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

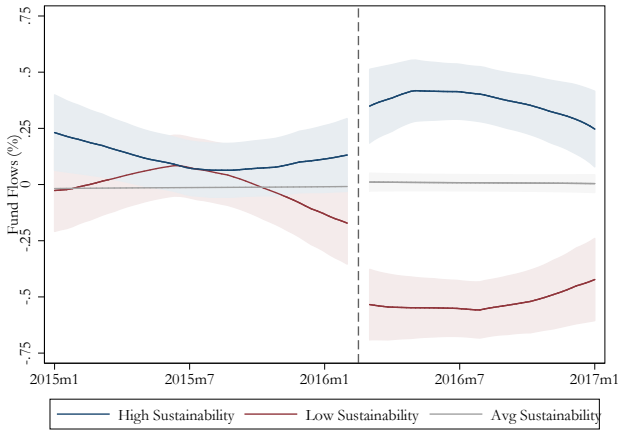
August 9, 2017

Abstract

Examining a shock to the salience of the sustainability of $8 trillion of mutual fund assets, we present evidence that investors value sustainability. Being categorized as a low sustainability fund resulted in outflows of more than $12 billion and an increased probability of liquidation. High sustainability funds received inflows greater than $22 billion. We can reject that mutual fund investors on average view sustainability negatively or do not care about sustainability. Investors reacted to the extreme categories, largely ignoring middle categories and detailed aspects of the ratings. Experimental evidence suggests that higher sustainability is viewed as a positive predictor of future performance, but we do not find evidence of such performance in the data.

*We are grateful to Sanjog Misra, Jacopo Ponticelli, Brad Shapiro, David Solomon, Kelly Shue and Eric Zwick for comments, and to Halley Bayer and Nathaniel Posner for excellent research assistance. We thank Ray Sin, Steve Wendel, and Sara Newcomb at Morningstar for providing the data.

Figure 1



Monthly percentage fund flows after removing year by month fixed effects by sustainability rating. The dashed vertical line denotes publication of the sustainability ratings. 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. 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 investorand 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 mutual fund investors in the US collectively put a positive value on sustainability. 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. 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 its 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. The Figure graphs monthly fund flows after controlling for monthly fixed effects. The dashed vertical line indicates the initial publication of the sustainability ratings. The navy line is flows to five globe funds, the maroon line is flows to one globe funds and the gray line is flows to 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 more than 0.3% per month or roughly 3.4% of fund size annualized. On the other hand, funds rated lowest in sustainability experienced outflows of more than 0.5% per month or over 5.6% of fund size per year. Over the 11 months after the sustainability ratings were published, we estimate between 12 and 22 billion dollars in assets left one globe funds and between 22 and 34 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

[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-

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 suffers 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 the assignment of globes was unforeseen and uncorrelated with other drivers of flows, then the simple difference in flows by globe rating after publication can be interpreted as causal. However, if the funds were systematically different before the publication of the ratings, then the change in flows could be reflecting this difference. The initial figure suggests this is not the case, and indeed, after controlling for pre-period trends for each fund, or for each rating percentile, we find similar results, suggesting that pre-period differences do not account for our results.

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 are responding to the five globe rating system rather than to other aspects of sustainability, we should find it is the globe category itself that is driving 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. When evaluating discrete categories, individuals often focus on extreme ranks (e.g. Hartzmark 2015; Feenberg et al. 2017). Consistent with this pattern, we find strong effects of being in one of the two extreme globe categories (i.e., one globe or five globe funds), relative to the three categories in the middle, but find insignificant differences across funds receiving two, three, or four globe ratings. In addition, due to a coding error, one percentile of funds was inadvertently miscategorized on the Morningstar website as one globe instead of two globes. The data suggests investors treated these funds as one globe funds even though along every dimension other than the rating it should not have been ranked as such.

---

funds-1475903728

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. While not the entire universe of stock market investors, investors in open-ended funds represent a large, diverse sample of the market suggesting the results are broadly applicable.

Sustainability is viewed as a positive attribute of funds, 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 institutions 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 do not 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 control for a variety of measures that could proxy for expectations of future returns, and do not find that these controls impact our results. We also examine whether funds experienced high returns after their high sustainability ratings and we do not find significant over or under-performance by one or five globe funds.[2]

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 for non-pecuniary motives (such as altruism, warm glow or social pressure), or an irrational belief that there is a positive correlation between future returns and sustainability. Unfortunately the data does not allow us to test between these two possibilities, so we run an experiment using MBA students and MTurk participants. We elicit their beliefs about future performance and investment decisions based on globe ratings, and we find a strong positive relation between globe ratings and expected future performance. While the full sample exhibits a strong positive correlation between

[2]The short time period means we cannot rule out the possibility that a belief in higher or lower returns was ex-ante rational, but simply did not manifest itself in the data.

expected performance and globe ratings, there is heterogeneity in these beliefs, with roughly 20% of participants believing that one globe funds will outperform five globe funds. We find evidence of non-pecuniary motives among investors who think five globe funds will outperform one globe funds, as they invest more money in five globe funds and less money in one globe funds than their expected future performance can account for. Investors believing one globe funds will have higher returns than five globe funds do not exhibit such behavior, as their portfolio allocations can be explained solely through their belief in expected performance.

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.

Additionally, our paper contributes to the literature on why firms invest in sustainability, and more broadly to investment in "doing well by doing good."[3] 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). While our paper does not break down the fraction of sustainability that is due to agency versus to 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. Instead, managers respond to the preferences of their investors.

The evidence highlights that the categorization and display of information can make certain

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

features salient, causing investors to alter their decisions in systematic ways with ramifications for entire financial markets. The information contained in the sustainability ratings was already publicly available. Furthermore, more detailed information on sustainability was presented alongside the globe ratings and was largely ignored. Thus, the results are a reflection of investors responding to the repackaging of this known information in a discrete, easy to understand measure. For example, funds on either side of a globe breakpoint are quite similar yet treated differently and the choice of breakpoint is an ad-hoc 10%. This means that investors are utilizing heuristics at the cost of ignoring some possibly useful information. Not only do investors underweight the continuous information underlying category rankings, but they also largely ignore sustainability for funds given intermediate ranks, instead reacting to funds that hold extreme category ranks.[4] This suggests that the construction of categories and display of information can have a significant impact on how investors decide to invest and on the market as a whole.

1 Data Sources and Summary Statistics

On March 1, 2016 Morningstar launched its sustainability rating system. They 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."[5]

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

[4] These results complement findings 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) as well as literature examining how the visual display of information impacts behavior, such as period of returns (Benartzi and Thaler 1999; Shaton 2017), earnings (Hirshleifer and Teoh 2003) or color of returns (Bazley et al. 2017).

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

this measure based on holdings to form a mutual fund specific sustainability score.[6] Each fund in a Morningstar category[7] 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 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.[8] 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 alongide the rating, see Figure 2 for an example.

All of the mutual fund data is provided by Morningstar and is at the monthly frequency. The sample includes all US based open-end funds rated by Morningstar at the share class level with a value greater than one million dollars. In addition to the sustainability rankings we examine a number of other fund attributes including fund flows, size, monthly returns, expense ratios, the Morningstar "star" fund ratings and fund web-traffic. We winsorize the continuous variables at the 1% level.

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 the globe breakdown is based on what each fund was 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.

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

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

[8] As discussed in Section 2.2, a coding error included 11% of the data in the one globe category.

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 from one globe category to another. 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.

2 Do Investors Value Sustainability?

2.1 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 had already incorporated 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 also 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."[9] The star rating is 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 by users of Google. The second, is that 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, 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.

2.2 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 shrinks. 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 reaction.[10]

In addition, our setting is rare in financial markets in that we examine an event that does not

[9] 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.

[10] Prior to the ratings publications it was difficult to ascertain a fund's sustainability without considerable effort. A small subset of funds, roughly 2% of our sample, had an explicit sustianability mandate. We do not focus on such funds due to the small sample size and because investors had sorted into these funds based on sustainability before the publication of the Morningstar ratings.

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 the 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 sustainability that were displayed together on a fund's page as shown in Figure 2. They released a fund's raw sustainability score, its percentile rank of that score within its Morningstar category and a picture of how many globes the fund was rated based on cutoffs of that percentile rank. If investors wanted to invest in the most sustainable fund in the market overall, then the raw sustainability score was the most informative measure, but it is difficult to interpret without a significant amount of effort to understand 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. This variable provides more granular detail than the globe rating. If investors wanted to invest in the most sustainable fund in a given Morningstar category, then the percentile rank was 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, so if investors are paying attention to the available information there would be 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 variables.

In Table 2 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 measure that appears 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 year by month fixed effects to control for aggregate market trends. In Column

1 we examine the raw sustainability score. If investors cared about how sustainable a fund was relative to the rest of the market, this is the most relevant measure of sustainability. Regressing fund flows on this measure we see an insignificant coefficient suggesting that it is not responsible for driving flows.[11] In Column 2 we examine the within category percentile rank of each fund. Adding the variable in Column 2 we see a strong inverse relation, consistent with investors valuing sustainability since poor sustainabality ranks correspond to higher percentiles. Finally, in Column 3 we include dummy variables for each of the globe ratings. One globe funds, the funds rated worst in terms of sustainability, experienced outflows roughly -0.6% per month lower than three globe funds, with a t-statistic of -7.83 clustered by month. Five globe funds, those rated highest in terms of sustainability experienced inflows of 0.3% per month, with a t-statistic of 3.80. These point estimates indicate that the lowest sustainability funds lost 6.6% of NAV per year while the highest rated funds gained about 3.7% of NAV per year. The globe ratings in the middle–two, three and four globes–are not statistically distinct from each other.

In Column 4 we include all three of the variables to try and understand which of the ratings seem to be driving mutual fund flows and find that it is the coarse globe ratings that investors appear to respond to, 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 roughly zero with t-statistics below 0.50. The coefficients on the globe rating variables 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 are responding to the globe ratings, not the other measures of sustainability. Further, regressions are focusing on the extreme ranks of globes and not the middle ranked positions.

The prior results may be due to globe ratings systematically varying with other variables associated with inflows. In column 4 we add a number of controls for such variables. We add dummy variables for the prior month's quintile of return relative to category benchmark to control for the fund-flow relation (Chevalier and Ellison 1997). To make sure the globe ratings aren't simply cap-

[11] In untabulated results we examine alternative transformations of this measure, such as shifting it into percentiles or dummy variables for certain levels, and find similar insignificant effects.

turing fund-flows based on fund size we add dummy variables for quintile of size the prior month. To make sure it is not capturing expense related motives we include dummy variables for quintiles of expense ratio. Another possibility is that there is a correlation between Morningstar's globe rating and their star ratings, so we include dummy variables for the Morningstar star rating. After including all of these controls, we find similar effects. One globe funds are associated with outflows of -0.41% with a t-statistic of -2.77, while five globe funds are associated with inflows of 0.36% with a t-statistic of 2.34.

Figure 4 takes a more detailed look at the relation between fund flows, the globe rating and the underlying percentile ranks. The Figure shows the average fund flow for each percentile rank from 1 through 100 after removing a year by month fixed effect. The dashed vertical lines indicate the globe cutoff levels with the category of globes listed at the top of the chart. So 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 1,000 observations. Examining the ten percentiles assigned to high sustainability funds (5 globes) we see that nine of the ten point estimates are positive and six 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 ten of eleven are negative and nine 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.

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 one exception where flows appear different based on percentile ranks, but not at globe cutoffs, appears to be the extreme low sustainability funds which received higher outflows when ranked 99th or 100th in terms of sustainability. It is plausible that investors who observed the 99th percentile rating assumed that the ratings went from 0 through 99 and thus this was in the worst ranked percentile, so it is not surprising that both the 99th percentile and the 100th percentile appear to induce a similar reaction. Comparing the average flow in percentiles 99 and 100 versus the other

one globe funds yields a difference of -0.51 with a t-statistic of 4.35. We do not see a similar effect of being in the top percentiles, with insignifant differences for funds ranked in the first or second percentile relative to the other five globe funds. Thus it appears that investors again pay attention to the extreme ranked funds by percentile, but only for those rated poorly in terms 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. These similar funds should receive different flows based on whether they are just over or just under a given cutoff. We examine this question more formally in Table 3 using regression discontinuity analysis. As the running variable, we examine the rank within each category. For example, in June of 2016 there were 265 funds ranked in US based Emerging Market funds 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 Imbens and Kalyanaraman (2012) as well as that from Calonico et al. (2014) 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 3 suggests that there are discontinuities surrounding the globe cutoffs. Examining the first two columns we see four estimates of roughly 0.3, with all four significant at the 10% level and three at the 5% level. This suggests that moving from a two globe rating to a one globe rating leads to a discontinuous change in flows of roughly 0.3% lower per month. Examining the five globe column we see coefficients ranging from -0.58% to -0.72%, 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.

One more interesting source of variation is due to a coding error by Morningstar. Morningstar meant to include 10 percentiles (91 through 100) as one globe funds, but mistakenly included 11 percentiles. Thus, the 868 observations in the 90th percentile should have been categorized as two globes, but were actually categorized as one globe. Given the small number of observations, this

serves as suggestive evidence, but lacks the power of a formal test. Figure 4 shows that the 90th percentile is negative and significant and roughly the magnitude of the one globe funds to the right of it. It is more negative than the two globe funds to its left. While not statistically significant, the 90th percentile is 0.08% different from the other one globe funds around it (91 through 93) and 0.24% more negative than the two globe funds in percentiles 87 through 89. These results are consistent with the 90th percentile funds being treated as one Globe funds even though the underlying methodology and corresponding percentile rank dictate that it should be a two globe fund. This should further mitigate concerns that we are capturing something other than the impact of the globe rating itself, as such a variable should be uncorrelated with the coding errors made by Morningstar.

2.3 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 the breakpoints, there was a jump when funds were assigned to one category or another. To the extent that we have controlled for the relevant variables and that the sustainability ratings represented an unanticipated shock, the prior results can be interpreted as the causal impact of the sustainability rating on flows. One still may be worried though that the prior section simply captured pre-period differences in funds that were not addressed by the controls in our 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 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. Before publication, to the left of the dashed line, there are not significant differences across the groups and the trends are roughly similar. The confidence intervals around the navy and maroon lines are significantly wider which is to be expected as each line represents 10% of observations while the gray line represents the other 80%. After the

publication, we see significant increases in flows to funds rated five globes and significant outflows from funds rated one globe.

We examine this pattern more formally in Table 4 by applying various controls for pre-period trends and find the results are materially similar. We again examine monthly fund flows, but now include data for the 12 months prior to the globe rating in the regression. In Columns 1 and 2 we run similar regressions to before, but include firm fixed effects. The fixed effects combined with the globe rating dummy variables in the post-period, means that we are examining variation of the impact of the globe rating relative to the pre-period fund specific average flow. Thus, the coefficient of -0.658 on one globe indicates that funds ranked one globe experienced flows 0.7% lower than their pre-period average with a t-statistic of -5.07. Five globe funds experienced inflows of 0.3% with a t-statistic of 3.74. After adding the additional controls the coefficients are similar, -0.5% outflows (with a t-statistic of -4.65) for one globe funds and 0.3% inflows for five globe funds (with a t-statistic of 3.26).

Columns 3 and 4 examine an alternative pre-period control based on percentile ranks and find similar results. Instead of a fund fixed effect, we control for the sustainability percentile rank of each fund. For the pre-period we assign each fund to the initial percentile rank that the fund receives. Thus we are examining the variation of the impact of post-publication globe rating relative to the pre-period average flows of each percentile rank category. Examining Column 3, we find the lowest ranked funds had outflows of -0.6% (with a t-statistic of -6.06) relative to their pre-period percentile rank average while those ranked highest on sustainability received monthly inflows of 0.4% (with a t-statistic of 4.85). The results suggest that pre-period differences do not account for the results.

2.4 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 returns.[12]

Table 5 examines such variation and finds that funds experience more extreme returns when they possess the extreme rank, relative to other periods. In Columns 1 and 2 the analysis is limited to only funds that are ranked 5 globes at some point in the sample period. Fund flows are regressed on a constant and a dummy variable equal to one if the fund is ranked five globes in that month. The constant is thus the average flows in months where the fund is not ranked five globes and the dummy variable is the difference in flows between the months the fund is not ranked five globes and the months the fund is ranked five globes. Column 1 shows that funds receive monthly inflows 0.24% higher (with a t-statistic of 2.39) in months when they are ranked five globes than months they are not. Column 2 shows similar results after controlling for year by month fixed effects. Column 3 shows funds ranked one globe experience outflows -0.18% lower (with a t-statistic of -1.96) in months when ranked one globe than in months they are not. Column 4 shows outflows of -0.24% (with a t-statistic of -2.31) in the one globe ranked months after controlling for year by month fixed effects. 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 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 globe 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 bin 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

[12]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.

and must coincide with the coding error mistakenly classified as one globe instead of two. While not impossible, we feel that the results strongly suggest that the most likely explanation is the parsimonious one that globe ratings had a causal impact on fund flows.

2.5 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 NAV 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 in a given month. For one globe funds our smallest regression coefficient is -0.354 while the largest is -0.658. Using these estimates we find that one globe funds lost between 12 and 22 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 0.253 and the largest coefficient is 0.394 we find that five globe funds received inflows of between 22 and 34 billion dollars as a result of their globe ratings.

Next we examine the impact of the sustainability rating on a given fund's Morningstar website traffic in Table 6. Columns 1 and 2 examine the total number of page views and finds they are 9 lower for one Globe funds and 24 higher for five Globe funds, relative to the average of 66 for three globe funds. Column 2 includes year by month fixed effects and finds similar results with t-statistics greater than 8 clustered by month for one globe and five globe funds. Columns 3 and 4 examine the number of unique visitors to a fund's Morningstar page. Relative to the three globe average of 52, one globe funds receive web traffic 17% lower and five globe funds receive web traffic 34% higher, again with t-statistics greater than 8 for both measures. Thus globe ratings seems to be an important driver of attention towards a fund, at least within Morningstar's website.

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 7 shows that funds rated with one globe are more likely to close during our sample period. We define a fund as closing if the last month a fund is in our data is prior to the last month of the sample and the fund is listed as liquidated by Morningstar. We examine the period after globe publication and exclude the final month of our data. The constant in our regression shows that, conditional on being open last month, a three globe fund has a 0.13% chance of closing the next month. A one globe fund is 0.15 percentage points more likely to close, more than twice as likely as the three globe fund. Columns 2 and 3 add additional controls and find similar results, that being rated one globe increases the probability that fund will cease to operate.

3 Why do investors value sustainability?

We now explore three separate hypothesis to further understand why it is that investors place a positive value on flows. The first hypothesis is that institutional investors value sustainability due to a constraint imposed by the institution that they work for. The second hypothesis is that investors do not care at all about sustainability, but (rightly or wrongly) view it as a signal of higher future returns. The third hypothesis is that there is simply a preference for sustainability as a characteristic. For whatever reason, investors view higher sustainability as a good thing, perhaps even at the expense of future returns. These hypothesis are not mutually exclusive and it is likely that each has a hand in our results to some degree. Thus in this section we attempt to understand the extent to which each is important, but we will not be able to offer definitive answers as to what the driving force behind the results is.

3.1 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 investments irrespective of maximizing returns. 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 proxy 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.[13]

Table 8 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. Examining the interaction terms in Table 8 we find in most cases insignificant effects. Examining the five globe interaction effect we see coefficients of between 0.28 and 0.37, though only the third column including both year by month fixed effects and the additional controls is statistically significant. The results are suggestive that institutional investors are more interested in buying high sustainability funds than non-institutional investors. However, institutional behavior cannot fully account for the results as the main effects are still present and significant in the regression.

3.2 Rational Performance Expectations

One possibility is that investors view sustainability as a predictor of future fund performance. For example, 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), or because corporate goodness could be used as a method for deterring regulation (Baron 2001) or broadly signal good governance (Ferrell et al. 2016). In our prior analysis, we attempted to control for such a motive by using a number of possible predictors of future returns. We included a variety of measures based on past performance, the expense ratio and the Morningstar star rating. None of these greatly impacted our estimate of the impact of the globe rating. Thus to the extent that globes were proxying for expectations of future returns they must be viewed as a signal largely unrelated to these measures.

We can also test whether such expectations were justified ex-post. In Table 9 we examine

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

whether high sustainability predicts high future returns in the subsequent month and do not find a strong relation. Returns over a month are regressed on the globe rating and a monthly fixed effect. If investors are correct in their belief that high sustainability is associated with high returns, we would expect to see a positive correlation between sustainability and returns. The results are generally insignificant, though the point estimates for high sustainability funds are 8-12 basis points per month below that of other funds. Equal weighted we see no discernible pattern for one globe funds, though value weighted we see positive point estimates of about 10 basis points, with t-statistics less than 0.50. The results do not present strong evidence that globe ratings are predictors of returns which suggests that investors should not rationally be updating expectations of future returns based on them, though it remains possible that such a belief was ex-ante justified and simply needs a longer time series to empirically identify such effects.

It is worth emphasizing that while we do not find evidence of high future returns for high sustainability funds, we also do not find significant under-performance. This might be expected if investors were correct in believing that sustainable investing was wasteful and indicative of poor corporate governance or if many investors demanded these shares for non-pecuniary motives. Our evidence does not suggest that globe ratings are strong predictors of either positive or negative future returns.

3.3 Naive Performance Expectations and Non-Pecuniary Motives

Thus, it is likely that either investors naively assumed that a high sustainability rating was predictive of high future fund returns or that the fund flows were driven by a simple preference for holding more sustainable mutual funds. It is possible that the fact that Morningstar chose to include globe ratings as part of their fund analysis leads investors to believe that sustainability is an important attribute to consider for fund evaluation when they would not have thought so in the absence of this rating. 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 and definitive answers regarding their relative importance is beyond the scope of this paper.

To begin understanding 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.[14] 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. Each participant was asked to report how well they thought the fund would perform over the next year on a seven point Likert scale and was also asked to allocate $1,000 between the fund and a savings account.[15] In one treatment we gave participants Morningstar sustainability information along with fund information related to past performance and other fund characteristics. In a second treatment we gave participants only the sustainability information.[16] We chose to examine MBA students at the University of Chicago Booth School of Business (112 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 217 participants on Amazon Mechanical Turk (MTurk) to see how decisions were made in a likely less financially sophisticated subject pool.

First, we examine whether investors generally associate globe ratings with higher performance and find that they do. In Figure 5 we graph the average performance rating for each of the three globe ratings, after removing an individual fixed effect. In Panel A, we examine the MBA students and see that one globe is rated roughly -0.5 less than an investor's average, and 5 globes is rated 0.5 more than average, with 3 globe funds receiving ratings in-between. Examining Panel B we see a similar pattern for MTurk participants. The scale is slightly wider, with one globe receiving ratings about -.8 and five globes receiving ratings about 1 greater than the individual average. 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 returns.[17]

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

[15] The order of these questions was counterbalanced across participants.

[16] Analyzing each treatment separately yield materially similar results. We therefore group our analysis across these two conditions. We also ran a control condition with no sustainability information included that we do not analyze.

[17] One alternative explanation of the higher expected performance for five globe funds than one globe funds is the belief that five globe funds are inherently more risky than one globe funds and thus have higher expected returns. While we did not ask participants about their expectation of risk, a variety of aspects seem inconsistent with a risk

While on average investors view globe ratings as predictive of future performance, the simple positive correlation does mask some heterogeneity. We find that over half of our MBA students (60 out of 112) and nearly 70% of our MTurk participants (150 out of 218) believe that five globe funds will have higher performance than one globe funds. Additionally, 21% of MBA students (24 out of 112) and 17% of MTurk participants (38 out of 218) believed that one globe funds would outperform five globe funds. For both of these groups the extreme expectations of performance were for the one and five globe funds with three globe funds on average receiving a rating in the middle.[18] Thus while on average investors view higher sustainability as positively predicting future performance, there is a significant subset of participants who view it as a negative predictor.

The results suggest that the globe rankings impact expectations of performance. However, we would also like to understand what role non-pecuniary motives are playing, if any. In other words, are investors 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 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

based explanation. Figure 6 shows that investors allocate more of their portfolio to positions with higher expected performance which need not be the case under a risk based explanation. Also, it seems possible that if any globe ranking should be viewed as the lowest risk it would be three globes as it is not at either extreme of sustainability. This would imply the lowest expected returns for three globe funds, but in the experiment we find only 6 MBA students and 12 MTurk participants who believed three globe expected returns were below both the one and five globe funds.

[18] The remaining 25% of MBA students and 14% of MTurk participants ranked one globe and five globes to have the same performance. We caution against interpreting this group too strongly as participants who marked the same answer for each globe may simply have been filling out the survey without thought.

performance.[19] Participants respond 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. Figure 6 displays the relation between performance expectations on the x-axis and dollars invested on the y-axis (both demeaned by an individual fixed effect). Examining all of the MBA students, for all three globes we see a similar strong positive relation where increased expectations of performance lead to higher allocations of funds. We do not find strong evidence of an altruistic motive influencing investment decisions for MBA students overall. In Panel B we see stronger evidence in favor of altruism by MTurk participants. While each globe rating has a somewhat similar slope, there appears to be a vertical jump from one globe to five globes which is consistent with altruism.

Next we extend the analysis by examining subsets of the data based on whether they viewed five globes as more positively or negatively predicting future returns. For MBA students, the right most graph of Panel A shows how study participants who have higher performance expectations for one globe than five globe funds decided to allocate their funds. Here, we see that dollars allocated across funds within each globe rating display a similar, positive and smooth relation between expectations of performance and dollars allocated to them. This suggests that these investment decisions are mainly based on expectations of performance and not on altruism. In the middle graph of Panel A, we examine MBA students who have higher performance expectations for five globe than one globe funds. Here, we see evidence of significant differences between each globe category. It appears that significantly less money is allocated to one globe funds than five globe funds conditional on the level of performance, consistent with an altruistic motive. Examining Panel B on the far right, we see that MTurk participants who rate one globe funds with higher expected performance than five globe funds make decisions that are consistent with performance as the driving force. For the observations with common support the three lines overlap, though the positive increase from one globe to five globes is larger than that for the MBA students. The middle graph of Panel B shows that, like the MBA students, the MTurk investors allocate significantly more money to five globe

[19] Interpreting the results in such a manner requires the assumption that portfolio weights for an investor who only cares about performance increase linearly in the expected performance measure of the Likert scale.

funds than to one globe funds for a given level of future performance. The evidence suggests that investors who rate five globes with higher expected returns than one globe exhibit more altruism in their investment decisions than investors who expect the opposite in return predictability.

Table 10 more formally explores this pattern by examining regressions of the dollars invested on expected performance, globe dummy variables, and subject fixed effects. If the globe dummy variables are significant, this indicates that after controlling for performance an investor is more or less likely to invest in the given globe level, relative to the omitted three globe group. Thus a positive dummy variable for a five globe fund and a negative dummy variable for a one globe fund is indicative of altruism. Examining Column 1 of Panel A, we see a positive and significant coefficient on performance and insignificant coefficients on the globe dummy variables. This suggests that, taken as a whole, the MBA students do not seem to exhibit an altruistic motive. Examining Column 1 of Panel B we again see evidence of performance based trade, but the MTurk participants allocate \$82 more towards five globe funds (with a t-statistic of 3.78) and \$64 less to one globe funds (with a t-statistic of -2.99), consistent with some altruism. Column 2 examines only participants who think five globe funds will outperform one globe funds. The MBA students in Panel A have investment decisions driven entirely by the globe levels. The performance variable is insignificant, but they allocate \$193 less to one globe funds (with a t-statistic of -3.92) and \$115 more to five globe funds (with a t-statistic of 2.21). MTurk subjects do respond to performance, but similar to the MBA students we see they allocate \$72 less to one globe funds (with a t-statistic of -2.15) and \$125 more to five globe funds (with a t-statistic of 3.47). Finally, Column 3 examines only investors who think one globe funds will outperform five globe funds. For both MBA students and MTurk participants we find that decisions among this group of investors are driven entirely by expected performance, with insignificant coefficients on both the one globe and five globe dummy variables. Thus for both groups we find that investors who think five globe funds will outperform one globe funds exhibit significant altruism, while those who think the opposite in terms of performance exhibit no altruism and appear to make portfolio allocations purely on the basis of expected returns.

4 Conclusion

We find that investors 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. By examining a shock to the salience of information about a fund's sustainability, we find that funds with the highest globe ratings face a $22 billion increase in fund flows while those with the lowest globe ratings face a $12 billion reduction in fund flows as well as an increased probability of liquidation. Although investors are presented with detailed information about the percentile rank of sustainabilty within Morningstar categories, they largely ignore this information and instead respond to the simpler and more salient globe ratings.

Our natural experiment where 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. The results allow us to demonstrate that the average market participant in US open-end mutual funds puts a positive value on sustainability. This suggests that a large portion of the market views sustainability as a positive company attribute.

We propose and find support for several explanations of this response. This pattern is present among institutional share classes, especially for high sustainability funds, suggesting that social constraints placed upon institutions is partially responsible for the effect. However, the pattern persists among individual investors as well. We do not find evidence supporting a rational belief that more sustainable funds perform better or worse. Our experimental evidence suggests that investors do believe that better globe ratings positively predict future returns and also find evidence of non-pecuniary motives, consistent with altruism or warm glow, among a subset of investors.

An additional question that emerges is how investors in our dataset and participants in our experiment are interpreting the sustainability ratings. For example, people may be considering the ratings 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, or even consider it as an additional signal of fund quality. The patterns we find in the experimental data are also consistent with perceptions of the sustainability ratings corresponding broadly to environmental or social goods rather than to fund quality. In particular, we would expect a higher portion of participants to rate the expected performance of five globe funds as higher than three globe funds, and three globe funds as higher than one globe funds. Furthermore, among participants who rate expected performance of five globe funds above that of one globe funds, we would expect a linear relationship between performance ratings and dollars invested to the extent that participants were interpreting the globes as being an additional measure of performance. 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 respond to the sustainability ratings, or any number of other socially responsible investment objectives is an interesting and open question for further study.

Investors may be responding to globe ratings rather than to the continuous measures of sustainability that underlie them as a result of the prominence of the globe ratings on the Morningstar website. Consistent with trusting Morningstar and valuing their decisions about how to display information, investors may be taking the salience of the globe ratings as a signal that this coarse metric is the most relevant and that the detailed information can be ignored. It is possible that if the display highlighted the percentile rank in category or raw sustainability score - with the globe measure in smaller font below- that investors would react to these highlighted measures instead.

However, Morningstar chose to highlight the globe rating likely because investors tend to respond to simple ratings (such as Morningstar star ratings) rather than to the continuous measures underlying them, consistent with the psychological literature on categorization. A key function of a 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 the current work,

each globe rating (1 globe, 2 globes, etc.) 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.[20] The results suggest that how these categories are constructed can have a significant impact on how decisions are made in a financial setting. While outside the boundaries of this paper, the optimal formation of such categories to induce better decision making could be an important area for future research.

[20]Our findings are consistent with a literature in psychology that focuses on rank dependent preferences (e.g., cumulative prospect theory; Tversky and Kahneman 1992) and with the corresponding intuition that extreme ranks are the most salient positions (Diecidue and Wakker 2001). Furthermore, it is consistent with existing literature showing that people overweight extreme attributes when making judgments (Skowronski and Carlston 1989) and make choices to avoid items with attributes ranked in extreme positions when confronted with tradeoffs (Simonson and Tversky 1992; Tversky and Simonson 1993).

References

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.

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

Bazley, William J, Henrik Cronqvist, and Milica Milosavljevic Mormann, 2017, In the red: The effects of color on investment behavior .

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.

Benartzi, Shlomo, and Richard H Thaler, 1999, Risk aversion or myopia? choices in repeated gambles and retirement investments, *Management science* 45, 364–381.

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

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

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.

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.

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

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

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.

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

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.

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.

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

Hirshleifer, David, and Siew Hong Teoh, 2003, Limited attention, information disclosure, and financial reporting, *Journal of accounting and economics* 36, 337–386.

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

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

Imbens, Guido, and Karthik Kalyanaraman, 2012, Optimal bandwidth choice for the regression discontinuity estimator, *The Review of economic studies* 79, 933–959.

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

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

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 .

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

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

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

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.

Shaton, Maya, 2017, The display of information and household investment behavior .

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.

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, 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.

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 page.

Figure 3
Google Search

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.

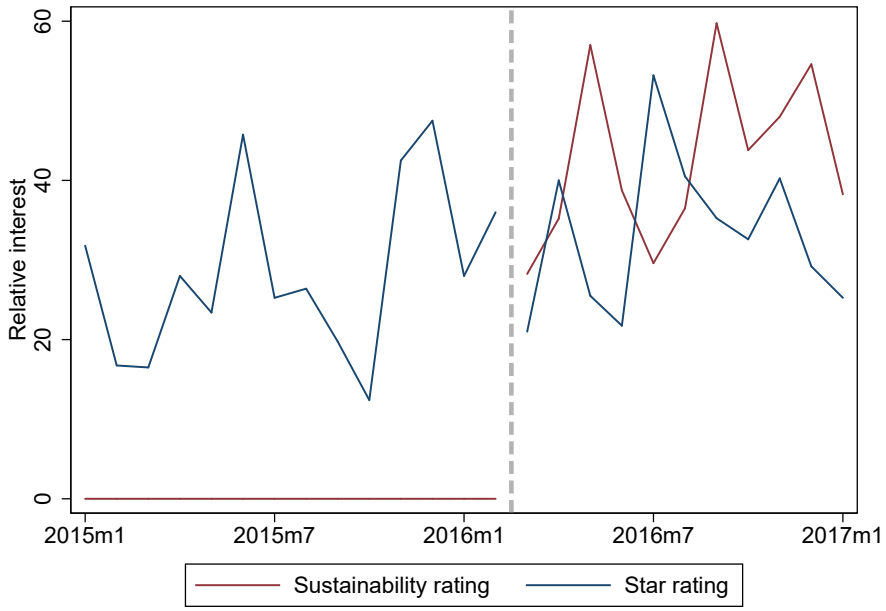


Figure 4

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. Significant indicates the average flow is significant at the 90% level.

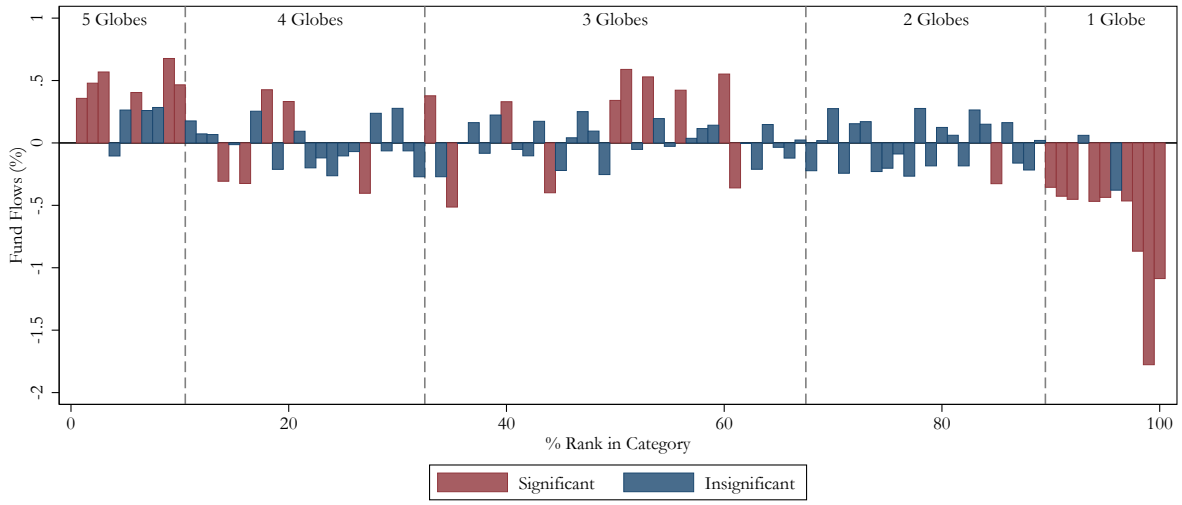
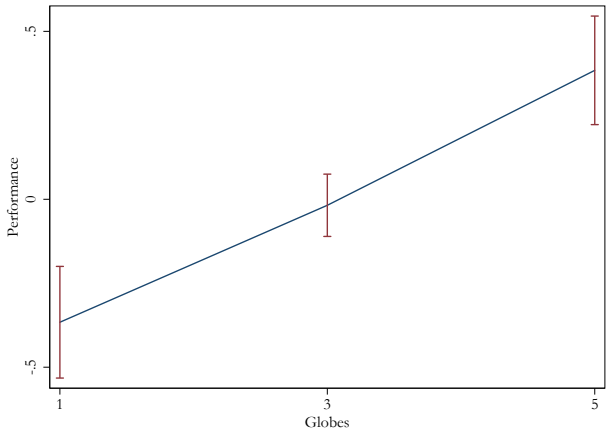


Figure 5

Experimental Expectations of Future Performance by Sustainability Rating

This graph shows the average performance rating after taking out an individual fixed effect by globe. Panel A examines MBA students while Panel B examines MTurk subjects. Maroon bars indicate the 90% confidence interval.

Panel A: MBA Students



Panel A: MTurk Subjects

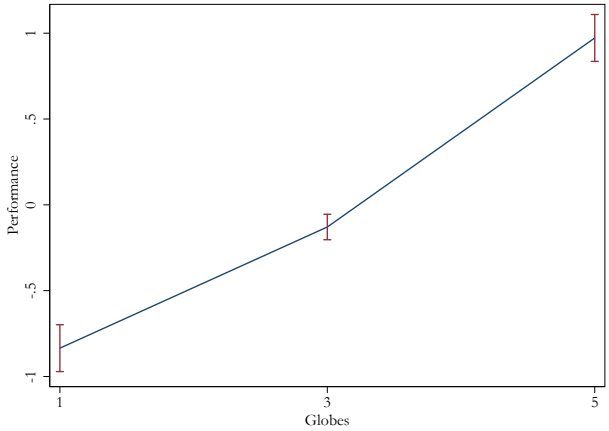
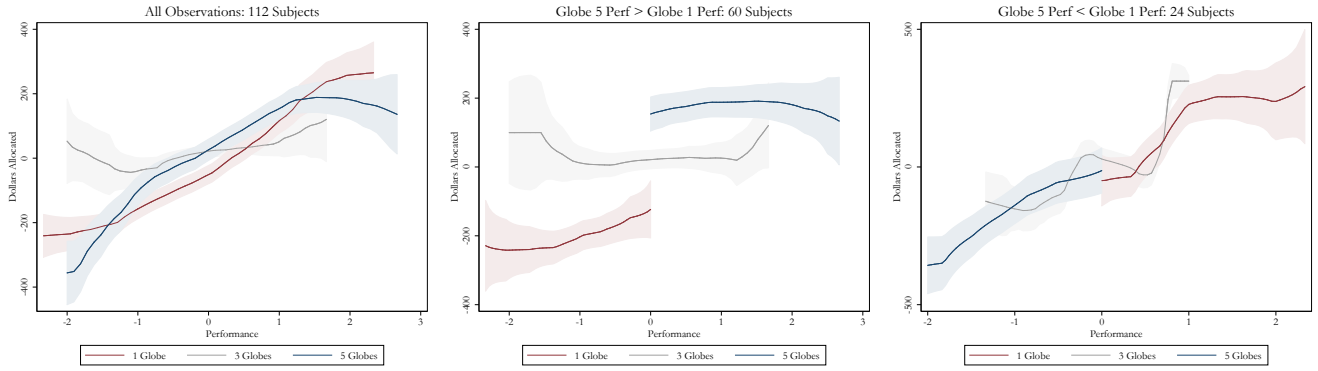


Figure 6
Experimental Investment Decisions and Expectations of Future Performance by Sustainability Rating

This graph shows the dollars allocated to a portfolio by expected performance after taking out individual fixed effects for both. The left graph includes all subjects. The middle graph includes only subjects where the performance of Globe 5 was greater than that of Globe 1 the right graph includes only subjects where the performance of Globe 1 was greater than or equal to that of Globe 5. Panel A examines MBA students while Panel B examines MTurk subjects. Shaded areas indicate the 90% confidence interval.

Panel A: MBA Students



Panel B: MTurk Subjects

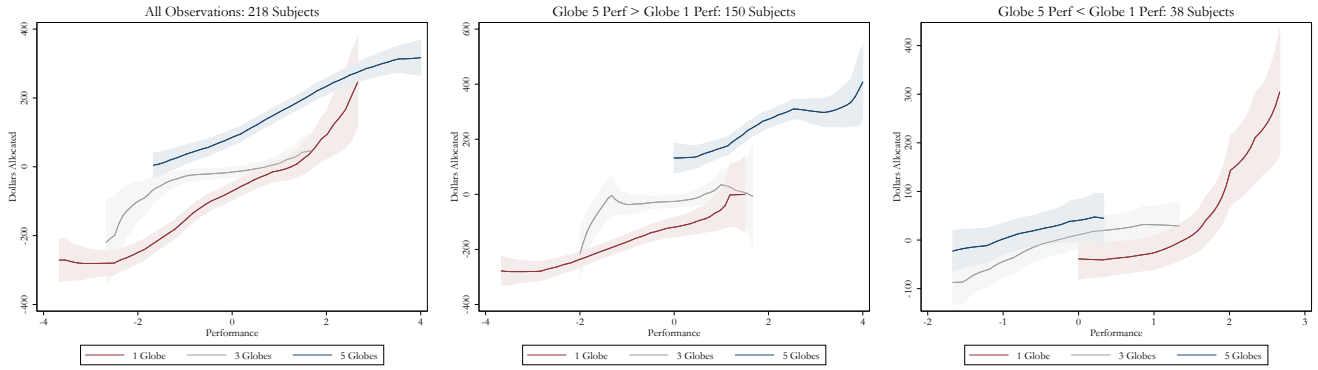


Table 1
Summary Statistics

This table shows summary statistics of the data. Panel A examines all funds post-publication, from March 2016 through January 2017. Panel B examines the data by Globe for the 12 months prior to publication. The Globes are the rating the fund receives in its first month of publication. Panel C examines the data by globe after publication. Panel D shows the transition matrix from month to month for each globe after publication.

Panel A: Post-Publication Summary Stats

	Mean	SD	p10	p25	p50	p75	p90
Flow	-0.424	6.279	-4.457	-2.003	-0.698	0.517	3.295
Visits	68.20	176.56	0	2	11	42	156
Size	857.31	4735.90	4	13	67	359	1413
Rating	3.09	1.02	2	2	3	4	4

Panel B: Pre-Publication Summary Stats By Globe

	Obs	Size	Flow	Visits	Rating
All	110672	838.97	0.345	56.91	3.09
1 Globes	10427	540.92	0.325	58.68	2.95
2 Globes	24838	911.26	0.631	56.87	3.13
3 Globes	38754	1018.16	0.256	58.08	3.16
4 Globes	26018	654.39	0.172	50.79	3.04
5 Globes	10635	761.25	0.445	65.99	2.97

Panel C: Post-Publication Summary Stats By Globe

	Obs	Size	Flow	Visits	Rating
All	100336	857.31	-0.424	68.20	3.09
1 Globes	8099	407.25	-0.911	57.67	2.78
2 Globes	22083	1028.31	-0.451	66.79	3.08
3 Globes	36727	917.16	-0.379	66.65	3.15
4 Globes	23586	754.49	-0.449	66.18	3.09
5 Globes	9841	867.06	-0.074	90.63	3.09

Panel D: Transition Probability

		Next Month				
		1	2	3	4	5
Current Month	1	5,729	1,511	96	18	0
		77.90%	20.55%	1.31%	0.24%	0.00%
	2	1,207	15,215	3,443	78	18
		6.05%	76.22%	17.25%	0.39%	0.09%
	3	158	2,911	26,304	3,654	72
		0.48%	8.79%	79.47%	11.04%	0.22%
	4	38	232	3,279	16,201	1,411
		0.18%	1.10%	15.50%	76.56%	6.67%
	5	8	32	163	1,320	7,389
		0.09%	0.36%	1.83%	14.81%	82.91%

Table 2
Fund Flows in Response to Sustainability Rating Publication

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 cateogry and dummy variables for globe rankings with 3 globes omitted. Column 5 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 quinitle of Morningstar star rating the prior month. All Columns include year by month fixed effects. Data is restricted to March 2016 and after, the period when the Globe ratings were published. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)	(4)	(5)
Sustainability Score	0.0202			-0.00747	-0.0168
	(1.25)			(-0.41)	(-0.90)
Category Percent Rank		-0.00438***		-0.000343	0.000267
		(-6.31)		(-0.13)	(0.11)
1 Globe			-0.565***	-0.568**	-0.408**
			(-7.83)	(-3.15)	(-2.77)
2 Globes			-0.0780	-0.0753	-0.0523
			(-1.48)	(-0.63)	(-0.51)
4 Globes			-0.0753	-0.0759	-0.00433
			(-1.69)	(-1.11)	(-0.06)
5 Globes			0.303***	0.308*	0.357**
			(3.80)	(2.07)	(2.34)
YM FE	Yes	Yes	Yes	Yes	Yes
Other Controls	No	No	No	No	Yes
R^2	0.00167	0.00197	0.00247	0.00248	0.0531
Observations	100336	100336	100336	100336	99461

Table 3
Fund Flows in Response to Sustainability Rating Around Rating Breakpoints

This table conducts regression discontinuity tests of mutual fund flows around Globe breakpoints. Optimal bandwidth is calculated using the procedure of Imbens and Kalyanaraman (2012) (IK) in Columns 1 and 3 and is selected using Calonico et al. (2014) (CCT) 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. Standard errors are clustered by month, and z-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	1 Globe		5 Globes	
	(1)	(2)	(3)	(4)
Conventional	-0.296**	-0.299*	-0.580***	-0.672***
	(-1.98)	(-1.94)	(-3.82)	(-3.36)
Bias-corrected	-0.311**	-0.325**	-0.608***	-0.720***
	(-2.08)	(-2.11)	(-4.00)	(-3.60)
IK Bandwidth	Yes	No	Yes	No
CCT Bandwidth	No	Yes	No	Yes
Bandwidth	40.82	37.15	38.67	19.32
Observations	30456	28280	32561	20012

Table 4

Fund Flows in Response to Sustainability Rating Publication Controlling for Pre-period

This table shows how mutual fund flows vary with Globe ratings. Fund flows are regressed on dummy variables for globe rankings with 3 globes omitted. 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 quinitle of Morningstar star rating the prior month. Data is restricted to March 2016 and after, the period when the Globe ratings were published. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Fund FE		Sustainability Pctile FE	
	(1)	(2)	(3)	(4)
1 Globe	-0.658***	-0.523***	-0.553***	-0.440***
	(-5.07)	(-4.65)	(-6.06)	(-5.40)
2 Globes	-0.141**	-0.128**	-0.0263	-0.00533
	(-2.35)	(-2.41)	(-0.56)	(-0.14)
4 Globes	-0.0237	-0.0477	-0.0293	0.0206
	(-0.28)	(-0.68)	(-0.62)	(0.38)
5 Globes	0.338***	0.253***	0.383***	0.394***
	(3.74)	(3.26)	(4.85)	(5.27)
Fund FE	Yes	Yes	No	No
Percentile FE	No	No	Yes	Yes
YM FE	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes
R^2	0.218	0.240	0.0438	0.0883
Observations	211008	208841	209404	207323

Table 5
Fund Flows when Ratings Change

This table shows how mutual fund flows vary with Globe ratings, limiting the sample to funds that are ever rated five globes (Columns 1 and 2) or ever rated one globe (Columns 3 and 4). In each subsample, flows are regressed on a constant and a dummy variable equal to one if the fund is 5 globes in that month (in Columns 1 and 2) or one globe in that month (in Columns 3 and 4). Columns 2 and 4 include year by month fixed effects. Data is restricted to March 2016 and after, the period when the Globe ratings were published. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Ever Rated 5 Globes		Ever Rated 1 Globe	
	(1)	(2)	(3)	(4)
1 Globe			-0.179*	-0.247**
			(-1.96)	(-2.31)
5 Globes	0.238**	0.242**		
	(2.39)	(2.48)		
Constant	-0.311***	-0.313***	-0.732***	-0.704***
	(-3.49)	(-6.42)	(-5.23)	(-15.85)
YM FE	No	Yes	No	Yes
R^2	0.000337	0.00206	0.000201	0.00372
Observations	19630	19630	19471	19471

Table 6
Change in Web Traffic Based on Globes

This Table shows how internet traffic varies with Globe ratings. Web traffic is regressed on dummy variables for globe rankings. In Columns 1 and 2 web traffic is measured as all visitors, while in Columns 3 and 4 it is measured by unique visitors. Year by month fixed effects are included in Columns 2 and 4. Data is restricted to March 2016 and after, the period when the Globe ratings were published. 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)
1 Globe	-8.984***	-10.70***	-7.581***	-8.703***
	(-4.82)	(-10.41)	(-5.75)	(-11.36)
2 Globes	0.135	-0.382	0.187	-0.172
	(0.16)	(-0.47)	(0.28)	(-0.27)
4 Globes	-0.472	-0.703	-0.646	-0.796
	(-0.31)	(-0.44)	(-0.55)	(-0.65)
5 Globes	23.98***	23.57***	17.94***	17.63***
	(8.92)	(8.34)	(8.89)	(8.31)
Constant	66.65***	67.00***	51.75***	51.99***
	(11.54)	(114.91)	(12.82)	(116.44)
YM FE	No	Yes	No	Yes
R^2	0.00194	0.0128	0.00200	0.0114
Observations	100342	100342	100342	100342

Table 7
Fund Death

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. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
1 Globe	0.00153**	0.00137*	0.000676*
	(2.28)	(1.99)	(2.02)
2 Globes	-0.000544	-0.000573	-0.000349
	(-1.50)	(-1.55)	(-1.05)
4 Globes	-0.000117	-0.000136	-0.000152
	(-0.79)	(-0.95)	(-0.84)
5 Globes	-0.000179	-0.000187	0.0000464
	(-0.29)	(-0.30)	(0.09)
Constant	0.00129**	0.00131***	0.000242
	(2.75)	(8.23)	(0.41)
YM FE	No	Yes	Yes
Other Controls	No	No	Yes
R^2	0.000209	0.00138	0.00184
Observations	91200	91200	90389

Table 8
Institutional Fund Flows in Response to Sustainability Rating Publication

This Table shows how mutual fund flows vary with Globe ratings examining only 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 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 quinitle of Morningstar star rating the prior month. Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1)	(2)	(3)
1 Globe*Institutional	0.0119	0.0149	-0.00742
	(0.07)	(0.09)	(-0.04)
2 Globes*Institutional	0.0296	0.0314	0.100
	(0.27)	(0.28)	(0.90)
4 Globes*Institutional	0.00769	0.0106	-0.00946
	(0.06)	(0.08)	(-0.07)
5 Globes*Institutional	0.281	0.278	0.372**
	(1.58)	(1.57)	(2.32)
1 Globe	-0.542***	-0.576***	-0.352***
	(-6.61)	(-7.22)	(-4.71)
2 Globes	-0.0739	-0.0799	-0.0535
	(-1.12)	(-1.27)	(-0.92)
4 Globes	-0.0720	-0.0776	-0.0298
	(-1.03)	(-1.13)	(-0.49)
5 Globes	0.232*	0.231*	0.204**
	(2.16)	(2.15)	(2.61)
YM FE	No	Yes	Yes
Other Controls	No	No	Yes
R^2	0.00324	0.00489	0.0531
Observations	100336	100336	99461

Table 9
Returns

This Table shows how mutual fund returns vary with Globe ratings. Fund flows are regressed on dummy variables for globe rankings. All Columns include year by month fixed effects. Columns 2 and 4 include 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.Standard errors are clustered by month, and t-statistics are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Equal Weighted		Value Weighted	
	(1)	(2)	(3)	(4)
1 Globe	-0.00923	-0.0340	0.0933	0.105
	(-0.08)	(-0.33)	(0.40)	(0.49)
2 Globes	0.0144	0.00502	0.0495	0.0521
	(0.26)	(0.09)	(0.39)	(0.51)
4 Globes	-0.107*	-0.111*	0.00108	0.0140
	(-2.04)	(-2.08)	(0.01)	(0.12)
5 Globes	-0.0783	-0.0824	-0.116	-0.0755
	(-1.28)	(-1.20)	(-1.51)	(-0.91)
YM FE	Yes	Yes	Yes	Yes
Other Controls	No	Yes	No	Yes
R^2	0.524	0.529	0.608	0.612
Observations	100322	99461	100320	99461

Table 10
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. Columns labeled "All" include all observations, "5 Perf > 1 Perf" contain only subjects who rated 5 globe funds with higher performance than 1 globe and "5 Perf < 1 Perf" includes only subjects rating the performance of one globe funds at higher than that of 5 globe funds. 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	5 Perf > 1 Perf	5 Perf < 1 Perf
	(1)	(2)	(3)
Performance	132.8***	30.86	139.2***
	(6.03)	(0.77)	(4.33)
1 Globe	-39.38	-193.0***	3.103
	(-1.51)	(-3.92)	(0.12)
5 Globes	-23.28	115.4**	-44.98
	(-0.71)	(2.21)	(-1.05)
Subject FE	Yes	Yes	Yes
R^2	0.786	0.745	0.856
Observations	336	180	156

Panel B: MTurk Subjects

	All	5 Perf > 1 Perf	5 Perf < 1 Perf
	(1)	(2)	(3)
Performance	85.11***	74.77***	53.34**
	(9.69)	(5.03)	(2.49)
1 Globe	-63.54***	-72.10**	-53.03
	(-2.99)	(-2.15)	(-1.58)
5 Globes	81.81***	125.2***	19.76
	(3.78)	(3.47)	(0.71)
Subject FE	Yes	Yes	Yes
R^2	0.796	0.776	0.851
Observations	651	447	204