Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows*

Samuel M. Hartzmark
University of Chicago
Booth School of Business

Abigail B. Sussman
University of Chicago
Booth School of Business

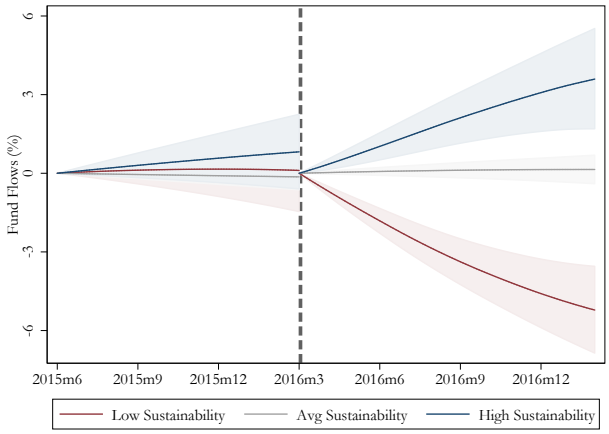
May 4, 2018

Abstract

Examining a shock to the salience of the sustainability of $8 trillion of mutual funds, we present causal evidence that investors marketwide value sustainability. Being categorized low sustainability resulted in net outflows of more than $12 billion while being categorized high sustainability led to net inflows greater than $24 billion. Investors reacted to extreme categories, ignoring middle categories and rating details, demonstrating that categorization makes extreme features salient, with marketwide impact. Experimental evidence suggests that sustainability is viewed as positively predicting future performance, but we do not find evidence that high sustainability funds outperform low sustainability funds.

*We are grateful to Jonathan Berk, Alex Edmans, Karl Lins, Vikas Mehrotra, Sanjog Misra, Jacopo Ponticelli, Brad Shapiro, David Solomon, Kelly Shue and Eric Zwick and seminar participants at Aalto, Emory, Cambridge, Chicago Booth, Warwick, London School of Economics, Bernstein Quantitative Finance Conference, Development Bank of Japan Conference, Texas Finance Festival for comments. We thank Halley Bayer, Nicholas Herzog, and Nathaniel Posner for excellent research assistance. We thank Ray Sin, Steve Wendel, and Sara Newcomb at Morningstar for providing the data. This work is supported by the True North Communications, Inc. Faculty Research Fund at the University of Chicago Booth School of Business.

Figure 1



Cumulative fund flows in percent by sustainability rating for 9 months before and 11 months after rating publication (denoted by the dashed vertical line). Estimates accumulated from local linear plot of monthly flows after removing year by month fixed effects. Shaded areas indicate the 90% confidence interval.

As firms invest more resources in sustainable and socially responsible endeavors, it is important to know whether such investments reflect investor's preferences marketwide. Some investors will believe that an increase in resources directed towards sustainability is costly and belies the primary goal of maximizing profits. Others will believe that a well run company should care about the environment or that companies should act for reasons beyond simple value maximization. Others still will value such an investment not because they inherently care about the environment, but because they view it as a sound way to maximize profit. And finally, some investors will be unaware that a firm is investing in sustainability or will not care. While surely the market contains examples of each of these investors, it remains unclear which type represents the average investor and thus it is unclear whether investments in sustainability are consistent with what investors want. Put simply, do investors collectively view sustainability as a positive, negative, or neutral attribute of a company?

This paper demonstrates that the universe of mutual fund investors in the US collectively put a positive value on sustainability by providing causal evidence that marketwide demand for funds varies as a function of their sustainability ratings. Directly addressing this question is difficult in most settings, as it is unclear how to identify the preferences of the average investor. Furthermore, market outcomes related to firm attributes, such as sustainability, are usually viewed in equilibrium

where analysis is by necessity indirect. Analysis of investment products with an explicit sustainability focus only reflects the preferences of the subset of investors holding those products, but does not speak to the average preferences of investors in the entire market.

We circumvent these challenges by examining a novel natural experiment where the salience of the sustainability of over $8 trillion of mutual fund assets experienced a large shock. Sustainability went from being difficult to understand to being clearly displayed and touted by one of the leading financial research websites, Morningstar. In March of 2016, Morningstar first published sustainability ratings where more than 20,000 mutual funds were ranked on a percentile basis and given a globe rating based on their holdings. The worst 10% of funds were rated one globe (low sustainability) while the best 10% were rated five globes (high sustainability). The publication was not expected and prior to it there was not an easy way for investors to judge the sustainability of most mutual funds without considerable effort.

Figure 1 illustrates the main finding of the paper: mutual fund investors collectively treat sustainability as a positive fund attribute, allocating more money to funds ranked five globes and less money to funds ranked one globe. Moderate ratings of either two, three, or four globes did not significantly affect fund flows. The dashed vertical line indicates the initial publication of the sustainability ratings. To the left of the line, fund flows after controlling for monthly fixed effects are accumulated over the 9 months prior to the rating publication and to the right of the line flows are accumulated for the 11 months post publication. The navy line represents five globe funds, the maroon line one globe funds and the gray line those rated in the middle (two to four globe funds). Prior to the rating publication, the funds were receiving similar levels of flows. After the publication, the funds rated highest in sustainability experienced substantial inflows of roughly 4% of fund size over the next 11 months. On the other hand, funds rated lowest in sustainability experienced outflows of about 6% of fund size. Over the 11 months after the sustainability ratings were published, we estimate between 12 and 15 billion dollars in assets left one globe funds and between 24 and 32 billion dollars in assets entered five globe funds as a result of their globe rating.

Our experiment is rare in financial markets in that it examines a large quasi-exogenous shock,

equivalent to approximately 40% of NYSE market cap, that does not directly impact fundamentals. The shock yields easy to understand measures of sustainability by simply repackaging publicly available information in a form that attracts attention and is easy to process. Further, the construction of the measure is based on within-category comparisons that rely on Morningstar's own classification of funds, so it is unlikely to be highly correlated with other general measures of sustainability.[1] Thus our measured response is to the rating itself, not to new information about fund fundamentals. In addition, examining mutual funds rather than individual stocks allows us to directly observe fund flows. This allows us to avoid focusing on indirect measures, such as prices, which suffer from the joint hypothesis problem that they could be capturing risk.

This shock allows us to identify the causal impact of the globe rating along a variety of different margins. If funds were systematically different before the publication of the ratings, then flows could be reflecting this difference. The initial figure suggests this is not the case, and indeed, a matching exercise based on fund characteristics before the ratings publication finds similar results, suggesting that pre-period differences do not account for our results. Further, as a placebo we construct pseudo globe ratings for funds in years prior to the Morningstar ratings publication and we do not find similar effects based on these pseudo ratings.

The globes are a discrete rating system of five categories, though Morningstar also released each fund's sustainability score and the within category percentile ranks underlying the ratings. If investors responded to the five globe rating system rather than to other aspects of sustainability, we should find it is the globe category itself that drove the mutual fund flows. Examining the percentile ranks that underlie the sustainability rating, we find evidence consistent with discontinuities at the extreme globe category edges, but find minimal impact of the percentiles themselves. This suggests that investors focused on the simple globe rating and ignored the more detailed sustainability information.

We find strong flow effects from being in the two extreme globe categories (i.e., one or five globe

[1] Put another way, *Barron's* noted that funds rated high sustainability by Morningstar were not "whom you'd associate with even a faint whiff of patchouli." http://www.barrons.com/articles/the-top-200-sustainable-mutual-funds-1475903728

funds) relative to the three categories in the middle, but find insignificant differences across funds receiving two, three, or four globe ratings. This is consistent with prior evidence that investors often focus on discrete rather than continuous measures and that when they do so they focus on extreme outcomes (e.g. Hartzmark 2015; Feenberg et al. 2017).[2] It underscores the general importance of salience on investment decisions (e.g. Bordalo et al. 2012; Bordalo et al. 2013a) as well as the impact of attributes that stand out in consumer choice (Bordalo et al. 2013b). These findings suggest that evaluating information based on extreme ranks reflects a fundamental cognitive process underlying decision making that impacts the market.

The large causal flow response we observe allows us to reject both the hypothesis that investors are indifferent to sustainability as well as the hypothesis that they view sustainability as a negative characteristic, but it remains unclear as to what specific aspect of sustainability drove investors to reallocate funds from one globe funds to five globe funds. While we are unable to definitively pinpoint the specific motive, we explore the importance of three possibilities. The first is that institutional pressure, either to hold high sustainability stocks or not to hold low sustainability stocks is responsible for the results. We find that fund flows from institutional share classes in response to the globe rating are similar to those from other share classes. This could be evidence that investors in institutional share classes face constraints that force them to behave like other investors, or that their preferences are similar to that of other investors. Since non-institutional share classes display a similar pattern, institutional constraints cannot fully account for the finding.

Another possible explanation is that investors rationally view a rating of high sustainability as a signal of high future returns. We examine whether funds experienced high returns after their high sustainability ratings relative to a variety of benchmarks and find evidence more consistent with the opposite or no relation. While it is difficult to make definitive statements using only 11 months of

[2]More broadly, our findings are consistent with literature in psychology and economics that model rank dependent preferences (e.g., cumulative prospect theory; Tversky and Kahneman 1992), and with the corresponding intuition that extreme ranks are the most perceptually salient positions (Diecidue and Wakker 2001; Tversky and Kahneman 1986). See also Quiggin (1982) and Schmeidler (1989) for early rank-dependent models of risk under uncertainty and Weber and Kirsner (1997)for an examination of why people rely on extreme rank in evaluations. Furthermore, it is consistent with existing literature showing that people overweight extreme attributes when making judgments about people (Skowronski and Carlston 1989) and make choices to avoid products with attributes ranked in extreme positions when confronted with tradeoffs (Simonson and Tversky 1992; Tversky and Simonson 1993).

data, we find marginally significant evidence suggesting that one globe funds outperform five globe funds after the publication of the sustainability ratings.

If the results are not driven purely by institutions or a rational belief in higher expected returns, then some investors want to hold high sustainability funds and avoid low sustainability investments either due to an irrational belief that there is a positive correlation between future returns and sustainability or for non-pecuniary motives (such as altruism, warm glow or social pressure). Unfortunately the data does not allow us to distinguish between these two possibilities, so we run an experiment using MBA students and MTurk participants. We elicit expectations about future performance, risk and investment decisions as a function of globe ratings. We find a strong positive relation between globe ratings and expected future performance and a strong negative relation between globe ratings and expected riskiness. We also find some evidence of non-pecuniary motives across both populations. Subjects considering environmental or social factors when making their decision invest more money in five globe funds and less money in one globe funds than their expectations for future performance and risk can account for, while those not considering such factors do not exhibit such a pattern. The results suggest that globe ratings impact expectations of future performance and also lead investors to make choices based on non-pecuniary motivations.

Our paper contributes to the literature that has examined how investors value non-financial aspects of stocks. While other studies have examined how subsets of investors value characteristics of securities, such as whether it is a "sin" (Hong and Kacperczyk 2009), local (Huberman 2001) or offers a certain dividend yield (Harris et al. 2015), our study has the benefit of examining a quasi-exogenous shock which means we can measure how all mutual fund investors collectively value the characteristic, rather than the subset that hold the security. Perhaps most closely related to our paper, Hong and Kacperczyk (2009) find that sin stocks yield higher returns, consistent with investors needing to receive a premium to hold these companies due to social norms. Our paper complements this finding by examining an exogenous shock to a significantly larger portion of the market with a more direct measure of demand.

A recent literature has examined the rapidly growing set of investment products with explicit

mandates of social responsibility (e.g. Bialkowski and Starks 2016; Barber et al. 2017; Benson and Humphrey 2008; Bollen 2007; Geczy et al. 2005; Riedl and Smeets 2017). While understanding the preferences underlying such investments represents an important area of research, it is only indicative of the investors selecting into this subset of products (roughly 2% of funds in our sample) and need not be representative of investors or funds marketwide. If a small subset of investors had strong preferences for sustainability while most investors in the market did not directly value sustainability, under standard models we would not expect to find an effect of the ratings on net flows.[3] The investors that value sustainability would move their investments into the high sustainability funds, this would push these funds above their optimal scale and the investors that did not value sustainability would move their investments to other funds. Thus our paper contributes to this literature by examining the preferences for sustainability of the universe of US mutual fund investors into products lacking explicit sustainability goals.

Additionally, our paper contributes to the literature on why firms invest in sustainability, and more broadly to investment in "doing well by doing good."[4] Some sustainable investing is clearly due to agency issues (Cheng et al. 2013) while others have argued that it is consistent with efficient investment, for example by improving morale (Edmans 2011). As emphasized by Hart and Zingales (2017), investments for non-pecuniary "pro-social" reasons, such as sustainability, are something that companies should engage in if they reflect the preferences of their shareholders. While our paper does not break down the fraction of sustainability that is due to agency versus appeasing shareholders, a general demand for sustainability from mutual fund investors suggests that a significant portion of the observed investment in sustainability is not purely due to agency issues.

Finally, the evidence highlights the potential role of emotion in guiding investment decisions. Specifically, although it may seem surprising that higher globe funds are associated with expectations of both higher returns and lower risk, this pattern is consistent with research on the affect heuristic

[3] E.g., under the assumptions of Berk and Green (2004) where funds were at their optimal scale prior to the ratings, the inflows would push high sustainability funds above that scale and the investors that did not value sustainability would reshuffle to the funds that the high sustainability investors vacated as they would be below their optimal scale.

[4] For recent overviews see: Bénabou and Tirole (2010); Heal (2005); Kitzmueller and Shimshack (2012); Margolis et al. (2009); Christensen et al. (2017); Chowdhry et al. (2017).

(e.g., Slovic et al. 2004, 2005, 2007; Finucane et al. 2000), which finds that feelings associated with a given stimulus often take the place of more reasoned analysis and guide subsequent judgments and decisions about the stimulus. While the affect heuristic has been prominent within psychology literature in discussions of risk evaluations, its role in decisions about financial products has received minimal attention in the context of financial products.[5] Thus, an additional contribution of the current work is to highlight the consequential role of affect versus analytic thought in financial decision making and financial markets as a whole.

1 Sustainability Ratings

On March 1, 2016 Morningstar launched the Morningstar Sustainability Rating. The company classified more than 20,000 mutual funds, representing over $8 trillion dollars in market value, into a simple rating between one and five globes. The rating system was designed to provide "a reliable, objective way to evaluate how investments are meeting environmental, social, and governance challenges. In short, it helps investors put their money where their values are."[6]

The classification system is based on the underlying holdings of a given mutual fund. Each holding is given a sustainability score based on research of public documents undertaken by the company Sustainalytics. This rating is related to how a firm scores on environmental, social and governance issues (ESG). At the end of each month, Morningstar takes the weighted average of this measure based on holdings to form a mutual fund specific sustainability score.[7] Each fund in a Morningstar category[8] is ranked based on their sustainability score and this ranking serves as the basis of the main measure of sustainability, the Morningstar globe ranking. According to the documentation, a fund is given five globes and rated as "High" if it is in the top 10% of funds in the category. It is given four globes and rated as "Above Average" if it is ranked between 10% and 32.5%. It is given three globes and rated "Average" if it is ranked between 32.5% and 67.5%. It is

[5] For an exception investigating the role of advertising for mutual funds see Jordan and Kaas, 2002.

[6] http://news.morningstar.com/articlenet/article.aspx?id=745467

[7] Complete details of the methodology can be found at: https://corporate1.morningstar.com/Morningstar-Sustainability-Rating-Methodology-2/

[8] For example, categories include Equity Large Growth, Equity Energy, and US Corporate Bond.

given two globes and rated "Below Average" if it is ranked between 67.5% and 90%. It is given one globe and rated "Low" if it is ranked in the bottom 10% of its fund category.[9] The globe ranking is prominently reported using pictures of one to five globes as well as the descriptive label (e.g., "High") on each fund's Morningstar page. The percentile rank in category and raw sustainability score are displayed in smaller text alongside the rating, see Figure 2 for an example.

While Morningstar's definition of sustainability is a precise formula transforming holdings and ESG ratings into a globe rating, "sustainability" has generally become a popular term that lacks a clear and consistent definition. An investor that wished to understand the details of Morningstar's system could easily do so, but it is likely that a number of investors responded not to the specific details of the rating methodology, but based on their preconceived notion of the meaning of sustainability. Thus it is useful to more precisely understand how investors interpret sustainability.

Therefore, we recruited 482 participants from an online sample and asked them which elements of a company's business practices they believe "sustainability" refers to.[10] The results are reported in Table 2. The dominant answer was that sustainability referred to a company's practices with regard to the environment, with 79% of participants including environmental issues in their definition of sustainability. Subjects included a number of other aspects of a company, but none other garnered more than 50% of responses. In total, participants listed 2.7 items on average, with less consistency in the selection of the additional items.[11] While the meaning of sustainability varied among participants, there was not confusion as to what their definition was. Only 2% of participants listed that they did not know what was meant when a company's business practices became more sustainable.

[9] A coding error included 11% of the data in the one globe category.

[10] Participants selected as many options as desired from the following list: Corporate Governance, Community, Diversity, Employee Relations, Environment, Human Rights, Products, Other, and I don't know. We chose these options because they are the dimensions by which KLD Research & Analytics, Inc, a leading provider of social investment research, evaluates companies on environmental, social, and governance issues.

[11] e.g., the next most popular item- product quality and safety- was listed by only 48% of people.

2 Data Sources and Summary Statistics

All of the mutual fund data is provided by Morningstar and is at the monthly frequency.[12] The sample includes all US based open-end funds with a sustainability rating from Morningstar. The data is provided at the share class level, but the analysis is conducted at the fund level and a number of fund attributes need to be calculated for the fund level from the share class data. Fund size (TNA), dollar flows and web traffic are calculated as the sum across share classes, while expense ratios and returns are the mean of these variables across share classes. Morningstar "star" fund ratings are the rating from the largest share class and fund age is calculated from the inception date of the earliest share class. Morningstar category names sometimes vary slightly within a fund across share classes, such as having one share class labeled "OE" and and another labeled "fund." We hand clean the share class data to form consistent categories within and across funds, removing these share class specific attributes.[13] We limit the sample to funds with a value greater than one million dollars. We winsorize the continuous variables at the 1% level.

Flows are the main variable of interest in the paper and are measured as the dollar flows in a given month divided by fund TNA as of the prior months end. Flows are noisy and may be systematically different based on characteristics, such as size. To make sure the results are not being driven by the distributional properties of flows, we also examine a normalized flow variable. To construct this variable we split firms into deciles based on size at the end of the prior month and then assign each fund to percentiles in a given month within each size decile. This normalized flow variable will be inoculated from differences in flow distribution across sizes as well as the impact from extreme observations.[14]

Table 1 Panel A shows summary statistics for the funds after the publication of the sustainability ratings, March of 2016 through January of 2017. In Table 1 Panel B we show the summary statistics prior to the globe publication for each globe ranking, where globe is what each fund was eventually

[12] The data was anonymized of fund specific identifiers by Morningstar.

[13] E.g. A given fund has shareclasses with the Morningstar category "US Fund Large Value" and "US OE Large Value" which we assign to the same category US Large Value.

[14] We thank an anonymous referee for suggesting this variable.

assigned in March 2016. Both one and five globe funds tend to be smaller, which could be due to the sustainability rating becoming less extreme for funds with more diversified holdings. Examining flows, web traffic and Morningstar star ratings, we see similar patterns across funds with each globe rating, with nothing suggesting that the one and five globe funds were distinct on dimensions other than size prior to the publication of the globe rating.

In Table 1 Panel C we examine the same variables during the publication period. Over this period mutual funds experienced outflows of -0.4% per month on average, but the funds rated lowest in sustainability experienced outflows of -0.9%, while those with inflows were nearly zero. Also, examining web visits, we see that the lowest amount of web traffic was received by funds rated one globe, while the highest rated funds in sustainability received substantially more traffic than the other funds. Finally, consistent with the flows, we see that one globe funds shrank while five globe funds grew relative to their pre-publication average.

In Table 1 Panel D we examine the probability of moving to a different globe category. The sample is restricted to the post-publication period, excluding the first month where no switching was possible. In general, if a fund is ranked as a given number of globes, there is a roughly 80% chance that it will have the same rating the next month. Funds that do change categories rarely change more than one category in a given month.

3 Do Investors Value Sustainability?

3.1 Attention to Ratings

While Morningstar created these ratings because they believed there would be investor interest in them, one reasonable hypothesis is that they did not receive attention when published and thus had no impact. This could be because investors did not care about the rating, did not know about the rating, or already were aware of the information contained in the rating. The Sustainalytics score for each stock was based on publicly available information and the Sustainalytics scores themselves were also publicly available, for example through Bloomberg. Further, fund holdings were publicly

reported. Thus all of the information used to construct the globe ratings was available before the publication of the ratings. Perhaps investors already understood the information that Morningstar aggregated into a globe rating and the ratings were simply ignored.

We provide evidence based on Google searches that the globe rating system attracted significant attention at its launch, but not prior to its launch. Figure 3 shows the relative interest of monthly Google searches using Google Trends data for "Morningstar star rating" versus "Morningstar sustainability rating."[15] The star rating refers to Morningstar's popular fund rating system. Its search intensity is represented by the navy line. The maroon line represents searches for "Morningstar sustainability rating" while the vertical gray line represents the first publication of those ratings.

There are two notable aspects of Figure 3. First, before their publication, there was no measurable volume of searches for the sustainability ratings. This suggests that their publication was not anticipated, at least not by Google users. Second, subsequent to their publication, there were roughly as many Google searches for the sustainability rating as there were for the star rating. This is consistent with there being significant interest in the sustainability ratings as indicators of ESG, which were publicized through white papers, traditional marketing campaigns, included as a search filter option for some Morningstar clients, covered by outside media outlets and included on every fund's Morningstar web page. The large search volume suggests many investors were aware of the existence of the rating and were likely interested in issues related to sustainable investing.

The validity of the experiment in the paper is based on investor perception of a fund's sustainability changing in response to the rating publication. The search frequency and subsequent findings suggest that it is the publication of the ratings that induced the flow response by investors. While investors did not respond to the ratings before their publication, it is possible that mutual funds predicted their publication and traded prior to the publication in an attempt to receive a high

[15]The monthly measure is the average of the weekly searches, where month is assigned based on the month that a given week ends. Although we often refer to the ratings as "globes" in this paper, this terminology is not widely used and the rating is typically referred to as the "Morningstar sustainability rating" by Morningstar and the media. Google trends normalizes the results of every search to a different scale with the maximum search volume in a week for the term with the highest intensity normalized to 100 at its maximum. The results in Figure 3 are from a search that included both terms so the magnitudes are comparable between the two measures.

globe rating.[16] If such behavior was widespread, this would potentially impact the interpretation of some of the results of the paper related to returns (which we discuss in Section 4.2), but would not change the core results related to fund flows and investor preferences. For our flow results, the key to interpreting them is that investors had not systematically sorted into funds based on their rating marketwide before publication.

3.2 Base Results

Did the publication of the sustainability ratings impact how investors decided to trade these mutual funds? To begin answering this question we examine the mutual fund flow reaction to the publication of the ratings. The ability to study flows makes mutual funds an ideal laboratory to examine the revealed preferences of investors. If a fund is generally viewed as more desirable after its rating becomes public, money will flow to it and it will grow. If it is viewed as less desirable than we will see money flow from it and it will shrink. This stands in contrast to studying individual stocks since a stock is in fixed supply in the short run, which would not allow for such a direct measure of investor response.[17]

In addition, our setting is rare in financial markets in that we examine an event that does not change fundamentals and is unexpected. Studies of socially conscious investing generally focus on fixed firm specific traits. For example, a tobacco company tends to remain a tobacco company, and any change to such a characteristic would represent a large shift in its business. Our study examines a shock to the salience of a characteristic, so while the characteristic is fixed, there is no change to the underlying business by the publication of the fund rating.

When Morningstar published their ratings, they released three separate measures of sustainabil-

[16] For example, sustainalytics announced that they had licensed their ratings to be used by Morningstar for sustainability prior to the ratings publication (https://www.sustainalytics.com/press-release/morningstar-to-launch-first-environmental-social-and-governance-esg-scores-for-funds-globally/).

[17] Prior to the ratings publications it was difficult to ascertain a fund's sustainability without considerable effort. An exception to this is the small subset of funds, roughly 2% of our sample, with an explicit sustainability mandate. 40% of these funds were rated 5 globes, 31% 4 globes, with the rest rated 3 globes and below. In our period there were inflows to these funds of roughly 0.7% per month higher than other funds. We do not see significant variation in fund flows for these funds based on globe ratings. We do not focus on such funds due to the small sample size and because investors had sorted into these funds based on sustainability prior to the Morningstar ratings. For papers examining these funds see Bialkowski and Starks (2016); Benson and Humphrey (2008); Bollen (2007); Geczy et al. (2005).

ity that were displayed together on a fund's page as shown in Figure 2. They released a fund's raw sustainability score, the percentile rank of that score within the fund's Morningstar category, and a picture of how many globes the fund was rated based on cutoffs of that percentile rank. If investors want to invest in the most sustainable fund in the market overall, then the raw sustainability score is the most informative measure, but it is difficult to interpret without a significant amount of effort dedicated to understanding the overall distribution of sustainability scores. The percentile rank variable yields a continuous measure of within Morningstar category rank available to investors that is easier to interpret and provides more granular detail than the globe rating. If investors want to invest in the most sustainable fund in a given Morningstar category, then the percentile rank is the most informative measure. As shown in Figure 2, the globe rating is given the most space on a fund's webpage and is presented as a large picture of the number of globes along with the name associated with that category (e.g. High, Average or Low) in a larger font than either of the two measures. All of the information needed to understand the globes is included in the percentile rank variable. If investors are paying attention to the available percentile information, there is no need to pay attention to the globe rating. If investors' attention is drawn to the globe rating itself, they may simply examine this salient measure and ignore the underlying percentiles.

In Table 3, we explore the reaction to each sustainability measure by regressing mutual fund flows on these measures and find that it is the globes, rather than the other available measures that appear to be the main driver of mutual fund flows. Fund flows are measured as the dollar flows for a fund in a given month scaled by the previous month's net asset value, multiplied by 100. All regressions include Morningstar category by year by month fixed effects to control for time variation by category. In Column 1, we examine the raw sustainability score and the percentile rank in category variables. If investors cared about how sustainable a fund was relative to the rest of the market, the raw score would be the most relevant measure of sustainability. If investors cared about the relative sustainability of funds within Morningstar category the percentile rank is the most informative measure. Regressing fund flows on these measure, we see an insignificant coefficients on both. In Column 2 we include dummy variables for each globe rating omitting the three globe

category. One globe funds, the funds rated worst in terms of sustainability, experienced outflows of roughly -0.44% per month lower than three globe funds, with a t-statistic of -3.57 clustered by month. Five globe funds, those rated highest in terms of sustainability, experienced inflows of 0.30% per month higher than three globe funds, with a t-statistic of 2.48. These point estimates indicate that the lowest sustainability funds lost 5.4% of TNA per year while the highest rated funds gained about 3.6% of TNA per year. Below the regression results is the difference between one and five globe funds, of 0.74 per month with the p-value on the test that they are equal, 0.0004, underneath. The globe ratings in the middle – two and four globes – are not statistically distinct from the omitted three globe funds.

The insignificance of the two and four globe funds suggests that investors focus on extreme one and five globe categories. If this is the case than the relevant test is how one and five globe funds compare against those rated in the middle. Column 3 conducts such a test, where two, three and four globe funds comprise the omitted category. One globe funds see outflows of -0.46% per month lower than middle ranked funds with a t-statistic of -4.17 while five globe funds see inflows of 0.28% higher than middle ranked funds with a t-statistic of 2.66.

The prior results may be due to globe ratings systematically varying with other variables associated with inflows so in Column 5 we add a number of controls. We include the prior month's return, the prior 12 month return and the prior 24 month return to control for the fund-flow relation (Chevalier and Ellison 1997). To make sure the globe ratings are not simply capturing fund-flows based on size, we control for the log of fund TNA the prior month. We also add controls for the expense ratio and for log of fund age. There could be a correlation between Morningstar's globe rating and their star ratings, so we also control for the star rating. After including these controls, we find similar effects. In Column 5, one globe funds are associated with outflows of -0.40% with a t-statistic of -4.32, while five globe funds had inflows of 0.33% with a t-statistic of 3.21.

In Column 6 we include all three of the variables to understand which of the ratings drive the mutual fund flows and find that investors respond to the coarse globe ratings, not the other two variables. After including the globe rating variables, the coefficients on both the category

percentile rank and the raw sustainability score are insignificant. The coefficients on globe ratings are materially unchanged. We see that the one globe variable is negative and significant while the five globe variable is positive and significant. The regression suggests that investors responded to the globe ratings, not the other measures of sustainability. In all specifications the shift in flows is above 0.7% per month moving from one to five globe funds.

One possible concern is that the results are driven by systematic noise over the short sample period. For example, perhaps small firms have more volatile flows which drive the results purely by chance. In Panel B we examine the normalized flow variable which should not be impacted by differences in flow across the size distribution. It will also be less influenced by general noise or distributional properties of the flow data. If the results were driven by these properties, rather than the sustainability ratings, we would expect the results to decrease, or disappear in this specification. If the measure is simply reducing noise that attenuated the estimates using raw flows, the relation will be stronger in this specification as the underlying relation is the same, but noise is decreased.

The first two columns of Panel B shows the results become statistically stronger when measured using the normalized flow variable. Examining Column 2, which includes additional controls, we find that one globe funds have flows 4.4 percentiles lower than middle ranked funds with a t-statistic of -7.50 while five globe funds have inflows 3.3 percentiles higher than middle ranked funds with a t-statistic of 5.37. The spread of 7.7 percentiles between one and five globe funds has a p-value of 0 to four decimal places. Reducing the noise in measuring flows using this normalization significantly increases the statistical significance of the results, consistent with a strong response by investors based on the globe ratings themselves.

Another possible concern is that the regressions are being driven by small, relatively economically unimportant funds. In columns 3 through 6 we repeat the analysis weighting the regressions based on the log of fund size the prior month. For both measures the results are similar and get slightly stronger in point estimates. For the flow measure, one globe funds underperform middle ranked funds by -0.39% with a t-statistic of -4.41 and five globe funds outperform middle ranked funds by 0.36% with a t-statistic of 3.67. The spread between the two of 0.74% has a p-value of 0.0004.

Examining the normalized measures in Panel D, one globe funds had outflows of -4.4 percentiles with a t-statistic of -8.03 while five globe funds received inflows of 3.5 percentiles with t-statistics of 5.61. The difference between the two of 7.9 percentiles has a p-value of 0 to four decimal places.

3.3 Within Globe Rating Analysis

The results suggest that investors focus on the extreme globe ratings and largely ignore both the middle globe ratings and the available underlying sustainability information. If so, funds within a globe rating should receive similar level of flows, regardless of how different they are based on the more detailed sustainability information. Further, investors should treat funds with similar sustainability characteristics that happen to fall on different sides of an ad-hoc globe rating breakpoint quite differently, leading to discontinuities in flows around the category edges. Finally these effects should be concentrated in the extreme one and five globe categories, not the three in the middle.

Figure 4 allows us to explore these hypotheses by taking a more detailed look at the relation between fund flows, the globe rating and the underlying percentile ranks. Panel A shows the average fund flow for each percentile rank from 1 through 100 after removing a year by month fixed effect. Panel B repeats the analysis using the normalized measure. The dashed vertical lines indicate the globe cutoff levels with the category of globes listed at the top of the chart. The bars to the extreme left are five globe rated funds while those to the extreme right are one globe funds. Examining each percentile separately limits our power as each bar is populated by roughly 350 observations. Examining the ten percentiles assigned to high sustainability funds (5 globes) we see that nine of the ten point estimates are positive and five of the ten are positive and significant at the 90% level. Examining the 11 percentiles assigned to low sustainability funds (1 globe) we see that all eleven are negative and five of the eleven are negative and significant at the 90% level. Looking at the two, three and four globe categories, there is a mix of positives and negatives throughout, with no discernible pattern. Of these 79 percentile ranks, only seven are significant at the 90% level, less than the ten significant percentiles in the 21 extreme percentile categories.

Panel B repeats the analysis with the percentile rank measures and the results are if anything

stronger. Six of the five globe percentiles are positive and significant while nine of the one globe percentiles are negative and significant. Across all other percentiles there are seven that are significant.The evidence suggests investors responded to the one and five globe categories, largely ignoring the 2, 3 and 4 globe categories.

While Figure 4 presents evidence suggesting that the extreme globe ratings are largely responsible for the observed flows, it also suggests that percentile ranks were not altogether ignored. The major exception where flows appear different based on percentile ranks, but not at globe cutoffs, is the extreme low sustainability funds which received higher outflows when ranked 98th and higher in terms of sustainability. Comparing the average flow in percentiles 98 and above versus the other one globe funds yields a difference of -0.51 with a t-statistic of 3.08. Examining the normalized measure yields an estimate -7.2 percentiles lower with a t-statistic of -8.37. We see a more muted effect of being in the top percentiles of high sustainability funds. The top 3 percentiles in the high sustainability category have inflows 0.35 higher with a t-statistic of 3.64, while the normalized measure shows these funds receive inflows 3.4 percentiles higher with a t-statistic of 2.51. Thus it appears that investors again pay attention to the extreme ranked funds by percentile, but only for the most extreme ratings of sustainability.

If investors are responding to the globe ratings, the ad-hoc choice of cutoff will leave very similar mutual funds receiving different ratings on either side of the cutoff. We examine this question more formally in Table 4 using regression discontinuity analysis. We use the rank within each category as the running variable. For example, in June of 2016, there were 265 funds ranked in the US based Emerging Market funds category, and the top 26 were ranked as 5 globes. Thus, we look at the break point of the five globe funds ranked just below 26 compared to the lower globe funds with ranks greater than 26 by running discontinuity tests (e.g. Thistlethwaite and Campbell 1960; Imbens and Lemieux 2008 and DiNardo and Lee 2011). We select the bandwidth using the method from Calonico et al. (2014) using uniform windows on both side of the cutoff and also allowing for different breakpoints on each side to show the results are robust to each. We present conventional estimates as well as the bias-corrected estimator from Calonico et al. (2014).

Table 4 suggests that there are discontinuities surrounding the globe cutoffs. Panel A examines flows and Panel B examines the normalized measure of flows. Examining the first two columns of Panel A we see four estimates of roughly -0.4, with all four significant at the 5%. This suggests that moving from a two globe rating to a one globe rating leads to a discontinuous decrease in flows of roughly 0.4% per month. Examining the five globe column we see coefficients ranging from -0.55% to -0.80%, each statistically significant. This suggests that moving from a five globe category to a four globe category results in monthly flows that are about 0.6% lower per month. Panel B repeats the results using the normalized variable. The results suggest that moving from two globes to one globe leads to a decrease in flows 1.6 to 3.4 per month while moving from five globes to four globes leads to flows 2.8 to 3.4 percentiles lower.

The results suggest that investors respond to the coarse globe ratings, largely ignoring the underlying information available to them. This is consistent with the psychological literature related to categorization. A key function of categories is to organize information in the world so as to provide the most information with the least amount of effort, thus allowing people to generalize information from a single example within a category to any other category member (e.g., Malt et al. 1995; Murphy and Ross 1994; Osherson et al. 1990; Rips 1975; Rosch and Mervis 1975; Rosch et al. 1976; Rosch 1978). In this setting, each globe rating functions as its own category, with each category ranked relative to the others. Thus, rather than looking to aggregate all possibly relevant details about each company's sustainability as a method of judgment formation, investors can generalize from each fund's ranked category membership (i.e. globe rating) to infer an overall level of sustainability. The results emphasize that the formation and display of information as categories can have a significant impact on investor decision making.

3.4 Controlling for pre-period

The prior section showed that there was a high correlation between globe ratings and flows. Further, when looking more finely around globe breakpoints, we observe discontinuities when funds were assigned to one category or another. One still may be worried though that the prior section simply

captured pre-period differences in funds that were not addressed by these specifications. In this section we examine whether the globe ratings were capturing such pre-period effects and find that it is unlikely to be the case.

Figure 1 examines cumulative flows based on globe ratings, both before and after their publication. The globe ratings did not exist before they were published, so for the period before their publication every fund is assigned their first globe rating from March 2016. Raw flows are regressed on year by month fixed effects to control for time trends. The estimates are from a local linear plot are accumulated to form the plot for the 9 months before and after the rating's publication. Before publication, to the left of the dashed line, there are not significant differences across the groups and the trends are roughly similar. After the publication, we see significant increases in flows to funds rated five globes and significant outflows from funds rated one globe.

Figure 5 examines this further presenting the raw averages for each month along with a version of the local linear plot figure without accumulating the flows. Examining the simple averages during the pre-period in Panel A, there is not a clear relation. Four of the nine pre-period months have funds that will be rated one globe with higher flows than funds that will be rated five globes with the other five having the opposite pattern. Examining the smoothed local linear plots in Panel B, we see evidence consistent with these patterns. In the pre-period there is not significant difference in the flow variable, consistent with Figure 1. The confidence intervals for all three categories are overlapping in each month. After publication, the pattern becomes stronger and less volatile. The gap between the blue dots and the red dots becomes more extreme and the white space between the red and blue lines becomes significantly greater. Every month post publication the five globe funds have higher inflows than the one globe funds. The results are consistent with flows being impacted by the ratings and the funds being broadly similar before the ratings were published.

We examine this pattern more formally in Table 5 by matching funds based on their characteristics in the period before rating publication. Funds are examined based on the intent to treat, so the globe category they were initially assigned to in March 2016 is assigned for all 11 months subsequent to publication. Funds in an extreme rank are matched to other funds that had the same

Morningstar star rating as of the month prior to the rating publication. A nearest neighbor match is used based on flows, size, age and return prior to the ratings. Using this method, the results suggests that one globe funds had outflows of -0.72% (t-statistic of -9.07) which were -6.7 percentiles lower (t-statistic of -11.60) using the normalized measure. Five globe funds had inflows of 0.21% (t-statistic of 2.60) or 3.8 percentiles higher (t-statistic of 7.44).

While we matched for major fund characteristics that could account for the results, clearly there is always a concern that we are omitting a relevant variable. Thus in Panel B we include the fund's loadings on orthogonal projections of vanguard benchmarks (see Section 4.2 for details of their estimation). To the extent that similar funds covary together on a wide variety of possible characteristics, this should also help to control for the characteristics not explicitly included. Results are similar after matching on these loadings. One globe funds experience outflows of -0.52% relative to the matched sample and five globe funds experience outflows of -0.19% per month. The results suggest that pre-period differences do not account for the results.

While the analysis suggests a reaction to the globes themselves and does not appear directly related to known pre-period characteristics, it remains possible that the results are due to some general trend related to sustainability. Perhaps companies that are extreme in their characteristics related to sustainability within a Morningstar category have been systematically receiving flows over time in ways our prior analysis did not account for.

To test for such a scenario we construct pseudo globe ratings based on what the funds would have been rated in the period before their actual publication.[18] We emulate the procedure used by Morningstar based on CRSP holdings data their ESG scores from KLD.[19] We first calculate the value weighted KLD score for each mutual fund. We then take the percentile of this measure within Lipper class. Using the same breakpoints as Morningstar, we then assign each fund to a "pseudo globe" rating of between one and five globes.

Table 6 examines this placebo analysis and does not find evidence suggesting strong pre-period effects based on these pseudo ratings. The first column looks at the four years before the analysis

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

[19] KLD scores are calculated as the number of strengths minus the number of weaknesses.

and finds insignificant effects of -0.02% for 1 pseudo globe and 0.08% for 5 pseudo globes. This difference of 0.1% is economically and statistically weaker than the roughly 0.7% found in Table 3. The next three columns repeat the analysis conducted in the paper assuming that the pseudo ratings were published in March of a prior year and then follows flows for the subsequent 11 months. All three years are insignificant, and in two of the years the point estimate of the difference is in the opposite direction of that found examining the actual ratings. Panel B repeats the analysis using the normalized measure, and finds weak evidence of higher flows into five globe funds with a spread of 1.5 percentiles. While significant, the spread of 8 percentiles found in Table 3 is more than five times larger than this value. Examining the years before the publication this difference does not seem to be driven by the period just before the ratings were published as the spread in 2015 is insignificant and negative. Instead, it seems driven by a pattern years earlier in 2013. The pseudo ratings are not a significant driver of mutual fund flows, further underscoring that the paper is examining a reaction to the publication of the globe ratings themselves.

3.5 Ratings Changes

Morningstar recalculates its sustainability ratings at the end of every month. Table 1 Panel D shows that ratings themselves are fairly sticky, with roughly 80% of funds remaining in the same category from month to month. Thus, while many funds remain in the same category throughout our sample, there are a number that receive different globe ratings in different months. This section examines how fund flows behave when a fund is rated either one or five globes compared to the months when it is not. If the globe rating itself is causing the flows, than we expect months where a fund is ranked either as one or five globes to experience more extreme flows.[20]

Table 7 examines such variation and finds that funds experience more extreme flows when they possess the extreme rank, relative to other periods. A dummy variable is included that is equal to one if a fund is rated 5 globes at some point in the sample period and a separate variable is formed

[20]While the Morningstar website is updated in response to new ratings, investors could still be responding to information from prior time periods. For example, if decisions are related to prior research, previously published articles, or press releases, then we would expect a muted impact to changes.

analogously for funds ever rated one globe. Fund flows are regressed on these variables along with variables for whether the fund is equal to one or five globes that particular month and category by month fixed effects. The coefficient on the one globe variable is thus the difference in flows for a fund in the month it is actually rated one globe relative to the months that it is not rated one globe with the same interpretation for the five globe dummy variable. Column 1 shows that funds ranked five globes receive inflows 0.28% higher (with a t-statistic of 2.35) than months they are not and funds rated one globe receive outflows of -0.20% lower (with a t-statistic of -1.43) than the months they are not. Column 2 adds the additional controls for size, age, return and star rating from Table 3 and shows materially similar results. Columns 3 and 4 repeat the analysis using the normalized flow variable. Column 4 shows inflows 2.4 percentiles higher (with a t-statistic of 2.59) for funds ranked five globes compared to months they are not and flows about -2.5 percentiles lower (with a t-statistic of -2.80) for funds ranked one globe compared to months that they are not. These results are another piece of evidence that the flow effects we are measuring are caused by the globe rating itself rather than some other related factor. The same fund receives more inflows in months when rated five globes than in months when it is not and more outflows when rated one globe.

In order for our results to be capturing something other than the impact of the globe ratings, the ratings would have to be correlated with some other variable which is accounting for flows. This variable would have to be related to the discrete globe ratings to account for the discontinuity analysis, but not the underlying sustainability score or more continuous percentile ranks. The alternate variable could not be capturing fixed fund attributes, as we find the effect is significantly stronger when funds are ranked high or low in sustainability than in months when they are not. The variable must also begin having its impact only when the ratings are published as the placebo analysis showed it was nor present before. While not impossible, we feel that the results strongly support the parsimonious explanation that the globe ratings had a causal impact on fund flows.

3.6 Economic Impact

The inflows to five globe funds and outflows from one globe funds provide evidence that investors on average view sustainability as a positive attribute. While statistically strong, how economically meaningful was the impact of the globe ratings?

We conduct a back of the envelope analysis to estimate the overall impact. We take all funds with a five globe or a one globe rating and multiply their prior month TNA by the regression coefficient. This serves as an estimate for how much higher or lower the flows to a fund were because of a globe rating. Examining Table 3, for one globe funds our smallest regression coefficient is -0.352 while the largest is -0.457. Using these estimates we find that one globe funds lost between 12 and 15 billion dollars in outflows in the 11 months after the globe publication. Using the range of estimates for five globe funds where the smallest coefficient is 0.281 and the largest coefficient is 0.379 we find that five globe funds received inflows of between 24 and 32 billion dollars as a result of their globe ratings.

These magnitudes are our estimate of the net-impact of the ratings publication and associated publicity and role out campaign by Morningstar. Thus, in some ways they are an overestimation of the impact of sustainability ratings as in the long run, once investors have sorted into various funds based on their characteristics we do not expect these effects to continue at the same magnitude without ratings changes. On the other hand, these are estimates of net flows which means they underestimate the number of investors that flowed into these funds based on sustainability ratings. On net investors flowed into high sustainability funds, but likely some investors flowed out as well. Thus the estimates represent a lower bar for the proportion of investors that value these sustainability ratings in the market as a whole.

Next, we examine the impact of the sustainability rating on a given fund's Morningstar website traffic in Table 8. Columns 1 through 4 examine the total number of page views divided by the number of page views in February 2016, the month before the ratings were published, and finds they are about 2% to 3% lower for one Globe funds and about 4% to 6% higher for five Globe funds,

compared to three globe funds in Columns 1 and 3 and all middle ranked funds in Columns 2 and 4. All regressions include category by month fixed effects and Columns 3 and 4 show similar results after including additional controls. The last four Columns examine the number of unique visitors to a fund's Morningstar page. It finds similar results of roughly 2% to 4% lower for one globe funds and about 3% to 5% for five globe funds compared to those in the middle. Thus globe ratings seem to be an important driver of attention towards a fund, at least within Morningstar's website.[21]

Increasing size is clearly an important aspect of overall fund health and as such the impact of the flows should be apparent in other fund attributes. One such attribute is the probability of a fund closing down. Table 9 examines the probability a fund shuts down based on its globe rating. We define a fund as closing if the final month a fund is present in our data occurs before the last month of the sample and Morningstar lists the fund as liquidated for each share class in our sample. Column 1 shows that 13 one globe funds shut down, while only 6 five globe funds did. The rate of closure of 0.41% is more than double that of any of the other globe categories. Column 2 uses linear probability models and shows that a one globe fund is 0.24 percentage points more likely to close (t-statistic of 2.50) than a three globe fund, and that the other ranked funds do not seem to close at a higher or lower rate. Column 3 shows that two globe funds are 0.25% more likely to close than all the other funds (with a t-statistic of 2.50). Columns 4 and 5 add category by year by month fixed effects and the additional controls respectively and finds similar results. Combining them all together in Column 6 the point estimate decreases to an insignificant 0.12%. The results are suggestive that being rated one globe leads to a higher probability of closing down, but given the rarity of the event we lack the statistical power to say for certain after including the full battery of controls and fixed effects.

[21] This estimate serves as a lowerbound as many investors only learn of the ratings upon visiting a fund's page. Thus, this likely captures the change in attention due to outside sources and the subset of investors who could filter their Morningstar searches based on globe ratings.

4 Why do investors value sustainability?

We now explore three separate hypothesis to examine why investors place a positive value on sustainability. The first hypothesis is that institutional investors value sustainability due to constraints imposed by their institution. The second hypothesis is that investors (rightly or wrongly) view sustainability as a signal of higher future returns. The third hypothesis is that investors have a preference for sustainability for non-pecuniary reasons, such as altruism. These hypothesis are not mutually exclusive and it is likely that each has a hand in our results to some degree. In this section we attempt to understand the extent to which each is important, but we are not be able to offer definitive answers as to the driving force for the demand for high sustainability rated mutual funds.

One remaining possibility that we cannot directly examine is that investors react to the globe rating as an arbitrary ranking without regard to the sustainability it is attempting to measure. This could occur either due to the salience of the image or because people believe that any rating Morningstar creates is a positive signal due to its reputation. While this is likely true for some investors, we believe it is unlikely to be the main driver of flows for several reasons. First, Morningstar spent significant resources attempting to make it clear to investors that the rating was measuring sustainability. Further, investors – especially institutional investors – presumably spent significant amounts of time and effort on their decisions, and they should therefore be likely to understand the globe ratings were constructed to capture a fund's sustainability. Finally, the Google search analysis shows that roughly as many people are searching directly for the phrase "Morningstar sustainability rating" as "Morningstar star ratings." This suggests there are a large number of individuals who are sufficiently knowledgeable to search directly for the sustainability rating and who are not simply responding to the globe image at the top of the Morningstar webpage. Thus, it seems reasonable to assume that the flows we observe are driven significantly by an aspect related to sustainability.

4.1 Institutional Constraints

We begin by examining the hypothesis based on institutional constraints. For example, a University endowment may impose implicit or explicit constraints on its managers to avoid or invest in certain types of funds irrespective of maximizing returns.[22] If the results are being driven by such constraints, then the reaction by institutions should be different from that of non-institutional investors who do not share the same constraints. The ideal analysis would be specifically examining institutions that we knew were subject to such constraints. While we do not have this exact data, we can isolate the flows into and out of institutional share classes based on sustainability ratings.[23]

The use of institutional share class warrants caution when interpreting the results. While flows in these share classes may represent the decisions from institutional investors, they may be capturing the behavior of participants in retirement plans with access to institutional share classes (e.g. Sialm et al. 2015). If the institutional share classes only represent these investors, this would indicate that institutional investors were absent from the US mutual fund market and by definition institutional investors could not be driving the effects we document. If institutions are present in these classes to some extent, and institutions are the main driver of such decisions, than even if the share classes include non-institutional investors we would expect the effect to be mainly present in these share classes and not the non-institutional share classes.

Table 10 repeats the analysis allowing for a differential impact of institutional funds based on globe ratings. Specifically, we include another set of dummy variables with globe ratings, but each is interacted with a dummy variable equal to one if the given fund is institutional. Analysis is run at the share-class level and standard errors are clustered by fund and date. Including the standard globe dummy variables and the interaction terms means that the coefficient on the institutional interaction represent how different the flows into the institutional share classes with a given globe rating compare to the non-institutional share classes of funds with the same globe rating. Examining

[22] Evidence supporting this hypothesis would be consistent with prior literature showing that institutional investors drive firms' environmental and social investments (e.g., Dyck et al. 2017) and the general importance of institutional investors more broadly (e.g. Gillan and Starks 2000; Gillan and Starks 2003).

[23] We use Morningstar's classification of institutional shares which typically require an investment of greater than $100,000.

these interaction terms in Table 10 we find insignificant effects.

While the institutions represent a portion of the effect that we observe, the effects are still present and significant in the non-institutional share classes, suggesting that institutional behavior cannot fully account for the results. One interpretation of these results is that institutions behave in a manner similar to non-institutional investors. This could be because institutions have similar preferences to the non-institutional investors, or it could be that they face constraints forcing them to behave as if their preferences were similar. Another interpretation is that this analysis does not reflect the preferences of institutional investors at all as the behavior represents individual investors trading in their retirement accounts. Under either of these interpretations, including the likely combination of both of them, the results suggest institutions are not the main driver of the results that we document.

4.2 Rational Performance Expectations

The pattern in fund flows could also have been due to investors rationally viewing sustainability as a positive predictor of future fund performance. While arguments have been made consistent with such a relation, there are a number of reasons why a rational investor might view sustainability as negatively predicting performance. If an investor believed that the sustainability rating would induce fund flows and that there was decreasing returns to scale for funds, consistent with the model of Berk and Green (2004), then observing the flow pattern we document would lead to a belief that one globe funds would outperform five globe funds.[24] A sustainability based explanation is related to Hong and Kacperczyk (2009) who argue that many investors are hesitant to hold "sin stocks," which leads these stocks to command higher returns. Applying this intuition to our setting, if investors believed that there was a hesitance to hold low sustainability stocks, then these investors might expect there to be an inverse relation between returns and globe ratings.

On the other hand, Edmans (2011) finds that employee satisfaction predicts positive returns, suggesting that socially responsible screens can positively predict future performance if the market is

[24]Empirically Grinblatt and Titman (1989), Chen et al. (2004), Pástor et al. (2015) find evidence consistent with an inverse fund flow relation, though Reuter and Zitzewitz (2010) do not find such an effect.

not taking such signals into account. Existing literature supports the possibility that sustainability could help a firm since it is well positioned to deliver warm-glow feelings to consumers (Becker 1974; Andreoni 1989; Cahan et al. 2015), or because corporate goodness could be used as a method for deterring harmful regulation or enforcement (Baron 2001; Hong and Liskovich 2015; Werner 2015) or broadly signal good governance (Deng et al. 2013; Dimson et al. 2015; Ferrell et al. 2016).[25] If an investor believed that the market was not correctly pricing positive attributes correlated with sustainability, then such an investor would be justified in expecting more sustainable funds to earn higher returns in the future.

The recent marketwide shift in attention towards sustainability suggests that it may be difficult to extrapolate past return patterns related to sustainability into the current market environment. Historically, it was difficult to ascertain information about a firm's sustainability and many investors did not consider it when making investment decisions. Thus, it is plausible that in the past the market price did not reflect a firm's sustainability and, to the extent it was an ignored positive attribute, sustainable companies may have earned high returns. The publication of the Morningstar ratings combined with the large market wide shift in attention towards sustainable investing suggests that it is unlikely that investors are still ignoring sustainability. Thus any past relations may no longer be relevant to predicting future performance. This suggests that the current environment may more closely resemble that of Hong and Kacperczyk (2009) where investors generally have a preference for holding certain stocks and against holding others which leads to predictable returns.

If investors had a rational belief that high sustainability funds would deliver high performance, we would hope that such out-performance would manifest itself in the data, but we find evidence more consistent with an inverse relation between globe ratings and returns. We examine raw excess returns, returns relative to Morningstar category (e.g. Pástor et al. 2015; Pástor et al. 2017), fund specific exposure to Vanguard indices and a 4-factor model (e.g. Berk and Van Binsbergen 2015). We measure excess returns by subtracting off the risk free rate. For Morningstar category, we subtract the value weighted return of funds in that category. For the Vanguard benchmark we first

[25]Other papers have found evidence of sustainable investments being negative for a firm, e.g. Di Giuli and Kostovetsky 2014; Dharmapala and Khanna 2016; Fernando et al. 2017.

follow Berk and Van Binsbergen (2015) to construct an orthogonal basis set of Vanguard index funds using data from 2014 to January 2017.[26] Fund specific betas on these projections are estimated in the period before the globe ratings are published and then these betas are used to construct a fund's Vanguard benchmark return in the post-publication period. The analyzed return is a fund's return minus the return of the Vanguard benchmark. A similar methodology is used to construct a fund's 4-factor benchmark, but beta estimates are on the factors of market, size, value and momentum rather than the Vanguard benchmark projections. Again, the return examined is a fund's return minus the four factor benchmark based on the estimated betas from the pre-publication period.

In Table 11, the returns in excess of each benchmark are regressed on globe ratings. Column 1 shows returns in excess of the risk free rate, Column 2 shows returns relative to the Morningstar benchmark, Column 3 shows returns relative to the Vanguard benchmark and Column 4 shows returns relative to the 4-factor benchmark. In Panel A regressions are value weighted and in Panel B regressions are equal weighted. Below the regression coefficients, the difference between the five globe coefficient and the one globe coefficient is reported with the p-value that the difference is zero reported underneath. For example, examining the excess returns in Column 1 Panel A we see that one globe funds outperformed their benchmark by 31 basis points and five globe funds underperformed by 25basis points. Below the regression, we display the 56 basis point difference along with the p-value that this difference is zero of 0.6.

Examining the eight point estimates, each one globe estimate is positive and each five globe estimate is negative. Five of the eight globe coefficients are significantly negative at the 10% level and two of the one globe coefficients are significantly positive at the 10% level. The point estimate of the spread between one and five globe funds is negative in each instance, ranging from 16 to 56 basis points per month with p-values on the difference ranging from 0.06 to 0.26. The lack of consistent significance, combined with the fact that we are examining only 11 months of returns calls for caution when interpreting these estimates.

[26] We utilize the same list of funds, though add the total bond market, short-term bond, intermediate-term bond and long-term bond. Our complete list (in order of inception date is thus): VFIAX, VBTLX, VEXAX, VSMAX, VEUSX, VPADX, VVIAX, VBIAX, VBIRX, VBILX, VBLLX, VEMAX, VIMAX, VSGAX and VSIAX.

Finally in Panel C we form portfolios that are long firms that are rated five globes and short firms that are rated one globe. We regress this portfolio on just the market factor in columns 1 and 3 and on the market, size, value and momentum factors in columns 2 and 4. We report the alpha from these regressions in basis points. Value weighted, the four factor alpha returns -48 basis points (with a t-statistic of -2.14) and equal weighted the alpha is -18 basis points (with a t-statistic of -1.33). The portfolio sorts thus yield a similar estimate.

The short time series and volatility of returns makes it difficult to make definitive statements on the relation between returns and globe ratings in this natural experiment. The evidence does not support higher performance of five-globe funds relative to one globe funds which is what would be necessary to explain the observed fund flows with a rational performance-based explanation, though it remains possible that such a belief was ex-ante justified and simply needs a longer time series to empirically identify such effects. The evidence is consistent with both the hypothesis that one and five globe funds performed similarly as well as the hypothesis that one globe fund outperformed five globe funds. The point estimate on five globes is lower then that for one globe in every specification suggesting the low sustainability funds outperformed the high sustainability, though the weak statistical significance in some specifications is also consistent with a lack of relation between globe ratings and performance. We leave it to future researchers with access to the underlying holdings data to further examine this issue, though it may simply be that the short time series makes it impossible to definitively say whether one globe funds outperformed over this period or not.

4.3 Naive Performance Expectations and Non-Pecuniary Motives

Thus, the remaining explanations are that investors either naively assumed that a high sustainability rating was predictive of high future fund returns or had a non-pecuniary preference for holding more sustainable mutual funds. Unfortunately, the natural experiment from Morningstar does not allow for testable predictions that distinguish between naive beliefs about future returns versus preferences for sustainable funds because under either hypothesis the prediction is that more money

would be allocated to high sustainability funds without observing higher subsequent performance. The difference between these two behaviors comes from the underlying motivation. Under the performance expectations hypothesis, the decision to invest more in high sustainability funds is driven by these performance expectations, while under the non-pecuniary motives hypothesis, the decision is driven by altruism, warm glow, or social motives. Thus differentiating between these two hypothesis requires a measure of expectations of future performance.

To obtain such a measure and begin to understand the source of the flows, we ran an experiment based on the Morningstar ratings to elicit the impact of the globe rating on expected future performance.[27] We gave participants information about three hypothetical mutual funds, derived from Morningstar's website. We picked three similar funds rated one globe, three globes and five globes, all with five star ratings on Morningstar's site. We randomized the sustainability ratings across these three funds in the experiment, and we gave participants Morningstar sustainability information along with fund information related to past performance and other fund characteristics. The display containing the globe ratings was taken directly from Morningstar's website to most closely simulate the information an investor would be seeing. However, it is possible that participants in the experiment did not understand the globe rating scale in the same way as a typical Morningstar investor. This would lead to a different motivation driving the responses of our experimental subjects than the Morningstar investors they are meant to represent. Thus, we replaced the text at the bottom of the Morningstar sustainability rating with a description of the globe ratings.[28]

Each participant was asked to (a) report how well she thought the fund would perform over the next year on a seven point Likert scale (b) report how risky she considered an investment in the fund to be on a seven point Likert scale and (c) allocate $1,000 between the fund and a savings account.[29] We chose to examine MBA students at the University of Chicago Booth School

[27] Additional details and survey materials are available in the online appendix.

[28] This text was taken from the Morningstar site and read, "This score provides a reliable, objective way to evaluate how investments are meeting environmental, social, and governance challenges." To avoid drawing additional attention to the globe ratings, this detail was designed to closely mimic text that appears in the globe display on the Morningstar site. Among the MTurk participants, half of participants the original text stating that the "Sustainability Mandate information is derived from the fund prospectus", and half saw the more informative message. We did not see meaningful differences in responses as a function of these messages and combine results for subsequent analysis.

[29] Participants responded to questions about performance for all three funds in one block, questions about risk

of Business (269 students participated) so that we could draw conclusions that would be more likely to be representative of market participants. In addition, we ran the experiment on 576 participants on Amazon Mechanical Turk (MTurk) to see how decisions were made in a likely less financially sophisticated subject pool.[30]

If flows to high sustainability funds are driven by increased performance expectations, then more globes will be positively correlated with these expectations. We first analyze whether people associate globe ratings with higher performance and find that they do. In Figure 6 Panel A, we graph the average performance rating for each of the three globe ratings, after removing an individual fixed effect. To the left, we examine the MBA students and see that moving from one globe to five globes is associated with an increase in expected performance of about 0.4, which is a statistically significant difference with a t-stat of 3.23 clustered by subject. To the right we see a similar, slightly stronger pattern for MTurk participants with a difference between extreme globe ratings of about 0.8 which is statistically significant with a t-statistic of 7.69. Thus the globes seem to have a slightly higher impact on MTurk participants than MBA students, but both groups strongly believe that higher globe ratings lead to higher future performance.

One possibility is that these participants expected a fund with a higher globe rating to have higher performance because they thought five globe funds were riskier. We plot the expectations of risk in Figure 6 Panel B and find a strong inverse correlation between perceptions of risk and globe ratings, the opposite of what would be necessary to explain the performance expectations with risk. MBA students rated 5 globe funds as about 0.6 points less risky than one globe funds, with a t-statistic on the difference of -4.67. MTurk participants exhibit similar, slightly stronger behavior with a difference of roughly 0.8, with a t-statistic of -6.86. Thus it is unlikely that the positive correlation between globe ratings and performance is due to compensation for risk. Participants believed that higher globe ratings would result in higher performance at lower risk.

for all three funds in one block, and questions about allocations for all three funds in one block. The order of these question blocks was counterbalanced across participants.

[30]Research examining this platform finds that participants recruited through MTurk tend to perform similarly on tasks (Casler et al., 2013) and better in attention checks (Hauser and Schwarz, 2016) than traditional participant pools recruited through labs, while representing a more diverse set of participants (Paolacci and Chandler, 2014).

Although the finding that investors believe both that performance will be superior and that risk will be lower for funds rated high in sustainability may appear surprising, it is consistent with existing research in psychology. Specifically, while risks and benefits may be positively correlated in the world, they have been shown to be negatively correlated in people's minds across a range of contexts (Fischoff and Lichtenstein, 1978; Slovic et al., 1991; McDaniels et al., 1997). The affect heuristic (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2004, 2005, 2007) as well as broader research examining the role of affect in decision making (Loewenstein et al. 2001; Nisbett and Wilson 1977; Klauer and Stern 1992) have been used to explain this pattern. This research posits that people rely on affect and emotion - rather than reasoned analysis - to assess attributes of a given stimulus and make subsequent decisions.[31] To the extent that the high sustainability rating causes positive affect towards a mutual fund, the affect heuristic would predict that they are likely to judge it to be both higher in returns and lower in risk.

While higher expected performance alone could account for the patterns we observe in Morningstar data, this does not rule out that non-pecuniary motives could also be playing a role. In other words, are people investing in highly sustainable funds only because they believe they will outperform, or also because they value sustainability and are willing to pay for it? This preference could derive from a number of non-economic motivations, and would be consistent with evidence and theorizing that people are concerned with increasing social welfare (Charness and Rabin 2002; Fehr and Schmidt 1999). For example, investors may experience altruism or warm glow (Andreoni 1989, 1990), in which case they would want to invest in sustainability because they derive value from the fact that others benefit, or feel good because they are responsible for benefiting others. Alternatively, it could stem from social motives and pressures such as the desire to impress others or to avoid contempt or social backlash (Becker 1974; DellaVigna et al. 2012; Olson 2009).

In the context of our experiment, one potential measure of non-pecuniary motives is the extent to which an investor allocates funds towards five globe funds or away from one globe funds that is not explained by their expectation of future performance or risk. If participants cared about

[31] For example, Finucane et al. (2000) experimentally manipulate participants' affective evaluations of items such as nuclear power and find that perceptions of both risks and benefits shift to be congruent with the overall evaluation.

the globe ratings solely as indicators of fund performance, we would expect the globes to impact expectations of future performance and risk.[32] Under such an explanation, after controlling for these expectations, the globe ratings would have no further explanatory power. In Table 12, we examine how dollars allocated to portfolios vary with expectations of risk, performance and globe ratings. Regressions include a subject fixed effect and a fund fixed effect. If there is a significant difference between the one and five globe dummy variables, this indicates that an investor is more or less likely to invest in the given globe level than can be accounted for by performance and risk expectations alone. Thus, a positive difference between the five globe and one globe dummy variables in this analysis is consistent with altruism. We do caution that interpreting the results in such a manner requires the assumption that the portfolio weights for an investor who only cares about performance and risk increase linearly in the measures based on a Likert scale. While not definitive, we believe that it offers insight into a question with little information currently available.

The first column of Table 12 shows that dollars allocated to a fund are strongly positively correlated with expected performance and strongly negatively correlated with expected risk. Column 2 shows that without controlling for either risk or performance, investors allocate more money to five globe funds and less to one. MBA students allocate \$108 more to five globe funds than to one globe funds (with a p-value of roughly 0 on the difference) and MTurk participants allocate about \$130 more (again with a p-value of roughly 0).

Column 3 includes risk, performance and the globe ratings to identify whether this difference in allocations can be explained by performance expectations alone or whether non-pecuniary motives also play a role. After including the controls for risk and performance, the difference between funds allocated by MBA students towards one versus five globe funds drops, but remains meaningful at \$48, with a p-value on the difference of 0.04. For MTurk participants this difference drops to \$71, with a p-value of roughly 0. The results suggest that slightly less than half of the difference in money allocated between one and five globe funds can be attributed to non-pecuniary motives for the MBA students, while non-pecuniary motives can account for slightly more than half of the difference for

[32]The same would be true if participants interpreted the globe ratings solely as indicators of performance.

MTurk participants.

If the difference in allocation is driven by non-pecuniary motives related to sustainability, then we would expect the effect of globe ratings to be concentrated among participants who considered these factors when making their decisions. After making their choices, we asked participants the extent to which they considered ESG factors when making their investment decisions. Investors who said they did not consider ESG factors have no reason to exhibit non-pecuniary motives, so to the extent the globe dummy variables are capturing such motives we would expect them to lose their explanatory power for such investors. This is what we find when we restrict the sample to such investors in Column 4. MBA students in this group exhibit only a $5 difference in allocation between 1 and 5 globe funds while MTurk subjects exhibit a marginally significant $41 difference. Examining investors who considered ESG factors in Column 5 we see strong evidence consistent with non-pecuniary motives. MBA students allocated a significant $79 more dollars towards five globe funds and MTurk participants allocated a significant $86 towards five globe funds. Thus we see evidence that dollar allocations are driven by expected performance and risk, but also by altruism (or other non-pecuniary motives) above and beyond these factors.

The results also suggest that the experiment is not capturing a pure attention effect induced by the ratings. Under such an explanation, any salient ranking we presented would induce the observed empirical pattern in allocations due to the picture itself, but not the underlying context of the rating. If this were the case, the amount that an investor considers environmental factors would be unlikely to influence investment decisions. This suggests that the difference in responses we observe in the experimental setting was largely due to considerations related to sustainability, and not simply an attention effect unrelated to sustainability.

This experiment provided evidence for some form of non-pecuniary motives, but was not able to tease apart whether this was an internally driven warm glow versus an externally driven social pressure. Participants responded to questions in our experiment privately and responses are shared only with the experimenter. Thus, it seems reasonable to interpret willingness to pay for sustainability in this context as altruism or warm glow rather than social motives. However, to examine

the role of social pressure (e.g., in comparison to warm glow), one-half of participants in the MTurk sample were randomly allocated to a "social pressure" condition that reminded participants that investment decisions are often not private.[33] Responses did not meaningfully differ based on experimental condition. While it is possible that participant responses were driven by warm glow and not by social pressure, leading to insensitivity to condition, the null results may also be driven by a weak manipulation. We are reluctant to test a stronger experimental manipulation out of concern that the manipulation itself would draw attention to the social component of investing and lead to experimenter demand (c.f., Orne 1962; Zizzo 2010), rather than measure a true reaction to social factors. We leave it to future researchers to disentangle the extent to which the non-pecuniary motives are being driven by social motives rather than internal drivers.

5 Conclusion

We present causal evidence that investors collectively value sustainability and rule out the possibility that investors are indifferent to this information or that they penalize a fund for maintaining a portfolio of sustainable investments. We find that funds with the highest globe ratings receive a more than $22 billion increase in fund flows while those with the lowest globe ratings face a more than $12 billion reduction in fund flows as well as an increased probability of liquidation. This suggests that a large portion of the market views sustainability as a positive company attribute.

Although investors are presented with detailed information about the percentile rank of sustainability within Morningstar categories, they largely ignore this information and instead respond to the simpler and more salient globe ratings, consistent with the psychological literature on categorization. They further respond mainly to the extreme ranked categories, largely ignoring the others, consistent with literature on the salience of extreme ranks. The results suggest that how categories are constructed, especially extreme categories, can have a significant impact on how decisions are

[33] Thus, the MTurk experiment used a 2 (globe description: present vs. absent) x 2 (social pressure: present vs. absent) between-subjects design. These instructions read: "When providing your responses, you should keep in mind that investment decisions people make are often not private. Many people tend to find out about your investment decisions, for example your family members, investment advisors, and friends."

made in a financial setting and impact marketwide variables such as fund flows.

Our natural experiment in which a large portion of the market experiences a quasi-exogenous shock that does not impact fundamentals is rare in financial markets. This allows us to cleanly identify the causal effect of the sustainability ratings on mutual fund flows. We propose and find support for several explanations of the response to the publication of the ratings. The flow pattern is present among institutional share classes, especially for high sustainability funds, consistent with social constraints placed upon institutions being partially responsible for the effect. However, the pattern persists among non-institutional investors as well. We do not find evidence supporting a rational belief that more sustainable funds perform better, instead the evidence is more consistent with the opposite. In spite of this, our experimental evidence suggests that investors have a strong belief that better globe ratings positively predict future returns. We also find suggestive evidence of non-pecuniary motives, consistent with altruism or warm glow.

Taken together, our experimental findings support the role of affect in investment decisions. Specifically, the finding that participants expect that funds rated high in sustainability will both perform better and have lower risk is consistent with prior research on the affect heuristic (Alhakami and Slovic 1994; Finucane et al. 2000; Slovic et al. 2004, 2005, 2007). The patterns we observe may be general, with investors generalizing from any positive fund rating to positive affect towards the fund. Alternatively, it may be specific to sustainability, with the positive sustainability rating leading to positive affect among investors who value the environment. Either response would be consistent with findings on halo effects, in which an impression formed in one area influences overall evaluations (Nisbett and Wilson, 1977; Klauer and Stern, 1992)

An additional question that emerges is how investors in our dataset and participants in our experiment are interpreting the sustainability ratings. For example, although we found that people tend to associate sustainability with the environment, people may be considering the Morningstar Sustainability Rating to be specific to environmental factors, or more broadly indicative of a fund's corporate social responsibility. It is also possible that due to Morningstar's reputation, investors trust that Morningstar has measured sustainability in the most sensible way and respond to it

without giving additional thought to what they are measuring. We have not attempted to define sustainability throughout this paper, instead simply using Morningstar's definition of the concept. What investors actually are responding to when they view the sustainability ratings, or any number of other socially responsible investment objectives, is an interesting and open question for further study.

References

Abadie, Alberto, and Guido W Imbens, 2006, Large sample properties of matching estimators for average treatment effects, *econometrica* 74, 235–267.

Abadie, Alberto, and Guido W Imbens, 2011, Bias-corrected matching estimators for average treatment effects, *Journal of Business & Economic Statistics* 29, 1–11.

Alhakami, Ali Siddiq, and Paul Slovic, 1994, A psychological study of the inverse relationship between perceived risk and perceived benefit, *Risk Analysis* 14, 1085–1096.

Andreoni, James, 1989, Giving with impure altruism: Applications to charity and ricardian equivalence, *Journal of political Economy* 97, 1447–1458.

Andreoni, James, 1990, Impure altruism and donations to public goods: A theory of warm-glow giving, *The economic journal* 100, 464–477.

Barber, Brad M, Adair Morse, and Ayako Yasuda, 2017, Impact investing.

Baron, David P, 2001, Private politics, corporate social responsibility, and integrated strategy, *Journal of Economics & Management Strategy* 10, 7–45.

Becker, Gary S, 1974, A theory of social interactions, *Journal of political economy* 82, 1063–1093.

Bénabou, Roland, and Jean Tirole, 2010, Individual and corporate social responsibility, *Economica* 77, 1–19.

Benson, Karen L, and Jacquelyn E Humphrey, 2008, Socially responsible investment funds: Investor reaction to current and past returns, *Journal of Banking & Finance* 32, 1850–1859.

Berk, Jonathan B, and Richard C Green, 2004, Mutual fund flows and performance in rational markets, *Journal of political economy* 112, 1269–1295.

Berk, Jonathan B, and Jules H Van Binsbergen, 2015, Measuring skill in the mutual fund industry, *Journal of Financial Economics* 118, 1–20.

Bialkowski, Jedrzej, and Laura T Starks, 2016, Sri funds: Investor demand, exogenous shocks and esg profiles, Technical report.

Bollen, Nicolas PB, 2007, Mutual fund attributes and investor behavior, *Journal of Financial and Quantitative Analysis* 42, 683–708.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2012, Salience theory of choice under risk, *The Quarterly journal of economics* qjs018.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013a, Salience and asset prices, *The American Economic Review* 103, 623–628.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013b, Salience and consumer choice, *Journal of Political Economy* 121, 803–843.

Cahan, Steven F, Chen Chen, Li Chen, and Nhut H Nguyen, 2015, Corporate social responsibility and media coverage, *Journal of Banking & Finance* 59, 409–422.

Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik, 2014, Robust nonparametric confidence intervals for regression-discontinuity designs, *Econometrica* 82, 2295–2326.

Casler, Krista, Lydia Bickel, and Elizabeth Hackett, 2013, Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing, *Computers in Human Behavior* 29, 2156–2160.

Charness, Gary, and Matthew Rabin, 2002, Understanding social preferences with simple tests, *The Quarterly Journal of Economics* 117, 817–869.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey D Kubik, 2004, Does fund size erode mutual fund performance? the role of liquidity and organization, *The American Economic Review* 94, 1276–1302.

Cheng, Ing-Haw, Harrison Hong, and Kelly Shue, 2013, Do managers do good with other people's money?, Technical report, National Bureau of Economic Research.

Chevalier, Judith, and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Chowdhry, Bhagwan, Shaun Davies, and Brian Waters, 2017, Investing for impact .

Christensen, Hans B, Eric Floyd, Lisa Yao Liu, and Mark Maffett, 2017, The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records, *Journal of Accounting and Economics* 64, 284–304.

DellaVigna, Stefano, John A List, and Ulrike Malmendier, 2012, Testing for altruism and social pressure in charitable giving, *The quarterly journal of economics* 127, 1–56.

Deng, Xin, Jun-koo Kang, and Buen Sin Low, 2013, Corporate social responsibility and stakeholder value maximization: Evidence from mergers, *Journal of Financial Economics* 110, 87–109.

Dharmapala, Dhammika, and Vikramaditya S Khanna, 2016, The impact of mandated corporate social responsibility: Evidence from india's companies act of 2013 .

Di Giuli, Alberta, and Leonard Kostovetsky, 2014, Are red or blue companies more likely to go green? politics and corporate social responsibility, *Journal of Financial Economics* 111, 158–180.

Diecidue, Enrico, and Peter P Wakker, 2001, On the intuition of rank-dependent utility, *Journal of Risk and Uncertainty* 23, 281–298.

Dimson, Elroy, Oğuzhan Karakaş, and Xi Li, 2015, Active ownership, *The Review of Financial Studies* 28, 3225–3268.

DiNardo, John, and David S Lee, 2011, Program evaluation and research designs, *Handbook of labor economics* 4, 463–536.

Dyck, IJ, Karl Lins, Lukas Roth, and Hannes Wagner, 2017, Do institutional investors drive corporate social responsibility? international evidence, Technical report.

Edmans, Alex, 2011, Does the stock market fully value intangibles? employee satisfaction and equity prices, *Journal of Financial Economics* 101, 621–640.

Feenberg, Daniel, Ina Ganguli, Patrick Gaule, and Jonathan Gruber, 2017, It's good to be first: Order bias in reading and citing nber working papers, *Review of Economics and Statistics* 99, 32–39.

Fehr, Ernst, and Klaus M Schmidt, 1999, A theory of fairness, competition, and cooperation, *The quarterly journal of economics* 114, 817–868.

Fernando, Chitru S, Mark P Sharfman, and Vahap B Uysal, 2017, Corporate environmental policy and shareholder value: Following the smart money, *Journal of Financial and Quantitative Analysis* 52, 2023–2051.

Ferrell, Allen, Hao Liang, and Luc Renneboog, 2016, Socially responsible firms, *Journal of Financial Economics* 122, 585–606.

Finucane, Melissa L, Ali Alhakami, Paul Slovic, and Stephen M Johnson, 2000, The affect heuristic in judgments of risks and benefits, *Journal of behavioral decision making* 13, 1.

Fischoff, Baruch, and Sarah Lichtenstein, 1978, Don't attribute this to reverend bayes., *Psychological Bulletin* 85, 239.

Geczy, Christopher, Robert F Stambaugh, and David Levin, 2005, Investing in socially responsible mutual funds, Technical report.

Gillan, Stuart, and Laura T Starks, 2003, Corporate governance, corporate ownership, and the role of institutional investors: A global perspective .

Gillan, Stuart L, and Laura T Starks, 2000, Corporate governance proposals and shareholder activism: The role of institutional investors, *Journal of financial Economics* 57, 275–305.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of business* 393–416.

Harris, Lawrence E, Samuel M Hartzmark, and David H Solomon, 2015, Juicing the dividend yield: Mutual funds and the demand for dividends, *Journal of Financial Economics* 116, 433–451.

Hart, Oliver, and Luigi Zingales, 2017, Companies should maximize shareholder welfare not market value .

Hartzmark, Samuel M., 2015, The worst, the best, ignoring all the rest: The rank effect and trading behavior, *Review of Financial Studies* 28, 1024–1059.

Hauser, David J, and Norbert Schwarz, 2016, Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants, *Behavior research methods* 48, 400–407.

Heal, Geoffrey, 2005, Corporate social responsibility: An economic and financial framework, *The Geneva papers on risk and insurance Issues and practice* 30, 387–409.

Hong, Harrison, and Marcin Kacperczyk, 2009, The price of sin: The effects of social norms on markets, *Journal of Financial Economics* 93, 15–36.

Hong, Harrison, and Inessa Liskovich, 2015, Crime, punishment and the halo effect of corporate social responsibility, Technical report, National Bureau of Economic Research.

Huberman, Gur, 2001, Familiarity breeds investment, *The Review of Financial Studies* 14, 659–680.

Imbens, Guido W, and Thomas Lemieux, 2008, Regression discontinuity designs: A guide to practice, *Journal of econometrics* 142, 615–635.

Jordan, Jenny, and Klaus P Kaas, 2002, Advertising in the mutual fund business: The role of judgmental heuristics in private investors' evaluation of risk and return, *Journal of Financial Services Marketing* 7, 129–140.

Kitzmueller, Markus, and Jay Shimshack, 2012, Economic perspectives on corporate social responsibility, *Journal of Economic Literature* 50, 51–84.

Klauer, Karl Christoph, and Elsbeth Stern, 1992, How attitudes guide memory-based judgments: A two-process model, *Journal of Experimental Social Psychology* 28, 186–206.

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

Malt, Barbara C, Brian H Ross, and Gregory L Murphy, 1995, Predicting features for members of natural categories when categorization is uncertain., *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21, 646.

Margolis, Joshua D, Hillary Anger Elfenbein, and James P Walsh, 2009, Does it pay to be good... and does it matter? a meta-analysis of the relationship between corporate social and financial performance .

McDaniels, Timothy L, Lawrence J Axelrod, Nigel S Cavanagh, and Paul Slovic, 1997, Perception of ecological risk to water environments, *Risk analysis* 17, 341–352.

Murphy, Gregory L, and Brian H Ross, 1994, Predictions from uncertain categorizations, *Cognitive psychology* 27, 148–193.

Nisbett, Richard E, and Timothy D Wilson, 1977, The halo effect: Evidence for unconscious alteration of judgments., *Journal of Personality and Social Psychology* 35, 250.

Olson, Mancur, 2009, *The logic of collective action*, volume 124 (Harvard University Press).

Orne, Martin T, 1962, On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications., *American psychologist* 17, 776.

Osherson, Daniel N, Edward E Smith, Ormond Wilkie, Alejandro Lopez, and Eldar Shafir, 1990, Category-based induction., *Psychological review* 97, 185.

Paolacci, Gabriele, and Jesse Chandler, 2014, Inside the turk: Understanding mechanical turk as a participant pool, *Current Directions in Psychological Science* 23, 184–188.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.

Pástor, L'uboš, Robert F Stambaugh, and Lucian A Taylor, 2017, Do funds make more when they trade more?, *The Journal of Finance* 72, 1483–1528.

Quiggin, John, 1982, A theory of anticipated utility, *Journal of Economic Behavior & Organization* 3, 323–343.

Reuter, Jonathan, and Eric Zitzewitz, 2010, How much does size erode mutual fund performance? a regression discontinuity approach, Technical report, National Bureau of Economic Research.

Riedl, Arno, and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *The Journal of Finance* .

Rips, Lance J, 1975, Inductive judgments about natural categories, *Journal of verbal learning and verbal behavior* 14, 665–681.

Rosch, Eleanor, 1978, Principles of categorization. cognition and categorization, ed. by eleanor rosch & barbara b. lloyd, 27-48.

Rosch, Eleanor, and Carolyn B Mervis, 1975, Family resemblances: Studies in the internal structure of categories, *Cognitive psychology* 7, 573–605.

Rosch, Eleanor, Carolyn B Mervis, Wayne D Gray, David M Johnson, and Penny Boyes-Braem, 1976, Basic objects in natural categories, *Cognitive psychology* 8, 382–439.

Schmeidler, David, 1989, Subjective probability and expected utility without additivity, *Econometrica: Journal of the Econometric Society* 571–587.

Sialm, Clemens, Laura T Starks, and Hanjiang Zhang, 2015, Defined contribution pension plans: Sticky or discerning money?, *The Journal of Finance* 70, 805–838.

Simonson, Itamar, and Amos Tversky, 1992, Choice in context: Tradeoff contrast and extremeness aversion, *Journal of marketing research* 29, 281.

Skowronski, John J, and Donal E Carlston, 1989, Negativity and extremity biases in impression formation: A review of explanations., *Psychological bulletin* 105, 131.

Slovic, Paul, Melissa L Finucane, Ellen Peters, and Donald G MacGregor, 2004, Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality, *Risk Analysis* 24, 311–322.

Slovic, Paul, Melissa L Finucane, Ellen Peters, and Donald G MacGregor, 2007, The affect heuristic, *European journal of operational research* 177, 1333–1352.

Slovic, Paul, Nancy Kraus, Henner Lappe, and Marilyn Major, 1991, Risk perception of prescription drugs: report on a survey in canada, *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique* 82, S15–S20.

Slovic, Paul, Ellen Peters, Melissa L Finucane, and Donald G MacGregor, 2005, Affect, risk, and decision making., *Health Psychology* 24, S35.

Thistlethwaite, Donald L, and Donald T Campbell, 1960, Regression-discontinuity analysis: An alternative to the ex post facto experiment., *Journal of Educational psychology* 51, 309.

Tversky, Amos, and Daniel Kahneman, 1986, Rational choice and the framing of decisions, *Journal of business* S251–S278.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and uncertainty* 5, 297–323.

Tversky, Amos, and Itamar Simonson, 1993, Context-dependent preferences, *Management science* 39, 1179–1189.

Weber, Elke, and Britt Kirsner, 1997, Reasons for rank-dependent utility evaluation, *Journal of Risk and Uncertainty* 14, 41–61.

Werner, Timothy, 2015, Gaining access by doing good: The effect of sociopolitical reputation on firm participation in public policy making, *Management Science* 61, 1989–2011.

Zizzo, Daniel John, 2010, Experimenter demand effects in economic experiments, *Experimental Economics* 13, 75–98.

Figure 2
Example of Globe Rating on Morningstar Website

This picture is an example from Morningstar's website of how sustainability information is displayed on a fund's webpage.

Figure 3
Google Search for Sustainability and Star Rating

This graph shows monthly google search volume based on sustainability rating and Morningstar star rating. The maroon line is based on searches for "Morningstar globe rating" while the navy line represents searches for "Morningstar star rating." The monthly measure is the average of the weekly measure where months are defined based on month ending period. Data cover January 2015 through January 2017.

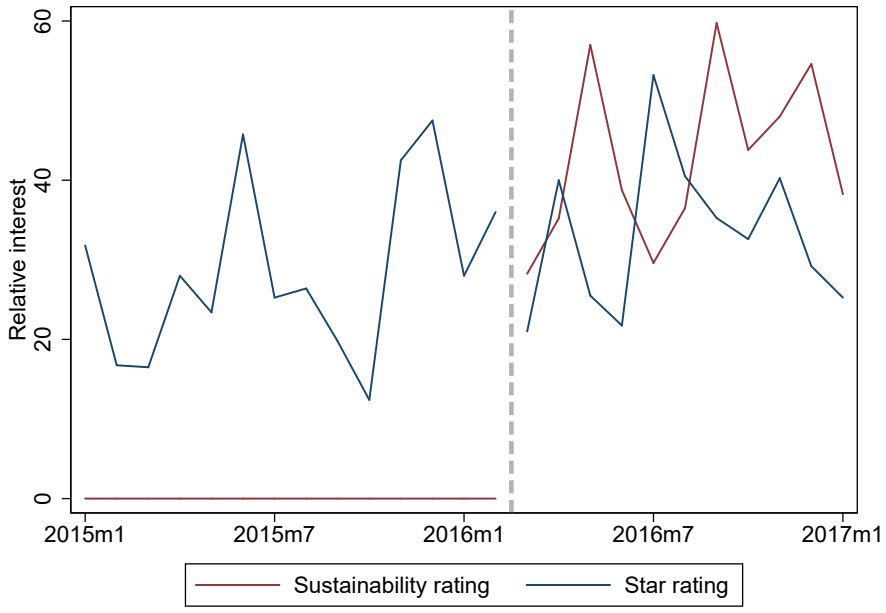
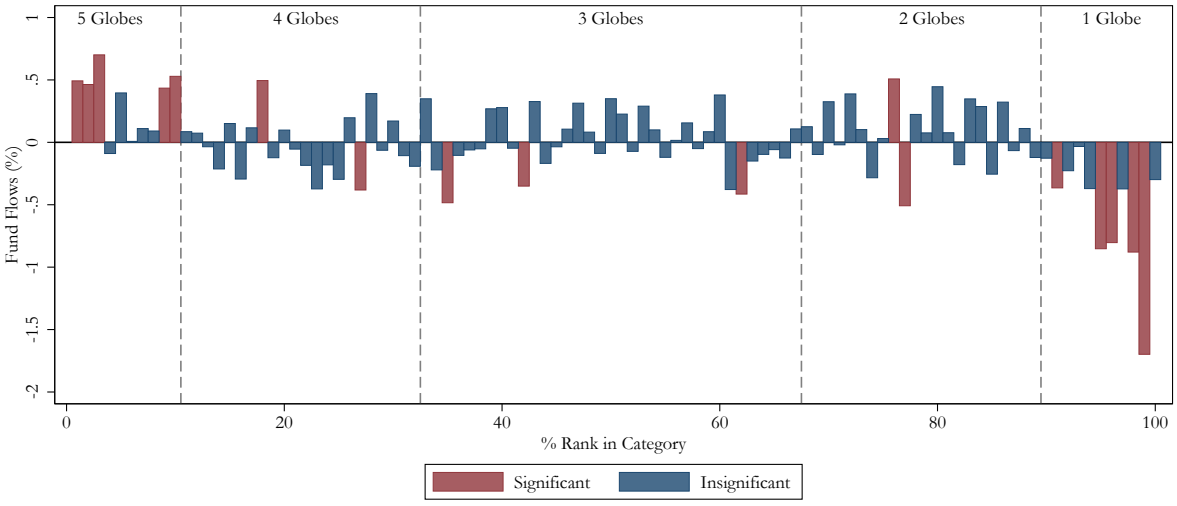


Figure 4
Flows by Percentile Rank of Sustainability

This graph shows average percentage flows for each sustainability percentile rank after controlling for year by month fixed effects. Panel A shows the averages of this variable. In Panel B, first flows are assigned a decile of size. Within each decile, for each month each fund is assigned to a percentile based on the flows they received. The average of this percentile rank (normalized to be mean 0 with a range from -49.5 to 49.5) for each sustainability percentile rank is graphed. Significant indicates the average flow is significant at the 90% level.

Panel A: Flows by Percentile Rank



Panel B: Normalized Flows Flows by Percentile Rank

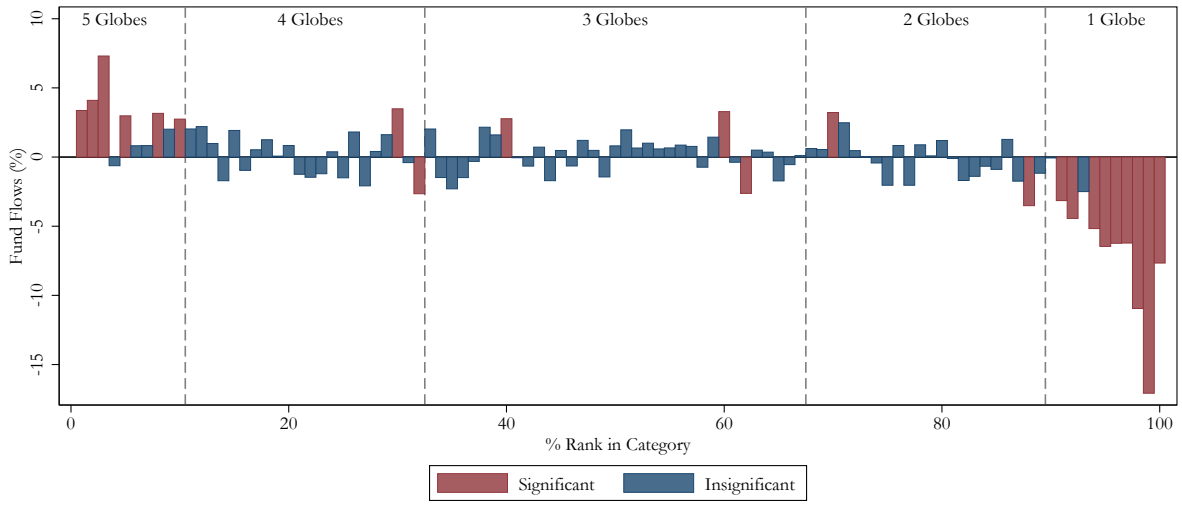
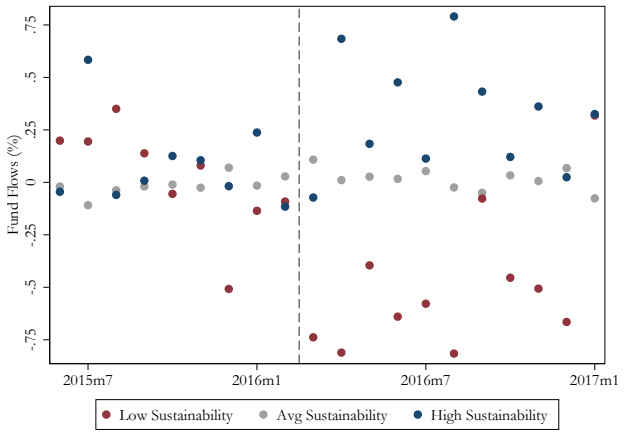


Figure 5
Flows by Month

This graph shows average percentage flows by month. Panel A shows the average of the variable for each month and Panel B shows a local linear plot with 90% confidence intervals.

Panel A: Summary Statistic Figure



Panel B: Local Linear Figure

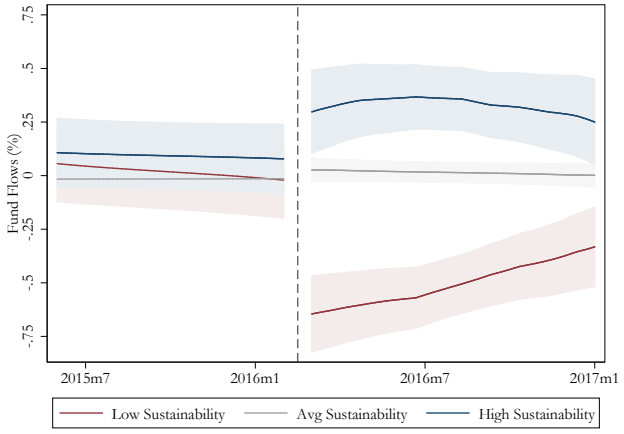
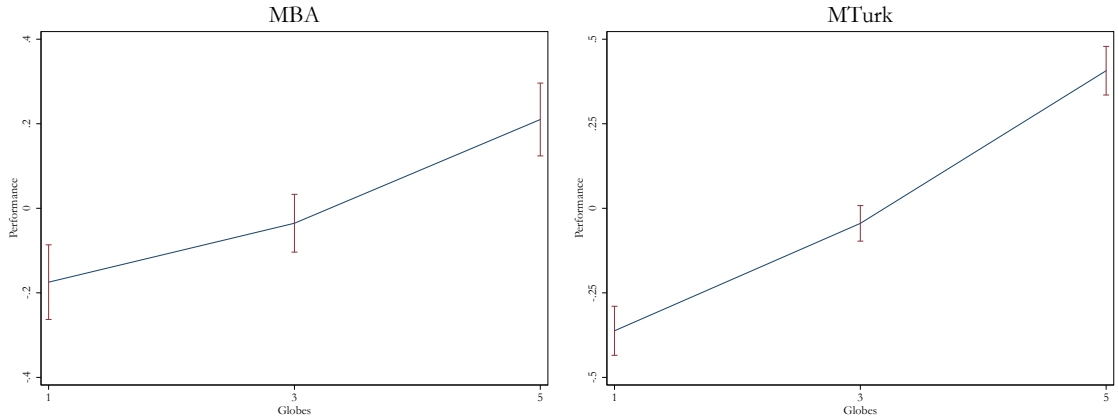


Figure 6

Experimental Expectations of Future Performance and Risk by Sustainability Rating

This graph shows the average performance rating in Panel A and risk rating in Panel B after taking out an individual fixed effect by globe rating. The left graphs are MBA students while the right graphs examine MTurk subjects. Maroon bars indicate the 90% confidence interval.

Panel A: Expectation of Performance



Panel B: Expectation of Risk

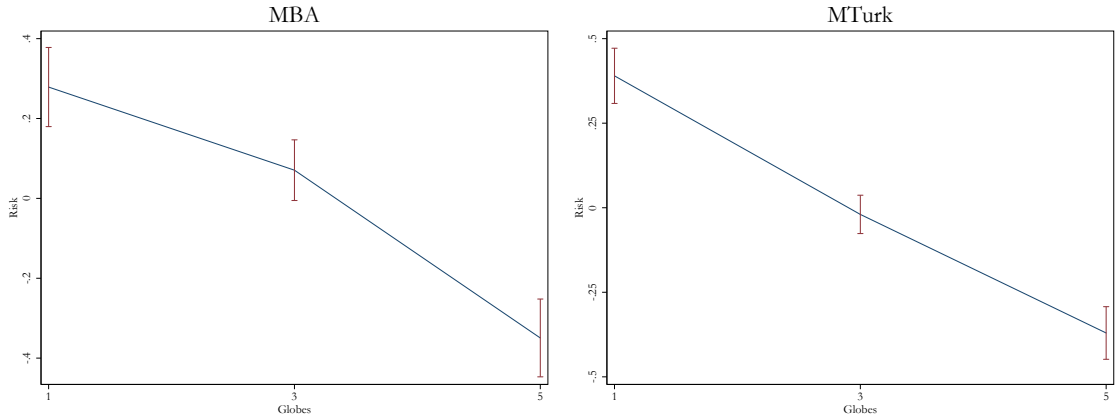


Table 1
Summary Statistics

This table shows summary statistics of the data. Panel A examines all funds post-publication, from March 2016 through January 2017. Statistics are at the share class level. Panel B examines the data by Globe for the 9 months prior to publication where Globes are defined as the rating the fund receives in March 2016. Panel C examines the data by globe after publication. Panel D shows the transition matrix from month to month for each globe rating after publication.

Panel A: Post-Publication Summary Stats

	Mean	SD	p10	p25	p50	p75	p90
Flow	-0.41	4.68	-3.43	-1.60	-0.60	0.37	2.46
Visits	209.45	474.48	1	14	44	159	521
Size	2184.33	8617.62	19.98	76.36	350.29	1370.35	4105.12
Rating	3.01	1.00	2	2	3	4	4

Panel B: Pre-Publication Summary Stats By Globe

	Obs	Size	Flows	Normalized Flows	Visits	Return	Rating
All	28713	2112.38	0.10	50.33	229.80	-1.27	3.03
1 Globe	2982	1392.05	0.12	48.12	235.64	-1.40	2.92
2 Globes	6215	2370.89	0.28	52.52	223.62	-1.23	3.10
3 Globes	9891	2353.41	0.01	50.10	229.37	-1.29	3.10
4 Globes	6422	1937.82	0.03	49.54	218.84	-1.24	3.02
5 Globes	3174	1885.88	0.19	50.43	259.84	-1.18	2.76

Panel C: Post-Publication Summary Stats By Globe

	Obs	Size	Flows	Normalized Flows	Visits	Return	Rating
All	34105	2184.32	-0.41	50.33	209.45	1.62	3.01
1 Globe	3170	1039.96	-0.90	44.69	164.34	1.72	2.74
2 Globes	7207	2438.41	-0.32	50.36	205.35	1.71	3.05
3 Globes	12183	2298.06	-0.41	50.70	201.52	1.62	3.10
4 Globes	7816	2197.10	-0.45	50.69	207.97	1.56	3.00
5 Globes	3730	2267.72	-0.10	53.13	284.43	1.53	2.83

Panel D: Transition Probability

		Next Month Rating				
		1 Globe	2 Globes	3 Globes	4 Globes	5 Globes
Current Month Rating	1 Globe	2297	539	37	8	0
		(79.73%)	(18.71%)	(1.28%)	(0.28%)	(0.00%)
	2 Globes	436	4869	1170	29	6
		(6.70%)	(74.79%)	(17.97%)	(0.45%)	(0.09%)
	3 Globes	64	983	8753	1185	28
		(0.58%)	(8.93%)	(79.48%)	(10.76%)	(0.25%)
	4 Globes	18	93	1032	5415	512
		(0.25%)	(1.32%)	(14.60%)	(76.59%)	(7.24%)
	5 Globes	4	14	61	467	2837
		(0.12%)	(0.41%)	(1.80%)	(13.80%)	(83.86%)

Table 2
Survey on the Meaning of Sustainability

This table shows summary statistics from a survey on the perceived meaning of sustainability. 482 participants on Amazon Mechanical Turk responded to the question "Recently, many companies have been trying to becoming more sustainable. Which of the following elements of a company's business practices do you think "sustainability" refers to?" Participants were given a list of categories with examples and were asked to select all categories that applied.

	Percent of People Selecting
Environment	79%
Products	48%
Human Rights	34%
Community	32%
Diversity	26%
Employee Relations	23%
Corporate Governance	22%
I do not Know	2%
Other	1%

Table 3
Fund Flows in Response to Sustainability Rating

This table shows how mutual fund flows vary with various measures of sustainability. The dependent variable is fund flows which are regressed on three proxies of sustainability, namely the raw sustainability score, the percentile rank within category and dummy variables for globe rankings. Columns 5 through 8 includes additional controls of return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating the prior month and the log of fund's age in years. Panel A does not weight regressions, while Panel B weights by log of TNA the month prior. All Columns include year by Morningstar category by month fixed effects. Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at the fund level. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Baseline Flow Regressions

	(1)	(2)	(3)	(4)	(5)	(6)
Sustainability Score	0.0744					0.0612
	(1.27)					(0.79)
Category Percent Rank	0.000983					0.00398
	(0.32)					(1.11)
1 Globe		-0.441***	-0.457***	-0.352***	-0.402***	-0.408**
		(-3.57)	(-4.17)	(-3.68)	(-4.32)	(-3.14)
2 Globes		0.0964		0.134		
		(1.17)		(1.59)		
4 Globes		-0.0353		0.0440		
		(-0.57)		(0.65)		
5 Globes		0.297**	0.281**	0.379**	0.331***	0.319**
		(2.48)	(2.66)	(3.13)	(3.21)	(2.27)
Diff: 5 Globe-1 Globe		0.737	0.738	0.731	0.733	0.727
P-value: 5 Globe=1 Globe		0.000370	0.000370	0.000759	0.000764	0.0138
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	Yes	Yes	Yes
R^2	0.0505	0.0513	0.0512	0.0911	0.0910	0.0911
Observations	34106	34106	34106	32475	32475	32475

Panel B: Normalized Flow and Size Weighted Regressions

	Normalized Flows		Size Weighted Flows		Size Weighted Normalized Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
1 Globe	-5.743***	-4.427***	-0.444***	-0.388***	-5.802***	-4.447***
	(-10.13)	(-7.50)	(-4.25)	(-4.41)	(-10.51)	(-8.03)
5 Globes	2.474***	3.253***	0.302**	0.355***	2.686***	3.460***
	(4.27)	(5.37)	(3.06)	(3.67)	(4.49)	(5.61)
Diff: 5 Globe-1 Globe	8.217	7.680	0.746	0.743	8.487	7.907
P-value: 5 Globe=1 Globe	0.00000701	0.0000165	0.000201	0.000362	0.00000498	0.00000929
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes	No	Yes
R^2	0.0724	0.158	0.0501	0.0902	0.0740	0.161
Observations	34106	32475	34106	32475	34106	32475

Table 4

Regression Discontinuity Tests of Fund Flows Around Sustainability Rating Breakpoints

This table conducts regression discontinuity tests of mutual fund flows around Globe breakpoints. Optimal bandwidth is calculated using one common mean-squared error for the treatment effect in Columns 1 and 3 and separate bandwidths for each group above and below the cutoff in Columns 2 and 4. The first row shows the conventional RD estimate while the second corrects for the bias described in Calonico et al. (2014). Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at fund level. z-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Flows

	1 Globe		5 Globes	
	(1)	(2)	(3)	(4)
Conventional	-0.427**	-0.366**	-0.727***	-0.484**
	(-2.40)	(-2.26)	(-2.91)	(-2.47)
Bias-corrected	-0.493***	-0.442***	-0.798***	-0.555***
	(-2.77)	(-2.73)	(-3.19)	(-2.84)
Common Cutoff	Yes	No	Yes	No
Separate Cutoff	No	Yes	No	Yes
Observations	31668	31668	32241	32241

Panel B: Normalized Flows

	1 Globe		5 Globes	
	(1)	(2)	(3)	(4)
Conventional	-2.834**	-1.648*	-3.184**	-2.777***
	(-2.40)	(-1.67)	(-2.33)	(-2.60)
Bias-corrected	-3.354***	-1.853*	-3.411**	-2.982***
	(-2.85)	(-1.88)	(-2.50)	(-2.79)
Common Cutoff	Yes	No	Yes	No
Separate Cutoff	No	Yes	No	Yes
Observations	31668	31668	32241	32241

Table 5
Fund Flows in Response to Sustainability Rating Matching on Pre-period Variables

This table reports the average treatment effect from nearest neighbor matching of Globe ratings on mutual fund flows. In Panel A funds are matched within Morningstar star rating based on flows, size, return over prior 12 months and age based on those values before the publication of the rating in February 2016. In Panel B funds are matched on these characteristics as well as their loadings on Vangaurd benchmark portfolios using the methodology of Berk and Van Binsbergen (2015). An extreme rated fund is matched to another fund, based on the intital rating in march of 2016. Matching is adjusted for the bias discussed in Abadie and Imbens (2006; 2011). Abadie-Imbens standard errors are used and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Matching on Characteristics

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
1 Globe	-0.715***		-6.700***	
	(-9.07)		(-11.60)	
5 Globes		0.206***		3.799***
		(2.60)		(7.44)
Observations	33262	33262	33262	33262

Panel B: Matching on Characteristics and Loadings

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
1 Globe	-0.523***		-4.391***	
	(-5.91)		(-6.36)	
5 Globes		0.185***		5.589***
		(2.68)		(9.75)
Observations	33232	33232	33232	33232

Table 6
Placebo Regressions of Flows on Pseudo Ratings using KLD Data

This Table shows how fund flows vary with placebo globe ratings formed using KLD data before the Morningstar ratings were published. KLD scores are calculated as the number of strengths minus the number of weaknesses for each stock. These are matched to mutual fund holdings and the value weighted average KLD score is a fund's pseudo sustainability score. Percentiles of this are calculated by Lipper class. Using the percentile breakpoint values used by Morningstar funds are categorized into five pseudo globe categories. Fund flows are regressed on dummy variables for these pseudo globe rankings. The first column includes data from 2012 through 2015. The subsequent three columns examine data from March of the indicated year for the next 11 months to emulate the analysis in the paper. All Columns include year by cateory by month fixed effects. All regressions include includes additional controls of return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio and the log of fund's age in years. Regressions are at the CRSP portno level and standard errors are clustered by month and fund, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Flow Regressions

	2012 to 2015	March 2015	March 2014	March 2013
	(1)	(2)	(3)	(4)
1 Pseudo Globe	-0.0261	-0.0303	0.0160	0.0858
	(-0.44)	(-0.27)	(0.12)	(0.84)
2 Pseudo Globes	-0.00618	-0.0205	0.0175	0.0732
	(-0.15)	(-0.27)	(0.21)	(1.15)
4 Pseudo Globes	0.122***	0.0105	0.125	0.172*
	(2.73)	(0.13)	(1.48)	(1.87)
5 Pseudo Globes	0.0794	-0.105	-0.0140	0.265**
	(1.15)	(-0.78)	(-0.10)	(2.76)
Diff: 5 Globe-1 Globe	0.105	-0.0751	-0.0301	0.179
P-value: 5 Globe=1 Globe	0.175	0.596	0.855	0.215
Cat by YM FE	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
R^2	0.0662	0.0793	0.0550	0.0652
Observations	238127	69906	48814	66022

Panel B: Percentile Regressions

	2012 to 2015	March 2015	March 2014	March 2013
	(1)	(2)	(3)	(4)
1 Pseudo Globe	-0.302	-0.676	0.352	1.202
	(-0.68)	(-0.99)	(0.41)	(1.56)
2 Pseudo Globes	-0.212	-0.500	-0.282	0.866
	(-0.69)	(-1.08)	(-0.48)	(1.77)
4 Pseudo Globes	0.854***	0.0434	0.775	1.586**
	(2.73)	(0.10)	(1.63)	(2.40)
5 Pseudo Globes	1.227**	-1.225	0.804	3.138***
	(2.21)	(-1.74)	(0.83)	(4.13)
Diff: 5 Globe-1 Globe	1.530	-0.549	0.452	1.936
P-value: 5 Globe=1 Globe	0.0187	0.552	0.674	0.0820
Cat by YM FE	Yes	Yes	Yes	Yes
Other Controls	Yes	Yes	Yes	Yes
R^2	0.117	0.123	0.106	0.122
Observations	238127	69906	48814	66022

Table 7
Fund Flows when Ratings Change

This table shows how mutual fund flows vary with Globe ratings, based on months funds were actually rated five or one globe relative to other months. Regressions include a five globe dummy, a one globe dummy along with variables equal to one if a fund was every rated five globes and a dummy equal to one if a fund was ever rated one globe. All columns include cateogry by month fixed effects and column 2 includes the additional controls of return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating the prior month and the fund's age in years. Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at the fund level. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
5 Globes	0.284**	0.304**	2.564**	2.397**
	(2.35)	(2.57)	(2.97)	(2.59)
1 Globe	-0.195	-0.200	-2.589**	-2.473**
	(-1.43)	(-1.62)	(-2.90)	(-2.80)
Cat by YM FE	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes
R^2	0.0516	0.0912	0.0741	0.159
Observations	34106	32475	34106	32475

Table 8
Change in Web Traffic Based on Globe Rating

This Table shows how internet traffic varies with Globe ratings. Web traffic is regressed on dummy variables for globe rankings. In Columns 1 through 4 web traffic is measured as all visitors divided by all visitors the month before the ratings were published, while in Columns 5 through 8 it is measured by unique visitors divided by unique visitors the month before the ratings were published. Category by year by month fixed effects are included in all Columns. Columns 3, 4, 7 and8 include the additional controls of return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating the prior month and the fund's age in years. Data is restricted to March 2016 and after, the period when the Globe ratings were published and analysis is at the share class level. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	All Visitors				Unique Visitors			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 Globe	-0.0317***	-0.0265***	-0.0246**	-0.0196*	-0.0376***	-0.0328***	-0.0281**	-0.0238*
	(-4.44)	(-3.92)	(-2.31)	(-1.89)	(-4.39)	(-4.48)	(-2.44)	(-2.18)
2 Globes	-0.00771		-0.00394		-0.00607		-0.00216	
	(-0.87)		(-0.42)		(-0.70)		(-0.24)	
4 Globes	-0.0109*		-0.0133**		-0.0108		-0.0127**	
	(-2.00)		(-2.83)		(-1.78)		(-2.42)	
5 Globes	0.0379***	0.0431***	0.0502***	0.0551***	0.0307***	0.0355***	0.0427***	0.0470***
	(3.18)	(3.76)	(3.70)	(4.60)	(3.48)	(4.35)	(4.18)	(5.25)
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	Yes	Yes	No	No	Yes	Yes
R^2	0.212	0.212	0.226	0.226	0.220	0.220	0.237	0.237
Observations	32422	32422	30861	30861	32422	32422	30861	30861

Table 9
Fund Liquidation Based on Globe Rating

This Table examines how the probability of mutual fund liquidation varies with Globe ratings. A dummy variable equal to one if a fund is liquidated is regressed on dummy variables for globe rankings. Columns 2 and 3 include year by month fixed effects. Column 3 includes additional controls of dummy variables for quintile of return in the prior month, dummy variable for quintile of size in the prior month, dummy variables for quintile of return in the prior month, dummy variables for quintile of expense ratio, dummy variable for quintile of Morningstar star rating the prior month. Analysis is at the share class level. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Summary Stats	Regressions				
	(1)	(2)	(3)	(4)	(5)	(6)
1 Globe	13	0.00237**	0.00248**	0.00222**	0.00236**	0.00121
	[0.41%]	(2.50)	(2.80)	(2.34)	(2.57)	(1.16)
2 Globes	8	-0.000614				
	[0.11%]	(-0.82)				
3 Globes	21					
	[0.17%]					
4 Globes	15	0.000191				
	[0.19%]	(0.45)				
5 Globes	6	-0.000123				
	[0.16%]	(-0.09)				
Cat by YM FE		No	No	Yes	No	Yes
Other Controls		No	No	No	Yes	Yes
R^2		0.000324	0.000281	0.0191	0.00545	0.144
Observations	34162	34162	34162	34162	32475	32475

Table 10
Institutional Fund Flows in Response to Sustainability Rating

This Table shows how mutual fund flows vary with Globe ratings comparing institutional to non-institutional share classes. Fund flows are regressed on dummy variables for globe rankings, a dummy variable equal to one if the share class is institutional and interactions of globe ratings and the institutional dummy variable. Columns 1 and 2 include category by year month fixed effects. Column 3 includes category by year month by fund level fixed effects. Additional controls in Column 2 and Column 3 include return in the prior month, return in the prior 12 months, return in the prior 24 months, log of size in the prior month, expense ratio, Morningstar star rating the prior month and the fund's age in years. Analysis is at the share class level. Standard errors are clustered by month and fund, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Flow		Normalized Flow	
	(1)	(2)	(3)	(4)
1 Globe*Institutional	0.0186	-0.0579	1.035	0.409
	(0.08)	(-0.23)	(0.79)	(0.31)
2 Globes*Institutional	0.00399	0.0111	-0.00732	-0.115
	(0.02)	(0.07)	(-0.01)	(-0.16)
4 Globes*Institutional	0.0528	-0.0924	1.219	0.512
	(0.32)	(-0.57)	(1.66)	(0.71)
5 Globes*Institutional	0.320	0.0970	1.524	1.190
	(1.13)	(0.36)	(1.27)	(1.08)
1 Globe	-0.484**	-0.219*	-5.311***	-2.940***
	(-3.10)	(-1.85)	(-5.24)	(-3.52)
2 Globes	-0.0430	0.0279	-0.668	-0.130
	(-0.43)	(0.34)	(-1.00)	(-0.23)
4 Globes	-0.0945	0.0594	-0.791	0.132
	(-0.92)	(0.63)	(-1.36)	(0.27)
5 Globes	0.190	0.347**	1.432	2.363**
	(1.03)	(2.28)	(1.17)	(2.43)
Diff: 5 Globe-1 Globe	0.674	0.565	6.743	5.303
P-value: 5 Globe=1 Globe	0.00781	0.0104	0.00134	0.00155
Cat by YM FE	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes
R^2	0.0335	0.0756	0.0582	0.161
Observations	100254	95136	100254	95136

Table 11
Returns Based on Globe Rating

This Table shows how mutual fund performance varies with Globe ratings. In Panel A regressions are value weighted based on the prior month's NAV and in Panel B regressions are equal weighted. Column 1 shows raw returns in excess of the risk free rate. Columns 2, 3 and 4 show returns in excess of a benchmark. The Benchmark in Column 2 is the value weighted average return in a Morningstar category. Column 3 measures returns in excess of Vanguard benchmarks using the methodology from Berk and Van Binsbergen (2015). Column 3 estimates returns in excess of a fund benchmark based on the market, SMB, HML and momentum. Both benchmark's are based on fund-specific beta estimates from the two years prior to the globe rating. Below the regression, the difference between five and one globe funds is reported along with the p-value for the test that they are equal. All regressions are at the fund level and all returns are measured in percentages. Standard errors are clustered by month, and t-statistics are in parentheses. In Panel C, portfolios are formed based on globe ratings. The difference portfolio long five globe stocks and short one globe stocks is regressed on the market in the Columns labeled "CAPM" and on the market, size, value and momentum in the Columns marked "4-factor". *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value Weighted

	Excess Return	Morningstar Benchmark	Vanguard Benchmark	4-Factor Benchmark
	(1)	(2)	(3)	(4)
1 Globe	0.311	0.0514	0.209**	0.159
	(1.78)	(0.44)	(2.52)	(1.08)
5 Globes	-0.252*	-0.158*	-0.0995	-0.193
	(-2.18)	(-2.06)	(-0.76)	(-1.33)
Diff: 5 Globe-1 Globe	-0.563	-0.209	-0.309	-0.351
P-value: 5 Globe=1 Globe	0.0645	0.256	0.137	0.211
R^2	0.00144	0.00224	0.000798	0.00149
Observations	34083	34083	33307	33307

Panel B: Equal Weighted

	Excess Return	Morningstar Benchmark	Vanguard Benchmark	4-Factor Benchmark
	(1)	(2)	(3)	(4)
1 Globe	0.0924	0.0812	0.113*	0.0782
	(0.83)	(0.97)	(1.84)	(0.88)
5 Globes	-0.0961*	-0.104*	-0.0494	-0.150**
	(-2.05)	(-2.13)	(-0.47)	(-2.47)
Diff: 5 Globe-1 Globe	-0.189	-0.185	-0.163	-0.228
P-value: 5 Globe=1 Globe	0.148	0.143	0.211	0.134
R^2	0.000152	0.000684	0.000129	0.000466
Observations	34095	34095	33319	33319

Panel C: Portfolios

	Value Weighted		Equal Weighted	
	(1) CAPM	(2) 4-Factor	(3) CAPM	(4) 4-Factor
Long 5 Globe - Short 1 Globes	-0.460*	-0.479*	-0.138	-0.173
	(-2.03)	(-2.14)	(-1.01)	(-1.33)
Observations	11	11	11	11

Table 12
Experimental Results

This Table shows how Globe ratings impact expectations of returns and portfolio allocations in an experimental setting. Panel A examines MBA students while Panel B examines MTurk subjects. Dollar allocation amounts are regressed on performance expectations and globe rating dummy variables. Below the regression, the difference between five and one globe funds is reported along with the p-value for the test that they are equal. Column 4 includes subjects indicating they did not consider environmental, social or governance (ESG) factors when making decisions while Column 5 includes subjects that indicated that they did consider ESG factors. All regressions include subject fixed effects. Standard errors are clustered by subjects, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: MBA Students

	All			No ESG Consideration	ESG Consideration
	(1)	(2)	(3)	(4)	(5)
Performance	75.14***		71.32***	92.04***	53.92***
	(5.44)		(5.22)	(3.81)	(3.44)
Risk	-54.83***		-49.73***	-32.67	-59.70***
	(-4.60)		(-3.99)	(-1.52)	(-4.20)
1 Globe		-50.56**	-27.99	-13.89	-30.82
		(-2.24)	(-1.32)	(-0.43)	(-1.13)
5 Globes		57.36***	20.11	-8.080	48.51*
		(2.78)	(1.00)	(-0.27)	(1.75)
Diff: 5 Globe-1 Globe		107.9	48.10	5.809	79.33
P-value: 5 Globe=1 Globe		0.0000329	0.0485	0.876	0.0140
Acct FE	Yes	Yes	Yes	Yes	Yes
R^2	0.767	0.718	0.770	0.770	0.773
Observations	807	807	807	354	450

Panel B: MTurk Subjects

	All			No ESG Consideration	ESG Consideration
	(1)	(2)	(3)	(4)	(5)
Performance	58.29***		51.43***	51.43***	50.54***
	(9.38)		(8.07)	(3.96)	(7.06)
Risk	-30.69***		-25.58***	-31.42***	-23.18***
	(-5.13)		(-4.31)	(-3.25)	(-3.06)
1 Globe		-65.69***	-39.28***	-30.29	-43.66***
		(-5.02)	(-3.15)	(-1.49)	(-2.73)
5 Globes		64.43***	31.74**	11.44	42.75***
		(4.89)	(2.48)	(0.53)	(2.68)
Diff: 5 Globe-1 Globe		130.1	71.03	41.73	86.42
P-value: 5 Globe=1 Globe		5.26e-16	0.00000210	0.103	0.00000283
Acct FE	Yes	Yes	Yes	Yes	Yes
R^2	0.755	0.719	0.763	0.812	0.725
Observations	1728	1728	1728	624	1101