informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–21
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Harnessing the Wisdom of Crowds

Zhi Da,[a] Xing Huang[b]

[a] Mendoza College of Business, University of Notre Dame, Notre Dame, Indiana 46556; [b] Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130

Contact: zda@nd.edu (ZD); xing.huang@wustl.edu, http://orcid.org/0000-0002-3911-5820 (XH)

Received: March 23, 2018
Revised: November 1, 2018
Accepted: November 7, 2018
Published Online in Articles in Advance: August 1, 2019

https://doi.org/10.1287/mnsc.2019.3294

Abstract. When will a large group provide an accurate answer to a question involving quantity estimation? We empirically examine this question on a crowd-based corporate earnings forecast platform (Estimize.com). By tracking user activities, we monitor the amount of public information a user views before making an earnings forecast. We find that the more public information users view, the less weight they put on their own private information. Although this improves the accuracy of individual forecasts, it reduces the accuracy of the group consensus forecast because useful private information is prevented from entering the consensus. To address endogeneity concerns related to a user's information acquisition choice, we collaborate with Estimize.com to run experiments that restrict the information available to randomly selected stocks and users. The experiments confirm that "independent" forecasts result in a more accurate consensus. Estimize.com was convinced to switch to a "blind" platform from November 2015 on. The findings suggest that the wisdom of crowds can be better harnessed by encouraging independent voices from among group members and that more public information disclosure may not always improve group decision making.

History: Accepted by Renee Adams, finance.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2019.3294.

Keywords: wisdom of crowds • herding • naïve learning • social learning • group decision making • earnings forecast

> The more influence we exert on each other, the more likely it is that we will believe the same things and make the same mistakes. That means it's possible that we could become individually smarter but collectively dumber.
> —James Surowiecki, *The Wisdom of Crowds*, p. 42

1. Introduction

Many important decisions are made in a group setting. Consider jury verdicts, the setting of interest rates by the Federal Open Market Committee, and the appointment of a CEO by a board of directors. Consequently, a crucial topic in social science is how to best elicit and aggregate information from individuals. A great deal of evidence suggests that, under certain conditions, the simple average of a large group's answers to a question involving quantity estimation is generally as good as, and often better than, the answer provided by any individual in that group.[1] This phenomenon is commonly referred to as the "wisdom of crowds." As long as individual estimates are unbiased and independent, the law of large numbers implies that the group consensus is very accurate.

In most social and economic settings, however, individual estimates are unlikely to be independent. The reason is that they are often issued sequentially, and individuals learn from observing other people's actions and beliefs, especially those of influential people. Does such herding behavior result in a better or poorer group consensus?

The answer is not always clear. On the one hand, there are several reasons why herding in a sequential setting may generate a more accurate group consensus. The sequential process may encourage discussion and the production of additional information. In addition, if a group member with a very precise signal speaks early, then subsequent herding by other members could improve the accuracy of the group consensus. On the other hand, herding results in correlated rather than independent estimates. In the extreme case of an information cascade, subsequent group members' private information is completely disregarded, so the group consensus is no more accurate than the estimate that started the cascade.[2]

We empirically examine the net impact of herding on the wisdom of crowds. We use a specific setting in which individuals make corporate earnings forecasts. Forecasting earnings is a suitable setting as both earnings forecasts and realizations are easily observable, and the forecast error can be clearly defined. Accurate earnings forecasts are moreover of crucial importance to investors and firms and to the functioning of the financial market in general. Not surprisingly, a wide range of market participants provide earnings forecasts. They include equity analysts from both the sell side and the buy side and, more recently, independent analysts.

We first study the impact of herding on the accuracy of consensus earnings forecasts in a simple theoretical setting in the online appendix.[3] Quantifying such an impact empirically is usually challenging as researchers are generally unable to observe the counterfactual, in which analysts make their forecasts independently. We tackle this challenge by taking advantage of a unique data set on user activities and by running randomized experiments on Estimize.com.

Estimize.com is an open web-based platform founded in 2011 on which users can make earnings forecasts. The resulting consensus forecasts are available on both the company's website and Bloomberg terminals. A diverse group of users make forecasts. Among the 2,516 users studied in our sample, one third are financial analysts coming from buy-side, sell-side, or independent research firms. The remaining users include working professionals from different industries and students. Both academic and practitioner studies have documented the value of the Estimize consensus forecasts. For example, Jame et al. (2016) document that the Estimize consensus forecasts are incrementally useful in predicting earnings and measuring the market's expectations of earnings. Adebambo and Bliss (2015) also find that Estimize consensuses are more accurate than the traditional Wall Street earnings consensus 58%–64% of the time.

Users on Estimize.com make their forecasts sequentially. Before making one's own forecast, users can view a default web page (the "release page") that provides information on past earnings, the current Estimize consensus forecast, and forecasts from other Estimize users. As a result, herding behavior likely emerges among Estimize users. The unique feature of our data is that we can observe the users' entire web activities on Estimize.com, which allows us to differentiate user forecasts made with and without viewing the release page. Forecasts made without a release-page view are less likely to be influenced by then-available Estimize forecasts.

For our main sample period from March 2012 through March 2015, we examine 2,147 quarterly firm earnings (releases) with at least 10 forecasts prior to the announcement. These releases come from 730 distinct firms in various sectors. Building on the empirical framework of Chen and Jiang (2006), we find the release-viewing activity to have a significant impact on the forecasts. First, release viewing is associated with less weighting on private information, consistent with herding behavior. Second, although release viewing improves the accuracy of an individual forecast, it makes the consensus less accurate. A forecast error decomposition exercise confirms that the consensus becomes less accurate not because the forecasts are individually less accurate, but because the consensus incorporates less-diverse opinions. Useful private information may be lost when a user places weight on the prior forecasts. In particular, errors in earlier forecasts are more likely to persist and to appear in the final consensus forecast, making it less efficient. An interesting implication of this finding is that overconfidence could counteract the negative impact of herding. Overconfident users who place more weight on their private information can make the consensus forecast more accurate.

However, our empirical tests may be affected by the endogeneity associated with viewing choice. One could argue that users may choose to view the release page only when they have little private information.[4] To address this endogeneity concern, we collaborate with Estimize.com to run experiments during the second and third quarters of 2015 to restrict the public information set for randomly selected stocks and users. Specifically, for randomly selected stock, we randomly select users, hide information on their release page, and ask them to make a blind forecast. Each blind forecast is then matched to a default forecast issued at about the same time by a user who could view the entire release page. Compared with the blind forecast, the default forecast uses significantly less private information and is more accurate on average. Nevertheless, the consensus computed from blind forecasts is significantly more accurate than that computed using matched default forecasts.

Immediately after the blind forecast is made, the release view is restored, and the user can choose to update the forecast. We then compare the accuracy of two consensus forecasts: (1) the blind consensus computed using all blind forecasts and (2) the revised consensus computed using all revised forecasts made when the release view is re-enabled. For the 13 stocks randomly selected in the pilot experiment, the blind consensus significantly outperforms the revised consensus 10 times, and the revised consensus outperforms the blind consensus only two times. They tie in the remaining case. These findings are so compelling that, in November 2015, Estimize.com decided to switch to a blind platform, on which users make forecasts without seeing the current consensus.[5] Our analysis of forecasts issued during the year after November 2015 confirms that the Estimize consensus indeed became more accurate following the switch.

Having confirmed that herding reduces the accuracy of the consensus, we then examine when herding behavior is predictably stronger. We find that herding behavior becomes more severe when the public information set includes the estimates of influential users.

We first define a novel measure of user influence on the Estimize network using users' viewing activities and the PageRank algorithm invented by Google to rank web pages. We keep track of how many times Estimize users view each other on the website. Intuitively, users with high PageRank measures are viewed more by other users (either directly or indirectly), so their forecasts have more influence on subsequent forecasts. We also attempt to identify influential users using three other criteria: the total number of their forecasts, the total number of times when their forecasts are viewed by other users, and whether their forecasts lead subsequent forecasts. Interestingly, we do not find influential users to provide more accurate forecasts. Hence, herding with influential users does not automatically improve the accuracy of subsequent forecasts.

We find very similar results whichever definition of influential user we adopt. First, users are more likely to underweight their private information when the releases they view include the forecasts of influential users. Second, when influential users issue forecasts that are higher (lower) than the current consensus, the final consensus moves up (down), consistent with the notion that subsequent users are herding with the influential users. Finally, this herding behavior predicts the accuracy of the final consensus forecasts. For example, when the contemporaneous stock return is negative and influential users nevertheless issue forecasts that are higher than the current consensus, the final consensus becomes less accurate. In this case, influential users' forecasts likely reflect positive sentiments that are propagated among subsequent users and drag the consensus in the wrong direction. In other words, because of herding, predictable errors made early by influential users are not offset by subsequent forecasts and, thus, persist in the consensus forecast.

Our paper contributes directly to work on herding, including the understanding of various mechanisms underlying herding behavior. Herding behavior has been documented in various laboratory settings [see Anderson and Holt (1997) and Kubler and Weizsacker (2004) among others]. Empirically, herding behavior has been found to be pervasive.[6] Our finding that blind forecasts produce a better consensus is broadly consistent with the view in Goldstein and Yang (2019) that too much public information disclosure may actually crowd out the use of private information. In Goldstein and Yang (2019), the information crowding out occurs in a setting featuring multiple sources of uncertainty, and in our setting, it operates through the herding behavior made possible by the release of public information (the current consensus).[7]

Our paper also contributes to an emerging literature in finance and accounting that extracts useful information from social media (Chen et al. 2014, Adebambo and Bliss 2015, Jame et al. 2016, Pelster et al. 2017, Bartov et al. 2018). Instead of passively learning information from the online platforms, we go one step further by actively changing their information structure to better harness the collective intelligence. Although Cote and Sanders (1997) have used field experiments to differentiate herding behaviors from correlated private signals, we push the research agenda further. By measuring and randomizing an individual's information set in a large crowd-based earnings forecast platform, we are able to isolate the net impact of herding behavior on the accuracy of consensus earnings forecasts with important financial market implications.

Our findings have broader implications regarding group judgment.[8] Our results confirm that independent views are crucial for reaching an efficient outcome in a group setting. This is relevant evidence on the dangers of groupthink, on a corporate board, for example. We focus on the simple arithmetic average in computing the group consensus estimate and find that this simple consensus can be significantly improved in a blind-forecasting environment in which herding is difficult. There are, of course, other ways to average individual estimates and to motivate users to voice independent opinions (see Glazer et al. 2017 for example). We leave these interesting mechanism design questions to future research.

2. Data and Sample Description

2.1. Estimize.com

Estimize.com is an open web-based platform that facilitates the aggregation of financial estimates from a diverse community of individuals. Because the firm was founded in 2011, increasing numbers of contributors have joined the platform, and the coverage of firms has also significantly expanded. As of December 2015, more than 10,000 regular users contributed on the platform, resulting in coverage of more than 1,500 stocks each quarter.

Unlike the Institutional Brokers' Estimate System (IBES), Estimize solicits contributions from a wide range of individuals, including both professionals, such as sell-side, buy-side, or independent analysts, and nonprofessionals, such as students, private investors, and industry experts. Because of the contributions of these individuals, who have diverse backgrounds and viewpoints, Estimize consensus is more accurate than the Wall Street consensus and provides incremental information for forecasting earnings and for measuring the market expectation of earnings as documented by Jame et al. (2016) and Adebambo and Bliss (2015).

There are several reasons why Estimize consensus can be more accurate than the Wall Street consensus in IBES. First, Wall Street forecasts are often subject to predictable biases whether driven by investment-banking relations (Lin and McNichols 1998, Michaely and Womack 1999) or by career concerns (Hong and Kubik 2003). Estimize users do not suffer from these biases. Nevertheless, the fact that the market still reacts to the earnings surprise computed using the Estimize consensus after controlling for the earnings surprise based on Wall Street consensus suggests that the Estimize consensus goes beyond simply debiasing the Wall Street consensus. Indeed, combining Estimize forecasts with the Wall Street consensus yields improvement in predicting earnings, which suggests that Estimize provides incremental value. For example, Estimize users may be the firm's employees or consumers, and they could have private information about the earnings of the firm that is not available to the Wall Street analysts.

Estimize users have several incentives to provide information and contribute to Estimize. First, many users (e.g., independent analysts and students) can create a verifiable track record of their accuracy and their ability to predict fundamental metrics.

Second, Estimize assigns points to its contributors' forecasts. Points winners are recognized on the website, featured in podcasts, and awarded prizes, such as an Apple watch. Recently, Estimize organized all-America student analyst competitions; winners received awards at *Institutional Investor*'s annual awards dinner. The point system rewards forecasts that are more accurate than the Wall Street consensus and penalizes forecasts that are less accurate than the Wall Street consensus. The system also motivates aggressive estimation by awarding points on an exponential scale to elicit more private information.[9] The point system also penalizes bold forecasts exponentially if they turn out to be incorrect, so deviating from the crowd systematically without private information should not be the optimal strategy in most cases.

Consistent with the incentive structure underlying the point system, our empirical analysis confirms that Estimize contributors, on average, overweight their private signals relative to a Bayesian benchmark even though they still put positive weights on the current consensus. Importantly, because the exact formula for computing points is never made public, it is not easy for users to game the scoring system or to compute the exact optimal forecasting strategy.

Third, a goodwill factor may motivate some user participation, especially during the site's early days, just for the sake of its success—the more contributions, the more valuable the data set is to everyone.

2.2. Data Set

We collected three sets of data from Estimize. The first data set includes information on the forecasts created by users in the Estimize community. The sample period is March 2012 through March 2015. The forecasted earnings per share (EPS) value and the time of the forecast are both provided.

The second data set includes background information on users in the Estimize community. Estimize uses a brief personal profile voluntarily provided by the users themselves to classify users into several career-biographical categories, such as buy-side and sell-side professionals, industry experts, or students.[10]

The third data set records users' entire activities on Estimize.com, including the pages they view and the actions they take (e.g., creating forecasts); the data include the time stamps of all activities. The detailed web activities are made available through Mixpanel, an advanced analytics platform for mobile and web. We focus mainly on how many times a user views the release page of a specific firm that the user covers.

Figure 1 gives an example of a typical release page. The figure presents a screenshot of the release page for the 2015 Q2 earnings of Facebook, Inc. The release page shows two charts. The chart on the left presents the actual EPS of the past eight quarters, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts for the current quarter and the past eight quarters. The chart on the right provides information on all individual forecasts created for the current quarter. The count of views on the release page could proxy for whether the user's information set includes information from other users on the platform. Users can click any individual listed in the right chart to access an estimate page that presents all forecasts created by that individual. We use the number of views of a user's estimates page to construct a measure of influence.

2.3. Sample Construction

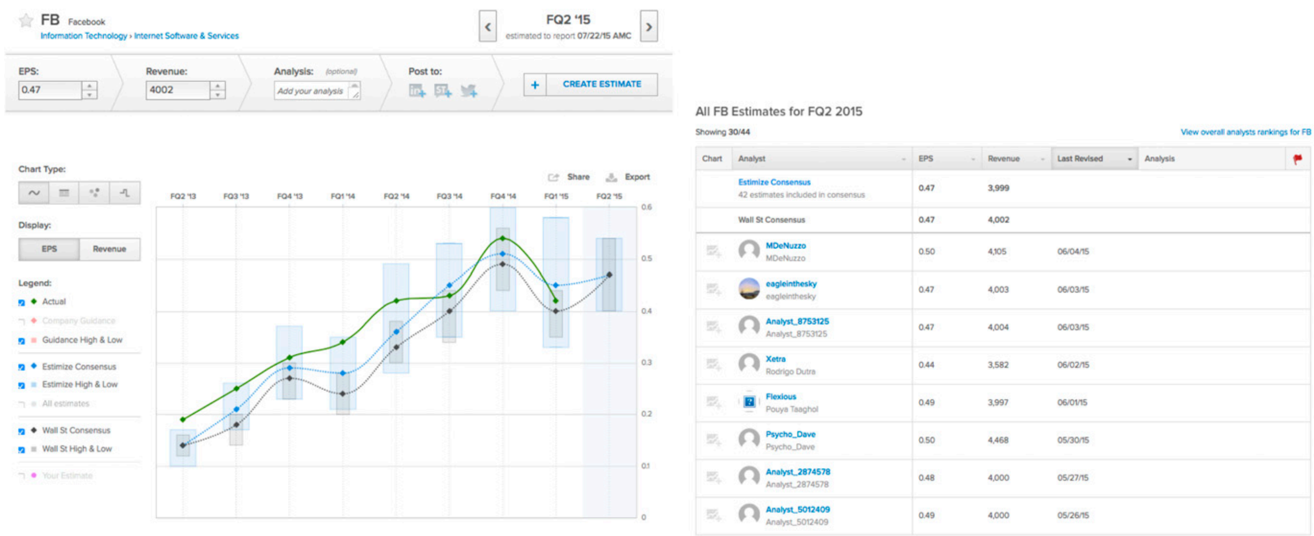
We match the information on forecasts and the information on web activities to form a comprehensive data set with forecast-level observations, covering the period from March 2012 through March 2015. For each forecast created, we track whether the user views the related release page for longer than five seconds.[11]

The initial sample includes 91,411 forecasts with 14,209 releases. We eliminate forecasts if the users cannot be successfully linked with an identifier in the activity data set. We also exclude forecasts that Estimize flags manually or algorithmically as unreliable.[12] Finally, to ensure a reasonably sized crowd for each release, we consider in our analysis only releases with at least 10 forecasts. The consensus forecast is always computed using the most recent forecast from a user.

2.4. Descriptive Statistics

Our final sample consists of 38,115 forecasts with 2,147 releases. Figure 2 presents the coverage of our

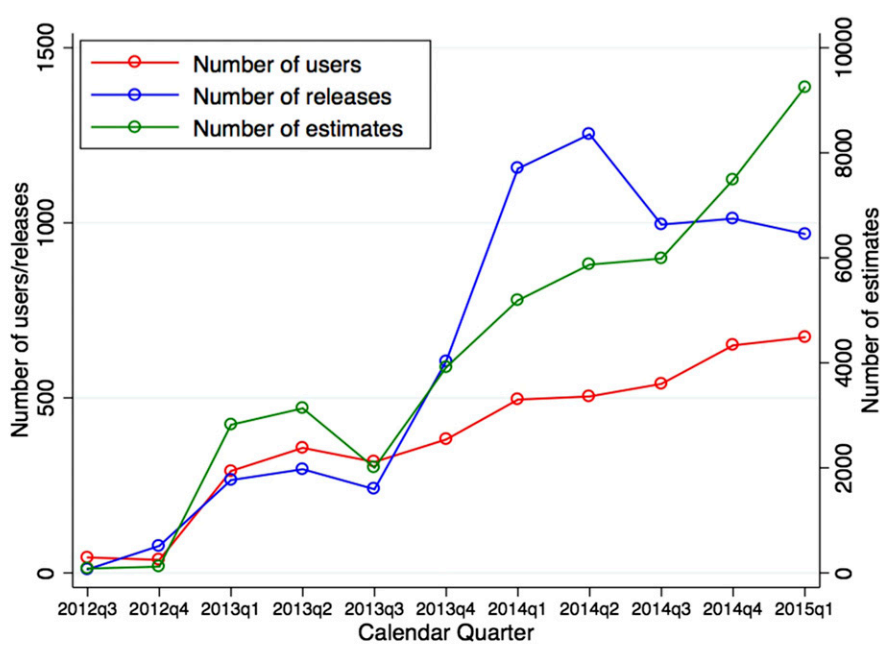
Figure 1. (Color online) Example of Release Page



Notes. The figure presents a screenshot of the release page for Facebook, Inc. (FB) for the second fiscal quarter of 2015. The chart on the left plots the historical data of actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. It also includes the current Wall Street and Estimize consensus. The chart on the right lists all Estimize estimates of Facebook's EPS underlying the current Estimize consensus.

sample over time. We can see the trend of increasing numbers of contributors and expanding coverage of firms, which is similar to the trend in the full sample. Table 1 provides descriptive statistics for the final sample. Panel A presents descriptive statistics for the release level. On average, about 16 users contribute 20 forecasts to a single release. The average release has around 19 views of the release page although the median count of release views is lower (12 views). It is worth noting that there is a wide range in the number of release views. Users may be very independent when making forecasts for some releases (e.g., only one release view) but check the release pages frequently for other releases (e.g., more than 114

Figure 2. (Color online) Coverage of Estimize Sample over Time



Notes. The figure plots the number of users, releases, and estimates in each quarter covered by our sample. Our main sample covers releases with at least 10 estimates from March 2012 through March 2015. The left axis represents the number of users and releases, and the right axis represents the number of estimates.

Table 1. Descriptive Statistics for the Estimize Sample

	Mean	Standard deviation	p1	p25	p50	p75	p99
Panel A: Release-level Estimize forecast characteristics (number of observations = 2,147)							
Number of forecasts	20.03	15.01	10.00	12.00	15.00	23.00	74.00
Number of distinct users	16.08	11.92	4.00	10.00	13.00	19.00	59.00
Number of release views	18.97	32.85	1.00	7.00	12.00	21.00	114.00
Consensus error (= consensus − actual)	−0.02	0.19	−0.51	−0.05	−0.01	0.02	0.31
Abs (consensus error)	0.08	0.18	0.00	0.01	0.03	0.08	0.63
Estimize abserr − WS abserr	−0.01	0.10	−0.17	−0.02	−0.01	0.01	0.14
Percentage of release view	35.60	17.20	0.00	23.08	35.29	46.94	76.47
Panel B: Release-level financial characteristics (number of observations = 1,953)							
Size (in millions)	24,512.35	46,548.06	430.66	2,896.06	7,635.54	21,862.95	241,171.05
B/M	0.40	0.40	0.03	0.18	0.31	0.49	2.08
Size group (= 1: bottom 20%; = 5, top 20%)	3.91	1.16	1.00	3.00	4.00	5.00	5.00
B/M group (= 1: bottom 20%; = 5: top 20%)	2.04	1.26	1.00	1.00	2.00	3.00	5.00
Panel C: User-level characteristics (number of observations = 2,516)							
Number of tickers covered	10.22	35.68	1.00	1.00	2.00	5.00	181.00
Number of forecasts submitted	17.10	78.43	1.00	1.00	2.00	6.50	320.00

Panel D: Distribution of stocks by sector

Sector	Frequency	Percentage
Consumer discretionary	146	20.00
Consumer staples	47	6.44
Energy	40	5.48
Financials	40	5.48
Health care	76	10.41
Industrials	95	13.01
Information technology	224	30.68
Materials	41	5.62
Telecommunication services	7	0.96
Utilities	14	1.92
Total	730	100.00

Panel E: Distribution of users by profession

	Frequency	Percentage
Financial professionals		
Buy-side	281	11.41
Sell-side	158	6.42
Independent	381	15.48
Nonprofessionals		
Information technology	519	21.08
Student	493	20.02
Financials	142	5.77
Consumer discretionary	110	4.47
Health care	94	3.82
Other	284	11.54
Total	2,462	100.00

Notes. The table presents descriptive statistics for the forecasts made on Estimize from March 2012 through March 2015. The sample covers 2,147 releases with at least 10 estimates. Panel A reports release-level forecast characteristics. Panel B reports release-level financial characteristics. The sample covers 1,953 releases with at least 10 estimates and matched financial data from Compustat. The size group and B/M group are obtained by matching each release with one of 25 size and B/M portfolios at the end of June based on market capitalization at the end of June and B/M, the book equity of the last fiscal year end in the prior calendar year divided by the market value of the equity at the end of December of the prior year. Panel C reports user-level characteristics. Panel D reports the sector distribution of the 730 distinct stocks in our sample. Panel E reports the distribution of users in our sample by profession.

release views). The wide range of release-viewing activities provides considerable variation across releases.

The average consensus on Estimize is slightly pessimistic with an average consensus error of −0.02. The average absolute value of the consensus error is

0.08, which is one cent more accurate than the average Wall Street consensus. When we examine a typical release in our sample, on average, 35.6% of all forecasts in that release are issued after viewing the release page. Across different releases, there is a lot of variation in the average release-viewing activity, which allows us to examine the impact of release-page viewing on forecast accuracy.

We also obtain financial characteristics data from Compustat. Panel B presents the size and book-to-market (B/M) statistics for release-level observations.[13] To compare the financial characteristics with NYSE stocks, we also report statistics on the size and B/M NYSE quintile group for firms in our sample.[14] The average firm size is $24.5 billion, and the median firm size is considerably smaller: about $7.6 billion. The average B/M ratio is 0.40, and the median B/M is 0.31. Our sample covers significantly larger firms than NYSE stocks with a strong growth tilt. These firms cover a wide range of sectors (panel D), such as information technology, consumer discretionary, industrials, healthcare, and consumer staples. Information technology and consumer discretionary are the two major sectors and account for more than 50% of our sample.

The forecasts covered in our sample are contributed by 2,516 users (panel C). The average user covers 10 firms and contributes 17 forecasts, and the distribution is strongly skewed to the right; there are many users contributing a moderate number of forecasts, and a few users frequently contribute on the platform. Estimize obtains contributions from individuals with remarkably diverse backgrounds. As panel E shows, 33.31% of the contributors studied in our sample are financial professionals, including sell-side (6.42%), buy-side (11.41%), and independent analysts (15.48%). The rest of the contributors are not professional analysts. Two major groups of nonprofessionals are in information technology (21.08%) or are students (20.02%).

3. Herding and Forecast Accuracy

In this section, we examine the impact of herding on the behavior and accuracy of individual and consensus earnings forecasts.

In our empirical analysis, we focus on the raw and unscaled earnings forecasts. Cheong and Thomas (2011) document that analysts' earnings forecast errors and dispersions do not actually vary with scale in the cross-section. Shue and Townsend (2018) also document that investors react to raw earnings surprise instead of scaled surprise. We find similar scale-invariance with the Estimize earnings forecasts. Robustness checks confirm that the results are qualitatively similar when we scale the earnings forecasts by the (split-adjusted) stock price at the end of the previous quarter. To save space, except for the main forecast accuracy analysis, these results are not reported.

We control for various fixed effects in our regressions. In our forecast-level regressions, release fixed effects subsume the need to control for stock characteristics and seasonality. Professional and individual fixed effects subsume the need to control for user characteristics. In our release-level regressions, we incorporate sector and quarter fixed effects.

Standard errors in our main regressions are double-clustered by sector and quarter. They are clustered by stock in regressions using our experimental data. In both cases, however, herding-induced correlations among different forecasts in the same release are accounted for because a release is nested in either the sector or the stock cluster. We confirm that the clustered standard errors are more conservative than those estimated from a random effect model, which represents an alternative way to deal with forecast error autocorrelation.

3.1. Release View and Weighting of Information

We first examine how release viewing affects the relative weighting between private and public information when a user makes a forecast. We follow the empirical framework of Chen and Jiang (2006).

Let z denote the true earnings and c denote the current market consensus about z. The user has a private signal y about z. Assume

$$c = z + \varepsilon_c,$$
$$y = z + \varepsilon_y,$$

where ε_c and ε_y are independent and normally distributed with zero means and precision of p_c and p_y, respectively. The user's best forecast according to Bayes' rule is

$$E[z|y, c] = hy + (1 - h)c,$$
$$h = \frac{p_y}{p_c + p_y}.$$

The user may not apply the most efficient weight h in reality. Instead, the actual forecast f could be $f = ky + (1 - k)c$. Chen and Jiang (2006) show that when regressing forecast error ($FE = f - z$) on a forecast's deviation from the consensus ($Dev = f - c$), the slope coefficient converges to $1 - \frac{h}{k}$. In other words, in the regression of

$$FE = \alpha + \beta_0 \cdot Dev + \varepsilon,$$

β_0 measures the actual weighting of private and public information compared with the optimal weighting. For example, a positive β_0 implies overweighting of private information ($k > h$).

Table 2 reports the regression results at the forecast level. In addition to Dev, we also include a release

Table 2. Release Views and Weighting of Information

	Forecast error (= forecast − actual)				
Dependent variable	(1)	(2)	(3)	(4)	(5)
Dev	0.424***	0.425***	0.489***	0.489***	0.470***
(= Forecast − pre consensus)	(0.087)	(0.087)	(0.077)	(0.077)	(0.057)
Dev × *view dummy*	−0.274***	−0.274***	−0.250**	−0.250**	−0.218***
	(0.090)	(0.090)	(0.110)	(0.110)	(0.062)
View dummy	0.00177	−0.00129	0.00102	0.00160	0.000262
	(0.002)	(0.005)	(0.001)	(0.001)	(0.001)
Release effect	No	No	Yes	Yes	Yes
Profession effect	No	Yes	No	Yes	No
Individual effect	No	No	No	No	Yes
Observations	30,429	30,429	30,429	30,429	30,429
R^2	0.034	0.035	0.917	0.918	0.934

Notes. The table presents the results of forecast-level weighting regressions. The dependent variable is *forecast error*, which is defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variables include (1) *Dev*: the forecast's distance from the consensus prior to the submitted forecast, (2) *View dummy*: a dummy variable for viewing the release page for longer than five seconds at least once, (3) the interaction term between *Dev* and *View dummy*. Standard errors are in parentheses and double-clustered by sector and quarter.

***, **, * = significant at the 1%, 5%, and 10% levels.

view dummy and its interaction with *Dev* as independent variables in the regressions. We find a significantly positive β_0, suggesting that Estimize users, on average, overweight their private signals.[15] Most important, we find a significant negative coefficient on the interaction term between *Dev* and the release-view dummy. For example, the coefficients reported in column (1) suggest that release viewing reduces the excessive weight on private information by 0.274 (from 0.424 to 0.150). In other words, viewing of the current consensus, not surprisingly, is associated with placing more weight on the consensus and less weight on the private signal, consistent with herding behavior. To rule out the possibility that our results are driven by a particular release or by a particular user type, we include firm-quarter (or release), profession, and individual fixed effects in columns (2) through (5). The results are very similar.

3.2. Release View and Forecast Accuracy

How does the viewing of public information affect the forecast accuracy? We first examine this question at the individual forecast level by regressing the absolute forecast error on the release-view dummy. We include release fixed effects. Effectively, we are comparing forecasts for the same release with and without release views. In addition, we include a close-to-announcement dummy variable that is equal to one if the forecast was issued during the last three days before the earnings announcement. This dummy variable controls for the fact that forecasts closer to the announcement should be more accurate.

In panel A of Table 3, we find a significant negative coefficient in column (1). Release viewing reduces the forecast error by more than 0.73 cents. In column (2), we further include user-profession fixed effects and again the result does not change much. In column (3), we replace user-profession fixed effects with individual fixed effects. We still find that viewing the release page reduces individual forecast error.

To further control for stock characteristics that may drive both the release-view activity and the forecast error, we adjust the Estimize absolute forecast error by that of the Wall Street analysts in columns (4)–(6). Results remain the same. Finally, columns (7)–(9) report similar results when we scale the forecast errors by the stock price at the end of the previous quarter. Overall, it is clear that viewing public information, including the current Estimize consensus, improves the accuracy of each individual forecast.

But what about the accuracy of the consensus forecast or the wisdom of the crowd? We examine this question at the release level in panel B. For each release, we measure the frequency of release viewing as the logarithm of one plus the ratio of the number of forecasts made by users who viewed the release for longer than five seconds to the number of total forecasts (LnNumView). In other words, if most users viewed the release page before making their forecasts for that release, LnNumView for that release will be higher.

Interestingly, when we regress absolute consensus forecast error on LnNumView, we find a significant positive coefficient on LnNumView, suggesting that the viewing of public information actually makes the consensus forecast less accurate. Compared with a release in which all forecasts are made without viewing the release page (LnNumView = 0), a release in which all forecasts are made after viewing the release

Table 3. Release Views and Forecast Accuracy

Panel A: Forecast-level analysis									
	Abs(FE = forecast − actual)			Abs(FE) relative to WS			Abs(FE = (Forecast − Actual)/Price)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
View dummy	−0.00737***	−0.00731***	−0.00267***	−0.00586***	−0.00582***	−0.00186**	−0.0146***	−0.0145***	−0.00285*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
CTA dummy	−0.0119***	−0.0116***	−0.00678***	−0.0111***	−0.0108***	−0.00631***	−0.0259***	−0.0252***	−0.0133***
	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.000)
Release effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Profession effect	No	Yes	No	No	Yes	No	No	Yes	No
Individual effect	No	No	Yes	No	No	Yes	No	No	Yes
Observations	37,674	37,674	37,674	37,674	37,674	37,674	33,622	33,622	33,622
R^2	0.824	0.824	0.835	0.494	0.495	0.529	0.819	0.820	0.832

Panel B: Release-level analysis									
	Abs(FE = consensus forecast − actual)			Abs(FE) relative to WS					
	(1)	(2)	(3)	(4)	(5)	(6)	Abs(FE = (Forecast − Actual)/Price)		
LnNumView	0.0450***	0.0401***	0.0551***	0.0106***	0.0111***	0.0116***	0.188***	0.154***	0.183***
	(0.008)	(0.006)	(0.010)	(0.003)	(0.003)	(0.003)	(0.050)	(0.046)	(0.046)
Std dev (FE)		0.0517**	0.0508**		−0.00564***	−0.00586***		0.646***	0.624***
/Abs (Median(FE))		(0.022)	(0.020)		(0.001)	(0.002)		(0.136)	(0.132)
Sector effect	No	No	Yes	No	No	Yes	No	No	Yes
Quarter effect	No	No	Yes	No	No	Yes	No	No	Yes
Observations	2,122	2,122	2,122	2,122	2,122	2,122	1,929	1,929	1,929
R^2	0.011	0.025	0.110	0.006	0.007	0.030	0.012	0.133	0.159

Panel C: Within-release horse race			
	Consensus	*avgFE*	*Diversity*
Forecasts with release views win, %	40.76	53.69	40.71
Forecasts without release views win, %	59.24	46.31	59.29
Binomial test p-value	< 0.0001	0.0004	< 0.0001

Notes. Panel A presents the results of forecast-level regressions. The dependent variable is the absolute value of raw forecast error (columns (1)–(3)), the absolute forecast error relative to that of the Wall Street consensus (columns (4)–(6)), or the absolute forecast error scaled by the (split-adjusted) stock price at the end of previous quarter (columns (7)–(9)). Forecast error is defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variable is *View dummy*, a dummy variable that is equal to one when the user views the release page for longer than five seconds at least once. The control variables include a close-to-announcement (CTA) dummy that is equal to one if the forecast was issued in the last three days before the announcement and various fixed effects. Panel B presents the results of release-level regressions. The dependent variable is the absolute value of consensus error (columns (1)–(3)), the absolute consensus error relative to that of the Wall Street forecasts (columns (4)–(6)), or the absolute forecast error scaled by the (split-adjusted) stock price at the end of previous quarter (columns (7)–(9)). The main independent variable is the logarithm of one plus the ratio of the number of forecasts made following release views longer than five seconds to the number of total forecasts. The control variables include the standard deviation of forecast error normalized by the absolute value of median forecast error and sector and quarter fixed effects. Standard errors are in parentheses and double-clustered by sector and quarter. Panel C presents comparisons between forecasts with release views and forecasts without release views within each release for three measures: (1) the consensus accuracy, (2) the average (squared) errors of individual forecasts, and (3) the diversity among individual forecasts.

***, **, * = significant at the 1%, 5%, and 10% levels.

page (LnNumView = ln(2) = 0.69) is 3.82 (= 0.0551 × 0.69 using the coefficient reported in column (3)) cents less accurate. This represents a significant reduction in accuracy as the median forecast error is only three cents in our sample (see Table 1, panel A).

In columns (2) and (3), we control for the forecast dispersion, so the result is not driven by a few hard-to-forecast stocks with lots of release views. We also evaluate the Estimize consensus error relative to that of the Wall Street consensus in columns (4)–(6) as another way to control for stock characteristics that may drive both the release-view activity and the consensus forecast error. We find qualitatively similar results. Scaling the forecast errors by the stock price at the end of the previous quarter in columns (7)–(9) does not change the conclusion either.

Another way of seeing this result is through a simple horse race, which we conduct in panel C. In each release, we separate all forecasts into two groups. The view group represents all forecasts made after viewing the release page. The no-view group represents the remaining forecasts, those made without first viewing the release page. We then compute two consensus forecasts using the forecasts from the two groups and compare which consensus is more accurate. In the 2,127 releases we studied, the no-view consensus wins 59.24% of the time, which is significantly more than 50%. Again, the viewing of public information makes the consensus forecast less accurate.

How can viewing a release page improve the accuracy of individual forecasts but, at the same time, make the consensus less accurate? The answer can be illustrated by decomposing the squared error in the consensus forecast into two parts:

$$\begin{aligned}(c-z)^2 &= \frac{1}{n}\sum_{i=1}^{n}(f_i-z)^2 - \frac{1}{n}\sum_{i=1}^{n}(f_i-c)^2\\ &= avgFE - diversity,\end{aligned}$$

where the first part measures average (squared) errors of individual forecasts and the second part measures the diversity among these individual forecasts.

According to panel C, if we focus on *avgFE*, forecasts with release views (i.e., with lower *avgFE*) win 53.69% of the time. In other words, these forecasts are individually more accurate, on average, consistent with the results in panel A. When we focus on *diversity*, however, forecasts with release views (i.e., with higher diversity) win only 40.71% of the time, which suggests these forecasts are less diverse on average. The reduced diversity more than offsets the higher accuracy in individual forecasts, making the consensus with views less accurate.

Herding prevents useful private information from entering the final consensus and makes forecasts less diverse. In the most extreme case, if all subsequent users completely herd on the first user, then the private information of the subsequent users is entirely absent, so the crowd consensus is no more accurate than the first forecast in that sequence. An interesting implication of this result is that overconfidence could counteract the negative impact of herding. Overconfident users who place more weight on their private information can make the consensus forecast !more accurate. The fact that Estimize users are overconfident, on average, could be one reason that their consensus is more accurate than the Wall Street consensus.

When herding happens, errors in earlier forecasts are more likely to persist and show up in the final consensus forecast. Table 4 examines one such persistent error at the release level. The dependent variable is a dummy variable that is equal to one if earlier and close-to-announcement estimates are biased in the same direction. The close-to-announcement window is defined as extending from five days before the announcement date through the announcement date ([−5, 0]). The early window is defined as any of the days prior to day −5. The consensus within the window is upwardly (downwardly) biased if the difference between the consensus and the actual EPS is above the Hth percentile (below the Lth percentile). The main independent variable is again LnNumView but measured using only forecasts in the close-to-announcement window. The control variables include the same measure of forecast uncertainty and sector and quarter fixed effects.

The results confirm a strong link between the persistence of error and release views. When more forecasts are made after viewing the release page, the initial error is more likely to persist and to show up in the final consensus, making it a less efficient forecast.

4. Blind Experiments

Our empirical tests so far are affected by the endogeneity associated with the choice of whether to view. One could argue that users may choose to view the release page only when they have little private information. To address the endogeneity concerning the information acquisition choice, we collaborate with Estimize.com to run randomized experiments during the second and third quarters of 2015. Note that the experiments take place after the sample period of our main analysis.

The stocks in our experiments are randomly selected to come from a wide range of industries. When users first arrive at the release page of a selected stock during the experiment period, some of them are randomly selected as "blind" users, and they see only a blind version of the release page for this stock.

Figure 3 gives one such example. The figure presents a screenshot of the blind release page for Lululemon Athletica, Inc., for the fourth quarter of 2015.

Table 4. Release Views and Lead-Lag Biases

	Consistent bias indicator		
Dependent variable	(1)	(2)	(3)
Bias is defined as (average forecasts − actual) above Hth percentile or below Lth percentile	$H = 60, L = 40$	$H = 70, L = 30$	$H = 80, L = 20$
LnNumView	0.221*	0.292***	0.466***
	(0.120)	(0.112)	(0.115)
Sector effect	Yes	Yes	Yes
Quarter effect	Yes	Yes	Yes
Observations	1,770	1,770	1,770
Pseudo R^2	0.0317	0.0359	0.0614

Notes. The table presents the results of forecast-level regressions. The dependent variable is a dummy variable that is equal to one if earlier and close-to-announcement estimates are biased in the same direction. The close-to-announcement window is defined as extending from five days before the announcement date through the announcement date ([−5, 0]). The early window is defined as days prior to day −5. The consensus within the window is upwardly (downwardly) biased if the difference between the consensus and the actual EPS is above the Hth percentile (below the Lth percentile). The main independent variable is the logarithm of one plus the ratio of the number of forecasts made following release views longer than five seconds to the number of total forecasts within the close-to-announcement window. The control variables include the standard deviation of forecast error normalized by the absolute value of median forecast error and sector and quarter fixed effects. Standard errors are in parentheses and double-clustered by sector and quarter.

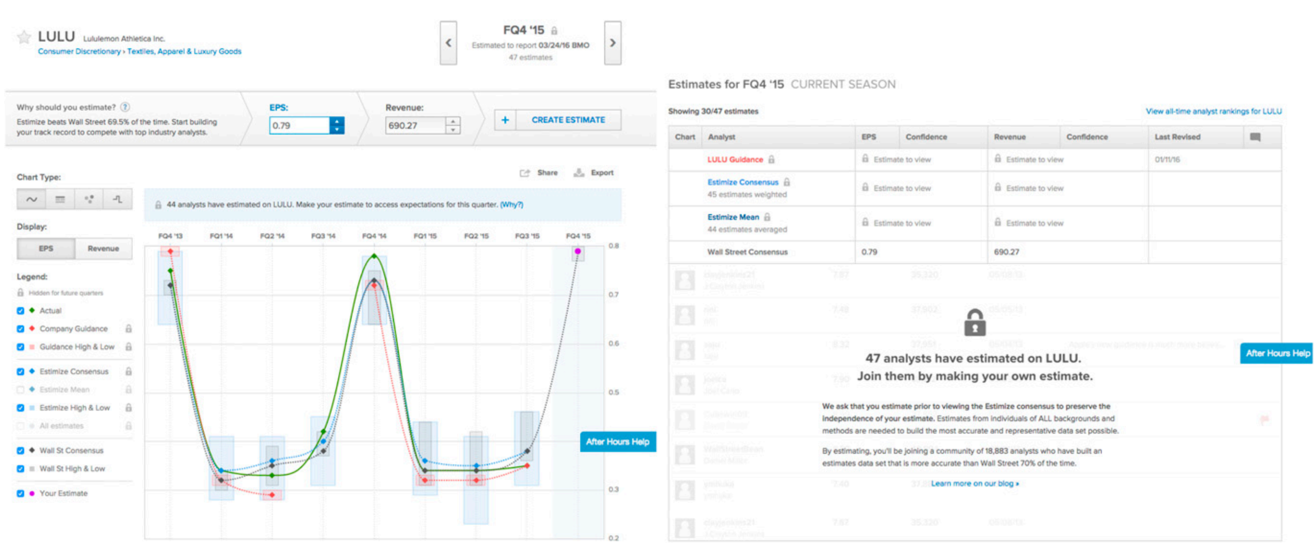
***, **, * = significant at the 1%, 5%, and 10% levels.

The chart on the left plots the historical data of the actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. Note that no information on the consensus is provided for the fourth quarter. The chart on the right shows that all Estimize EPS estimates for Lululemon Athletica, Inc., including the current Estimize consensus, are hidden.

Importantly, the current Wall Street consensus is still available on the blind release page. Even if the selected blind users have no private information about the earnings, they can always use the current Wall Street consensus as their default forecast and revise it later when the release page is restored. In addition, making the current Wall Street consensus always available also limits the downside associated with the blind forecasting environment by eliminating completely uninformed forecasts.

After viewing the blind release page, a blind user may choose to issue a forecast, which we label as the

Figure 3. (Color online) Example of Blind View



Notes. The figure presents a screenshot of a blind release page for Lululemon Athletica. Inc. (LULU) for the fourth fiscal quarter of 2015. The chart on the left plots the historical data of actual EPS, the range and consensus of Wall Street forecasts, and the range and consensus of Estimize forecasts. Note no information on the Estimize consensus is provided for the fourth quarter. The chart on the right shows that all individual Estimize estimates of Lululemon Athletica's EPS are also hidden.

blind forecast (f_b). Immediately after the blind forecast is issued, the release page is re-enabled so that the blind users can view the Estimize forecasts and consensus and they can immediately revise the blind forecast if desired. In other words, each blind user can issue at most one blind forecast for that release in our experiment.

Each blind forecast is matched with the closest forecast in the sequence made by a different user (default user) who could view the release page. The matched estimate is labeled the default forecast. That pair is eliminated if the time difference between the blind estimate and the default estimate exceeds 24 hours.[16] The final sample includes releases with at least 15 matched pairs. There are 103 releases in the final sample, 13 from the first-round pilot experiment and 90 from the second-round experiment.

Although blind users are randomly selected to view the blind release page, not all blind users will issue blind forecasts. Put differently, the decision to participate by issuing a blind forecast after viewing the blind page may not be random. For example, one may worry that blind users are more reluctant to issue forecasts, and maybe only the skilled blind users choose to issue blind forecasts. This endogenous choice may drive the accuracy of the blind consensus. Empirically, we observed the opposite. Conditioning on arriving at the release page, blind users are more likely to issue forecasts than default users.

To ensure no systematic bias is introduced, in panel A of Table 5, we first confirm that blind users who actually issued blind forecasts (the blind forecasters) are similar to the matched default users who issued default forecasts (the default forecasters). Blind and default forecasters are equally likely to be professional analysts (27.2% versus 27.1%) and to be highly viewed by other users. They have covered similar numbers of stocks (55 versus 53) and have participated in similar numbers of releases (102 versus 94) on average. Their past forecast accuracy is also similar.

4.1. Blind vs. Default

Panel B of Table 5 repeats the analysis in panel A of Table 3 for the experimental sample. We again find the blind forecast to be individually less accurate than the matching default forecast. Blind users do not seem to make forecasts only when they have precise private signals.

We also examine the dispersion in the blind forecasts versus that in the default forecasts. We find that, on average, the standard deviation of the default forecasts is 11.09% lower than that of the blind forecasts (t-value = 1.95). In other words, the ability to view the current Estimize consensus and other individual users' forecasts reduces the forecast dispersion. This finding is more consistent with herding behavior than with antiherding, in which a winner-takes-all payoff scheme induces the user to deviate from the crowd strategically. Nevertheless, for a small set of releases with fewer than 15 contributing users, the default forecasts can have wider dispersions, suggesting that strategic behavior can be relevant when there are fewer players. Eliminating such strategic behavior offers another channel for blind forecasts to improve the accuracy of the consensus forecast.

We then compare the blind forecasts to their matching default forecasts in terms of information weighting. As in panel A of Table 2, we regress forecast errors (*FE*) on *Dev* and its interaction with the default forecast dummy (*Default*) with release fixed effects. The results are reported in panel C of Table 5. The regression in column (1) does not include profession fixed effects.

First, the large, positive, and significant coefficient on *Dev* (0.670) confirms that blind forecasts are made almost exclusively with private information. The coefficient is higher than the corresponding number (0.489) in panel A of Table 2, suggesting that the blind forecasts in the experiment rely more on private information than forecasts from the full sample made without viewing the release page. Second, the significant negative coefficient of −0.113 on *Dev* × *Default* indicates that the ability to view public information results in less overweighting of private information and more reliance on public information. Importantly, because both experiment participants and stocks are randomly selected, the difference between the blind forecast and the default forecast cannot be driven by the endogenous decision to view the release page. The results with profession fixed effects in column (2) are very similar.

Because users can always see the Wall Street consensus, we consider a placebo test that replaces the Estimize consensus (c) with the Wall Street consensus in the regression (c_{ws}). We find a small and insignificant coefficient of less than 0.1 on *Dev* × *Default*. This is not surprising as both blind and default forecasts are made with c_{ws} included in the information set.

The more interesting question is whether blind forecasts result in a more accurate consensus than the default forecasts. We examine this question with a simple horse race. For each release, we compute two consensus forecasts. The blind consensus is computed as the average of all blind forecasts, and the default consensus is computed as the average of all default forecasts. By construction, the two consensuses are computed using the same number of forecasts. In the 103 releases examined, the blind consensus is more accurate 62 times. The associated one-tail p-value is lower than 0.0001 in rejecting the hypothesis that the blind and default consensuses are equally accurate.

Table 5. Blind Experiment: Blind vs. Default

Panel A: Blind vs. default forecasters

	Professional (1)	Highly viewed (2)	Number of releases (3)	Number of tickers (4)	Abs(FE) (5)
Blind	27.23%	19.96%	102	55	0.111
Default	27.10%	17.88%	94	53	0.115
Diff (blind − default)	0.13%	2.08%	8	2	−0.004
T-statistic	0.07	0.69	0.29	0.16	0.32
p-value	0.947	0.49	0.770	0.871	0.750

Panel B: Release view and forecast accuracy

	Abs (forecast error)	
Dependent variable	(1)	(2)
Default	−0.00221**	−0.00253**
	(0.001)	(0.001)
CTA dummy	−0.00648***	−0.00647***
	(0.001)	(0.001)
Release effect	Yes	Yes
Profession effect	Yes	Yes
Individual effect	No	Yes
Observations	8,630	8,630
R-squared	0.825	0.825

Panel C: Release view and information weighting

	Forecast error (= forecast − actual)	
Dependent variable	(1)	(2)
Dev	0.670***	0.670***
(= forecast − preconsensus)	(0.068)	(0.068)
Dev × *default*	−0.113**	−0.113*
	(0.059)	(0.058)
Default	−0.00311**	−0.00299*
	(0.001)	(0.002)
Release effect	Yes	Yes
Profession effect	No	Yes
Observations	8,198	8,198
R^2	0.956	0.956

Notes. When users are randomly selected to participate in the experiment, they are asked to make an earnings forecast while the release page is disabled. The resulting forecast is labeled the blind forecast (f_b). Each blind forecast is matched with the closest estimate in the sequence made by a different user who could view the release page. The matched estimate is labeled the default forecast. The pair is removed if the time difference between the blind estimate and the default estimate exceeds 24 hours. The final sample includes releases with at least 15 matched pairs. Blind and default forecasts are pooled in the regression. In panel A, we compare user characteristics between the blind forecasters (blind users who issued the blind forecasts) and the default forecasters (default users who issued the default forecasts). Column (1) represents the percentage of professional users. Column (2) represents the percentage of users with numbers of estimate view over the 80th percentile. Column (3) represents the average number of releases submitted by the users in each group. Column (4) represents the average number of firms covered by the users in each group. Column (5) represents the average absolute value of forecast errors of users in each group. Columns (2)–(5) are based on users' forecasts before the experiment. *T*-statistics and *p*-values of the difference between these two groups are also reported. In panel B, the dependent variable is the absolute forecast error. The control variables include a close-to-announcement (CTA) dummy that is equal to one if the forecast was issued in the last three days before the announcement and release, individual, and user profession fixed effects. In panel C, the dependent variable is the forecast error defined as the difference between the blind forecast and the actual EPS. Independent variables include (1) *Dev*: the forecast distance from the consensus prior to the submitted forecast, (2) *Default*: a dummy variable that is equal to one if it is a default forecast and zero if it is a blind forecast, and (3) the interaction term between *Dev* and *Default*. Standard errors are in parentheses and clustered by ticker.

***, **, * = significant at the 1%, 5%, and 10% levels.

To gauge the statistical significance in each pairwise comparison, we also conduct jackknife resampling. Take the Q1 earnings for Facebook as an example. Twenty-four distinct users are randomly selected to participate in the experiment and issued blind forecasts, which are, in turn, matched to 24 default forecasts. In each resample, we remove one user and compute the blind and default consensus using the remaining 23 forecasts and check which is more accurate. We find the blind consensus to beat the revised consensus in all 24 resamples, resulting in a *p*-value of zero. In the 103 releases examined, the blind consensus significantly beats the default consensus 58 times with a *p*-value of less than 10%, and the default consensus wins significantly only 38 times.

4.2. Blind vs. Revised

The experimental evidence so far confirms that limiting information access may encourage the user to express more independent opinions and, therefore, improve the accuracy of the consensus forecast. So far, we have compared the forecasts from two different groups of users (blind and default). Next, we compare two different forecasts from the same user.

Recall that, in our experiment, immediately after the blind forecast (f_b) is issued, the release page is reenabled so the user can view the Estimize forecasts and consensus and choose to revise the forecast. The new forecast is labeled the revised forecast (f_r). Users can, of course, choose not to change their forecasts, in which case the revised forecast is the same as the blind forecast. We focus this part of our analysis on the pilot experiment in which many users immediately revised their forecasts after issuing the blind forecast.[17]

In this case, we could interpret both f_b and f_r as the combination of the same private signal y and the Estimize consensus: $f_b = w_b y + (1 - w_b)c$ and $f_r = w_r y + (1 - w_r)c$. It can then be shown that

$$f_b - f_r = \frac{w_b - w_r}{w_b}(f_b - c).$$

In other words, if we regress $f_b - f_r$ on $f_b - c$ and obtain a positive slope coefficient, this means that the blind forecast places more weight on the private signal than the revised forecast does ($w_b > w_r$). When we run the regression in panel A of Table 6, we indeed find a positive and significant coefficient of about 0.534 (column 2). In a placebo test, we replace the Estimize consensus (c) with the Wall Street consensus in the regression (c_{ws}). We find a small and insignificant coefficient on $f_b - c_{ws}$.

In panel B, we compare the accuracy of two consensus forecasts: (1) the blind consensus computed using all blind forecasts and (2) the revised consensus computed using all revised forecasts. In the 13 randomly selected releases in the pilot experiment, the blind consensus significantly outperforms the revised consensus 10 times, and the revised consensus wins only two times. They tie in the remaining one case. The statistical inference is again conducted using jackknife resampling.

To examine the reason behind the improvement of the blind consensus over the revised consensus, we again decompose the squared consensus error into the *avgFE* component and the *diversity* component. When we focus on the *avgFE* component, we find that individual forecast errors are actually higher among blind forecasts than among revised forecasts for 10 of the 13 stocks. When we focus on the *diversity* component, we find that blind forecasts are overwhelmingly more diverse. Overall, it is clear that blind consensus is more accurate, not because blind forecasts are individually more accurate, but because the blind consensus incorporates more diverse opinions.

4.3. Event Study

So far, our experiment results suggest that the wisdom of crowds can be better harnessed by encouraging independent voices among participants. Motivated by our findings, Estimize.com decided to switch to the blind forecast platform; since November 2015, forecasts from other users are always blocked initially. Estimize states in their announcement of the switch: "[Consensus] only gets better with a greater number of independent opinions. . .While your estimate for a given stock may be less accurate than the average of your peers, it is an important part of building a better consensus."

A natural question is whether the blind platform indeed improves the accuracy of the Estimize consensus. Our analysis of the Estimize forecasts during the four quarters immediately after the switch shows that the answer is yes. We compare the Estimize consensuses during the before-experiment period (before March 2015) to those during the after-experiment period (November 2015–October 2016). We limit the comparison with the same set of stocks and the same quarter of the year. For example, we compare the consensus of Facebook's 2016 Q3 earnings to other Facebook Q3 consensuses before the experiment. There are 1,641 stocks that are covered by Estimize in both the before- and the after-experiment periods. To control for marketwide variations during our sample period, we always compare the Estimize consensus against the corresponding Wall Street consensus.

Column (1) in Table 7 shows that Estimize consensus accuracy improved after the blind platform was adopted. In the before-experiment period, the Estimize consensus beat the Wall Street consensus 56.67% of the time. In the after-experiment period, the Estimize consensus is more accurate 64.11% of the time. The increase in the winning percentage of 7.44 percentage points is highly significant (*t*-value of 6.28). We also

Table 6. Blind Experiment: Blind vs. Revised

Panel A: Forecast-level weighting regression

	Forecast (blind) − forecast (revised)	
Dependent variable	(1)	(2)
Forecast (blind) − preconsensus	0.523***	0.534***
	(0.050)	(0.051)
Sector effect	No	Yes
Observations	104	104
R^2	0.466	0.481

Panel B: Within-release horse race: Blind consensus versus revised consensus

Ticker	*N*	Blind FE	Revised FE	*p*-value	Consensus	*avgFE*	*diversity*
WFM	20	−0.0020	−0.0030	0.05	Y	N	Y
UA	40	0.0170	0.0173	0.05	Y	N	Y
F	24	0.0342	0.0392	0.00	Y	Y	Y
CMG	35	−0.2114	−0.2086	1.00	N	N	Y
AA	22	−0.0059	−0.0059	1.00	-	-	-
XOM	19	−0.3195	−0.3232	0.05	Y	N	Y
BAC	16	0.0263	0.0306	0.06	Y	N	Y
GS	17	−1.6682	−1.6812	0.00	Y	Y	Y
GILD	58	−0.5341	−0.5047	1.00	N	N	Y
JNJ	17	−0.0318	−0.0329	0.06	Y	N	Y
AAPL	133	−0.0744	−0.0745	0.06	Y	N	Y
FB	91	0.0227	0.0236	0.00	Y	N	Y
FEYE	16	0.0344	0.0369	0.00	Y	N	Y

Notes. The blind and revised forecasts are from the pilot experiment. Panel A presents the results of forecast-level weighting regressions. The dependent variable is the difference between the blind forecast and the revised forecast from the same user in the blind experiment. The main independent variable is the blind forecast's distance from the consensus prior to the submitted forecast. Standard errors are in parentheses and clustered by ticker. Panel B presents the results of the within-release horse race between the blind consensus and the revised consensus. When calculating the revised consensus, we fill the forecast with the initial one for users who choose not to revise their forecasts. The forecast error (FE) is defined as the difference between the consensus and the actual EPS. The last three columns report whether the blind forecast beats the revised forecast in the consensus error and its two components. The answer is yes if the blind forecast has a lower (squared) consensus error or a lower *avgFE* term or a higher *diversity* term.

***, **, * = significant at the 1%, 5%, and 10% levels.

confirm that the improvement is not driven by increasing participation of Estimize users during the after-experiment period so that the Estimize consensus is computed using more individual forecasts. Columns (2)–(4) confirm the improvement even when we restrict the comparisons to consensuses computed using similar numbers of individual forecasts. When there are more than 30 individual forecasts, there is less of an improvement in the Estimize consensus. As the online appendix shows, the variance of sequential forecast consensus declines exponentially as the number of forecasts increases.

Table 7. Estimize Consensus: Before March 2015 and After November 2015

	(1)	(2)	(3)	(4)
Pr(Estimize consensus beats WS consensus)	All	Number of estimates (0,10]	Number of estimates (10,30]	Number of estimates > 30
N	1,641	660	315	210
Before experiment, %	56.67	54.85	56.51	62.38
After experiment, %	64.11	64.24	64.44	66.67
Z-test (after − before)	6.28	5.03	2.94	1.32
p-value	0.000	0.000	0.002	0.090

Notes. The table presents the comparison of the probability that the Estimize consensus beats the Wall Street consensus between the before-experiment period (before March 2015) and the after-experiment period (November 2015–February 2016). The sample is limited to the same set of stocks before and after the experiment. Column (1) include all releases. Columns (2)–(4) include releases with the number of estimates below 10, between 10 and 30, and greater than 30, respectively.

5. Influential Users

So far, we have confirmed that herding, although it improves the accuracy of individual forecasts, reduces the accuracy of the consensus forecast; interestingly, withholding certain information from individual users actually improves the average forecast. A natural question is when is herding behavior more severe and results in predictable errors in the consensus forecast?

5.1. The Role of "Influential" Users in Herding

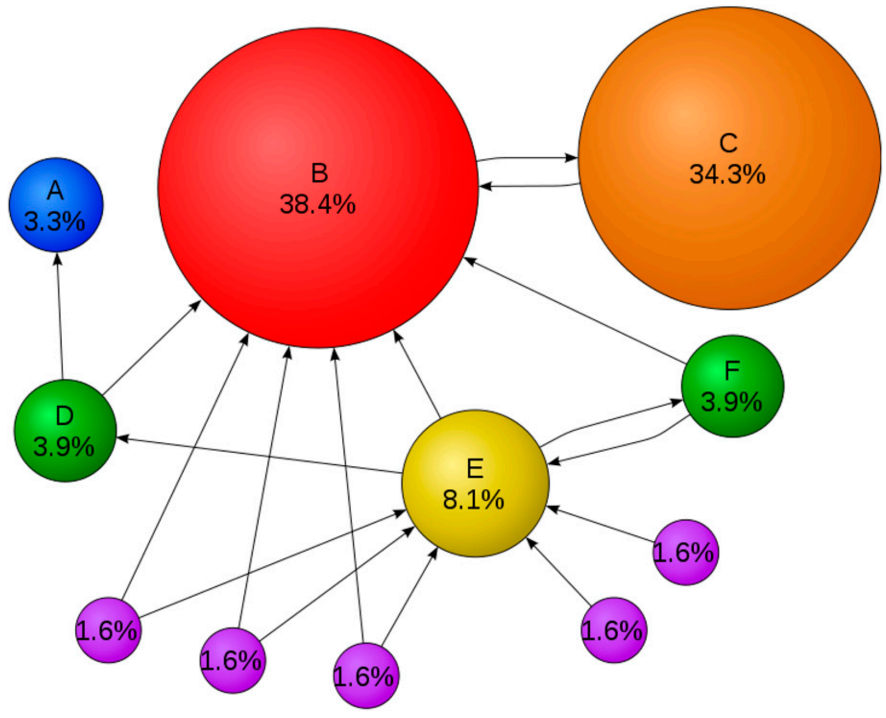
The evidence using the unique release-view information suggests that the influence we exert on each other can make the crowd's average estimate less accurate. Of course, not all users are created equal. Some users can potentially exert a stronger influence on others. We would, therefore, expect more pronounced herding behavior when more influential users are in the crowd.

To measure the influence of Estimize users, we use the user viewing-activity data and the PageRank algorithm developed by Google for ranking web pages. Figure 4 illustrates an example. Different Estimize users are represented by different circles that are linked by arrows that capture viewing activities. For example, when user D views user A, we see an arrow going from user D to user A. Such viewing activities provide direct observations of person-to-person influence, just like insider-trading legal documents provide direct observations of person-to-person communication as in Ahern (2017).

An influential user (represented by a larger circle) either receives more incoming arrows (as in the case of user B) or receives an arrow from another influential user (as in the case of user C). The user influence is measured by the PageRank measure, which is reported inside the circle. Intuitively, users with high PageRank measures are viewed more by other users (either directly or indirectly), so their forecasts are more influential; they have more impact on subsequent forecasts.

In computing the PageRank measure for each Estimize user, we also account for the number of times user A viewed user B. When we regress PageRank scores on user characteristics across different Estimize users, we find users to be more influential if they make more forecasts and if their forecasts are more often viewed by other users. Interestingly, the user's average forecast accuracy and average forecast bias do not affect the PageRank measure. In two simple alternative measures of user influence, we, therefore, also consider the total number of forecasts made by the user and the total number of times the user has been viewed by others.

Figure 4. (Color online) Measuring Estimize User Influence Using PageRank



Notes. The figure presents an example of how we measure user influence on the Estimize network using the PageRank method. Users are represented by circles that are linked by arrows that capture viewing activities. For example, when user D views user A, an arrow goes from user D to user A. An influential user (represented by a larger circle) either receives more incoming arrows (as in the case of user B) or receives an arrow from another influential user (as in the case of user C). User influence is measured by the PageRank, reported inside the circle.

Our fourth measure of user influence attempts to capture the extent to which a user's forecasts lead subsequent forecasts. For each estimate in a release, we measure the ratio of the distance of subsequent estimates from the current estimate over the distance of subsequent estimates from the consensus of previous estimates. A lower ratio means subsequent estimates are dragged toward the current estimate. In other words, a lower ratio indicates a leading estimate. Then we count the number of times each user's estimates are identified as leading (i.e., the ratio of the estimate is among the lowest three for that release) and normalize the count by the total number of submitted estimates by the user as the probability of being a leader.

The measures for users who submit fewer than 20 forecasts are assigned to the lowest value. Users who rank above the 80th percentile on the measure are identified as influential users. None of the four criteria gives a complete description of an influential user, yet when we find consistent results across all four criteria, we are confident that we are indeed capturing many influential users.

Table 8 examines how influential users affect subsequent users in their relative weighting of public and private information at the forecast level. The key independent variable of interest is a triple interaction term among *Dev*, the release-view dummy, and an influence dummy variable that equals one when a large number of influential users have made forecasts. As in Table 2, we find a negative coefficient on the interaction term between *Dev* and the release-view dummy so that viewing of the release page is associated with more weight on the consensus and less weight on the private signal. More importantly, the coefficient on the triple interaction term is negative and significant. In other words, when the current release page includes the forecasts of influential users, viewing of this page is associated with placing even more weight on the consensus and less weight on the private signal. Simply put, users herd more with influential users.

5.2. Predictable Forecast Error

Because influential users issue more accurate earnings estimates on average, herding with influential users may not always result in a less accurate consensus forecast. Given that influential users' forecasts strongly swing subsequent forecasts, we conjecture that, if influential users' early forecasts are inaccurate, this is likely to drag the consensus in the wrong

Table 8. Impact of Influential Users on the Weighting of Information

Dependent variable	Forecast error (= forecast − actual)			
Measure of influential users	PageRank (1)	Number of releases (2)	Number of releases viewed (3)	Probability of being leader (4)
Dev	0.496***	0.496***	0.485***	0.498***
(= forecast − preconsensus)	(73.26)	(74.60)	(71.10)	(75.61)
Dev × *view dummy*	−0.206***	−0.214***	−0.200***	−0.210***
	(−28.31)	(−29.92)	(−27.28)	(−29.53)
Dev × *view dummy* × *influenced*	−0.134***	−0.125***	−0.153***	−0.140***
	(−10.95)	(−9.93)	(−12.57)	(−10.97)
Dev × *influenced*	0.0940***	0.108***	0.128***	0.107***
	(8.13)	(8.99)	(11.07)	(8.73)
View dummy × *influenced*	−0.00341***	−0.00249***	−0.00193**	−0.00192**
	(−3.74)	(−2.72)	(−2.11)	(−2.09)
Influenced	0.00202**	0.000808	0.00106	0.000873
	(2.26)	(0.89)	(1.21)	(0.97)
View dummy	0.00262***	0.00214***	0.00184***	0.00185***
	(4.17)	(3.41)	(2.95)	(2.97)
Release effect	Yes	Yes	Yes	Yes
Profession effect	Yes	Yes	Yes	Yes
Observations	33,264	33,264	33,264	33,264
R^2	0.920	0.920	0.920	0.920

Notes. The table presents the results of forecast-level weighting regressions. The dependent variable is forecast error, defined as the difference between a user's forecasted EPS and the actual EPS. The main independent variables include (1) *Dev*: forecast's distance from the consensus prior to the submitted forecast, (2) *View dummy*: dummy variable for forecasts made after viewing the release page for longer than five seconds at least once, (3) *Influenced*: dummy variable that is equal to one if the number of influential users ahead of the observed user is above the 80th percentile across all observations and the interaction terms among these three variables. To identify influential users, we consider four measures: (1) PageRank, (2) number of releases, (3) number of releases viewed, (4) probability of being a leader. The measures for users who submit fewer than 20 forecasts are assigned to the lowest value. The users who rank above the 80th percentile on the measure are identified as influential users. Standard errors are in parentheses and double-clustered by sector and quarter.

***, **, * = significant at the 1%, 5%, and 10% levels.

direction. To identify such a forecasting error ex ante, we use the contemporaneous stock return as a proxy for the information content and compare the direction of influential users' forecast revisions against the sign of the contemporaneous stock return. If their signs are consistent, then the revision is likely to be informative; if they are opposite one another, then the revision is likely to contain an error.

To examine directly how influential users' forecasts affect subsequent forecasts, we again separate the forecasting period into earlier and close-to-announcement periods as in Table 4. In panel A of Table 9, we then regress the consensus forecast revisions in the later period (the close-to-announcement period) on influential users' forecast revisions in the earlier period. Across all four definitions of influential users, we find very consistent results. If influential users issued forecasts that are higher (lower) than the current consensus in the earlier period, the consensus moves up (down) in the later period, confirming that influential users' forecasts strongly swing subsequent forecasts.

In panel B, we find that, when the contemporaneous stock return is negative and influential users issue forecasts that are lower than the current consensus, the final consensus becomes more accurate, consistent with the notion that influential users facilitate the incorporation of negative information. On the other hand, when the contemporaneous stock return is negative and influential users nevertheless issue forecasts that are higher than the current consensus, the final consensus becomes less accurate. In this case,

Table 9. Predicting the Change in Consensus Error and Change in Consensus Accuracy from Early to Close-to-Announcement Period

Panel A: Predicting change in consensus error				
Dependent variable	Change in consensus error			
Measure of influential users	PageRank (1)	Number of releases (2)	Number of releases viewed (3)	Probability of being leader (4)
ln(1 + num of upward revision, neg CAR)	0.003**	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of upward revision, pos CAR)	0.006***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of downward revision, pos CAR)	−0.005***	−0.005***	−0.005***	−0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of downward revision, neg CAR)	−0.007***	−0.007***	−0.007***	−0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,988	2,004	2,004	2,004
R^2	0.098	0.102	0.100	0.102
Panel B: Predicting change in accuracy				
Dependent variable	Change in Abs(consensus error)			
ln(1 + num of upward revision, neg CAR)	0.002*	0.003**	0.003**	0.003**
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of upward revision, pos CAR)	−0.001	−0.001	−0.000	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of downward revision, pos CAR)	−0.001	−0.001	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
ln(1 + num of downward revision, neg CAR)	−0.002*	−0.002*	−0.002*	−0.002*
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	1,988	2,004	2,004	2,004
R^2	0.031	0.028	0.028	0.028

Notes. The table presents the results of release-level regressions. Panel A regresses the change in the consensus error from the early to the close-to-announcement period on four variables constructed on forecasts made by influential users in the early period. All forecasts made by influential users in the early period are sorted into four groups by two dimensions: (1) whether the forecast leads to an upward or a downward revision of the consensus, and (2) whether the cumulative abnormal returns (CAR) on the corresponding day of the forecast are positive or negative. The main independent variables are the logarithm of one plus the number of forecasts in each group. Panel B uses the same set of independent variables, and the dependent variable is the change in the absolute value of the consensus error. The close-to-announcement window is defined as from five days before the announcement date through the announcement date ([−5, 0]). The early window is defined as days prior to day −5. Standard errors are in parentheses and double-clustered by sector and quarter.

***, **, * = significant at the 1%, 5%, and 10% levels.

influential users' forecasts likely reflect positive sentiments that spread to subsequent users and drag the consensus in the wrong direction.

6. Conclusion

As the idea of the wisdom of crowds has infused new life into the investment research industry over the past few years and provides incrementally useful information, we take a further step to shed light on how collective intelligence could be harnessed to provide better earnings forecast consensus. The wisdom of crowds hinges on each crowd member making independent estimates. In most earnings forecast settings, however, estimates are elicited in a sequential basis. Because participants learn from observing each other, they also exert influence on each other, and herding behavior arises, resulting in the loss of useful private information.

In this paper, we empirically examine the impact of herding on the accuracy of earnings forecast consensuses from a crowdsourcing corporate earnings forecast platform, Estimize.com. By monitoring user information sets and tracking user viewing activities, we find that the more public information users view, the more they underweight their private information. Although this improves the accuracy of the individual's forecast, it reduces the accuracy of a consensus forecast, as useful private information is prevented from entering the consensus, consistent with herding. We also find that herding behavior becomes more severe if the public information set includes the estimates of more influential users.

A randomized experiment offers clean evidence that the wisdom of crowds can be better harnessed by encouraging independent voices among participants. Ironically, by limiting the crowd's access to information, we can actually improve the accuracy of their consensus forecast. Motivated by our findings, Estimize.com decided to switch to a blind forecast platform, in which forecasts from all other users are always blocked initially. Results show that, after switching to the blind platform, the Estimize consensus beats the Wall Street consensus with an even higher probability than the before-experiment period. We are confident that, by adopting such a blind forecast platform, Estimize.com will continue to generate more accurate corporate earnings forecasts, which are crucial for the efficiency and function of the financial market.

Acknowledgments

The authors thank Renee Adams, Kenneth Ahern (discussant), Qi Chen, Stefano DellaVigna, Erik Eyster (discussant), Cary Frydman, Itay Goldstein, Umit Gurun, David Hirshleifer, Harrison Hong, Byoung-Hyoun Hwang, Russell Jame (discussant), Petri Jylha (discussant), Peter Kelly, Tse-Chun Lin (discussant), Yin Luo, Davud Rostam-Afschar (discussant), Jacob Sagi (discussant), Johan Sulaeman (discussant), Adam Szeidl, Baolian Wang (discussant), Chishen Wei (discussant), Holly Yang (discussant), Liyan Yang, and Ezra Zuckerman; as well as seminar participants at Arizona State University, Cambridge University, Emory University, HEC-Paris, London School of Economics, Shanghai Advanced Institute of Finance, Tsinghua University, the University of Arizona, the University of Edinburgh, the University of Georgia, the University of Notre Dame, Villanova University, Washington University in St. Louis, the 2016 FSU SunTrust Beach Conference, the 2016 SFS Cavalcade, the 2016 ABFER 4th Annual Conference in Singapore, the 2016 CEIBS Finance Conference, the 2016 WFA, the 2016 CICF, the 2016 Early Career Behavioral Economics Conference, the NBER PRIT workshop, the 2016 Helsinki Finance Submit, the 2016 European Finance Association meeting, the 27th Annual Conference on Financial Economics and Accounting, the 2016 CalTech Behavioral Finance Junior Faculty Conference, the 2017 Utah Winter Finance Conference, the 2017 Finance Down Under Conference, and the Second Annual News and Financial Markets Conference at Columbia Business School for their helpful comments and suggestions. The authors thank Leigh Drogen and Josh Dulberger of Estimize for their generous support.

Endnotes

[1] See Sunstein (2005) for a general survey of group judgments.

[2] As forcefully put by Bikhchandani et al. (1992, p. 1009), "The social cost of cascades is that the benefit of diverse information sources is lost. Thus a cascade regime may be inferior to a regime in which the actions of the first n individuals are observed only after stage $n + 1$."

[3] Throughout the paper, we follow Hirshleifer and Teoh (2003) and use the term "herding" in a broad sense to refer to situations in which individuals place positive weights on other people's estimates when forming their own estimates. This behavior can be completely rational when the individual computes the weights using Bayes' rule (see Banerjee (1992) and Bikhchandani et al. (1992), among others). Individuals can also herd by underweighting their private signals (see Scharfstein and Stein (1990), Banerjee (1992), Bikhchandani et al. (1992), Trueman (1994), Hong et al. (2000), Welch (2000), and Clement and Tse (2005) among others). Alternatively, they can antiherd by overweighting their private signals (see Ehrbeck and Waldmann (1996), Bernhardt et al. (2006), and Ottaviani and Sorensen (2006) among others). Finally, they can apply other naive weights (see Eyster and Rabin (2010, 2014) among others) or be subject to persuasion bias as in DeMarzo et al. (2003). We leave the interesting question of differentiating among these various forms of herding behaviors to future research.

[4] See Trueman (1994) and Graham (1999) among others.

[5] See http://blog.estimize.com/post/133094378977/why-the-estimize-platform-is-blind, accessed November 12, 2015.

[6] Hirshleifer and Teoh (2003) review several possible sources, including (1) payoff externalities, (2) sanctions upon deviants, (3) preference interactions, (4) direct communication, and (5) observational influence as well as empirical evidence for herding behavior in securities trading, security analysis, firm investment, financing, and reporting decisions.

[7] The blind forecasting environment may also improve group judgment by eliminating other inefficient strategic behavior. For example, it has been shown that, with a convex payoff, individuals may even antiherd. In other words, they may exaggerate their private signals to stand out from the crowd (see Ehrbeck and Waldmann (1996), Bernhardt et al. (2006), and Ottaviani and Sorensen (2006) among others). Because the Estimize scoring method also penalizes a bold

forecast exponentially if it turns out to deviate in the wrong direction, antiherding behavior is not prevalent on Estimize.com. Nevertheless, the blind forecasting environment also prevents antiherding by hiding information about the crowd.

[8] Field studies by Barber et al. (2003), Adams and Ferreira (2010), Charness et al. (2011), and Charness and Sutter (2012), among others, all demonstrate that group decisions are moderate and reason-based.

[9] According to Estimize.com, "The number of points received is determined by the distance of your estimate to the reported results of the company and the distribution of all other estimates for that earnings release. The system incentivizes aggressive estimation by awarding points on an exponential scale. While being a little more accurate than Wall Street may score you a few points, an aggressive estimate well outside the mean will have both a higher risk, and a far higher reward."

[10] The profile information, although voluntarily provided, should be reasonably reliable. When new analysts contribute to Estimize, they are put through a manual review process that considers the depth of their biographical information and the reliability of their first five estimates.

[11] The two data sets exploit different identifiers for users. We first use the time stamp of forecast creation activities in both data sets to construct a table to link the two identifiers. We set a five-second cutoff because we want to exclude instances when a user just passes a page to access the next one. We obtain similar results when we use other cutoff points and when we do not use a cutoff.

[12] According to Estimize.com, it flags forecasts and does not include them in the Estimize consensus if they have been manually or algorithmically unreliable or if they have not been revised within the past 60 days and fall well outside the current consensus. About 2.5% of all estimates made on the platform are determined unreliable.

[13] Only 1,953 of 2,147 release-level observations are successfully matched with data from Compustat.

[14] The size group and B/M group are obtained by matching each release with one of 25 size and B/M portfolios at the end of June based on the market capitalization at the end of June and B/M, the book equity of the last fiscal year end in the prior calendar year divided by the market value of equity at the end of December of the prior year.

[15] Without the interaction terms, β_0 is 0.18, similar to the value reported by Chen and Jiang (2006) who examine sell-side equity analysts.

[16] We also examined a more conservative matching procedure in which the default estimate is always chosen during the 24 hours *after* the blind estimate. To the extent that a more recent estimate is usually more accurate, this match procedure is biased against the blind estimate. We find similar results under this alternative approach.

[17] Users selected to participate in the experiment see a pop-up message box on the blind release page explaining the purpose and details of the experiment. In the pilot experiment, half of the users in the blind group revised their forecasts within 10 minutes after the release page was re-enabled, whereas, in the second experiment, the revised forecast lags the blind forecast by two days on average. Because new information may have arrived during that gap, the revised forecast may be less comparable to the blind forecast in the second experiment. As a result, the data from the pilot experiment provide a cleaner setting for the blind versus revised analysis.

References

Adams R, Ferreira D (2010) Moderation in groups: Evidence from betting on ice break-ups in Alaska. *Rev. Econom. Stud.* 77(3): 882–913.

Adebambo BN, Bliss B (2015) The value of crowdsourcing: Evidence from earnings forecasts. Working paper, University of California, San Diego.

Ahern K (2017) Information networks: Evidence from illegal insider trading tips. *J. Financial Econom.* 125(1):26–47.

Anderson LR, Holt CA (1997) Information cascades in the laboratory. *Amer. Econom. Rev.* 87(5):847–862.

Banerjee AV (1992) A simple model of herd behavior. *Quart. J. Econom.* 107(3):797–818.

Barber B, Heath C, Odean T (2003) Good reasons sell: Reason-based choice among group and individual investors in the stock market. *Management Sci.* 49(12):1636–1652.

Bartov E, Faurel L, Mohanram PS (2018) Can Twitter help predict firm-level earnings and stock returns? *Accounting Rev.* 93(3):25–57.

Bernhardt D, Campello M, Kutsoati E (2006) Who herds? *J. Financial Econom.* 80(3):657–675.

Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J. Political Econom.* 100(5):992–1026.

Charness G, Sutter M (2012) Groups make better self-interested decisions. *J. Econom. Perspect.* 26(2):157–176.

Charness G, Karni E, Levin D (2011) Individual and group decision making under risk: An experimental study of Bayesian updating and violations of first-order stochastic dominance. *J. Risk Uncertainty* 35(2):129–148.

Chen H, De P, Hu Y, Hwang B-H (2014) Wisdom of crowds: The value of stock opinions transmitted through social media. *Rev. Financial Stud.* 27(5):1367–1403.

Chen Q, Jiang W (2006) Analysts' weighting of private and public information. *Rev. Financial Stud.* 19(1):319–355.

Cheong FS, Thomas J (2011) Why do EPS forecast error and dispersion not vary with scale? Implications for analyst and managerial behavior. *J. Accounting Res.* 49(2):359–401.

Clement MB, Tse SY (2005) Financial analyst characteristics and herding behavior in forecasting. *J. Finance* 60(1):307–341.

Cote JM, Sanders DL (1997) Herding behavior: Explanations and implications. *Behav. Res. Accounting* 9(1):20–45.

DeMarzo P, Vayanos D, Zwiebel J (2003) Persuasion bias, social influence, and uni-dimensional opinions. *Quart. J. Econom.* 118(3): 909–968.

Ehrbeck T, Waldmann R (1996) Why are professional forecasters biased? Agency versus behavioral explanations. *Quart. J. Econom.* 111(1):21–40.

Eyster E, Rabin M (2010) Naive herding in rich-information settings. *Amer. Econom. J. Microeconomics* 2(4):221–243.

Eyster E, Rabin M (2014) Extensive imitation is irrational and harmful. *Quart. J. Econom.* 129(4):1861–1898.

Glazer J, Kremer I, Perry M (2017) Crowd learning without herding: A mechanism design approach. Working paper, University of Warwick, Coventry, UK.

Goldstein I, Yang L (2019) Good disclosure, bad disclosure. *J. Financial Econom.* 131(1):118–138.

Graham J (1999) Herding among investment newsletters: Theory and evidence. *J. Finance* 54(1):237–268.

Hirshleifer D, Teoh SH (2003) Herd behaviour and cascading in capital markets: A review and synthesis. *Eur. Financial Management* 9(1):25–66.

Hong H, Kubik JD (2003) Analyzing the analysts: Career concerns and biased earnings forecasts. *J. Finance* 58(1):313–351.

Hong H, Kubik JD, Amit S (2000) Security analysts' career concerns and herding of earnings forecasts. *RAND J. Econom.* 31(1): 121–144.

Jame R, Johnston R, Markov S, Wolfe MC (2016) The value of crowdsourced earnings forecasts. *J. Accounting Res.* 54(4):1077–1110.

Kubler D, Weizsacker G (2004) Limited depth of reasoning and failure of cascade formation in the laboratory. *Rev. Econom. Stud.* 71(2): 425–441.

Lin H-W, McNichols MF (1998) Underwriting relationships, analysts' earnings forecasts and investment recommendations. *J. Accounting Econom.* 25(1):101–127.

Michaely R, Womack KL (1999) Conflict of interest and the credibility of underwriter analyst recommendations. *Rev. Financial Stud.* 12(4):653–686.

Ottaviani M, Sorensen PN (2006) The strategy of professional forecasting. *J. Financial Econom.* 81(2):441–466.

Pelster M, Breitmayer B, Massari F (2017) Swarm intelligence? Stock opinions of the crowd and stock returns. Working paper, Poderborn University, Poderborn, Germany.

Scharfstein DS, Stein JC (1990) Herd behavior and investment. *Amer. Econom. Rev.* 80(3):465–479.

Shue K, Townsend R (2018) Can the market multiply and divide? Non-proportional thinking in financial markets. Working paper, Yale University, New Haven, CT.

Sunstein CR (2005) Group judgments: Statistical means, deliberation, and information markets. *New York University aw Rev.* 80(3): 962–1049.

Surowiecki J (2005) *The Wisdom of Crowds* (Anchor Books, New York).

Trueman B (1994) Analyst forecasts and herding behavior. *Rev. Financial Stud.* 7(1):97–124.

Welch I (2000) Herding among security analysts. *J. Financial Econom.* 58(3):369–396.