, and -5 points for each startup that fails. The subjects are also incentivized to make an accurate response to the beliefs question—additional investment performance points are added to account for the distance between the realized outcome and the expectation. An example experiment and summary statistics of the subjects and their responses are provided in Appendix F.

C. Results

C.1. Interaction Features and (Inaccurate) Beliefs. We first test the beliefs channel by estimating ψ_μ and ψ_σ from Eq. (6a) and (6b). We estimate the model using OLS

[19] Out of the 1,020 experimental investment rounds (10 videos $\times$ 102 participants), 68 were incomplete due to an integration glitch between the survey software and video platforms, leaving 952 experimental rounds. Our results are robust to an alternative approach of dropping all subjects with incomplete experimental rounds.

with fixed effects at the subject j level. For the belief measures, we use μ and σ for three different probabilities—$alive|invested$ ("|" means conditional on), $success|invested$, and $alive|notinvested$.

[Insert Table 9 Here.]

Table 9 shows the relation between beliefs and Pitch Factor θ. A more positive Pitch Factor is associated with a higher expectation of the success probability of the startup venture, yet only a mild and statistically weak decrease in the variance—in other words, a weak increase of investor confidence. Regarding the economic magnitude, a one-standard-deviation increase of the Pitch Factor is associated with an increase of 2 percentage points in $P(alive|invested)$, which is a 6.5 percent increase from the baseline expectation of 31 percent. These results confirm the existence of a channel through beliefs.

We next quantify whether the relation between Pitch Factor and beliefs is due to accurate or inaccurate belief updating. Note that Table 7 finds that realized performance often negatively correlates with the Pitch Factor. This means that the positive coefficients shown thus far in Table 9 are a sign of miscalibrated beliefs. The level of inaccuracy relies on the gap between the coefficients on μs in Table 9 and the realized outcome. We present the relation between realized outcome and the Pitch Factor in column (5). Our key measure of performance is on survival since "success" is difficult to code and maps to the questions asked in the experiment. Consistent with Table 7, actual survival probability negatively correlates with the Pitch Factor conditional on investment, in contrast to subjects' beliefs that Pitch Factor would positively predict survival. The miscalibration of beliefs has a magnitude of 0.137 (= 0.020 − (−0.117)).

C.2. Decomposing Inaccurate Beliefs and Preferences. We next estimate the full model, using a logit framework based on Eq. (5):

$$I_{ij} = \underbrace{\kappa \cdot \theta_i}_{\text{Taste}} + \underbrace{\gamma_\mu \cdot \mu_{ij} + \gamma_\sigma \cdot \sigma_{ij}}_{\text{Beliefs}} + \delta_j + \varepsilon_{ij}. \tag{7}$$

Table 10 shows the results. We first confirm in column (1) that, in our experimental sample, the Pitch Factor is positively associated with the probability that the startup company is chosen to be invested. Comparing this experimental estimate with the real-world estimate, the experimental estimate is larger, but this is partially because the video sample used in the experiment (described above) has more higher-quality videos for standardization purposes. After taking this into consideration, economic magnitudes are comparable. Not surprisingly, in columns (2) and (3), when the subject thinks that the company has a higher μ or lower σ, the subject is more likely to make an investment.

[Insert Table 10 Here.]

Column (4) is the key test of model (7). The Pitch Factor strongly correlates with the investment decision after controlling beliefs, and this suggests that there exists a taste/preference channel through which the pitch features affect investment decisions. In other words, the model supports scenario 3 in Section 4.A above.

With both channels present, this estimation provides a way to decompose the relative contributions of the two channels to the overall effect of the Pitch Factor. To map the estimated parameters to the framework, $\kappa = 0.067$ and $\gamma_\mu = 2.208$, assuming away the impact of Pitch Factor on σ as supported in Table 9. This means that the taste channel leads to an increase (bias) in investment probability by 0.067. The inaccurate beliefs channel leads to an increase in investment by 0.302 ($= 0.137 \times 2.208$). So, the beliefs channel and the taste channel contribute to the bias by 81.8 percent ($= \frac{0.067}{0.067+0.302}$) and 18.2 percent ($= \frac{0.302}{0.067+0.302}$), respectively.[20] This quantitative decomposition may vary from setting to setting, and further explorations of this exercise may be a fruitful path for researchers and practitioners interested in different contexts. However, it is likely that in other financial investment settings, especially those involving professional investors, the inaccurate beliefs channel is more important than the taste-based channel.

[20] One caveat is that the 81.8 percent attributed to the beliefs channel could still be an underestimate if solicited beliefs from the experiment are measured with error, attenuating the importance of the belief-based channel. Alternatively, this could be overestimated since the preference-based channel is less salient when the subjects would not interact with the startup founders while the VCs would.

5. Conclusion

It is widely speculated that the delivery of a persuasion matters for the final outcome—sales agents achieve different results selling the same product using the same standard pitch; researchers of the same team convince peers to a different level when presenting the same paper using the same slides. Yet there is little evidence on how much and why the delivery features matter, especially in a real-world investment setting.

We shed light on this issue using a novel video-based method applied to a classic setting of persuading investors. We find that non-content delivery features in persuasive interactions have statistically significant and economically sizable effects on investors' decisions. These features do not seem to help investors to make better investment decisions. Instead, our evidence using both archival data and an experiment suggests a bias induced by those features, particularly through leading investors to form inaccurate beliefs.

The results leave many questions unanswered and suggest directions for future research. Conceptually, it will be a fruitful path to explore further the root of the inaccurate beliefs by connecting more closely to behavioral models on persuasion. Among many models, two prominent candidates in persuading investors include categorical and coarse thinking (Fryer and Jackson, 2008; Mullainathan et al., 2008), and failure to account for repeated information (DeMarzo et al., 2003). We also believe the literature on emotions and affects in behavioral economics could be useful in considering the impact of different pitch styles.

Empirically, our video-based approach is extendable to accommodate more complex settings and measures. The extensions could be along several dimensions. Researchers can track multiple players who sequentially send and receive signals via the three-V dimensions. Moreover, the method can be extended to capture more behaviors—such as gestures, speech fluency, etc. We are hopeful that this paper lays the groundwork for such future research.

List of Metrics from ML-Based Video Processing and Collecting

Variable	Definition and Construction
A. Visual Metrics	
Visual-Positive	Probability that the facial emotion is happiness by Face++ emotion recognition API
Visual-Negative	Sum of the probabilities that the facial emotion is sadness, anger, fear, and disgust by Face++ emotion recognition API
Visual-Beauty	Beauty scores for the faces in videos by Face++ beauty score API
B. Vocal Metrics	
Vocal-Positive	Probability that the vocal emotion is happiness by the LSTM model in `speechemotionrecognition`
Vocal-Negative	Probability that the vocal emotion is sadness by the LSTM model in `speechemotionrecognition`
Vocal-Arousal	Degree of vocal arousal by the SVM model in `pyAudioAnalysis`
Vocal-Valence	Degree of vocal valence by the SVM model in `pyAudioAnalysis`
C. Verbal Metrics	
Verbal-Positive	Whether a word is included in the positive category of the LM Master Dictionary (Loughran and McDonald, 2011)
Verbal-Negative	Whether a word is included in the negative category of the LM Master Dictionary (Loughran and McDonald, 2011)
Verbal-Ability	The direction (-1 or $+1$) of a word if it is included in the ability category of the NBF dictionary (Nicolas et al., 2019)
Verbal-Warmth	The direction (-1 or $+1$) of a word if it is included in the warmth category of the NBF dictionary (Nicolas et al., 2019)
D. Startup-level Variables (as of July 2019)	
I(*Invested*)	Whether the startup team receives funding from the accelerator
Employment	The inverse hyperbolic sine of the number of employees recorded in Crunchbase
Raised VC	Whether the startup team raised another round of VC investment after being funded by the accelerator
VC Amount	The inverse hyperbolic sine of the total amount of raised funding recorded in Crunchbase
Startup Alive	Whether the startup's website is still operational

References

Abrams, Eliot, 2019, Heterogeneous effects of bank and mortgage company advertising: New evidence from television commercial content .

Adelino, Manuel, Song Ma, and David Robinson, 2017, Firm age, investment opportunities, and job creation, *Journal of Finance* 72, 999–1038.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber, 2005, An evaluation of instrumental variable strategies for estimating the effects of catholic schooling, *Journal of Human resources* 40, 791–821.

Ambady, Nalini, and Robert Rosenthal, 1992, Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis., *Psychological Bulletin* 111, 256.

Ambady, Nalini, and Robert Rosenthal, 1993, Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness., *Journal of Personality and Social Psychology* 64, 431.

Angrist, Joshua D, and Jörn-Steffen Pischke, 2010, The credibility revolution in empirical economics: How better research design is taking the con out of econometrics, *Journal of Economic Perspectives* 24, 3–30.

Antonakis, John, Marika Fenley, and Sue Liechti, 2011, Can charisma be taught? tests of two interventions, *Academy of Management Learning & Education* 10, 374–396.

Antweiler, Werner, and Murray Z Frank, 2004, Is all that talk just noise? the information content of internet stock message boards, *Journal of Finance* 59, 1259–1294.

Arkes, Hal R, Lisa Tandy Herren, and Alice M Isen, 1988, The role of potential loss in the influence of affect on risk-taking behavior, *Organizational behavior and human decision processes* 42, 181–193.

Åstebro, Thomas, Holger Herz, Ramana Nanda, and Roberto A. Weber, 2014, Seeking the roots of entrepreneurship: Insights from behavioral economics, *Journal of Economic Perspectives* 28, 49–70.

Awamleh, Raed, and William L Gardner, 1999, Perceptions of leader charisma and effectiveness: The effects of vision content, delivery, and organizational performance, *The Leadership Quarterly* 10, 345–373.

Bagues, Manuel F, and Berta Esteve-Volart, 2010, Can gender parity break the glass ceiling? evidence from a repeated randomized experiment, *Review of Economic Studies* 77, 1301–1328.

Barsade, Sigal G, 2002, The ripple effect: Emotional contagion and its influence on group behavior, *Administrative science quarterly* 47, 644–675.

Benjamin, Daniel J, and Jesse M Shapiro, 2009, Thin-slice forecasts of gubernatorial elections, *Review of Economics and Statistics* 91, 523–536.

Berggren, Niclas, Henrik Jordahl, and Panu Poutvaara, 2010, The looks of a winner: Beauty and electoral success, *Journal of Public Economics* 94, 8–15.

Bernstein, Shai, Arthur Korteweg, and Kevin Laws, 2017, Attracting early-stage investors: Evidence from a randomized field experiment, *Journal of Finance* 72, 509–538.

Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman, 2010, What's advertising content worth? evidence from a consumer credit marketing field experiment, *Quarterly Journal of Economics* 125, 263–306.

Bestelmeyer, Patricia EG, Sonja A Kotz, and Pascal Belin, 2017, Effects of emotional valence and arousal on the voice perception network, *Social cognitive and affective neuroscience* 12, 1351–1358.

Biasi, Barbara, and Heather Sarsons, 2021, Flexible wages, bargaining, and the gender gap, *Quarterly Journal of Economics* Forthcoming.

Blankespoor, Elizabeth, Bradley E Hendricks, and Gregory S Miller, 2017, Perceptions and price: Evidence from ceo presentations at ipo roadshows, *Journal of Accounting Research* 55, 275–327.

Bohren, J Aislinn, Kareem Haggag, Alex Imas, and Devin G Pope, 2019, Inaccurate statistical discrimination, *NBER Working Paper* .

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, 2019, Beliefs about gender, *American Economic Review* 109, 739–73.

Boxell, Levi, 2018, Slanted images: Measuring nonverbal media bias .

Brooks, Alison Wood, Laura Huang, Sarah Wood Kearney, and Fiona E Murray, 2014, Investors prefer entrepreneurial ventures pitched by attractive men, *Proceedings of the National Academy of Sciences* 111, 4427–4431.

Burkhardt, Felix, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss, 2005, A database of german emotional speech, in *Ninth European Conference on Speech Communication and Technology*.

Clore, Gerald L, Norbert Schwarz, and Michael Conway, 1994, Affective causes and consequences of social information processing., *Handbook of social cognition* 1, 323–417.

Cullen, Zoë B, and Ricardo Perez-Truglia, 2019, The old boys' club: Schmoozing and the gender gap, *NBER Working Paper* .

Dahl, Gordon, and Stefano DellaVigna, 2009, Does movie violence increase violent crime?, *Quarterly Journal of Economics* 124, 677–734.

Dana, Jason, Robyn Dawes, and Nathanial Peterson, 2013, Belief in the unstructured interview: The persistence of an illusion, *Judgment and Decision making* 8, 512.

DellaVigna, Stefano, 2009, Psychology and economics: Evidence from the field, *Journal of Economic literature* 47, 315–72.

DellaVigna, Stefano, and Matthew Gentzkow, 2010, Persuasion: Empirical evidence, *Annual Review of Economics* 2, 643–669.

DellaVigna, Stefano, and Ethan Kaplan, 2007, The fox news effect: Media bias and voting, *Quarterly Journal of Economics* 122, 1187–1234.

DellaVigna, Stefano, and Devin Pope, 2018, What motivates effort? evidence and expert forecasts, *Review of Economic Studies* 85, 1029–1069.

DeMarzo, Peter M, Dimitri Vayanos, and Jeffrey Zwiebel, 2003, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics* 118, 909–968.

Dewatripont, Mathias, and Jean Tirole, 1999, Advocates, *Journal of Political Economy* 107, 1–39.

Eagly, Alice H, and Steven J Karau, 2002, Role congruity theory of prejudice toward female leaders., *Psychological review* 109, 573.

Eckel, Catherine C, and Ragan Petrie, 2011, Face value, *American Economic Review* 101, 1497–1513.

Ellemers, Naomi, 2018, Gender stereotypes, *Annual review of psychology* 69, 275–298.

Engelberg, Joseph E, and Christopher A Parsons, 2011, The causal impact of media in financial markets, *Journal of Finance* 66, 67–97.

Ewens, Michael, and Joan Farre-Mensa, 2020, The deregulation of the private equity markets and the decline in ipos, *Review of Financial Studies* Forthcoming.

Ewens, Michael, and Richard R Townsend, 2020, Are early stage investors biased against women?, *Journal of Financial Economics* .

Fiske, Susan T, 2010, Venus and mars or down to earth: Stereotypes and realities of gender differences, *Perspectives on Psychological Science* 5, 688–692.

Fiske, Susan T, Amy JC Cuddy, and Peter Glick, 2007, Universal dimensions of social cognition: Warmth and competence, *Trends in cognitive sciences* 11, 77–83.

Fisman, Raymond, Daniel Paravisini, and Vikrant Vig, 2017, Cultural proximity and loan outcomes, *American Economic Review* 107, 457–92.

Fredrickson, Barbara L, and Tomi-Ann Roberts, 1997, Objectification theory: Toward understanding women's lived experiences and mental health risks, *Psychology of women quarterly* 21, 173–206.

Fryer, Roland, and Matthew O. Jackson, 2008, A categorical model of cognition and biased decision making, *The B.E. Journal of Theoretical Economics* 8.

Gandomi, Amir, and Murtaza Haider, 2015, Beyond the hype: Big data concepts, methods, and analytics, *International journal of information management* 35, 137–144.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy, 2019, Text as data, *Journal of Economic Literature* 57, 535–74.

Giannakopoulos, Theodoros, 2015, pyaudioanalysis: An open-source python library for audio signal analysis, *PloS one* 10, e0144610.

Goldin, Claudia, and Cecilia Rouse, 2000, Orchestrating impartiality: The impact of" blind" auditions on female musicians, *American Economic Review* 90, 715–741.

Gompers, Paul A, Will Gornall, Steven N Kaplan, and Ilya A Strebulaev, 2020, How do venture capitalists make decisions?, *Journal of Financial Economics* 135, 169–190.

Gompers, Paul A, and Sophie Q Wang, 2017, And the children shall lead: Gender diversity and performance in venture capital, *NBER Working Paper w23459* .

Gornall, Will, and Ilya A Strebulaev, 2019, Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels .

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera, 2021, The voice of monetary policy, *NBER Working Paper* .

Graham, John R, Campbell R Harvey, and Manju Puri, 2016, A corporate beauty contest, *Management Science* 63, 3044–3056.

Haltiwanger, John, Ron S Jarmin, and Javier Miranda, 2013, Who creates jobs? small versus large versus young, *Review of Economics and Statistics* 95, 347–361.

Hamermesh, Daniel S, and Jeff E Biddle, 1994, Beauty and the labor market, *American Economic Review* 1174–1194.

Hatfield, Elaine, John T Cacioppo, and Richard L Rapson, 1993, Emotional contagion, *Current directions in psychological science* 2, 96–100.

Hebert, Camille, 2020, Gender stereotypes and entrepreneur financing .

Hentschel, Tanja, Madeline E Heilman, and Claudia V Peus, 2019, The multiple dimensions of gender stereotypes: a current look at men's and women's characterizations of others and themselves, *Frontiers in psychology* 10, 11.

Hess, Ursula, Reginald B Adams, Karl Grammer, and Robert E Kleck, 2009, Face gender and emotion expression: Are angry women more like men?, *Journal of Vision* 9, 19–19.

Heyes, Anthony, and John A List, 2016, Supply and demand for discrimination: Strategic revelation of own characteristics in a trust game, *American Economic Review* 106, 319–23.

Hirshleifer, David, and Tyler Shumway, 2003, Good day sunshine: Stock returns and the weather, *Journal of Finance* 58, 1009–1032.

Hoberg, Gerard, and Gordon Phillips, 2016, Text-based network industries and endogenous product differentiation, *Journal of Political Economy* 124, 1423–1465.

Hoberg, Gerard, and Gordon M Phillips, 2010, Product market synergies and competition in mergers and acquisitions: A text-based analysis, *Review of Financial Studies* 23, 3773–3811.

Hobson, Jessen L, William J Mayew, and Mohan Venkatachalam, 2012, Analyzing speech to detect financial misreporting, *Journal of Accounting Research* 50, 349–392.

Hochberg, Yael V, 2016, Accelerating entrepreneurs and ecosystems: The seed accelerator model, *Innovation Policy and the Economy* 16, 25–51.

Howell, Sabrina T, 2019, Reducing information frictions in venture capital: The role of new venture competitions, *Journal of Financial Economics* .

Huang, Xing, Zoran Ivković, J Jiang, and I Wang, 2018, Swimming with the sharks: Entrepreneurial investing decisions and first impression .

Johnson, Eric J, and Amos Tversky, 1983, Affect, generalization, and the perception of risk., *Journal of personality and social psychology* 45, 20.

Joo, Jungseock, and Zachary C Steinert-Threlkeld, 2018, Image as data: Automated visual content analysis for political science, *arXiv preprint arXiv:1810.01544* .

Kamenica, Emir, and Matthew Gentzkow, 2011, Bayesian persuasion, *American Economic Review* 101, 2590–2615.

Kite, Mary E, Kay Deaux, and Elizabeth L Haines, 2008, Gender stereotypes, *Psychology of women: A handbook of issues and theories* 2, 205–236.

Knapp, Mark L, Judith A Hall, and Terrence G Horgan, 2013, *Nonverbal communication in human interaction* (Cengage Learning).

Kortum, Samuel, and Josh Lerner, 2000, Assessing the contribution of venture capital to innovation, *The RAND Journal of Economics* 31, 674.

Krauss, Robert M, William Apple, Nancy Morency, Charlotte Wenzel, and Ward Winton, 1981, Verbal, vocal, and visible factors in judgments of another's affect., *Journal of Personality and Social Psychology* 40, 312.

Kuhnen, Camelia M, and Brian Knutson, 2011, The influence of affect on beliefs, preferences, and financial decisions, *Journal of Financial and Quantitative Analysis* 46, 605–626.

Lerner, Josh, Antoinette Schoar, Stanislav Sokolinski, and Karen Wilson, 2018, The globalization of angel investments: Evidence across countries, *Journal of Financial Economics* 127, 1–20.

Lian, Chen, Yueran Ma, and Carmen Wang, 2019, Low interest rates and risk-taking: Evidence from individual investment decisions, *Review of Financial Studies* 32, 2107–2148.

List, John A, 2004, The nature and extent of discrimination in the marketplace: Evidence from the field, *Quarterly Journal of Economics* 119, 49–89.

Loewenstein, George F, Elke U Weber, Christopher K Hsee, and Ned Welch, 2001, Risk as feelings., *Psychological bulletin* 127, 267.

Loughran, Tim, and Bill McDonald, 2011, When is a liability not a liability? textual analysis, dictionaries, and 10-ks, *Journal of Finance* 66, 35–65.

Loughran, Tim, and Bill McDonald, 2016, Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* 54, 1187–1230.

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers, 2011, Big data: The next frontier for innovation, competition, *Washington, DC: McKinsey Global Institute* .

Mayew, William J, and Mohan Venkatachalam, 2012, The power of voice: Managerial affective states and future firm performance, *Journal of Finance* 67, 1–43.

McCloskey, Donald, and Arjo Klamer, 1995, One quarter of gdp is persuasion, *American Economic Review* 85, 191–195.

McNamara, Quinten, Alejandro De La Vega, and Tal Yarkoni, 2017, Developing a comprehensive framework for multimodal feature extraction, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1567–1574.

Mehrabian, Albert, 1972, *Nonverbal Communication* (Transaction Publishers).

Mian, Atif, and Amir Sufi, 2014, What explains the 2007–2009 drop in employment?, *Econometrica* 82, 2197–2223.

Milgrom, Paul, and John Roberts, 1986, Relying on the information of interested parties, *The RAND Journal of Economics* 18–32.

Mobius, Markus M, and Tanya S Rosenblat, 2006, Why beauty matters, *American Economic Review* 96, 222–235.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer, 2008, Coarse thinking and persuasion, *Quarterly Journal of Economics* 123, 577–619.

Nicolas, Gandalf, Xuechunzi Bai, and Susan T Fiske, 2019, Automated dictionary creation for analyzing text: An illustration from stereotype content .

Oster, Emily, 2019, Unobservable selection and coefficient stability: Theory and evidence, *Journal of Business & Economic Statistics* 37, 187–204.

Peng, Lin, Siew Hong Teoh, Yakun Wang, and Jiawen Yan, 2020, Face value: Trait inference, performance characteristics, and market outcomes for financial analysts, *Working Paper* .

Peng, Yilang, 2018, Same candidates, different faces: Uncovering media bias in visual portrayals of presidential candidates with computer vision, *Journal of Communication* 68, 920–941.

Puri, Manju, and Rebecca Zarutskie, 2012, On the lifecycle dynamics of venture-capital- and non-venture-capital-financed firms, *Journal of Finance* 67, 2247–2293.

Rosenberg, Shawn W, Lisa Bohan, Patrick McCafferty, and Kevin Harris, 1986, The image and the vote: The effect of candidate presentation on voter preference, *American Journal of Political Science* 108–127.

Rudman, Laurie A, and Peter Glick, 2001, Prescriptive gender stereotypes and backlash toward agentic women, *Journal of social issues* 57, 743–762.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram, 2020, Gender differences in recognition for group work, *Journal of Political Economy* 000–000.

Schubert, James N, Carmen Strungaru, Margaret Curren, and Wulf Schiefenhovel, 1998, Physische erscheinung und die einschätzung von politischen kandidatinnen und kandidaten, *Biopolitics: Politikwissenschaft jensets des Kulturismus. Nomos Verlagsgesellschaft, Baden-Baden* .

Schwartzstein, Joshua, and Adi Sunderam, 2020, Using models to persuade, *American Economic Review* .

Stigler, George J, 1961, The economics of information, *Journal of Political Economy* 69, 213–225.

Strahan, Carole, and Donald G Zytowski, 1976, Impact of visual, vocal, and lexical cues on judgments of counselor qualities., *Journal of Counseling Psychology* 23, 387.

Sun, Tony, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang, 2019, Mitigating gender bias in natural language processing: Literature review, *arXiv preprint arXiv:1906.08976* .

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

Tetlock, Paul C, Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More than words: Quantifying language to measure firms' fundamentals, *Journal of Finance* 63, 1437–1467.

Todorov, Alexander, Anesu N Mandisodza, Amir Goren, and Crystal C Hall, 2005, Inferences of competence from faces predict election outcomes, *Science* 308, 1623–1626.

Tskhay, Konstantin O, Rebecca Zhu, Christopher Zou, and Nicholas O Rule, 2018, Charisma in everyday life: Conceptualization and validation of the general charisma inventory., *Journal of Personality and Social Psychology* 114, 131.

Figure 1. Examples of Positive and Negative Visual Features

(a) Example of High-Positivity Visual Features

(b) Example of Low-Positivity Visual Features

Notes. This figure presents examples of a frame showing positive facial expressions (Panel (a)) and less-positive facial expressions (Panel (b)).

Figure 2. Examples of High-Ability and Warmth Script

(a) Example of High-Ability Pitch Script

Hi, I'm Vitali CEO of Fitness Lab. There are a lot of fitness apps in the world, but they all have the same problems. They offer their users random workout and diet plans. They use just do it marketing to push sales and they have low retention because people leave after three months. So we decided to fix it and make the best fitness app in the world and power it with artificial intelligence. We're going to offer our users highly personalized workout and diet plans by now. We have launched MVP. We have 50,000 installations and five thousand active users. We have received several design award including the one for the best interaction design by American Institute of graphic arts. So we have awesome design and very efficient technology. This is going to be more efficient than any living coach. We know that Y Combinator has a lot of connections with artificial intelligence businesses. That's why we're looking forward to your support. Thank you.

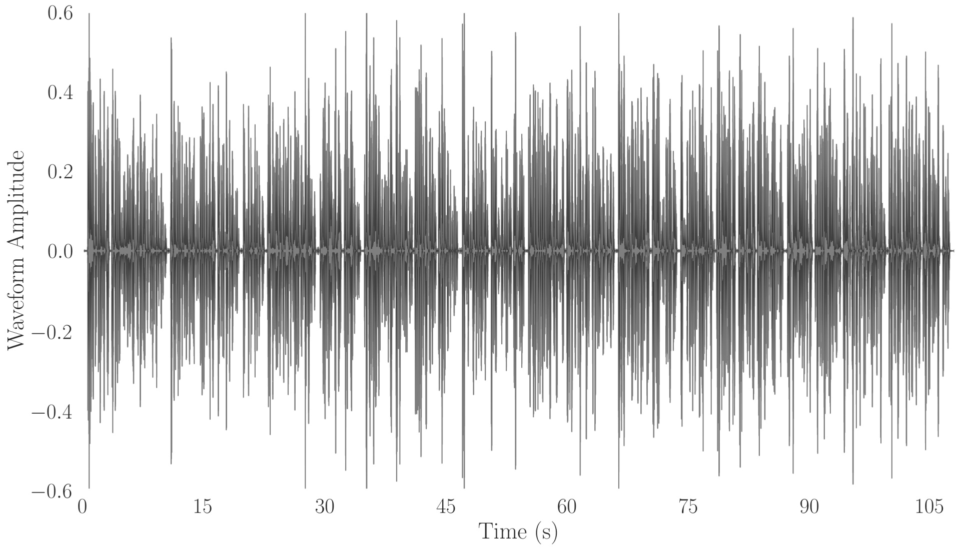
(b) Example of High-Warmth Pitch Script

Hello, I'm Marcus and I'm Rebecca and together we're the proud founders of Fine Print Fighters LLC. We help expose small and misleading content in contracts. We help consumers make much more informed decisions during the purchasing process both pre and post purchase. We like to help the consumers gain back control of the purchasing process, and we like to create value well through our pleasing personalities as you can tell. Well, we look forward to working with angel pad, and we appreciate the opportunity in advance. Look forward to working with the staff and the rest of the constituents and hopefully be a good representation of what angel angel pad represents. So we thank you again in advance, and we look forward to speaking with you all and seeing you all soon. Thank you.

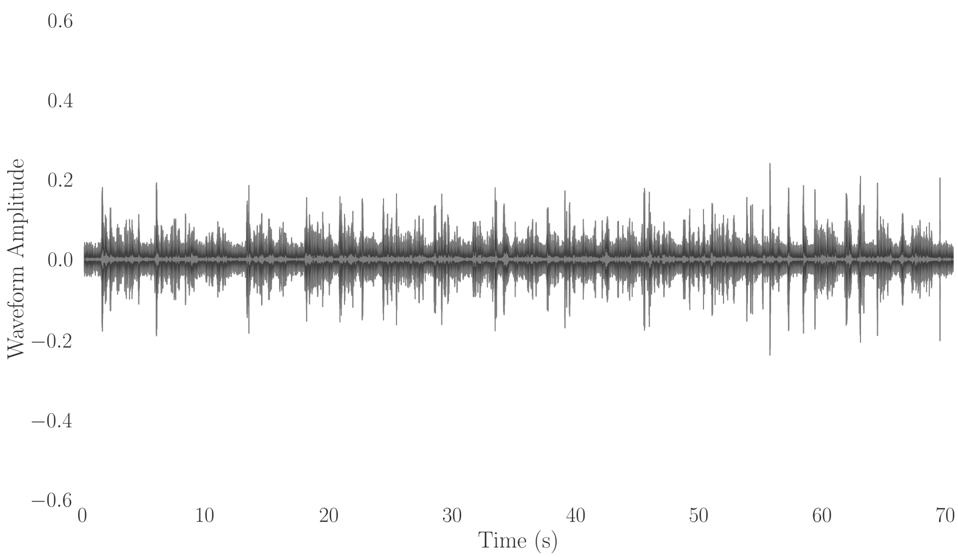
Notes. This figure presents examples of startup pitch scripts with high-ability (Panel (a)) and high-warmth (Panel (b)) verbal features. The key ability words are highlighted in Panel (a), and key warmth words are highlighted in Panel (b).

Figure 3. Visualized Examples of High- and Low- Arousal Vocal Features

(a) Visualized Example of High-Arousal Pitch

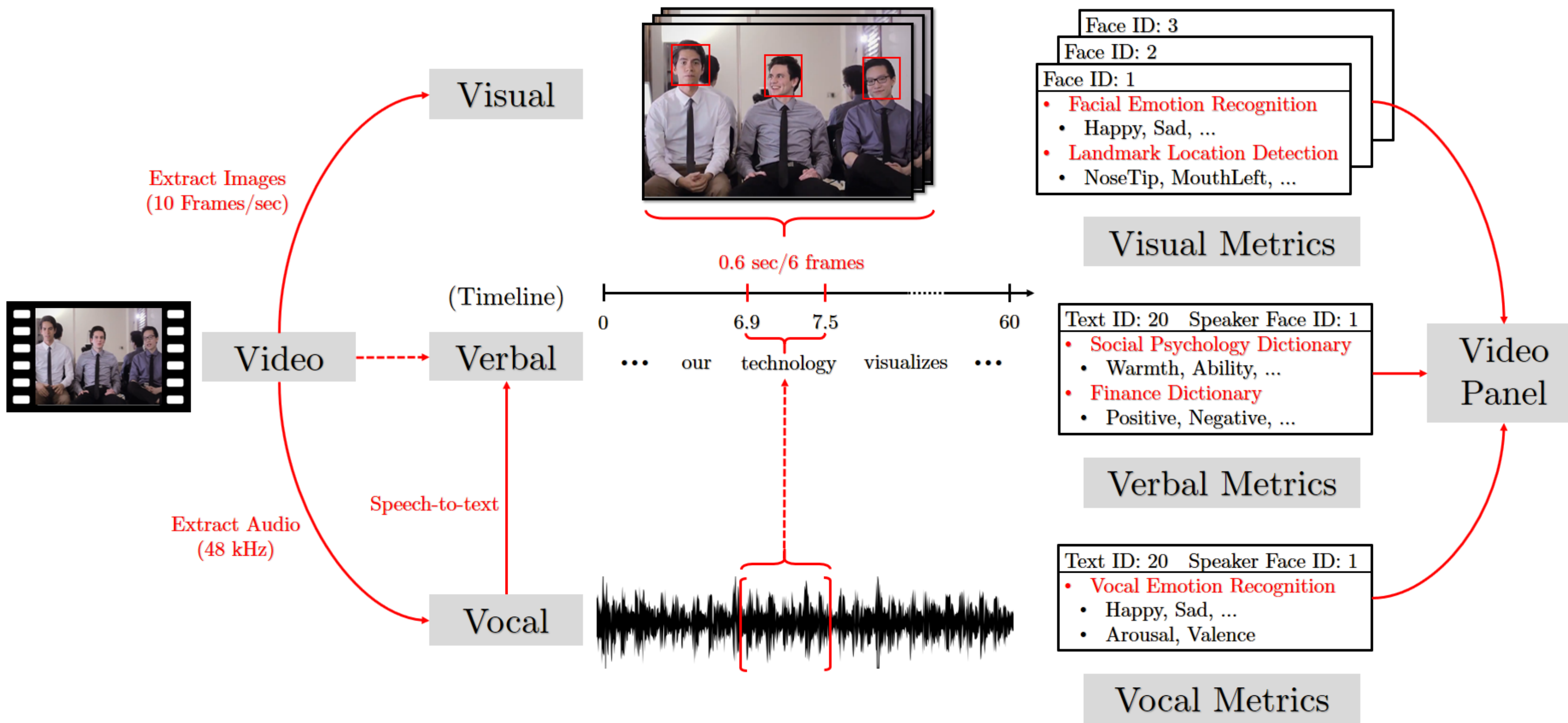


(b) Visaulized Example of Low-Arousal Pitch



Notes. This figure presents visualized examples of startup pitches with high-arousal (Panel (a)) and low-arousal (Panel (b)) vocal features. The visualization uses the waveform amplitude of those pitches. The high-arousal pitch can be downloaded from `https://www.dropbox.com/s/ipluo2w9tsszu2m/High%20Arousal%20Example.wav?dl=0`, and the low-arousal pitch can be downloaded from `https://www.dropbox.com/s/7igoqkjla72usdc/Low%20Arousal%20Example.wav?dl=0`.

Figure 4. Data Structure and Processing Procedure



Notes. This figure illustrates the flow of processing video data using a real video example.

Table 1. Summary Statistics of Videos and Startups

Panel A: Summary Statistics of Video Pitches

	N	Mean	STD	25%	50%	75%
Duration (second)	1,139	83.43	39.51	60.00	68.00	97.00
Video Size (MB)	1,139	18.16	18.37	6.88	12.86	22.64
Number of Words	1,139	228.53	107.76	163.00	201.00	262.00
Number of Sentences	1,139	15.91	7.33	11.00	14.00	19.00
Number of Views (YouTube)	1,139	764.37	6956.02	31.00	79.00	197.00
Number of Likes (YouTube)	1,139	1.51	6.60	0.00	0.00	1.00
Number of Dislikes (YouTube)	1,139	0.15	0.65	0.00	0.00	0.00

Panel B: Summary Statistics of Startups (as of July 2019)

	N	Mean	STD	25%	50%	75%
Startup Alive	1,139	0.53	0.50	0.00	1.00	1.00
Firm Age	1,139	3.20	2.01	2.00	3.00	5.00
Invested by Accelerator	1,139	0.09	0.28	0.00	0.00	0.00
Raised VC	194	0.37	0.48	0.00	0.00	1.00
Total Funding Amount ($000)	194	12,292	69,031	103	365	2,200
Total Funding Rounds	264	2.04	1.44	1.00	1.50	3.00
Number of Employees	388	21.77	73.13	5.00	5.00	30.00

Panel C: Summary Statistics of Teams

	N	Mean	STD	25%	50%	75%
Number of People	1,139	1.74	0.84	1.00	2.00	2.00
Single-Member	1,139	0.46	0.50	0.00	0.00	1.00
Multi-Member	1,139	0.54	0.50	0.00	1.00	1.00
Men-Only	1,139	0.49	0.50	0.00	0.00	1.00
Women-Only	1,139	0.27	0.45	0.00	0.00	1.00
Mixed Gender	1,139	0.24	0.43	0.00	0.00	0.00
Prior Senior Position	1,139	0.47	0.50	0.00	0.00	1.00
Prior Startup Experience	1,139	0.30	0.46	0.00	0.00	1.00
Elite University	1,139	0.06	0.24	0.00	0.00	0.00
Master's Degree	1,139	0.19	0.40	0.00	0.00	0.00
PhD Degree	1,139	0.03	0.17	0.00	0.00	0.00

Notes. This table provides descriptive statistics of pitch videos and the underlying startups in our sample. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles. Panel A reports basic information of the pitch videos. Panel B reports characteristics of startups measured as of July 2019 from Crunchbase and PitchBook. Panel C reports the summary statistics of the startup teams. Team member background information is collected from LinkedIn.

Table 2. Summary Statistics of Pitching Behavior Metrics

Panel A: Summary Statistics of Unstandardized Features

	N	Mean	STD	25%	50%	75%	Pitch Factor Loading	Uniqueness
Visual (Facial)								
Visual-Positive	1,139	0.17	0.16	0.05	0.12	0.25	0.08	0.29
Visual-Negative	1,139	0.15	0.14	0.06	0.11	0.20	−0.14	0.27
Visual-Beauty	1,139	0.58	0.08	0.54	0.59	0.64		
Vocal (Audio)								
Vocal-Positive	1,139	0.09	0.05	0.06	0.08	0.12	0.39	0.41
Vocal-Negative	1,139	0.01	0.01	0.01	0.01	0.02	−0.30	0.43
Vocal-Arousal	1,139	0.55	0.27	0.39	0.58	0.76	0.91	0.15
Vocal-Valence	1,139	0.44	0.22	0.31	0.46	0.59	0.88	0.18
Verbal (Text)								
Verbal-Positive	1,139	0.01	0.01	0.01	0.01	0.02	0.03	0.35
Verbal-Negative	1,139	0.01	0.01	0.00	0.01	0.01	−0.14	0.42
Verbal-Warmth	1,139	0.02	0.01	0.01	0.01	0.02	0.06	0.62
Verbal-Ability	1,139	0.03	0.02	0.02	0.03	0.05	0.06	0.56

Panel B: Correlations of the Features

	(1)	(2)	(3)	(4)	(5)
(1) Visual-Positive	1.00				
(2) Visual-Negative	−0.12***	1.00			
(3) Visual-Beauty	−0.02	−0.20***	1.00		
(4) Vocal-Positive	0.16***	0.07**	−0.05*	1.00	
(5) Vocal-Negative	0.05*	0.06**	0.01	−0.07**	1.00
(6) Vocal-Arousal	0.02	−0.07**	0.05*	0.24***	−0.15***
(7) Vocal-Valence	−0.02	−0.07**	0.09***	0.13***	−0.12***
(8) Verbal-Positive	0.01	0.03	−0.01	0.02	−0.06*
(9) Verbal-Negative	−0.10***	0.04	−0.01	0.02	−0.04
(10) Verbal-Warmth	−0.05	0.00	0.01	−0.01	−0.01
(11) Verbal-Ability	0.00	0.02	−0.02	0.04	0.02
Continued	(6)	(7)	(8)	(9)	(10)
(6) Vocal-Arousal	1.00				
(7) Vocal-Valence	0.75***	1.00			
(8) Verbal-Positive	−0.01	0.01	1.00		
(9) Verbal-Negative	−0.08***	−0.07**	0.00	1.00	
(10) Verbal-Warmth	0.02	0.04	0.03	−0.05*	1.00
(11) Verbal-Ability	0.01	0.04	0.08***	−0.03	−0.02

Notes. This table provides summary statistics of the pitch delivery features. In Panel A, for each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles. Variables are categorized into vocal, video, and verbal. The last two columns in Panel A report the factor loading and uniqueness of each feature when performing the principal component factor analysis to generate the single Pitch Factor that captures the maximum variance in the set of pitch features (Visual-Beauty is excluded as it is a static appearance feature). Panel B provides correlations of the features extracted from the pitches. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 3. Features in Pitch Delivery and Investment Decisions

Dependent Var: I(*Invested*)	Logit without Controls			Logit with Startup/Team Controls		
	Marginal Effect	S.E.	Pseudo R^2	Marginal Effect	S.E.	Pseudo R^2
Pitch Factor	0.030***	(0.007)	0.193	0.026***	(0.007)	0.253
Visual (Facial)						
Visual-Positive	0.015***	(0.005)	0.178	0.012**	(0.006)	0.240
Visual-Negative	−0.027***	(0.007)	0.187	−0.029***	(0.007)	0.253
Visual-Beauty	0.015**	(0.006)	0.178	0.015**	(0.007)	0.242
Vocal (Audio)						
Vocal-Positive	0.009**	(0.005)	0.174	0.011*	(0.006)	0.239
Vocal-Negative	−0.045***	(0.016)	0.183	−0.047***	(0.017)	0.248
Vocal-Arousal	0.023***	(0.009)	0.184	0.019**	(0.008)	0.245
Vocal-Valence	0.023***	(0.006)	0.185	0.020***	(0.007)	0.246
Verbal (Text)						
Verbal-Positive	−0.010	(0.009)	0.174	−0.011	(0.009)	0.239
Verbal-Negative	−0.026***	(0.007)	0.186	−0.022***	(0.008)	0.246
Verbal-Warmth	0.026***	(0.008)	0.190	0.028***	(0.008)	0.256
Verbal-Ability	−0.049***	(0.009)	0.243	−0.043***	(0.007)	0.298

Notes. Logit regressions, marginal effect, $N = 1,139$. The analysis is obtained using the following model:

$$I(Invested) = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon.$$

I(*Invested*) takes a value of one if the startup team was chosen by the accelerator, and zero otherwise. All pitch feature variables are standardized into a zero-mean variable with a standard deviation of one. All regressions include Accelerator FE. Control variables include founders' education background (whether they have a master's degree or a PhD degree, whether they attended an elite university, defined as the U.S. News & World Report's Top 10), founders' prior work experience (whether they have prior entrepreneurship experience, whether they ever held a senior position in prior employment), team size, and video resolution. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4. Sample Selection of Available Videos

	(1)	(2)	(3)	(4)	(5)	(6)
	Video Selected Out = 1					
Pitch Factor	0.006	0.015				
	(0.022)	(0.023)				
I(*Invested*)			−0.042	−0.044		
			(0.183)	(0.172)		
VC Invested					−0.011	−0.034
					(0.064)	(0.054)
Observations	527	527	527	527	527	527
Pseudo R^2	0.000	0.047	0.000	0.046	0.000	0.046
Startup/Team Controls		Y		Y		Y
Accelerator FE		Y		Y		Y

Notes. Logit regressions, marginal effect. This table investigates the sample selection issue of the video sample. The analysis restricts to videos that were uploaded between 2018 and July 2019. By the end of March 2020, 126 videos, or 23.9 percent, were selected out (unlisted, privatized, or removed) from the hosting platforms. The analysis investigates the relation between a video being "selected out" (made private, unlisted, or completely removed) and pitch delivery features and the outcomes of the startup. The set of startup/team control variables is identical to that in Table 3. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5. Measure Construction—Full Video and Full Channels

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent Var: I*(Invested)*					
	First Slice		Random Slice		Individual Channels	
Pitch Factor		0.026***		0.035**		0.064**
		(0.008)		(0.014)		(0.027)
Pitch Factor (First Slice)	0.015*	0.001				
	(0.008)	(0.008)				
Pitch Factor (Random Slice)			0.018***	−0.011		
			(0.005)	(0.013)		
Vocal Factor					0.023***	−0.040
					(0.007)	(0.028)
Visual Factor					0.025***	0.019***
					(0.006)	(0.006)
Verbal Factor					0.000	−0.009
					(0.008)	(0.009)
Observations	1,139	1,139	1,139	1,139	1,139	1,139
Pseudo R^2	0.241	0.253	0.243	0.254	0.263	0.269
Startup/Team Controls	Y	Y	Y	Y	Y	Y
Accelerator FE	Y	Y	Y	Y	Y	Y

Notes. Logit regressions, marginal effect. This table uses horse-race regressions to compare Pitch Factor (constructed from three-V channels jointly, using complete videos) with Pitch Factor First Slice (constructed from three-V channels jointly, using the slice of the first word time interval in each video), Pitch Factor Random Slice (constructed from three-V channels jointly, using the slice of a random word time interval in each video), and Visual/Vocal/Verbal Factor (constructed from three-V channels separately, using complete videos). I*(Invested)* takes a value of one if the startup team was chosen by the accelerator, and zero otherwise. All pitch feature variables are standardized into a zero-mean variable with a standard deviation of one. The construction of Pitch Factor is identical to that in Table 3. Control variables include founders' education background (whether they have a master's degree or a PhD degree, whether they attended an elite university, defined as the U.S. News & World Report's Top 10), founders' prior work experience (whether they have prior entrepreneurship experience, whether they ever held a senior position in prior employment), team size, and video resolution. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 6. Features in Pitches and Investment Decisions—Oster Test

$R^2_{max} = \min(2.2R^2_c, 1)$						
$\delta = 1$			$\delta = 2$			δ s.t. $\beta_{adj} = 0$
β_{adj}	Identified Set	Reject Null?	β_{adj}	Identified Set	Reject Null?	
0.023	[0.023,0.026]	Y	0.019	[0.019,0.026]	Y	6.157

$R^2_{max} = \min(3R^2_c, 1)$						
$\delta = 1$			$\delta = 2$			δ s.t. $\beta_{adj} = 0$
β_{adj}	Identified Set	Reject Null?	β_{adj}	Identified Set	Reject Null?	
0.021	[0.021,0.026]	Y	0.014	[0.014,0.026]	Y	3.752

$R^2_{max} = 1$						
$\delta = 1$			$\delta = 2$			δ s.t. $\beta_{adj} = 0$
β_{adj}	Identified Set	Reject Null?	β_{adj}	Identified Set	Reject Null?	
0.012	[0.012,0.026]	Y	−0.007	[−0.007,0.026]	N	1.678

Notes. This table tests the role of omitted and unobservable control variables in explaining the relation between the Pitch Factor and the venture investment decision, using the test designed in Oster (2019). To implement, we estimate a linear model of

$$I(Invested) = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon,$$

first without any control variables through which we obtain β_u and R^2_u, and then with the added startup/team control variable, through which we obtain β_c and R^2_c. The set of startup/team control variables is identical to that in Table 3. The raw OLS estimates used in this test are provided in Appendix Table A.5.

For any given test parameter combination δ and R^2_{max}, Oster (2019) defines the bias-adjusted coefficient, denoted as β_{adj} that is determined by parameters δ and R^2_{max}, to be closely approximated by (strictly equal to when $\delta = 1$)

$$\beta_{adj} \approx \beta_c - \delta \frac{(\beta_u - \beta_c)(R^2_{max} - R^2_c)}{R^2_c - R^2_u}.$$

With this adjusted coefficient β_{adj}, the recommended identified set is the interval between β_{adj} and β_c. In the table, we report the adjusted β and identified set for different combinations of parameters, and we also report whether the identified set rejects the null of $\beta = 0$ and the δ value to make certain R^2_{max} reach zero.

Table 7. Long-Term Performance of Startups and Features in Pitches

	(1) Employment	(2) Raised VC	(3) VC Amount	(4) Startup Alive
Pitch Factor	−0.166**	−0.089***	−0.168*	−0.043**
	(0.050)	(0.018)	(0.086)	(0.021)
Observations	150	132	132	174
(Pseudo) R^2	0.267	0.257	0.306	0.290
Age Controls	Y	Y	Y	Y
Startup/Team Controls	Y	Y	Y	Y
Accelerator FE	Y	Y	Y	Y
Region FE	Y	Y	Y	Y

Notes. OLS regressions (columns (1) and (3)) and Logit regressions, marginal effect (columns (2) and (4)). The analysis is obtained using the following model conditional on receiving funding from a VC:

$$Performance = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon.$$

Employment is the inverse hyperbolic sine of number of employees. Raised VC is a dummy variable that takes a value of one if a startup raised another round of VC investment after receiving funding from accelerators. VC Amount is the inverse hyperbolic sine of total amount of VC investment that a startup has raised. Startup Alive is a dummy variable that takes a value of one if a startup's website is still operational. All performance variables are as of July 2019. Pitch Factor is standardized into a zero-mean variable with a standard deviation of one. The set of startup/team control variables is identical to that in Table 3. Standard errors clustered at the industry-level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 8. Gender Differences in the Pitch-Investment Relation

	(1)	(2)	(3)	(4)
	Dependent Var: I(*Invested*)			
	Single-Gender Teams			Mixed-Gender Teams
	Men	Women	Pooled	Pooled
Pitch Factor (Men)	0.018**		0.018**	0.048*
	(0.008)		(0.008)	(0.026)
Pitch Factor (Women)		0.170***	0.077**	0.019
		(0.051)	(0.031)	(0.042)
p-value of Men vs. Women Test			0.079*	0.661
Observations	559	310	869	270
Pseudo R2	0.194	0.334	0.217	0.653
Startup/Team Controls	Y	Y	Y	Y
Accelerator FE	Y	Y	Y	Y
Industry FE	Y	Y	Y	

Notes. Logit regressions, marginal effect. The analysis is obtained using the following model:

$$I(Invested) = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon.$$

I(*Invested*) takes value of one if the startup team was chosen by the accelerator, and zero otherwise. Pitch Factor is standardized into a zero-mean variable with a standard deviation of one. Columns (1) and (2) separately analyze for men-only and women-only startup teams and Pitch Factor is standardized separately. Column (3) pools the sample of men-only and women-only startup teams. Pitch Factor (Men) is Pitch Factor times Pure Men Team dummy. Pitch Factor (Women) is Pitch Factor times Pure Women Team dummy. Column (4) analyzes mixed-gender teams, but Pitch Factors are calculated for men and women separately in each video. The set of startup/team control variables is identical to that in Table 3. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 9. Experiment Results: Pitch Factor and Investor Beliefs

	(1)	(2)	(3)	(4)	(5)
	P($alive \mid invested$)		P($success \mid invested$)		$alive \mid invested$
	μ	σ	μ	σ	Realized
Pitch Factor (θ)	0.020**	−0.020	0.016**	−0.030	−0.117**
	(0.009)	(0.027)	(0.007)	(0.028)	(0.053)
Observations	952	952	952	952	495
R^2	0.569	0.545	0.565	0.519	0.673
Startup/Team Controls	Y	Y	Y	Y	Y
Subject FE	Y	Y	Y	Y	Y

Notes. OLS regressions. This table investigates the relation between the Pitch Factor and investor beliefs in an experiment setting (columns (1) to (4)) and realized startup performance conditional on being invested (column (5)). P($alive \mid invested$) μ and P($alive \mid invested$) σ are subjects' beliefs and precision of beliefs on the probability of a startup to be alive three years later conditional on raising funding. P($success \mid invested$) μ and P($success \mid invested$) σ are subjects' beliefs and precision of beliefs on the probability that a startup will be a success conditional on raising funding. In column (5), $alive \mid invested$ is a dummy variable which takes a value of one if a startup's website is still operational as of July 2019. The sample focuses on those startups that obtained venture funding, i.e., conditional on a startup being invested. The Pitch Factor is standardized into a zero-mean variable with a standard deviation of one. All analysis controls for subject individual fixed effects and controls at the startup/team level as in Table 3. Standard errors two-way clustered at the startup and subject levels are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 10. Experiment Results: Inaccurate Beliefs, Tastes, and Investment

	(1)	(2)	(3)	(4)
	Dependent Var: I(*Invested*)			
Pitch Factor (θ)	0.125***			0.067***
	(0.037)			(0.022)
$\mu(alive\|invested)$		2.309***		2.208***
		(0.120)		(0.132)
$\sigma(alive\|invested)$			−0.171***	−0.054**
			(0.041)	(0.026)
Observations	952	952	952	952
Pseudo R^2	0.157	0.423	0.135	0.436
Startup/Team Controls	Y	Y	Y	Y
Subject FE	Y	Y	Y	Y

Notes. Logit regressions, marginal effect. This table estimates the model of investment decisions in an experiment setting, as in Eq. (7). I(*Invested*) takes a value of one if a subject decides to invest in a startup team in the experiment, and zero otherwise. $\mu(alive|invested)$ and $\sigma(alive|invested)$ are subjects' beliefs and precision of beliefs on the probability that a startup will be alive three years later conditional on receiving funding. The Pitch Factor is standardized into a zero-mean variable with a standard deviation of one. All analysis controls for subject individual fixed effects and control at the startup/team level as in Table 3. Standard errors two-way clustered at the startup and subject levels are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Online Appendix (Not For Publication)

A. Collecting Video Data from Video Platforms

When startups apply to accelerator programs, they are required (or highly recommended) to record and submit a standardized self-introductory pitch video as part of the application process. Figure A.1 shows such examples from those accelerators' application systems. These videos, rather than being submitted to the accelerators directly, are submitted through uploading to a public multimedia platform, such as YouTube or Vimeo, and then providing the url links to these videos in application forms.

We use an automatic searching script for two public video-sharing websites, YouTube and Vimeo. Integrated with query APIs, our web crawler returns a list of video indices according to a set of predefined keywords, which include but are not limited to the names of these accelerators, "startup accelerator application video","accelerator application videos" and so on. We first obtain the full list of potential videos returned by each keyword search (there is a limit of returned videos by YouTube), and then filter the potential videos by a combination of different conditions on video info obtained along with the video itself. Filtering variables include but are not limit to data format, duration, title, and annotation.

Table A.1. List of Searching Keywords for Collecting Videos

Keywords
YC Application Videos
Y Combinator Application Videos
MassChallenge Application Videos
500 Startups Application Videos
Techstars Application Videos
AngelPad Application Videos
Y Combinator Application Videos + *YEAR*
Techstars Application Videos + *YEAR*
500 Startups Application Videos + *YEAR*
AngelPad Application Videos + *YEAR*

Notes. This table shows the list of keywords we use for searching and collecting the pitch videos from Youtube and Vimeo. The *YEAR* takes values from 2005 to 2019.

We also employ startup names listed on accelerators' web pages to expand our video data set. Specifically, we first obtain the full list of startups accelerated by the accelerator each year if such a list is published on the accelerator's website. Then our script automatically searches these startup names and checks the first three results returned by the search API. A match is defined as having both the startup name and the accelerator name appear in the video title or annotation.

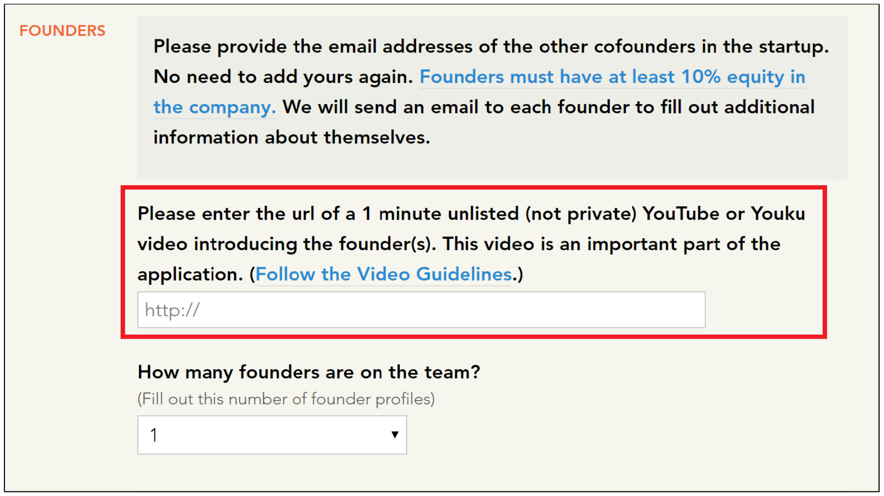
It is worth noting that if one company has more than one video in our sample, we only keep the video recorded first. There are 33 such firms in our analysis, which make up only 2.90% of our sample. These firms have multiple videos because of the following reasons. First, there are some entrepreneur teams applying to different accelerators. Second, there are some teams that applied to the same accelerator multiple times. For these firms, we only keep their videos and outcomes in the first application.

In total we obtain 1,139 videos. Table A.2 describes the sample, in which the number of videos is reported by accelerator (Panel A) and by year (Panel B). Y Combinator contributes the largest number of application videos, followed by MassChallenge and Techstars. Among all the companies that applied, 97 (8.52%) were chosen by the accelerator program, and 248 (21.77%) were invested by any venture investor (accelerator or angels/VCs). The videos are more available for recent years due to the increase in video requirements in the application.

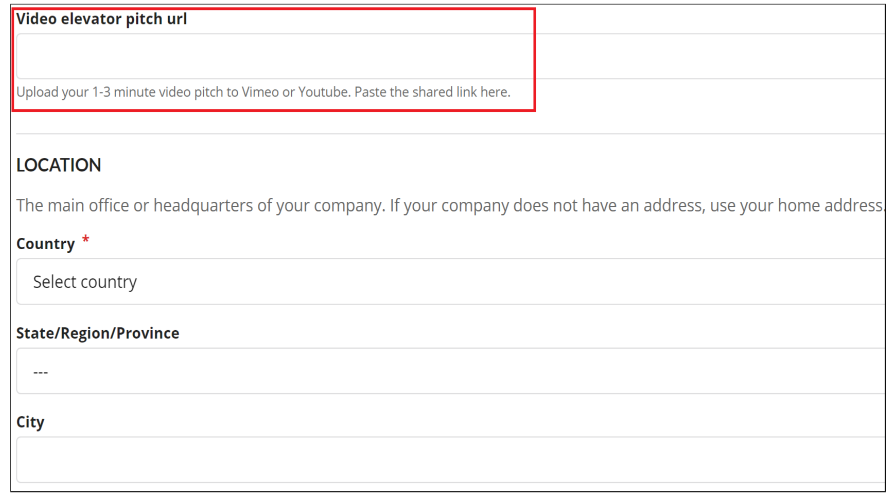
After collecting the videos, we parse each video web page to collect other relevant information. This includes the video's duration, upload date, title, annotation, subtitle, and uploader ID. This set of information also allows us to identify the startup almost perfectly. Specifically, by scrutinizing video titles and annotations, we double-check names of the startups and names of the accelerators they are applying for. If the startup name cannot be identified from these items, we search the uploader name on LinkedIn and back out the company information. It is common that many people have the same name on LinkedIn, so to verify that the person on Linkedin is the founder, we also double-check the name, background, experience, and even photos.

Figure A.1. Examples of Accelerator Online Application Forms

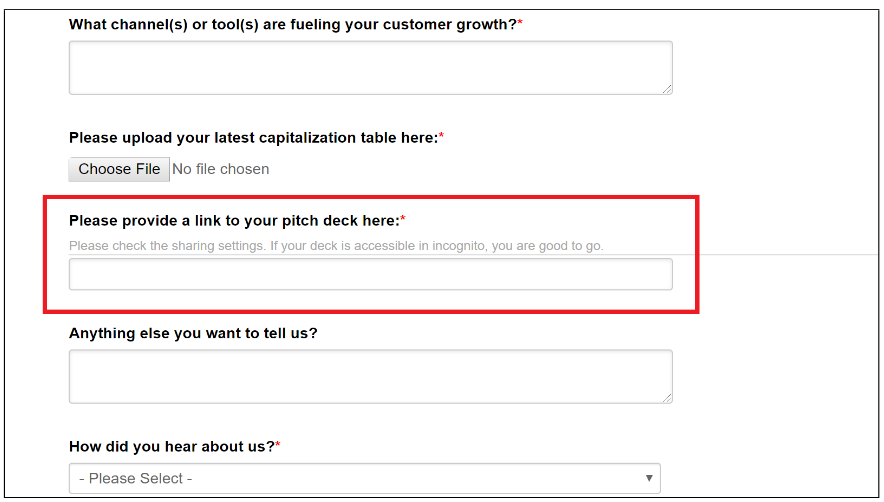
(a) Y Combinator



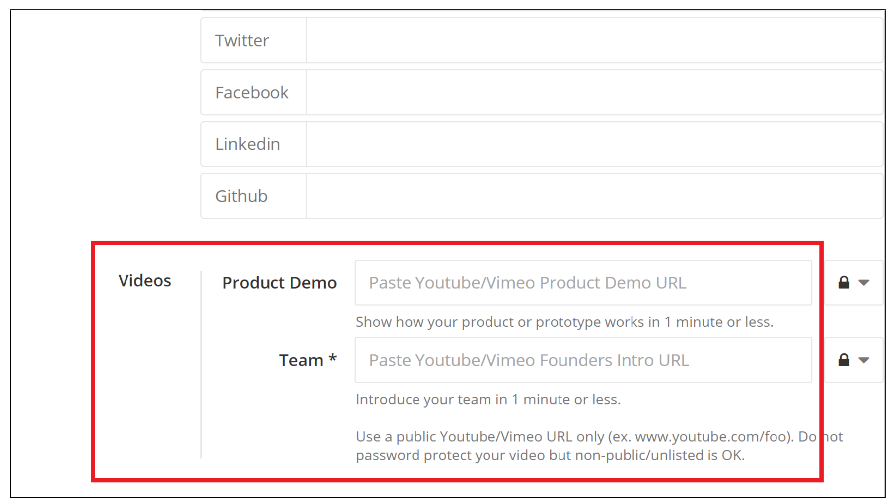
(b) MassChallenge



(c) 500 Startups



(d) Techstars



Notes. This figure shows screenshots of accelerators' online application forms. We show forms for Y Combinator, MassChallenge, 500 Startups, and Techstars. On each online application form, we highlight the question specifically asking for uploading pitch videos.

Figure A.2. Screenshot of Search Results from YouTube

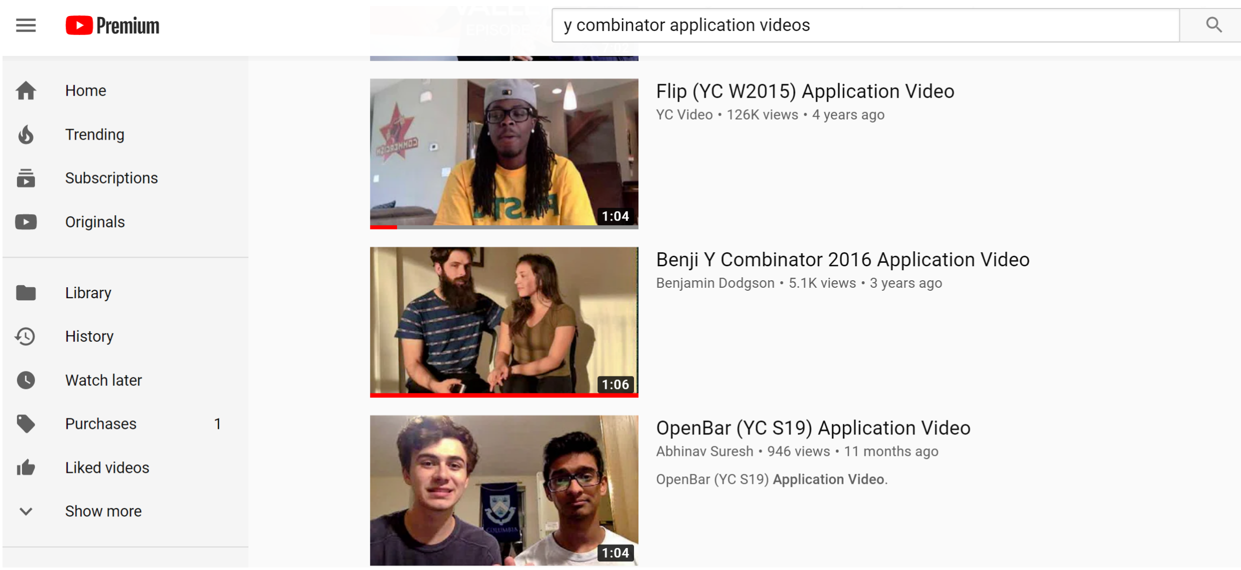


Table A.2. Sample Description of Pitch Videos

Panel A: Breakdown by Accelerators and Investment Status

Accelerator	Videos #	Accelerator Invested	Website Active	In Crunchbase	In Pitchbook
500 Startups	33	1	15	19	8
AngelPad	83	2	33	36	18
MassChallenge	166	56	129	113	79
Techstars	136	3	67	53	21
Y Combinator	713	35	363	238	91
YC Fellowship	8	0	2	3	0
Total	1,139	97	609	462	217
% of Full Sample	100.00%	8.52%	53.47%	40.56%	19.05%

Panel B: Breakdown by Years

Accelerator	<=2012	2013	2014	2015	2016	2017	2018	2019
500 Startups	1	1	7	7	2	8	5	2
AngelPad	11	7	13	4	12	14	21	1
MassChallenge	4	9	4	13	34	33	34	35
Techstars	9	17	12	15	8	30	32	13
Y Combinator	10	31	29	82	67	110	164	220
YC Fellowship	0	0	0	8	0	0	0	0
Total	35	65	65	129	123	195	256	271
% of Full Sample	3.07%	5.71%	5.71%	11.33%	10.80%	17.12%	22.48%	23.79%

Notes: This table provides descriptive statistics on collected videos by accelerators that the applications are made to (Panel A) and by year (Panel B). We obtain pitch videos using an automatic searching script for two public video-sharing websites, YouTube and Vimeo. Integrated with query APIs, our web crawler returns a list of video indices according to a set of predefined keywords, which include but are not limit to the names of these accelerators, "startup accelerator application video","accelerator application videos" and so on. We first obtain the full list of potential videos returned by each keyword search (there is a limitation of returned videos by YouTube), and then filter the potential videos by a combination of different conditions on video info obtained along with the video itself. Filtering variables include but are not limit to data format, duration, title, and annotation. We also obtain additional videos from accelerators' websites. Panel A reports the number of videos submitted to each accelerator and the proportion of each accelarator in the full sample. Panel B reports the breakdown by application year (typically the year of video uploading).

B. Method Appendix

This appendix provides more details on the steps to perform video analysis used in our paper. Compared to the more theoretical descriptions provided in Section 2 of the paper, this appendix proceeds with a more practical approach with information on our code structure, key functions, and notes on important steps.

Video Processing Example

This example shows how to use `interactionvideo` package to process a video for studies in human interactions. Please also refer to our research paper: Hu and Ma (2020), "Pursuading Investors: A Video-Based Study", available at: https://songma.github.io/files/hm_video.pdf.

Overview

The video processing involves the following steps:

1. Set up folders and check dependencies (requirements)
2. Extract images and audios from a video using `pliers`
3. Extract text from audios using Google Speech2Text API
4. Process images(faces) using Face++ API
5. Process text using Loughran and McDonald (2011) Finance Dictionary and Nicolas, Bai, and Fiske (2019) Social Psychology Dictionary
6. Process audios using pre-trained ML models in `pyAudioAnalysis and speechemotionrecognition`
7. Aggregate information from 3V (visual, vocal, and verbal) to video level

Structure

```
├── interactionvideo
│   ├── __pycache__
│   ├── prepare.py
│   ├── decompose.py
│   ├── faceppml.py
│   ├── googleml.py
│   ├── textualanalysis.py
│   ├── audioml.py
│   ├── aggregate.py
│   └── utils.py
├── data
│   ├── example_video.mp4
│   └── VideoDictionary.csv
├── mlmodel
│   ├── pyAudioAnalysis
│   └── speechemotionrecognition
├── output
│   ├── audio_temp
│   ├── image_temp
│   └── result_temp
├── PythonSDK
├── example.py
├── Video Processing Example.ipynb
├── README.md
└── requirement.txt
```

Dependencies

- pandas
- tqdm
- codecs
- pliers
- pydub
- PIL
- google-cloud-speech
- google-cloud-storage
- speechemotionrecognition
- pyAudioAnalysis

1. Set up folders and check dependencies (requirements)

In [1]:
```python
from os.path import join
# Set your root path here
RootPath = r''
# Set your video file path here
VideoFilePath = join(RootPath,'data','example_video.mp4')
# Set your work path here
# Work path is where to store meta files and output files
WorkPath = join(RootPath,'output')
```

In [2]:
```python
# Set up the folders
from interactionvideo.prepare import setup_folder
setup_folder(WorkPath)

# check the requirements for interactionvideo
from interactionvideo.prepare import check_requirements
check_requirements()
```

```
decompose.py requirements satisfied.

faceppml.py requirements satisfied.

googleml.py requirements satisfied.

audioml.py requirements satisfied.
```

Out[2]:
```
True
```

2. Extract images and audios from video

In [3]:
```python
from interactionvideo.decompose import convert_video_to_images

# Decompose the video into a stream of images
# The default sampling rate is 10 frames per second
# Find the output at WorkPath\image_temp
convert_video_to_images(VideoFilePath, WorkPath)
```

```
Video is 70.12 seconds long.

100%|████████████████████████████████████████████████████████████████| 7
02/702 [06:03<00:00,  1.86it/s]


Video is sampled to 702 images.

Video to images finished.
```

Out[3]:
```
True
```

In [4]:
```python
from interactionvideo.decompose import convert_video_to_audios

# Decompose the video into audios
# Find the output at WorkPath\audio_temp
convert_video_to_audios(VideoFilePath, WorkPath)
```

```
MoviePy - Writing audio in %s

MoviePy - Done.
Video to audios finished.
```

Out[4]:
```
True
```

3. Extract text from audios using Google Speech2Text API

Set up your Google Cloud environment following

- https://cloud.google.com/python (https://cloud.google.com/python)
- https://cloud.google.com/storage/docs/quickstart-console (https://cloud.google.com/storage/docs/quickstart-console)
- https://cloud.google.com/speech-to-text (https://cloud.google.com/speech-to-text)

Create a Google Cloud Storage bucket.

In [5]:
```python
from interactionvideo.googleml import upload_audio_to_googlecloud

# Set your Google Cloud Storage bucket name here
GoogleBucketName = ''

# Upload audio file to Google Cloud Storage
upload_audio_to_googlecloud(WorkPath, GoogleBucketName)
```

```
Uploaded the audio file to Google Cloud.
```

Out[5]:
```
True
```

In [6]:
```python
from interactionvideo.googleml import convert_audio_to_text_by_google

# Use Google Speech2Text API to convert audio to text
# Return a txt file of full speech script and a csv file of text and punctuation
# Find the output at
# - WorkPath\result_temp\script_google.txt (full speech script)
# - WorkPath\result_temp\text_panel_google.csv (text panel from Google)
google_result_text, google_result_df = convert_audio_to_text_by_google(WorkPath, GoogleB
ucketName)
```

```
Google Speech2Text begins. 70.12 seconds audio to process.

Google Speech2Text ends. 70.12 seconds audio processed.
```

In [7]:
```python
# Check full speech script from Google
print(google_result_text)
```

```
Hello, everyone. First of all, we will like to thank you for your interest in our resear
ch in this paper. We try to understand how human interaction features such as facial exp
ressions vocal emotions and word choices might influence economic agents decision making
in order to study this question empirically, we build an empirical approach that uses vi
deos of human interactions as data input and and machine learning based algorithms as th
e tool. We apply an empirical approach in a setting where early stage Turn up Pitch Vent
ure capitalists for early-stage funding. We find that pitch features along visual vocal
and verbal damages all matter for the probability of receiving funding and we also show
that this event impact is largely due to interaction induced biases rather than that int
eractions provide additional valuable information the empirical structure that you see i
n this code example can hopefully help you to get started with using video in other impo
rtant settings such as As interviews classroom recordings among many other exciting thin
gs. We look forward to hearing your feedback and reading about your research. Thank you.
```

In [8]:
```python
# Check text panel from Google
google_result_df.head(10)
```

Out[8]:

	Text	Onset	Offset	Duration	Sentence End
0	Hello,	0.1	0.7	0.6	True
1	everyone.	0.7	1.1	0.4	True
2	First	1.1	1.5	0.4	False
3	of	1.5	1.6	0.1	False
4	all,	1.6	1.9	0.3	True
5	we	1.9	2.0	0.1	False
6	will	2.0	2.2	0.2	False
7	like	2.2	2.3	0.1	False
8	to	2.3	2.4	0.1	False
9	thank	2.4	2.7	0.3	False

4. Process images(faces) using Face++ API

Get your key and secret from https://www.faceplusplus.com (https://www.faceplusplus.com).

If you register at https://console.faceplusplus.com/register (https://console.faceplusplus.com/register), use https://api-us.faceplusplus.com (https://api-us.faceplusplus.com) as the server.

If you register at https://console.faceplusplus.com.cn/register (https://console.faceplusplus.com.cn/register), use https://api-cn.faceplusplus.com (https://api-cn.faceplusplus.com) as the server.

The `Python SDK` of Face++ is included in this package.

You can also download it from https://github.com/FacePlusPlus/facepp-python-sdk (https://github.com/FacePlusPlus/facepp-python-sdk).

```
In [9]: from interactionvideo.faceppml import process_image_by_facepp

        # Use Face++ ML API to process images
        # Return csv files of facial emotion, gender, predicted age
        # Find the output
        # - WorkPath\result_temp\face_panel_facepp.csv (full returns from Face++)
        # - WorkPath\result_temp\face_panel.csv (clean results)

        # Set your key, secret, and server here
        FaceppKey = ''
        FaceppSecret = ''
        FaceppServer = 'https://api-us.faceplusplus.com'

        facepp_result_df, facepp_result_clean_df = process_image_by_facepp(VideoFilePath, WorkPa
        th,\
                                                                   FaceppKey, FaceppSecret, FaceppServer)
```

```
Face++ API begins. 702 images to process.

100%|██████████████████████████████████████████████████| 70
2/702 [1:13:21<00:06,  6.47s/it]


Face++ API ends. 702 images processed.
```

```
In [10]: # Check full returns from Face++
         facepp_result_df.head(10)
```

Out[10]:

	ImageName	Onset	Offset	Duration	face_rectangle#top	face_rectangle#left	face_rectangle#width
0	frame[0]	0.0	0.1	0.1	405	868	249
1	frame[3]	0.1	0.2	0.1	406	867	250
2	frame[6]	0.2	0.3	0.1	404	866	252
3	frame[9]	0.3	0.4	0.1	403	867	253
4	frame[12]	0.4	0.5	0.1	401	866	258
5	frame[15]	0.5	0.6	0.1	405	867	261
6	frame[18]	0.6	0.7	0.1	407	867	261
7	frame[21]	0.7	0.8	0.1	404	869	258
8	frame[24]	0.8	0.9	0.1	403	867	262
9	frame[27]	0.9	1.0	0.1	402	868	262

10 rows × 193 columns

In [11]:
```
# Check clean results
facepp_result_clean_df.head(10)
```

Out[11]:

	Onset	Offset	Duration	Number of Faces	Gender	Age	Visual-Positive	Visual-Negative	Visual-Beauty
0	0.0	0.1	0.1	1	Male	31	0.00007	0.26876	0.430900
1	0.1	0.2	0.1	1	Male	33	0.00008	0.22857	0.406690
2	0.2	0.3	0.1	1	Male	30	0.00115	0.33071	0.413915
3	0.3	0.4	0.1	1	Male	28	0.00152	0.33477	0.402910
4	0.4	0.5	0.1	1	Male	28	0.00040	0.92615	0.415210
5	0.5	0.6	0.1	1	Male	26	0.00734	0.98612	0.447690
6	0.6	0.7	0.1	1	Male	30	0.00196	0.80259	0.449480
7	0.7	0.8	0.1	1	Male	32	0.00021	0.09574	0.449665
8	0.8	0.9	0.1	1	Male	29	0.00095	0.60956	0.451470
9	0.9	1.0	0.1	1	Male	29	0.00046	0.05656	0.468895

5. Process text using LM and NBF Dictionaries

Use Loughran-McDonald (2011) Finance Dictionary (LM) to construct verbal positive and negative.

For more details, please check [https://sraf.nd.edu/textual-analysis/resources (https://sraf.nd.edu/textual-analysis/resources)](https://sraf.nd.edu/textual-analysis/resources).

Use Nicolas, Bai, and Fiske (2019) Social Psychology Dictionary (NBF) to construct verbal warmth and ability.

For more details, please check [https://psyarxiv.com/afm8k (https://psyarxiv.com/afm8k)](https://psyarxiv.com/afm8k).

In [12]:
```
from interactionvideo.textualanalysis import process_text_by_dict

# Set LM-NBF dictionary path
DictionaryPath = join(RootPath,'data','VideoDictionary.csv')

# Dictionary-based textual analysis to get verbal measures
# (e.g., verbal positive, negative, warmth, ability)
# Return csv files of verbal positive, negative, warmth, and ability
# Find the output at
# - WorkPath\result_temp\text_panel.csv
text_result_df = process_text_by_dict(WorkPath, DictionaryPath)
```

```
LM and NBF Dictionaries loaded.

Dictionary-based textual analysis begins. 183 words to process.

Dictionary-based textual analysis ends. 183 words processed.
```

In [13]:
```
# Check text panel from Dictionary
text_result_df.head(10)
```

Out[13]:

	Text	Onset	Offset	Duration	Sentence End	Verbal-Negative	Verbal-Positive	Verbal-Warmth	Verbal-Ability
0	Hello,	0.1	0.7	0.6	True	0.0	0.0	0.0	0.0
1	everyone.	0.7	1.1	0.4	True	0.0	0.0	0.0	0.0
2	First	1.1	1.5	0.4	False	0.0	0.0	0.0	0.0
3	of	1.5	1.6	0.1	False	0.0	0.0	0.0	0.0
4	all,	1.6	1.9	0.3	True	0.0	0.0	0.0	0.0
5	we	1.9	2.0	0.1	False	0.0	0.0	0.0	0.0
6	will	2.0	2.2	0.2	False	0.0	0.0	0.0	0.0
7	like	2.2	2.3	0.1	False	0.0	0.0	1.0	0.0
8	to	2.3	2.4	0.1	False	0.0	0.0	0.0	0.0
9	thank	2.4	2.7	0.3	False	0.0	0.0	1.0	0.0

6. Process audios by pre-trained ML models

Construct vocal arousal and vocal valence from pre-trained SVM ML models in `pyAudioAnalysis` .

The pre-trained models are located at mlmodel\pyAudioAnalysis

- svmSpeechEmotion_arousal
- svmSpeechEmotion_arousalMEANS
- svmSpeechEmotion_valence
- svmSpeechEmotion_valenceMEANS

For more details, please check https://github.com/tyiannak/pyAudioAnalysis/wiki/4.-Classification-and-Regression (https://github.com/tyiannak/pyAudioAnalysis/wiki/4.-Classification-and-Regression).

Construct vocal positive and vocal negative from pre-trained LSTM ML models in `speechemotionrecognition` .

The pre-trained models are located at mlmodel\speechemotionrecognition

- best_model_LSTM_39.h5

For more details, please check https://github.com/harry-7/speech-emotion-recognition (https://github.com/harry-7/speech-emotion-recognition).

Note: speechemotionrecognition requires tensorflow and Keras.

In [14]:
```
from interactionvideo.audioml import process_audio_by_pyAudioAnalysis

# Set the model path
pyAudioAnalysisModelPath = join(RootPath,'mlmodel','pyAudioAnalysis')

# Construct vocal arousal and vocal valence
# Find the output at
# - WorkPath\result_temp\audio_panel_pyAudioAnalysis.csv
audio_result_df1 = process_audio_by_pyAudioAnalysis(WorkPath, pyAudioAnalysisModelPath)
```

```
pyAudioAnalysis vocal emotion analysis begins. 70.12 seconds audio to process.

pyAudioAnalysis ML model loaded.

pyAudioAnalysis vocal emotion analysis ends. 70.12 seconds audio processed.
```

In [15]:
```
# Check audio panel from pyAudioAnalysis
audio_result_df1.head()
```

Out[15]:

	Onset	Offset	Duration	Vocal-Arousal	Vocal-Valence
0	0	70.12	70.12	0.404089	-0.01519

In [16]:
```
from interactionvideo.audioml import process_audio_by_speechemotionrecognition

# Set the model path
speechemotionrecognitionModelPath = join(RootPath,'mlmodel','speechemotionrecognition')

# Construct vocal positive and vocal negative
# Find the output at
# - WorkPath\result_temp\audio_panel_speechemotionrecognition.csv
audio_result_df2 = process_audio_by_speechemotionrecognition(WorkPath, speechemotionreco
gnitionModelPath)
```

```
speechemotionrecognition vocal emotion analysis begins. 70.12 seconds audio to process.


Using TensorFlow backend.


_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 128)               86016
_________________________________________________________________
dropout_1 (Dropout)          (None, 128)               0
_________________________________________________________________
dense_1 (Dense)              (None, 32)                4128
_________________________________________________________________
dense_2 (Dense)              (None, 16)                528
_________________________________________________________________
dense_3 (Dense)              (None, 4)                 68
=================================================================
Total params: 90,740
Trainable params: 90,740
Non-trainable params: 0
_________________________________________________________________

speechemotionrecognition ML model loaded.

speechemotionrecognition vocal emotion analysis ends. 70.12 seconds audio processed.
```

In [17]:
```
# Check audio panel from speechemotionrecognition
audio_result_df2.head()
```

Out[17]:

	Onset	Offset	Duration	Vocal-Positive	Vocal-Negative
0	0	70.12	70.12	0.459319	0.006388

7. Aggregate information from 3V to video level

In [18]:
```
from interactionvideo.aggregate import aggregate_3v_to_video

# Aggregate 3V information
# Find the output at
# - WorkPath\result_temp\video_panel.csv
video_result_df = aggregate_3v_to_video(WorkPath)
```

```
3V to video aggregation finished.
```

In [19]:
```
# Check video panel
video_result_df.T
```

Out[19]:

	0
Number of Faces	1
Gender	Male
Age	32
Visual-Positive	0.0142308
Visual-Negative	0.443333
Visual-Beauty	0.450598
Vocal-Positive	0.46
Vocal-Negative	0.01
Vocal-Arousal	0.4
Vocal-Valence	-0.02
Verbal-Positive	0.010929
Verbal-Negative	0.010929
Verbal-Warmth	0.0327869
Verbal-Ability	0.0382514

C. Appendix: MTurk Rating Survey

This appendix presents details of our survey designs. The goal of these exercises is to bridge our ML-algorithm that rates pitch videos with the traditional approach of using human raters.

Both exercises take the form of an online survey that participants complete using their own electronic devices (e.g., computers and tablets), and they are distributed through Amazon Mechanical Turk (MTurk). In both surveys, we require the participants to be located in the U.S. and to be identified as masters at completing our types of tasks by the MTurk platform through its statistical performance monitoring. The experiments recruit 115 and 100 participants respectively. Our experiments on MTurk provide relatively high payments compared to the MTurk average to ensure quality responses.

Sample survey designs are attached toward the end of this appendix.

C.1. Survey 1: Rating on Pitch Positivity

In the first survey, we elicit ratings of positivity from MTurkers. In each survey, a respondent is allocated a random set of six pitch videos. For each video, we first mandate the completion of watching the full video, and the respondent is not able to skip the video before answering the rating questions. Then, on the next survey screen, we elicit the rating of positivity, defined as passion, enthusiasms, based on the video just watched. The rating is on a 1-9 scale with nine choices. The evaluations of the videos are completed one by one, and ratings may not be revised after moving to the next video.

We then compare the ratings from MTurkers with the Pitch Factor. Figure A.3 shows the binned scatter plot of the relation between the two variables. The clearly positive correlation provides the first assurance of the validity of the ML-generated measure. In a regression analysis, as shown in Table A.3, we also show a strong correlation between the two variables.

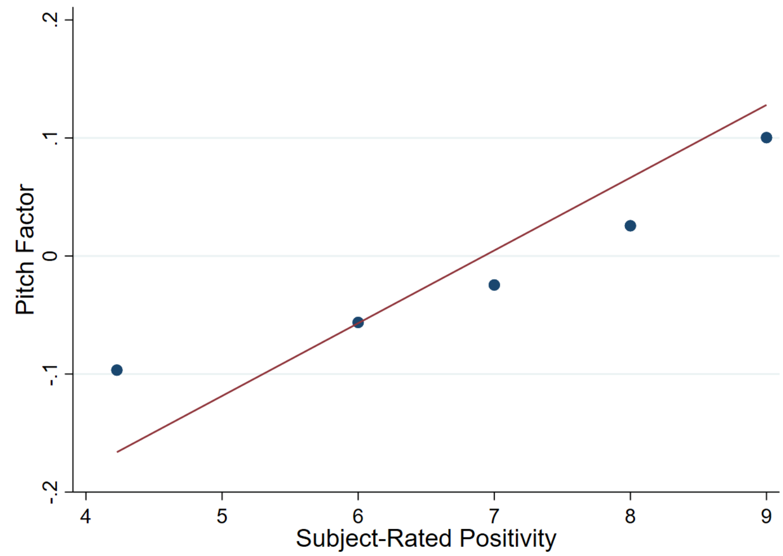


Figure A.3. Pitch Factor and Respondent-Rated Positivity

Table A.3. Pitch Factor and Respondent-Rated Positivity

	(1)	(2)
	Pitch Factor	
Respondent-Rated Positivity	0.062**	0.088***
	(0.028)	(0.034)
Observations	690	690
R^2	0.011	0.167
Respondent FE	No	Yes

C.2. Survey 2: Comparing Pitches

In our second and separate survey, we ask MTurker respondents to compare pitch positivity in pairs of randomly-drawn videos. By asking respondents to directly compare pitches, we mitigate noise that could arise from the rating survey in Survey 1 due to the small sample—such as the impact of the order of videos and individual fixed effects in interpreting scales, among others.

In this survey, each respondent is allocated four pairs of videos. For each of these random pairs, we require both videos, clearly labeled as "Video 1" and "Video 2," to be completely watched. Then on the next screen, the respondents are asked to choose the pitch video that gives them the more positive impression (passionate, enthusiastic). Finally, we evaluate the consistency between our ML-based ranking and the human ranking. In other words, does the algorithm pick the same winners as the raters?

We find that the same winner is picked with nearly 89.5% consistency. Interestingly, we also find strong disagreement among MTurker raters themselves when the two videos in the same pair have close algorithm-generated Pitch Factors. In other words, our method seems to be able to provide a more decisive ranking when there are high levels of noise.

Yale University

Video Pitch Experiment Introduction

This survey will take you about 10 minutes. You will get a base payment of $3 as long as you finish this survey. We will also award you bonus payment (up to $3), which is determined by how well you did in the survey.

During the survey, you are going to watch 6 videos where company founders are describing their startup. You will then rate how positive (e.g. passionate, happy, enthusiastic) each video is on four dimensions: **facial expressions, voices, word choices, and overall**.

Please get your audio device (e.g., earphone and computer speaker) ready now.

Note: The submission button will appear only after you watch the video. If the submission button does not appear even after **you watch the video*****, please wait several seconds and do not reload the web page.***

Video Pitch -Kru865yB-M (Example)

Please watch the video. You will then rate how positive (e.g. passionate, happy, enthusiastic) this video is on four dimensions: facial expressions, voices, word choices, and overall.
(The submission button will appear after the video is played.)

Video Pitch -Kru865yB-M Question (Example)

Which of the following industry or industries best describe the business of this startup?

- ○ Consumption Goods
- ○ Health Care
- ○ Information Technology
- ○ Consumer Services
- ○ Industrials

What is your rating for the **overall positivity** of this video?

	1	2	3	4	5	6	7	8	9	
Most negative	○	○	○	○	○	○	○	○	○	Most positive

What is your rating for the **visual positivity** of this video?

	1	2	3	4	5	6	7	8	9	
Most negative	○	○	○	○	○	○	○	○	○	Most positive

What is your rating for the **vocal positivity** of this video?

	1	2	3	4	5	6	7	8	9	
Most negative	○	○	○	○	○	○	○	○	○	Most positive

What is your rating for the **verbal positivity** of this video?

	1	2	3	4	5	6	7	8	9	
Most negative	○	○	○	○	○	○	○	○	○	Most positive

Questions on Basic Information

What is your year of birth? (e.g., 1990)

Choose one or more races that you consider yourself to be:

- ☐ White
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Other

What is your gender?

- ◯ Male
- ◯ Female
- ◯ Other

What is the highest level of school you have completed or the highest degree you have received?

- ◯ Less than High School
- ◯ High School
- ◯ College
- ◯ Graduate or Professional (JD, MD)

Ending

This is the end of the survey. Thank you for your valuable time.

To obtain your payment, please input your unique ID below to MTurk.

Here is your unique ID: ${e://Field/Random%20ID}. Copy this value to paste into MTurk.

When you have copied this ID, please click the Submit button to submit your answers.

Yale University

Video Pitch Experiment Introduction

This survey will take you about 15 minutes. You will get a base payment of $3 as long as you finish this survey. We will also award you bonus payment (up to $3), which is determined by how well you did in the survey.

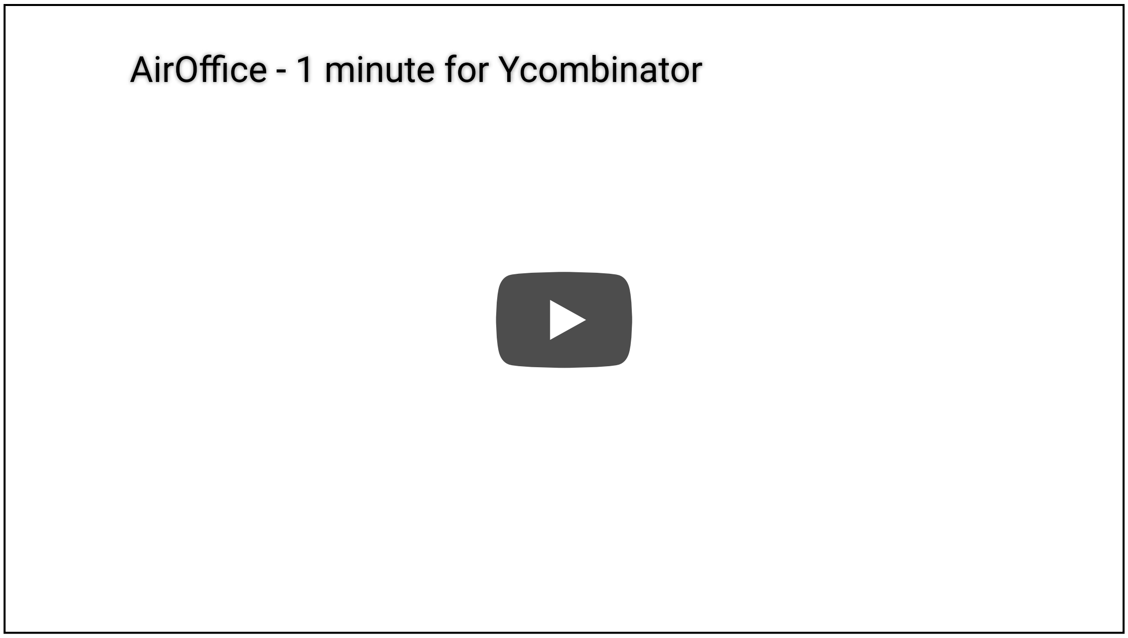
During the survey, you are going to watch 4 pairs of videos where company founders are describing their startup. You will then select which video is more positive (e.g. passionate, happy, enthusiastic) on four dimensions: **facial expressions, voices, word choices, and overall**.

Please get your audio device (e.g., earphone and computer speaker) ready now.

Note: The submission button will appear only after you watch both videos. If the submission button does not appear even after **you watch the video*****, please wait several seconds and do not reload the web page.***

Video Pitch Ie3s6qSV1Ck and n4d1TXm-RUk (Example)

Video 1



Video 2

Please watch both videos. You will then select which video is more positive (e.g. passionate, happy, enthusiastic) on four dimensions: facial expressions, voices, word choices, and overall.
(The submission button will appear after both videos are played.)

Video Pitch le3s6qSV1Ck and n4d1TXm-RUk Question (Example)

Which of the following industry or industries best describe the business of the startup in video 1?

- ○ Consumption Goods
- ○ Health Care
- ○ Information Technology
- ○ Consumer Services
- ○ Industrials

Which of the following industry or industries best describe the business of the startup in video 2?

- ○ Consumption Goods
- ○ Health Care
- ○ Information Technology
- ○ Consumer Services
- ○ Industrials

Which video is more positive in terms of **overall positivity**?

	Video 1	Video 2
Overall Positivity	○	○

Which video is more positive in terms of **visual positivity**?

	Video 1	Video 2
Visual Positivity	○	○

Which video is more positive in terms of **vocal positivity**?

	Video 1	Video 2
Vocal Positivity	○	○

Which video is more positive in terms of **verbal positivity**?

	Video 1	Video 2
Verbal Positivity	○	○

Questions on Basic Information

What is your year of birth? (e.g., 1990)

Choose one or more races that you consider yourself to be:

- ☐ White
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Other

What is your gender?

- ○ Male
- ○ Female
- ○ Other

What is the highest level of school you have completed or the highest degree you have received?

- ○ Less than High School
- ○ High School
- ○ College
- ○ Graduate or Professional (JD, MD)

Ending

This is the end of the survey. Thank you for your valuable time.

To obtain your payment, please input your unique ID below to MTurk.

Here is your unique ID: ${e://Field/Random%20ID}. Copy this value to paste into MTurk.

When you have copied this ID, please click the Submit button to submit your answers.

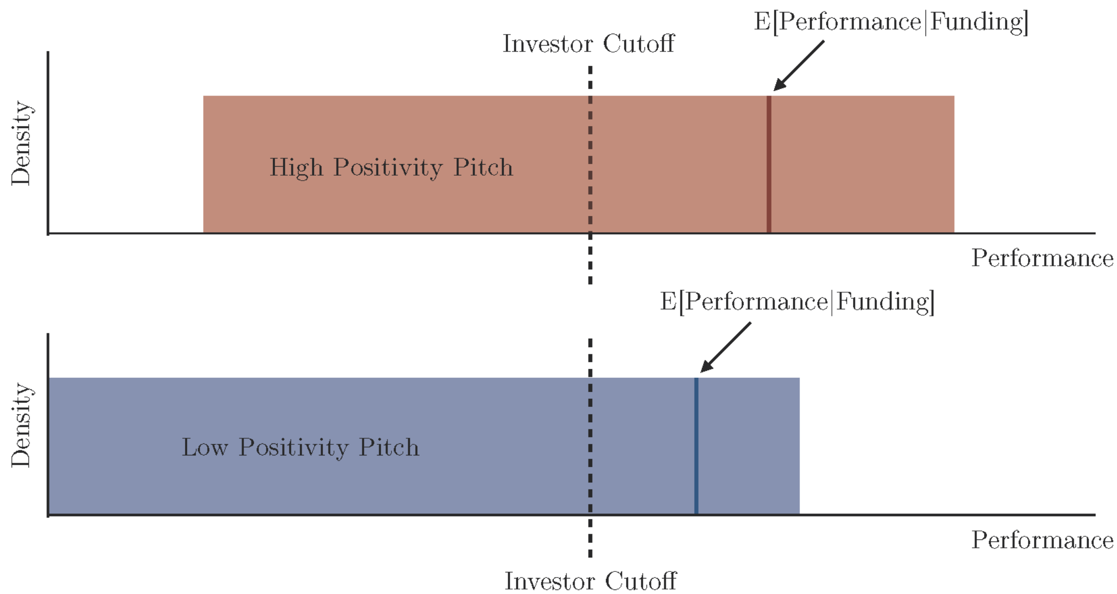
D. Appendix: Performance Analysis and Sources of Bias

This appendix presents a simple conceptual framework, visualized in Figure A.4, to illustrate how pitch deliveries could introduce investment bias that then leads to poorer startup performance. Panel (a), presenting the no-bias scenario, shows hypothetical performance/quality distributions for startups that an investor may be considering funding. Separate overlapping distributions are assumed for startups with high- versus low-positivity pitches. The distributions shown are identical, except that the high-positivity distribution is shifted to the right of the low-positivity distribution. In other words, the high-positivity teams first-order stochastically dominates the low-positivity distribution. We assume the investor funds startups according to a simple cutoff rule, offering funding to all startups above a certain threshold. Since the investor is unbiased, he or she applies the same cutoff rule to all startups, regardless of the pitch positivity. In this case, because the high-positivity distribution first-order stochastically dominates the low-positivity distribution, the investor will invest in startups with high-positivity pitches with greater probability. In addition, expected performance, conditional on funding, will be higher for high-positivity startups.

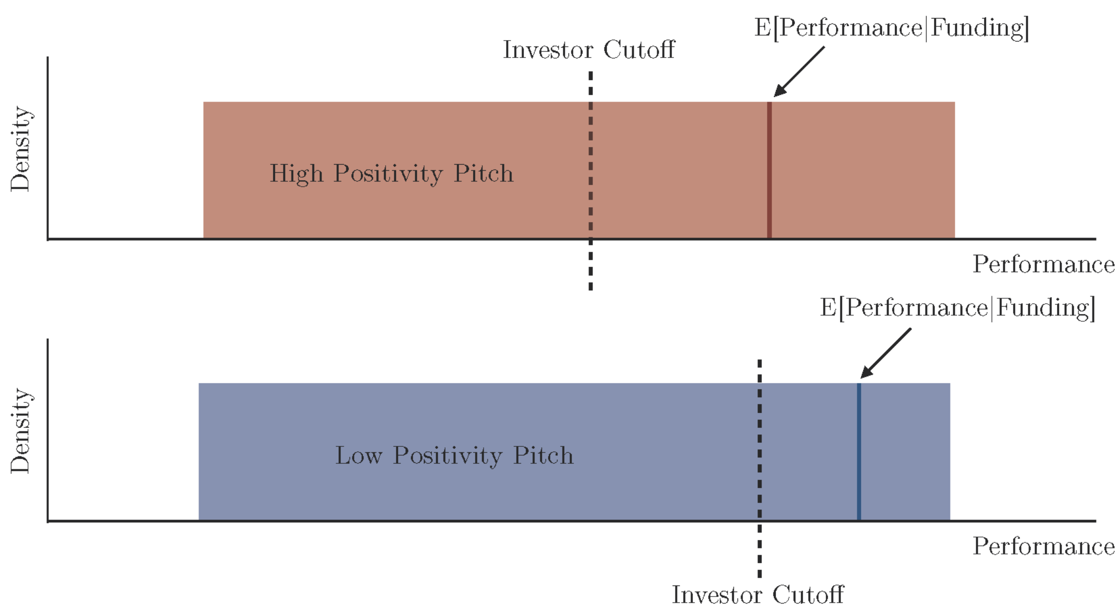
In contrast, if investors are biased, either due to a taste-based channel or inaccurate beliefs, it is possible that high-positivity startups may underperform. Figure A.4 Panel (b) illustrates taste-based bias. In the example, the performance distributions of high- and low-positivity teams are assumed to be the same. The investor continues to derive utility from startup performance. But she or he now also derives disutility from investing in startups with low positivity pitches—as a result, the investor sets a higher cutoff for them. With a taste-based channel, the investor will again fund founders with more positive pitches with greater probability. However, now expected performance, conditional on funding, will be lower for these investments. Figure A.4 Panel (c) illustrates the case of inaccurate beliefs. Inaccurate beliefs imply a gap between the investor's perceived performance distribution for low-positivity (or high-positivity) startups and the true performance distribution. In the example shown, the investor acts exactly like an investor with no bias according to the investor's *perceived* performance distribution. Inaccurate beliefs can also lead investors to fund founders of high-positivity with greater probability while having lower (true) expected performance for those investments.

Figure A.4. Startup Performance Under Different Investment Models

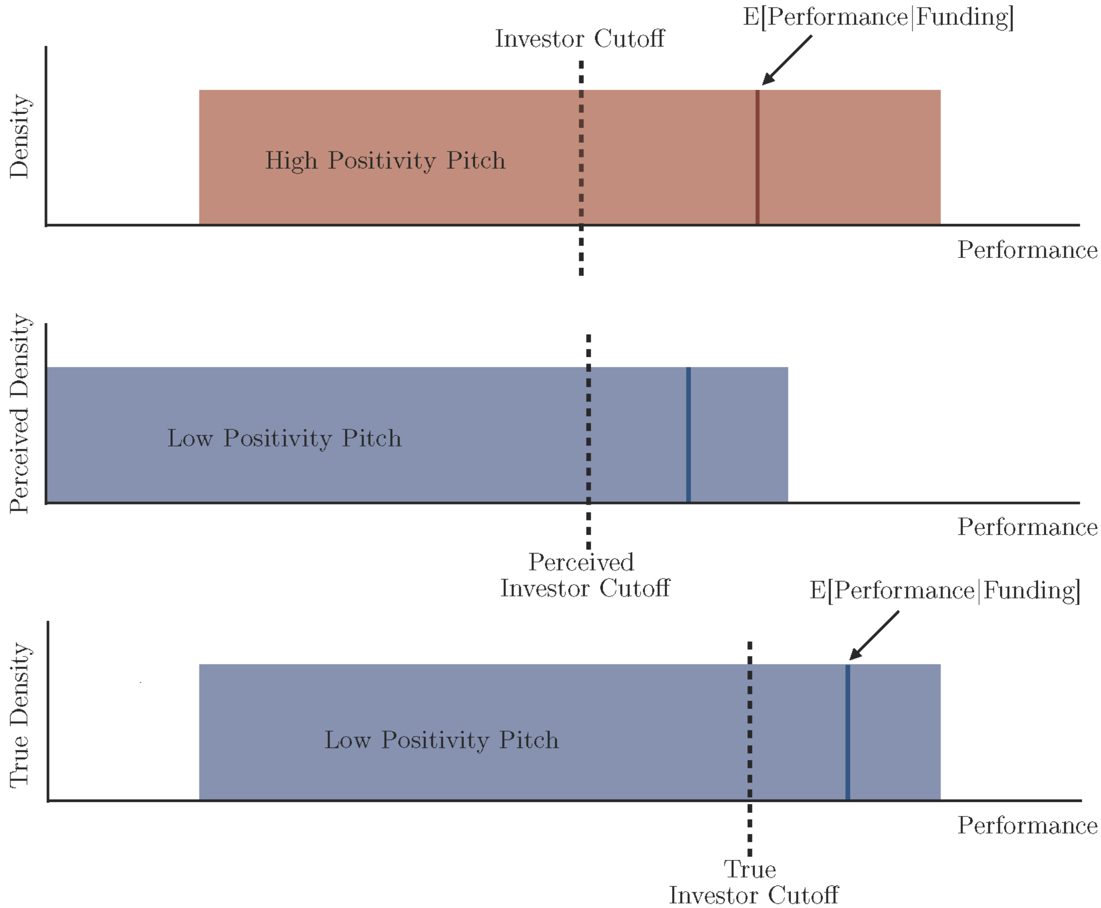
(a) No Bias



(b) Taste-Based Bias



(c) Inaccurate Beliefs



Notes. These figures present hypothetical startup performance distributions combined with investor decision rules. Panel (a) considers the situation where the investors have no bias and startups with low-positivity pitches underperform high-positivity startups. Investors use the same performance cutoff rule (the vertical dashed line) and the solid vertical lines represent the expected performance conditional on the funding decision. Panel (b) considers the situation where investors exhibit taste-based bias and founders of both high- and low-positivity have the same performance distribution. The taste-based bias leads investors to have a higher cutoff rule (the vertical dashed line) for low-positivity startups. This, in turn, leads to higher performance outcomes conditional on funding. Panel (c) presents the situation where investors have inaccurate beliefs about startups with different pitch features. The low-positivity startups' distribution is shifted to the left because of the miscalibration, which has the effect of increasing the expected performance conditional on funding.

E. Appendix Figures and Tables

Table A.4. Features in Pitch Delivery and Investment Decisions: MSFT Azure

Dependent Var: I(*Invested*)	Logit without Controls			Logit with Startup/Team Controls		
	Marginal Effect	S.E.	Pseudo R^2	Marginal Effect	S.E.	Pseudo R^2
Pitch Factor	0.028***	(0.007)	0.191	0.025***	(0.007)	0.251
Visual (Facial)						
Visual-Positive	0.012**	(0.006)	0.176	0.012*	(0.007)	0.239
Visual-Negative	−0.013*	(0.007)	0.176	−0.012	(0.008)	0.253
Visual-Beauty	0.015**	(0.006)	0.178	0.015**	(0.007)	0.242
Vocal (Audio)						
Vocal-Positive	0.009**	(0.005)	0.174	0.011*	(0.006)	0.239
Vocal-Negative	−0.045***	(0.016)	0.183	−0.047***	(0.017)	0.248
Vocal-Arousal	0.023***	(0.009)	0.184	0.019**	(0.008)	0.245
Vocal-Valence	0.023***	(0.006)	0.185	0.020***	(0.007)	0.246
Verbal (Text)						
Verbal-Positive	−0.010	(0.009)	0.174	−0.011	(0.009)	0.239
Verbal-Negative	−0.026***	(0.007)	0.186	−0.022***	(0.008)	0.246
Verbal-Warmth	0.026***	(0.008)	0.190	0.028***	(0.008)	0.256
Verbal-Ability	−0.049***	(0.009)	0.243	−0.043***	(0.007)	0.298

Notes. Logit regressions, marginal effect, $N = 1,139$. The analysis is obtained using the following model:

$$I(Invested) = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon.$$

I(*Invested*) takes a value of one if the startup team was chosen by the accelerator, and zero otherwise. All pitch feature variables are standardized into a zero-mean variable with a standard deviation of one. Visual variables are constructed by Microsoft Azure APIs. Vocal and verbal variables are identical to those in Table 3. Pitch Factor is constructed by the same method as in Table 3. All regressions include Accelerator FE. Control variables include founders' education background (whether they have a Master's or a PhD degree, whether they attended an elite university, defined as the U.S. News & World Report's Top 10), founders' prior work experience (whether they have prior entrepreneurship experience, whether they ever held a senior position in prior employment), team size, and video resolution. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A.5. Features in Pitches and Investment Decisions: Robustness Checks

	(1)	(2)	(3)	(4)
	Dependent Var: I(*Invested*)			
Pitch-Factor	0.028***	0.026**	0.016***	0.027***
	(0.010)	(0.010)	(0.006)	(0.007)
Observations	1,139	1,139	1,139	1,139
Specification	OLS	OLS	Logit	Logit
R^2/Pseudo R^2	0.151	0.181	0.402	0.261
Accelerator FE	Y	Y		Y
Startup/Team Controls		Y	Y	Y
Accelerator-Year FE			Y	
Industry FE				Y

Notes. The analysis is obtained using the following model:

$$I(Invested) = \alpha + \beta \cdot X + \delta_{FE} + \varepsilon.$$

I(*Invested*) takes a value of one if the startup team was chosen by the accelerator, and zero otherwise. All pitch feature variables are standardized into a zero-mean variable with a standard deviation of one. All variables are identical to those in Table 3. Control variables include founders' education background (whether they have a Master's or a PhD degree, whether they attended an elite university, defined as the U.S. News & World Report's Top 10), founders' prior work experience (whether they have prior entrepreneurship experience, whether they ever held a senior position in prior employment), team size, and video resolution. Standard errors clustered at the accelerator-year level are displayed in parentheses. ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table A.6. Gender Breakdown by Industry

	Men-Only	Women-Only	Mixed-Gender
Communication Service	4.83	7.10	4.81
Consumer Discretionary	20.57	21.94	15.19
Consumer Staples	2.50	6.13	2.59
Energy	0.36	0.65	0.00
Financials	5.19	5.16	4.07
Health Care	6.62	8.06	10.00
Industrials	7.69	8.39	9.63
Information Technology	48.12	37.42	50.00
Materials	0.18	0.65	0.00
Real Estate	1.97	0.97	1.48
Unclear	1.97	3.55	2.22
Total Observation	559	310	270
Total %	100.00	100.00	100.00

Notes: This table provides industry (GICS) distributions of collected videos across different team gender compositions.

F. Experiment: Summary Statistics and Sample

Table A.6. Summary Statistics of Subjects in Experiments

	N	Mean	STD	25%	50%	75%
Age	102	28.35	3.31	25.00	28.00	31.00
Man	102	0.60	0.49	0.00	1.00	1.00
Woman	102	0.40	0.49	0.00	0.00	1.00
White	102	0.45	0.50	0.00	0.00	1.00
Black or African American	102	0.03	0.17	0.00	0.00	0.00
Asian	102	0.42	0.50	0.00	0.00	1.00
Hispanic or Latino	102	0.05	0.22	0.00	0.00	0.00
Mixed Race	102	0.03	0.17	0.00	0.00	0.00
Other Race	102	0.02	0.14	0.00	0.00	0.00

Notes. This table provides descriptive statistics of demographic information of subjects in our experiment sample. The demographic information is collected during the experiment. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles.

Table A.7. Summary Statistics of Unstandardized Features in Experiments

	N	Mean	STD	25%	50%	75%
Visual (Facial)						
Visual-Positive	62	0.18	0.17	0.06	0.13	0.30
Visual-Negative	62	0.17	0.18	0.05	0.10	0.24
Visual-Beauty	62	0.59	0.09	0.52	0.60	0.64
Vocal (Audio)						
Vocal-Positive	62	0.08	0.04	0.04	0.07	0.09
Vocal-Negative	62	0.02	0.01	0.01	0.01	0.02
Vocal-Arousal	62	0.35	0.35	0.09	0.23	0.67
Vocal-Valence	62	0.28	0.26	0.08	0.22	0.49
Verbal (Text)						
Verbal-Positive	62	0.02	0.01	0.01	0.01	0.02
Verbal-Negative	62	0.01	0.01	0.00	0.01	0.02
Verbal-Warmth	62	0.02	0.02	0.01	0.02	0.02
Verbal-Ability	62	0.03	0.03	0.01	0.03	0.04

Notes. This table provides descriptive statistics of pitch features. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles. Variables are categorized into vocal, video, and verbal.

Table A.8. Summary Statistics of Video Pitches in Experiments

	N	Mean	STD	25%	50%	75%
Duration (second)	62	61.76	4.88	58.00	61.00	66.00
Video Size (MB)	62	12.79	10.22	4.55	9.10	17.06
Number of Words	62	174.74	39.34	149.00	176.00	199.00
Number of Sentences	62	11.65	3.53	9.00	11.50	13.00
Number of Views	62	2,742.06	14,558.83	65.00	149.50	327.00
Number of Likes	62	3.03	7.53	0.00	0.00	2.00
Number of Dislikes	62	0.24	0.97	0.00	0.00	0.00

Notes. This table provides descriptive statistics of basic information of the pitch videos. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles.

Table A.9. Summary Statistics of Startups in Experiments (as of July 2019)

	N	Mean	STD	25%	50%	75%
Startup Alive	62	0.68	0.47	0.00	1.00	1.00
Firm Age	62	3.44	1.71	2.00	3.00	5.00
Invested by Accelerator	62	0.44	0.50	0.00	0.00	1.00
Raised VC	25	0.60	0.50	0.00	1.00	1.00
Total Funding Amount ($000)	25	16,432	52,817	250	1,500	3,000
Total Funding Rounds	31	2.90	2.30	1.00	2.00	4.00
Number of Employees	30	29.67	72.53	5.00	5.00	30.00

Notes. This table provides descriptive statistics of characteristics of startups all measured as of July 2019 from Crunchbase and PitchBook. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles.

Table A.10. Summary Statistics of Teams in Experiments

	N	Mean	STD	25%	50%	75%
Number of People	62	2.10	1.20	1.00	2.00	3.00
Single-Member	62	0.34	0.48	0.00	0.00	1.00
Multi-Member	62	0.66	0.48	0.00	1.00	1.00
Men-Only	62	0.52	0.50	0.00	1.00	1.00
Women-Only	62	0.32	0.47	0.00	0.00	1.00
Mixed Gender	62	0.16	0.37	0.00	0.00	0.00
Prior Senior Position	62	0.60	0.49	0.00	1.00	1.00
Prior Startup Experience	62	0.42	0.50	0.00	0.00	1.00
Elite University	62	0.10	0.30	0.00	0.00	0.00
Master Degree	62	0.24	0.43	0.00	0.00	0.00
PhD Degree	62	0.10	0.30	0.00	0.00	0.00

Notes. This table provides descriptive statistics of the startup teams. Team member background information is collected from LinkedIn. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles.

Table A.11. Summary Statistics of Beliefs and Investment Decisions in Experiments

	N	Mean	STD	25%	50%	75%
Belief (μ)						
Baseline P(*invested*)	952	0.20	0.17	0.08	0.15	0.29
Baseline P(*alive*|*invested*)	952	0.25	0.15	0.12	0.20	0.32
I(*invested*)	952	0.46	0.50	0.00	0.00	1.00
P(*alive*|*invested*)	952	0.31	0.23	0.14	0.26	0.45
P(*alive*|*not invested*)	952	0.17	0.18	0.05	0.10	0.24
P(*success*|*invested*)	952	0.13	0.18	0.02	0.05	0.17
Precision of Belief (σ)						
Baseline P(*invested*)	952	3.30	0.79	3.00	3.00	4.00
Baseline P(*alive*|*invested*)	952	3.24	0.69	3.00	3.00	4.00
I(*invested*)	952	2.60	0.90	2.00	3.00	3.00
P(*alive*|*invested*)	952	2.74	0.85	2.00	3.00	3.00
P(*alive*|*not invested*)	952	2.74	0.86	2.00	3.00	3.00
P(*success*|*invested*)	952	2.73	0.88	2.00	3.00	3.00

Notes. This table provides descriptive statistics of beliefs and investment decisions elicited in the experiment. For each variable, we report the number of observations, mean, standard deviation, and 25th, 50th, and 75th percentiles.

Yale University

Consent Form

Hi, this is a survey designed by the research team of Song Ma (Assistant Professor of Finance at Yale School of Management). We are conducting research to examine the relation between entrepreneurs' performance in video pitching and their outcomes in obtaining venture investment.

We are inviting you to participate in this study by completing this short survey. This survey will take you around 20 minutes. The results of the survey will be used for research purposes only. All of your responses will be held in confidence.

This survey is also a required assignment of MGT 897 - Entrepreneurial Finance. You will get a base point of 5 as long as you finished this survey. In addition to your base point, we will award you bonus credits. The bonus credit (up to 3 points) is determined by how well you did in the survey (e.g., you choose to invest in an entrepreneur team that later became more successful.)

Would you like to participate in the study?

- ○ Yes
- ○ No

Basic Information Section

What is your Yale NetID?

What is your year of birth? (e.g., 1990)

Which program do you currently entroll at Yale University?

- ◯ Undergraduate
- ◯ Master at Yale SOM (e.g., MBA, EMBA, MAM, and MMS)
- ◯ PhD
- ◯ Other

Which year are you in the current program at Yale University?

- ◯ First year
- ◯ Second year
- ◯ Third year and above

Choose one or more races that you consider yourself to be:

- ☐ White
- ☐ Asian
- ☐ Black or African American
- ☐ Hispanic or Latino
- ☐ Other

What is your gender?

- ◯ Male
- ◯ Female

○ Other

Which of the following categories best describes your previous occupation? (Choose at least one and no more than four)

☐ Student
☐ Asset Management and Banking
☐ Consulting
☐ Education
☐ Energy/Healthcare/Manufacturing
☐ Entrepreneur
☐ Technology
☐ Venture Capital and Private Equity
☐ No Full-time Work Experience
☐ Other

Benchmark Belief Section

On average, **what percentage** of startups do you think can successfully raise Series A financing from VC conditional on trying?

	0	10	20	30	40	50	60	70	80	90	100
Percentage of Obtaining Fundings											

How confident are you with your answers to the question about the probability of obtaining the investment that your were just asked?

○ Extremely confident
○ Very confident
○ Somewhat confident
○ Not very confident
○ Not at all confident

If a startup has already been invested by a venture capital, what do you think is the **average** successful rate of a startup to survive in the following three years?

	0	10	20	30	40	50	60	70	80	90	100
Probability of Surviving											

How confident are you with your answers to the question about the surviving probability that your were just asked?

- Extremely confident
- Very confident
- Somewhat confident
- Not very confident
- Not at all confident

Video Pitch Experiment Introduction

Now, imagine that you are a venture investor. **You are going to decide whether to invest in a given startup after watching its one-minute video pitch.** If you decide to invest in this startup, the contract will be the same – you will invest $150K in this startup team for 7% share of the company.

In the following part of the survey, you are going to watch 10 video pitches and decide whether to invest in these startups.

Note: The submission button for each page will appear only after the video is watched and all questions are answered. If the submission button does not

appear even after all questions are answered, please wait several seconds and do not reload the web page. (Reloading will only reset the your answers.)

Video Pitch IY3hoi1eizM (Example)

These page timer metrics will not be displayed to the recipient.

First Click: *0 seconds*

Last Click: *0 seconds*

Page Submit: *0 seconds*

Click Count: *0 clicks*

Please watch the video. All survey questions are related to this video. (The submission button will appear after the video is played and questions are answered.)

If you were an investor, are you willing to invest $150K in this startup team for 7% share of the company?

- Yes
- No

How confident are you with your answers to the question about the investment decision that your were just asked?

- Extremely confident
- Very confident
- Somewhat confident
- Not very confident
- Not at all confident

If this startup **was able to raise Series A financing from VC**, what is the probability that you think this startup will still be alive (including being acquired) three years later?

0 10 20 30 40 50 60 70 80 90 100

Probability of Surviving

How confident are you with your answers to the question about the surviving probability that your were just asked?

- Extremely confident
- Very confident
- Somewhat confident
- Not very confident
- Not at all confident

If this startup **was not chosen by a VC**, what is the probability that you think this startup will still be alive (including being acquired) three years later?

	0	10	20	30	40	50	60	70	80	90	100
Probability of Surviving											

How confident are you with your answers to the question about the surviving probability that your were just asked?

- Extremely confident
- Very confident
- Somewhat confident
- Not very confident
- Not at all confident

What is the probability that you think this startup will become a huge success (e.g., a Unicorn)?

	0	10	20	30	40	50	60	70	80	90	100
Probability of Huge Success											

How confident are you with your answers to the question about the huge success probability that your were just asked?

- Extremely confident
- Very confident
- Somewhat confident

- Not very confident
- Not at all confident

What are the most important factors in your decision of whether to invest in **this startup**?

	Extremely important	Very important	Somewhat important	Not very important	Not at all important
Team's pitching traits (e.g., facial expression, passionate voice, beauty)	○	○	○	○	○
Team's general ability	○	○	○	○	○
Team's general sociability	○	○	○	○	○
Products, business model, industry, and market	○	○	○	○	○
Team's previous industry experience	○	○	○	○	○
Team's previous entrepreneurial experience	○	○	○	○	○
Team's education background	○	○	○	○	○

Ending

This is the end of the survey. Thanks for your valuable time.

If you have any additional comments about this survey, please provide them below. (Optional)

Powered by Qualtrics