Internet Appendix for
Attracting Early Stage Investors: Evidence from a Randomized Field Experiment

Shai Bernstein, Arthur Korteweg, and Kevin Laws*

This appendix contains additional descriptions, results, and robustness tests to supplement the analyses in the paper. Section A describes AngelList's information disclosure algorithm in detail. Section B shows robustness to alternative ways of clustering standard errors. Section C shows robustness to dropping connected investors. Section D provides further tests of the investor specialization results. Section E describes the activity of inexperienced investors. Section F analyzes the ordering of the information presented in the randomized emails. Section G discusses combinations of information categories. Section H shows regression results using information categories disaggregated to a higher level of detail.

* Shai Bernstein (shaib@stanford.edu) is from Stanford Graduate School of Business, Arthur Korteweg (korteweg@marshall.usc.edu) is from the University of Southern California Marshall School of Business, and Kevin Laws is from AngelList, LLC.

A. Information Disclosure Algorithm

Featured start-ups are chosen by AngelList for their potential appeal to a broad set of investors who have indicated an interest in the industry or the location of the start-up. The featured email sent by AngelList includes a description of the start-up and its product and business idea, the amount of money that the company aims to raise, and how much the company raised so far (see the example email in Figure 1 of the paper). In addition, the email lists up to three categories of information regarding the start-up team's background (which includes information about the founders' academic institutions, and previous employment), current investors (investors who have previously invested in the startup, or incubators from which the startup graduated), and traction (revenue, growth, or active users).

Outside of the experiment, a category is shown if it passes a threshold defined by AngelList. These thresholds are set independently from each other, and are the same across all industries and sectors. AngelList determines the thresholds with the aim of showing only information that investors might be interested in, and therefore, only information in the far right tail of its population distribution is disclosed. Specifically, team information is displayed if a founder's alma mater is in the top 3.5% of all academic institutions in the AngelList population (using an internal ranking of institutions), or if a founder's prior employer is in the top 3.5% of all employers. The investors' information category is shown only if the investor ranking is in the top 5% of all investors on AngelList. The ranking of employers and investors is determined by the signal value calculated for each investor and company, as described in Section IIIC of the paper. To illustrate, Figure A1 shows the distribution of the signal for the investors in the randomized experiment.

The traction category is shown if the company discloses any traction. According to AngelList, companies do not add traction information to their profile if it portrays them in an unflattering light. Moreover, fraud and misreporting of traction is unlikely, particularly for fundraising promising and highly growing start-ups. Therefore, when traction is included it is likely to be very good, and therefore AngelList will report it. Consequently, AngelList assumes that if firms do not report traction, it is not notable. Indeed, only 7.5% of fundraising companies have filled the traction field.

To give a better sense of exactly what kind of information is displayed, Table AI shows the information that passed the disclosure threshold for the 21 start-ups in the randomized experiment of the paper.

B. Standard Error Clustering

In the paper we cluster standard errors by investor, because unobserved investor characteristics could generate correlated responses across emails. In this section we show that the results are robust to other potential sources of correlation in the residuals. First, we consider clustering by treatment (i.e., by unique featured email). This controls for unobserved correlation in responses to the same email by different investors, which is amplified by the absence of variation in the treatment variables within each unique email (Moulton (1986, 1990)). Second, we double-cluster by both treatment and investor, to allow for unobserved correlations in both dimensions (such that only residuals for different emails received by different investors are assumed to be uncorrelated).[1]

[1] We are grateful to Mitchell Petersen for posting Stata code for estimating OLS and logit models with double-clustered standard errors on his website.

The asymptotic properties of the cluster-robust standard errors rely on the number of clusters going to infinity (in both dimensions in two-way clustering). We have thousands of unique investors but only 58 unique emails. With a small number of clusters, standard errors tend to be biased downward (Rothemberg (1988), Kauermann and Carroll (2001), Petersen (2009), Thompson (2011)), so this is a potential problem for both alternative clustering schemes considered here. To avoid the resulting over-rejection problem, we use the paired-sample bootstrap-t method. Cameron et al. (2008) find that this method performs well in samples with few clusters, both in absolute terms as well as relative to other bootstrap methods. We use 5,000 bootstrap samples to approximate the distribution of the pivotal quantities.

Columns 1 and 4 of Table AII replicate the base case ordinary least squares (OLS) and logit specifications, respectively, from Table IV in the paper (clustering standard errors by investor). Columns 2 and 5 show the results when clustering by treatment, and columns 3 and 6 show the regressions that double-cluster by both treatment and investor. For the team and investors information the average bootstrapped standard errors in columns 2, 3, 5, and 6 are slightly higher than the standard errors of the base case specifications in columns 1 and 4, whereas they are slightly lower for the traction information. The p-values for the treatment-clustered or double-clustered models - which are based on the bootstrapped pivotal quantities - are slightly higher than the base case investor-clustered regressions for nearly all variables. Most importantly, however, the statistical significance of the information categories does not change. The team information is still significant at the 5% level across all specifications, and the investor and traction information remains strongly insignificant across all specifications.

C. Dropping Connected Investors

Some investors may have prior knowledge of the start-up, and will therefore be less sensitive to the information provided in the featured emails. In the paper we add controls for the number of prior connections between the start-up and investors, and whether investors already follow the start-up on the AngelList platform. An alternative strategy is to drop the connected investors from the regressions altogether. This has the advantage of robustness to any non-linear effects of connectedness, but has the drawback of reducing the sample size from 8,189 observations to 6,394. Table AIII columns (3) and (6) show that dropping the connected investors strengthens the results, if anything: team becomes significant at the 1% level, and the other information categories remain insignificant. For ease of comparison, the remaining columns in Table AIII replicate the results of Table IV of the paper.

D. Investor Specialization Metric

In the paper we measure the degree of investor specialization in the start-up's market using the cosine similarity between the sectors that an investor has tagged him or herself with and the tags that the start-up uses to describe the sectors in which it operates. This allows us to capture the extent to which an investor focuses specifically on the start-up's market environment. The cosine similarity measure in the paper uses the standard TF-IDF (Term Frequency and Inverse Document Frequency) vectorizer, and is calculated in three steps:

1. Create the Term Frequency (TF) vectors

   We first convert the lists of sector tags provided by the investors and start-ups into TF vectors, one for each investor and start-up. The common length of these vectors is the number of sector tags used on the AngelList platform. Each element in an investor's (start-up's) vector equals one if the investor (start-up) included the corresponding sector in their list, and zero otherwise.

2. Weigh the TF vectors with the Inverse Document Frequency (IDF)

   We weigh the TF vectors by the IDF vector, which is a measure of how rare sector tags are across all investors and start-ups. In other words, sector tags that are rarely used receive higher weight than commonly used tags. For example, suppose an investor uses two sector tags, "Information Technology" and "Personal Safety Wearables". The former tag does not receive much weight, as it is extremely broad and frequently used, while the latter is quite narrow and rare, and receives higher weight. The reason for this weighting scheme is that it is considerably more informative if an investor and a start-up both include a rare tag in their lists than if they both include a common tag.

3. Calculate cosine similarity

   The final step is to compute the cosine similarity measure for each possible pair of start-up and investor. The measure is calculated as the cosine of the angle between the weighted TF vectors of the investor and the start-up. Each combination of investor and start-up thus has its own similarity number.

The similarity measure is highest if an investor and start-up have identical sets of tags, and lowest if they have no tags in common.

A potential concern with the specialization measure is that it may be correlated with investor experience. For example, if experienced investors tend to use more tags (reflecting their higher experience), then they may mechanically appear less specialized. This would change the interpretation of the investor specialization results in the paper.

Though we find a positive association between investor experience and the number of tags used, the correlation is very low. In a regression of the number of tags on several measures of experience, we find that a one standard deviation increase in the number of past investments corresponds to 0.45 more tags in the list. A one standard deviation increase in investor signal is associated with 0.62 more tags. And finally, an increase of one standard deviation in the weighted number of followers is associated with 0.18 more tags in the list. Relative to the average of 15 tags per investor, these magnitudes correspond to increases of 3%, 4%, and 1%, respectively.

The weighing scheme in step 2 also helps to mitigate this problem. Longer vectors are more likely to include rare tags (this is often the reason why they are longer), which can boost the similarity measure. Short vectors, on the other hand, tend to stick to broad market tags and can therefore result in quite low similarity measures.

To explore this concern further, Panel A of Table AIV adds two measures of experience (the natural logarithm of the number of prior investments, and the natural logarithm of the number of followers on the AngelList platform) as controls to the regressions of Table VI in the paper. Columns 1 through 4 replicate the first four columns of Table VI, for ease of comparison. Columns 5 through 8 show that the specialization results in the paper go through unchanged, due to the low correlation between the specialization measure and investor experience.

Panel B checks robustness to alternative similarity metrics. First, we use the cosine similarity without the weighing step 2. Second, we replace the cosine similarity with the negative of the Euclidean distance (we use the negative so that the signs of the coefficients have the same interpretation as for the cosine similarity measures). The results are robust for both measures, with higher similarity corresponding to a higher investor click rate, while the interaction with the information categories is small and insignificant.

E. Activity of Inexperienced Investors

It is possible that some inexperienced investors are on the AngelList platform not to invest, but to observe and learn from the actions of experienced investors. To mitigate this concern in the paper, we dropped investors who had not yet made any introduction requests before the experiment took place. To further address this question, Table AV analyzes the rates of introduction requests by experience, where we define experience in the same way as we do in the paper (i.e., by number of investments, signal, and weighted number of followers). Though the rate of introduction requests by inexperienced investors is lower than for experienced investors, it is clear that both groups are quite active on the platform. Even the least experienced investors are actively requesting introductions.

F. Ordering of Information Categories

In the randomized experiment, the ordering of the information categories in the featured emails is always the same. Conditional on passing the threshold, team is shown first, current investors second, and traction third (as illustrated in Figure 1 in the paper). In this section we consider the importance of this ordering for the interpretation of the results.

The psychology literature has identified two common types of *response order effects*, called *primacy* and *recency* effects. Primacy effects occur when response options are more likely to be chosen when presented at the beginning of a list of options. In contrast, recency effects occur when response options are more likely to be chosen when presented at the end of the list.

While recency effects are most prevalent in vocal surveys, primacy effects are the main concern in visual presentations such as ours (see, for example, Krosnick (1999)). The primacy effect is likely to occur because of "satisficing", in which respondents select the first alternative that is reasonable, without taking the time to consider the full choice set (Krosnick (1991)). In our context, the concern is that investors click on the email once they come across the first reasonable category, ignoring the remaining information presented in the email.

The satisficing effect, however, has been documented to be problematic among respondents with relatively limited cognitive skills (Krosnick (1991), Krosnick and Alwin (1987), Krosnick et al. (1996), McClendon (1986), Narayan and Krosnick (1996)), and when questions are particularly difficult such that respondents became fatigued (Mathews (1927), Payne (1949), Malhotra (2008)). Moreover, primacy effects are less likely to occur when the list of alternatives is short (Schuman and Presser (1981), Krosnick and Alwin (1987), and Sekely and Blakney (1994)). In fact, Eveland and Sekely (2001) argue that "primacy effects are not of a concern in lists of four or fewer, even with very general questions." This empirical evidence suggest that primacy effect is unlikely to occur in our setting. The investors participating in our experiment are highly sophisticated, and were not aware that they were participating in an experiment (rather, they were searching for actual investment opportunities). Moreover, the information provided in our information categories was simple and concise (see Figure 1 in the paper and Table AI below). Finally, and perhaps most importantly, the list in our experiment design is short (having only three information categories).

We can test for primacy effects more directly by taking advantage of the experiment's design. Investors looking at the same start-up are exposed to different information sets. For example, when information about the team category is not shown, then the current investors information becomes the first category to be displayed in the email. If investor clicks are driven by primacy effects, then we expect to see a stronger reaction to the investors information category when it is presented first. We test this hypothesis by simply comparing average click rates based on the presence or absence of the information categories. This works because the information categories are randomly assigned and thus orthogonal, and the results are easier to interpret. To illustrate, Panel A of Table AVI uses average click rates to replicate the regression results in the paper. The first row shows that the average click rate when the team information category is displayed is 17.4%, compared to 15.0% when this information is absent. The 2.3% difference in the click rate is statistically significant at the 1% level, and nearly identical to the regression results in Table IV of the paper. The current investors and traction results, shown in the second and third rows of Panel A, respectively, are also equivalent to the results in the paper.

Panel B of Table AVI shows the impact of investors information on average click rates when the investors information is shown as the second item (line 4, conditioning on team information being shown) versus when it is shown first (line 5, conditioning on team information not being shown). In both cases the effect of investor information is insignificant. If anything, the investor category seems to have a slightly larger effect when it is not first (i.e., in line 4), relative to it appearing first in the email (line 5). This result is inconsistent with the primacy hypothesis that information appearing first in the list triggers a greater click rate response.

The identifying assumption in the above test is that there are no interaction effects between information categories. As reported in section G below, we do not find evidence of any interactions, though this is difficult to rule out completely.

Note that we cannot repeat the same exercise for the traction information category. In order for the traction category to appear first in an email, it is necessary to condition on the absence of both team and investor information. But it is impossible to estimate the traction category effect in this case, because the experiment never includes empty emails (to be consistent with AngelList's email policy outside of the experiment).

We also run a set of simulation exercises to assess the potential effects of primacy and satisficing on the regression results. We simulate featured emails and investor responses in the random experiment for a start-up for which team and investors information pass the disclosure threshold (while traction does not, for simplicity but without loss of generality). Therefore, the three unique versions of emails show: 1) only team information; 2) only investors information; 3) both team and investors information.

To generate investor responses under primacy, suppose that investors read through the disclosed information categories in order, until they reach a "bar". The bar represents the point where investors are satisficed, and they are likely to click on the email (with 80% probability in our simulations)[2], ignoring any information further down in the email. If the bar is not reached after seeing all information, then investors surely will not click.

In the first exercise, which we call scenario A, suppose that both team and investors information separately pass the bar. Hence, investors are likely to click as soon as they see the team information in email versions 1 and 3, and as soon as they see the investors information in email 2. Clicks thus happen with equal probability in every version of the email. As a consequence, the disclosure of information in the randomized experiment is insignificant for all categories, as seen in column A of Table AVII.

In scenario B only the team information passes the bar for investors. Investors are likely to click on emails 1 and 3, as they both show the team information, but not on email 2, which

[2] Making clicks probabilistic upon reaching the bar yields regression results that are more interesting, but the intuition behind the results is exactly the same if clicking occurs with certainty.

only shows the investors information. Column B of Table AVII shows that in this case the team information is significant while the investors information is insignificant. Conversely, in scenario C only the investors information passes the bar, so that email recipients only respond to emails 2 and 3, and only investors information is significant in the regressions (column C). Importantly, despite the primacy effect, the point estimates for team in scenario B and investors in scenario C (and their statistical significance) are the same, even though team is always shown first when disclosed, while investors is sometimes shown first and sometimes second.

The simulations reveal that if primacy is the dominant effect and investors care about only one information category, the regression results do not depend on whether that information is presented first or second. If they care about both information categories, then neither is significant, because investors click on all emails with equal probability. This suggest that the finding in the paper that team is significant is not due to the team information appearing first. Even under primacy it means that investors care about team and not investors. In comparison, if investors instead consider each information category separately and click probabilities rise with the disclosure of each piece of information, then both categories are significant, as shown in the last column of Table AVII.

G. Combinations of Information Categories

In this section we explore whether information categories have different effects on click rates when they are combined with other information categories. As in section F of this appendix, we tackle this question by comparing average click rates based on whether or not information categories are displayed. The results are reported in Table AVIII. The first rows in panels A through C replicate the first three rows in Table AVI, and show that the effect of each information category not conditioned on the appearance or absence of other information is consistent with the regressions in Table IV of the paper.

In the second line of Panel A of Table AVIII, we estimate the effect of showing team information when we condition on the appearance of the current investors information. We find a statistically significant 3.1% increase in the click rate when team is shown, which is roughly the same increase as when we do not condition on the appearance or absence of investor information in the first row. The results are similar when we condition on the appearance of traction information instead, or on both the current investors and traction information categories (rows 3 and 4, respectively). In Panel B we show that the effect of the current investors information also does not change when combined with other information, and similarly, Panel C reports that the impact of the traction information remains the same whether or not it is shown in combination with the other information categories.

One caveat to this analysis is that we have fairly limited cross section variation because only 21 start-ups participated in the experiment. Therefore, we cannot rule out that the lack of interaction effects between the information categories is due to lack of power.

H. Detailed Information Categories

Table AI shows that the featured emails contain information at a more disaggregated level than the team, investors, and traction categories used in the paper. Defining information categories at a more refined level may provide insight into the channel through which they operate. For team we consider the disclosure of information about the educational background (*Education*) and prior work experience of the founders (*Work*), and whether the founder(s) have experience in the same industry as the start-up (*Industry*). For the investors information we look at whether existing investors are individuals (*Individual*) or venture capitalists (*VC*), and whether the start-up had previously gone through an incubator or accelerator program (*Incubated*). We disaggregate traction information into information about revenues (*Revenue*), customer base (*Customer*), and growth (*Growth*). All of the above variables equal one if the information is disclosed in an email, and zero otherwise.

Panel A of Table AIX contains descriptive statistics of the detailed information categories. The least frequently disclosed information is on incubation, which is present in 24 percent of start-ups and disclosed in 18 percent of unique emails. The most frequently occurring information is on founders' prior work experience and customer base, which shows up in 67 percent of start-ups and is disclosed in 48 percent of unique emails.

Panel B shows OLS regressions of investor clicks on the detailed information categories. The logit results are qualitatively the same and are not reported. The first column replicates the results in Table IV of the paper. Columns 2 through 5 use the detailed team information. Only founders' education shows up weakly statistically significant at the 10% level, while work history and same-industry experience are insignificant. Columns 6 through 9 use the detailed investor information, and columns 10 through 13 use the breakdown of traction. None of the refined investor or traction information is statistically significant.

Generally, the issue with the detailed information categories is that the tests are weak, as indicated by the larger standard errors compared to the broad categories used in the paper. The lack of power is due to there being only 21 start-ups in the experiment, combined with not having any randomization on the more detailed information within each of the three broad categories. This limits the degree of useful cross-sectional variation in the detailed information categories. The evidence provided in this section should therefore be interpreted as suggestive only.

References

Cameron, A. Colin, Jonah B Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90: 414-427.

Eveland, Vicki Blakney, and William S. Sekely. 2001. "Effect of Response Position and Number of Responses on Response Selection." *Business Quest*, http://www.westga.edu/~bquest/2001/response.htm, retrieved May 25, 2015.

Kauermann, Göran, and Raymond J. Carroll. 2001. "A Note on the Efficiency of Sandwich Covariance Matrix Estimation." *Journal of the American Statistical Association* 96: 1387-1396.

Krosnick, Jon A. 1991. "Response Strategies for Coping with the Cognitive Demands of Attitude Measures in Surveys." *Applied Cognitive Psychology* 5: 213-236.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537-567.

Krosnick, Jon A., and Duane F. Alwin. 1987. "An Evaluation of a Cognitive Theory of Response-Order Effects in Survey Measurement." *Public Opinion Quarterly* 51: 201-219.

Krosnick, Jon A., Sowmya Narayan, and Wendy R. Smith. 1996. "Satisficing in Surveys: Initial Evidence." *New Directions for Evaluation* 70: 29-44.

Malhotra, Neil. 2008. "Completion Time and Response Order Effects in Web Surveys." *Public Opinion Quarterly* 72: 914-934.

Matthews, C. O. 1927. “The Effect of Position of Printed Response Words upon Children's Answers to Questions in Two-response Types of Tests.” *Journal of Educational Psychology* 18: 445.

McClendon, McKee J. 1986. “Response-order Effects for Dichotomous Questions.” *Social Science Quarterly* 67: 205-211.

Moulton, Brent R. 1986. “Random Group Effects and the Precision of Regression Estimates.” *Journal of Econometrics* 32: 385-397.

Moulton, Brent R. 1990. “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units.” *Review of Economics and Statistics* 72: 334-338.

Narayan, Sowmya, and Jon A. Krosnick. 1996. “Education Moderates Some Response Effects in Attitude Measurement.” *Public Opinion Quarterly* 60: 58-88.

Payne, Stanley L. 1949. “Case Study in Question Complexity.” *Public Opinion Quarterly* 13: 653-658.

Petersen, Mitchell A. 2009. “Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches.” *Review of Financial Studies* 22: 435-480.

Rothemberg, Thomas J. 1988. “Approximate Power Functions for Some Robust Tests of Regression Coefficients.” *Econometrica* 56: 997-1019.

Schuman, Howard, and Stanley Presser. 1981. “Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.” *New York: Academic Press*.

Sekely, William S., and Vicki L. Blakney. 1994. “The Effect of Response Position on Trade Magazine Readership and Usage.” *Journal of Advertising Research* 34: 53-60.

Thompson, Samuel B. 2011. “Simple Formulas for Standard Errors that Cluster by both Firm and Time.” *Journal of Financial Economics* 99: 1-10.

Table AI: Information Disclosed in Featured Emails

For each start-up in the randomized experiment, we show the information that passed AngelList's disclosure threshold. This information would be shown in the featured emails outside of the experiment. To protect the companies' identities each column is shown in a different order, so that rows do not correspond to companies. The team, investors, and traction information passed AngelList's disclosure threshold for 19, 17, and 18 start-ups, respectively.

Team information	Investor information	Traction information
Team worked at Microsoft, Google, and Ask.com.	Great Oaks and Josh Abramowitz are investing in this round.	$125K revenue in first 5 months, 15 companies, 1.7K testers.
Team worked at Starbucks and Nabisco.	500 Startups is investing. Incubated by Startmate.	$30M in transaction volume.
Team worked at Royal Bank of Canada and went to University of Toronto.	Summit Partners are investing in this round.	3.2K health providers, 15% monthly growth.
Team worked at IBM and went to the University of Waterloo.	Hadi Partovi, Keith Rabois, and Tony Hsieh are investing in this round.	$1M revenue/year, 70K users, 20% monthly growth.
Team worked at Accel. Went to Cambridge and Oxford.	Quest Venture Partners are investing in this round. Incubated by Y Combinator.	$800K revenue/year, 60% annual growth, 1K customers.
Team went to Stanford and Berkeley.	Laurent Drion is investing $500K in this round.	$250K in pre-sales, 960 pre-orders.
Team worked at Microsoft, Groupon, and went to Stanford GSB.	Grishin Robotics is investing $500K in this round.	40 vending machines, 3 pilot contracts.
Team members worked at Accenture.	Lightbank is investing in this round.	350 subscribing businesses, 700 active users, 125% monthly growth.
Team worked at E*TRADE and studied at Stanford.	Adventure Capital is investing in this round.	$10K revenue/month, 120K users, 7.5K courses.
Team includes the founder of SIMMS - radiology software used by 2 million patients.	Incubated by AngelPad.	$20K revenue/month, 25% monthly growth, 12K monthly active users.
Team worked at Intel and went to The University of Chicago.	SoftTech VC and Matt Mullenweg are investing in this round.	$1K revenue, 12 customers.
Team worked at JPMorgan	Jeff Fluhr and Great Oaks are	130K users, 10% monthly

and went to MIT.	investing in this round.	growth, $6K revenue/month.
Team worked at Yahoo!, Oracle and went to Stanford.	Dave McClure is investing in this round.	80 users. Waiting list includes BHP Billiton, USGS and the WWF.
Team members went to Harvard.	Sandbox Industries are investing.	90K items for sale, 10K monthly active users, 30% monthly growth.
Team worked at Google and went to The University of Cambridge. Includes 2 Artificial Intelligence PhDs.	Golden Gate Ventures is investing in this round.	$20K revenue/month, 10K engineers.
Team worked at Microsoft, GE and went to Cornell.	Patrick Condon (co-founder of Rackspace) investing. Incubated by TechStars.	$1.4M revenue/year, 10K units sold.
Team founded well.ca ($40M/year revenue), worked at IBM and RIM.	Boris Wertz is investing in this round. Incubated by Y Combinator.	$10M revenue/year, 60% annual growth, 13 stores, 25% profit margin.
Team went to University of British Columbia.		$70K revenue/month, 500K monthly active users, 100K daily active users.
The founder's last company designed and built 12 composting facilities in the US.		

Table AII: Alternative Standard Error Clustering Methods

This table shows the regressions of Table IV of the paper under various error clustering specifications. The dependent variable equals one if an investor clicked on the "View" button in the featured email, and zero otherwise. The explanatory variables are indicators that equal one when information about the start-up's team, investors, or traction is disclosed in the email. Columns 1 and 4 are ordinary least squares (OLS) and logit regressions, respectively, with error clustering by investor, replicating Table IV of the paper. Columns 2 and 5 cluster by unique featured email (the treatment in the randomized experiment), and columns 3 and 6 double-cluster by both investor and treatment. The statistics in columns 2, 3, 5, and 6 were computed from the paired-sample t-bootstrap, using 5,000 bootstrap samples due to the low number of treatment clusters. *Controls* are *Connections, Prior follow=1,* and *Prior emails,* as defined in Table IV in the paper, but not reported here for brevity. Standard errors are in round brackets. For the specifications that use bootstrapping, the average standard errors across bootstrap samples is shown. The p-values for the hypothesis tests that the regression coefficients equal zero are in square brackets. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Model	OLS	OLS	OLS	Logit	Logit	Logit
Team = 1	0.023**	0.023**	0.023**	0.172**	0.172**	0.172**
	(0.010)	(0.012)	(0.011)	(0.074)	(0.083)	(0.083)
	[0.020]	[0.039]	[0.024]	[0.019]	[0.026]	[0.026]
Investors = 1	0.009	0.009	0.009	0.067	0.067	0.067
	(0.013)	(0.014)	(0.014)	(0.097)	(0.108)	(0.107)
	[0.472]	[0.570]	[0.468]	[0.491]	[0.492]	[0.492]
Traction = 1	0.017	0.017	0.017	0.123	0.123	0.123
	(0.014)	(0.011)	(0.013)	(0.106)	(0.098)	(0.096)
	[0.244]	[0.232]	[0.260]	[0.246]	[0.264]	[0.263]
Controls	Y	Y	Y	Y	Y	Y
Start-up fixed effects	Y	Y	Y	Y	Y	Y
Clusters	Investor	Treatment	Investor x Treatment	Investor	Treatment	Investor x Treatment
# clusters	2,925	58	2,925 x 58	2,621	58	2,925 x 58
Number of observations	8,189	8,189	8,189	8,189	8,189	8,189

Table AIII: Dropping Connected Investors

This table reports regression results of investor responses to the featured emails in the randomized field experiment. The dependent variable is one when an investor clicked on the "View" button in the featured email, and zero otherwise. Only opened emails are included in the sample. *Team = 1* is an indicator variable that equals one if the team information is shown in the email, and zero otherwise. Similarly, *Investors = 1* and *Traction = 1* are indicator variables for the current investors, and traction information, respectively. *Connections* counts the number of people on the start-up's profile (in any role) that the investor already follows prior to receiving the email. *Prior follow = 1* is an indicator variable that equals one if the investor was already following the start-up on AngelList prior to receiving the featured email. *Prior emails* is the number of emails that the investor has received in the experiment prior to the present email. Columns 1, 2, 5, and 6 replicate Table IV of the paper. The regressions in columns (3) and (6) drop investors with Connections > 0 or Prior follow = 1. R2 is the adjusted R^2 for OLS regressions, and pseudo R^2 for logit models. Standard errors are in parentheses, and are clustered at the investor level. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

Model	(1) OLS	(2) OLS	(3) OLS	(4) Logit	(5) Logit	(6) Logit
Team = 1	0.022**	0.023**	0.028***	0.162**	0.172**	0.222***
	(0.010)	(0.010)	(0.011)	(0.073)	(0.074)	(0.085)
Investors = 1	0.010	0.009	0.016	0.070	0.067	0.128
	(0.013)	(0.013)	(0.014)	(0.097)	(0.097)	(0.108)
Traction = 1	0.016	0.017	0.018	0.122	0.123	0.143
	(0.014)	(0.014)	(0.015)	(0.106)	(0.106)	(0.131)
Connections		0.010			0.064*	
		(0.006)			(0.038)	
Prior follow = 1		0.143***			0.835***	
		(0.033)			(0.166)	
Prior emails		0.001			0.006	
		(0.003)			(0.022)	
Start-up fixed effects	Y	Y	Y	Y	Y	Y
Number of observations	8,189	8,189	6,394	8,189	8,189	6,394
R2	0.001	0.005	0.001	0.028	0.033	0.021

Table AIV: Robustness of Investor Specialization Results

This table reports regression results with experience controls and alternative similarity metrics. The dependent variable is one when an investor clicked on the "View" button in the featured email. In Panel A, *Sector similarity = 1* is an indicator that equals one if the cosine similarity between investor sector interests and the start-up's sectors is in the top 25%. *Log(#investments)* is the log of the number of past investments by the investor, and *Log(#followers)* is the log of an investor's number of followers. Other variables and controls are as defined in Table AII. Panel B uses the cosine similarity without weighing the tags, and the negative of the Euclidean distance as alternative metrics. R2 is the adjusted R^2. Standard errors are in parentheses, and clustered at the investor level. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

Panel A: Experience Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	OLS	OLS	OLS	OLS	OLS
Team shown = 1	0.023**	0.024**	0.024**	0.023**	0.023**	0.024**	0.025**	0.023**
	(0.010)	(0.010)	(0.011)	(0.011)	(0.010)	(0.010)	(0.011)	(0.011)
Investors shown = 1	0.009	0.009	0.009	0.003	0.011	0.011	0.011	0.004
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.013)	(0.014)
Traction shown = 1	0.017	0.016	0.016	0.011	0.016	0.016	0.016	0.011
	(0.014)	(0.014)	(0.014)	(0.016)	(0.014)	(0.014)	(0.014)	(0.016)
Sector similarity = 1		0.039***	0.041***	0.006		0.037***	0.039***	0.001
		(0.011)	(0.015)	(0.035)		(0.011)	(0.015)	(0.032)
Sector similarity = 1			-0.002	0.004			-0.004	0.002
x Team shown = 1			(0.019)	(0.020)			(0.019)	(0.020)
Sector similarity = 1				0.026				0.026
x Investors shown = 1				(0.023)				(0.023)
Sector similarity = 1				0.017				0.019
x Traction shown = 1				(0.025)				(0.025)
Log(# investments)					0.002	0.002	0.002	0.002
					(0.007)	(0.007)	(0.007)	(0.007)
Log(# followers)					-0.017***	-0.016***	-0.016***	-0.016***
					(0.005)	(0.005)	(0.005)	(0.005)
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Start-up fixed effects	Y	Y	Y	Y	Y	Y	Y	Y
Number of observations	8,189	8,189	8,189	8189	8,189	8,189	8,189	8189
R2	0.005	0.007	0.007	0.007	0.033	0.035	0.035	0.035

Panel B: Alternative Similarity Metrics

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
Sector similarity metric	Unweighted cosine similarity	Unweighted cosine similarity	Unweighted cosine similarity	Negative Euclidean distance	Negative Euclidean distance	Negative Euclidean distance
Team shown = 1	0.024**	0.022**	0.021*	0.024**	0.022**	0.022*
	(0.010)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)
Investors shown = 1	0.010	0.010	0.002	0.009	0.009	0.006
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Traction shown = 1	0.016	0.016	0.014	0.017	0.016	0.021
	(0.014)	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)
Sector similarity = 1	0.038***	0.035**	-0.001	0.045***	0.043***	0.056
	(0.011)	(0.016)	(0.036)	(0.012)	(0.016)	(0.036)
Sector similarity = 1		0.005	0.011		0.005	0.009
x Team shown = 1		(0.020)	(0.021)		(0.020)	(0.021)
Sector similarity = 1			0.037			0.015
x Investors shown = 1			(0.026)			(0.025)
Sector similarity = 1			0.009			-0.035
x Traction shown = 1			(0.026)			(0.027)
Start-up fixed effects	Y	Y	Y	Y	Y	Y
Other controls	Y	Y	Y	Y	Y	Y
Number of observations	8,189	8,189	8,189	8189	8,189	8,189
R2	0.007	0.007	0.007	0.008	0.008	0.008

Table AV: Number of Introduction Requests by Investor Experience

This table shows the mean and median number of introduction requests per year for the investors in our experiment, separated by their experience as measured by the prior number of investments (*# Investments*), the quality signal as described in the paper (*Signal*), and the weighted number of followers (*Wtd # followers*). As in the paper, we use different cutoffs for each experience measure based on percentiles of the measure's distribution (*Cutoff*). Columns 3 and 4 show the mean number of introduction requests scaled by the number of years an investor has been on AngelList, for investors below and above the cutoff, respectively. The fifth column shows the p-value of the difference in means test. The final three columns show the median number of yearly introduction requests by experience for investors below and above the cutoff, and the p-value of the non-parametric Wilcoxon-Mann-Whitney rank-sum test.

Measure	Cutoff	Mean # intros/year			Median # intros/year		
		<= cutoff	> cutoff	p-value	<= cutoff	> cutoff	p-value
# Investments	Zero	4.4	5.8	0.100	1.6	1.8	0.000
# Investments	25%	4.3	6.0	0.017	1.5	1.9	0.000
# Investments	50%	4.4	6.9	0.000	1.4	2.5	0.000
Signal	25%	4.2	5.9	0.022	1.6	1.8	0.032
Signal	50%	4.5	6.5	0.001	1.5	2.2	0.000
Signal	75%	4.7	8.0	0.000	1.5	2.9	0.000
Wtd # followers	25%	3.6	6.2	0.000	1.6	1.9	0.132
Wtd # followers	50%	4.2	6.9	0.000	1.5	2.2	0.000
Wtd # followers	75%	4.8	7.6	0.000	1.5	3.1	0.000

Table AVI: Ordering Effect for Investors Information Category

This table reports average click rates conditional on the display of various information categories. Row 1 compares the average click rate when Team information is shown in the email (column 1) to the average click rate when the Team information is excluded (column 2). Column 3 reports the difference between the average click rates. Rows 2 and 3 are constructed analogously. In Panel B, row 4 reports the average click rates when investors information is shown (column 1) or excluded (column 2), but conditioning on the team information being shown in both cases. Conversely, row 5 columns 1 and 2 report average click rates with and without investors information, respectively, but conditioning on team information not being displayed in the email. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

		Included	Excluded	Difference
Panel A		(1)	(2)	(3)
(1)	Team	0.174	0.150	0.023***
(2)	Investors	0.164	0.163	0.001
(3)	Traction	0.165	0.161	0.004
Panel B				
(4)	Investors, cond. on team	0.180	0.165	0.014
(5)	Investors, cond. on no team	0.149	0.157	-0.008

Table AVII: Simulations of Information Category Order

This table shows regression results from simulated featured emails and investor responses. In the simulated featured emails, only team and investors information categories passed the disclosure threshold. Therefore, the three unique versions of emails sent to investors show: 1) only team information; 2) only investors information; 3) both team and investors information. A total of 5,000 investors receive each version of the emails, for a total of 15,000 observations. To generate investor responses under primacy, investors read through the disclosed information categories in order, until they reach a "bar". The bar represents the point where investors are "satisficed", and they are likely to click on the email (with 80% probability), ignoring any information further down in the email. If the bar is not reached after seeing all information, then investors surely will not click. The first set of simulations (column A) assumes that both team and investors information pass the bar. The second scenario (column B) assumes only the team information passes the bar, and in scenario C only the investors information passes the bar. The dependent variable is one when an investor clicked on the "View" button in the featured email, and zero otherwise. *Team = 1*, *Investors = 1*, and *Traction = 1* are indicator variables that equal one if the team, current investors, or traction information, respectively, are shown in the email. Standard errors are in parentheses, and are clustered at the investor level. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

	Primacy scenarios			All information considered
	A	B	C	
Team = 1	0.005	0.805***	0.005	0.404***
	(0.008)	(0.007)	(0.007)	(0.009)
Investors = 1	0.008	0.008	0.805***	0.410***
	(0.008)	(0.007)	(0.007)	(0.009)
Intercept	0.793***	-0.008	-0.005	-0.009
	(0.010)	(0.008)	(0.008)	(0.011)
Number of observations	15,000	15,000	15,000	15,000

Table AVIII: Combinations of Information Categories

This table reports the average click rate conditional on the display of various information categories. Row 1 compares the average click rate when Team information is shown in the email (column 1) to the average click rate when the Team information is excluded (column 2). Column 3 reports the difference between the average click rates. Row 2 reports the average click rate conditional on the appearance of the team and investors information category (column 1) and the average click rate when information about the team is excluded, but investors information is shown (column 2). The other rows are constructed analogously. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

Dependent variable: Click rates	Included	Excluded	Difference
Panel A	(1)	(2)	(3)
(1) Team	0.174	0.150	0.023***
(2) Team, cond. on investors	0.180	0.149	0.031***
(3) Team, cond. on traction	0.175	0.152	0.023***
(4) Team, cond. on investors and traction	0.186	0.150	0.036***
Panel B			
(5) Investors	0.164	0.163	0.001
(6) Investors, cond. on team	0.180	0.165	0.014
(7) Investors, cond. on traction	0.166	0.163	0.003
(8) Investors, cond. on team and traction	0.176	0.165	0.011
Panel C			
(9) Traction	0.165	0.161	0.004
(10) Traction, cond. on team	0.175	0.170	0.004
(11) Traction, cond. on investors	0.166	0.161	0.005
(12) Traction, cond. on team and investors	0.176	0.170	0.006

Table AIX: Detailed Information Categories

Panel A shows descriptive statistics of the detailed information categories. The indicator variables, *Team*, *Investors*, and *Traction* are the information categories used in the paper. The team information is decomposed into information about the educational background of the founders (*Education*), their prior work experience (*Work*), and whether the founder(s) have experience in the same industry as the start-up (*Industry*). The detailed investor information consists of indicators whether the start-up had previously gone through an incubator or accelerator program (*Incubated*), whether existing investors are venture capitalists (*VC*) or individuals (*Individual*). Traction information is disaggregated into information about revenues (*Revenue*), customer base (*Customer*), and growth (*Growth*). All variables equal one if the information is disclosed in the email, and zero otherwise. Panel B shows OLS regressions of the click indicator variable (which equals one when an investor clicked on the "View" button in the featured email, and zero otherwise) on the information variables. Only opened emails by investors without prior connections to the start-up are included in the sample. R2 is the adjusted R^2. Standard errors are in parentheses, and are clustered at the investor level. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level, respectively.

Panel A: Descriptive Statistics

Information	Passed AL threshold (% of start-ups)	Shown in experiment (% of unique emails)
Team information (*Team = 1*)	90.48	73.24
Education (*Education = 1*)	61.90	43.66
Prior work experience (*Work = 1*)	66.67	47.89
Experience in start-up's industry (*Industry = 1*)	33.33	23.94
Investor information (*Investors = 1*)	80.95	73.02
Incubated (*Incubated = 1*)	23.81	18.31
VC investor (*VC = 1*)	52.38	43.66
Individual investor (*Individual = 1*)	38.10	28.16
Traction information (*Traction = 1*)	85.71	72.06
Revenues (*Revenue = 1*)	57.14	43.66
Customer base (*Customer = 1*)	66.67	47.89
Growth (*Growth = 1*)	38.10	29.58

Panel B: Ordinary Least Squares Regressions (Dependent Variable = Click)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Team information:													
Team = 1	0.023**					0.021**	0.021**	0.022**	0.021**	0.022**	0.022***	0.021**	0.023***
	(0.010)					(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Education = 1		0.020*			0.028*								
		(0.011)			(0.015)								
Work = 1			0.009		-0.023								
			(0.012)		(0.019)								
Industry = 1				0.018	0.024								
				(0.014)	(0.019)								
Investors information:													
Investors = 1	0.009	0.006	0.002	0.002	0.005					0.008	0.008	0.006	0.008
	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)					(0.012)	(0.013)	(0.012)	(0.013)
Incubated = 1						-0.001			-0.007				
						(0.017)			(0.020)				
VC = 1							0.000		0.002				
							(0.015)		(0.016)				
Individual = 1								0.015	0.019				
								(0.026)	(0.029)				
Traction information:													
Traction = 1	0.017	0.011	0.010	0.012	0.012	0.014	0.014	0.014	0.013				
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)				
Revenue = 1										0.023			0.014
										(0.021)			(0.031)
Customer = 1											0.012		0.005
											(0.014)		(0.018)
Growth = 1												0.026	0.009
												(0.031)	(0.041)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Start-up FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189	8,189
R2	0.005	0.005	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005

Figure A1: Distribution of Investor Signal Measure

This figure shows the histogram of the investor signal for the 2,925 active investors who received emails about featured start-ups in the randomized field experiment, and opened at least one such email. The signal ranges from zero to ten. See the paper for a description of the algorithm used to compute the signal.