The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXIV, NO. 6 • DECEMBER 2019

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

SAMUEL M. HARTZMARK and ABIGAIL B. SUSSMAN*

ABSTRACT

Examining a shock to the salience of the sustainability of the U.S. mutual fund market, we present causal evidence that investors marketwide value sustainability: being categorized as low sustainability resulted in net outflows of more than $12 billion while being categorized as high sustainability led to net inflows of more than $24 billion. 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. The evidence is consistent with positive affect influencing expectations of sustainable fund performance and nonpecuniary motives influencing investment decisions.

AS FIRMS INVEST MORE RESOURCES in sustainable and socially responsible endeavors, it is important to know whether such investments reflect investors' preferences marketwide. Some investors will believe that an increase in resources directed toward 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 pursue goals beyond simple value maximization. Others still will value such investments not because they care about the environment per se, but because they view such investments as a sound way to maximize profits. Finally, some investors will be unaware

*Samuel M. Hartzmark and Abigail B. Sussman are with the University of Chicago Booth School of Business. We are grateful to Jonathan Berk, Anat Bracha, Alex Edmans, Max Farrell, Mariassunta Giannetti, Matti Keloharju, Karl Lins, Vikas Mehrotra, Sanjog Misra, Giovanna Nicodano, Jacopo Ponticelli, Antonino Rizzo, Brad Shapiro, David Solomon, Kelly Shue, Paul Smeets, Oleg Urminksy, and Eric Zwick, as well as to seminar participants at Aalto, Emory, Cambridge, Chicago Booth, Warwick, London School of Economics, Imperial College, Northwestern, Bernstein Quantitative Finance Conference, Boulder Summer Conference on Financial Decision Making, UVA Darden Symposium on Mutual Funds and ETFs, Harvard Global Corporate Governance Colloquia, Development Bank of Japan Conference, Texas Finance Festival, European Finance Association Conference, Global Research Alliance for Sustainable Finance Conference, Swedish House of Finance Conference on Sustainable Finance, and Research Affiliates Investment Research Retreat 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. Sussman is a member of the Morningstar Behavioral Science Advisory Board. The authors have nothing further to disclose with respect to the *Journal of Finance*'s Submission Guidelines and Conflict of Interest Disclosure Policy.

DOI: 10.1111/jofi.12841

whether a firm is investing in sustainability or will not care. While the market surely contains examples of each of these investors, it remains unclear which type represents the average investor and thus 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?

In this paper, we demonstrate that the universe of mutual fund investors in the United States collectively puts a positive value on sustainability by providing causal evidence that marketwide demand for funds varies as a function of their sustainability ratings. Addressing this question directly is difficult in most settings, as it is not clear how to identify the preferences of the average investor. Analysis of investment products with an explicit sustainability focus reflects the preferences only of the subset of investors holding those products, and hence does not speak to the average preferences of all investors in the market. Furthermore, market outcomes, such as prices, related to firm attributes, such as sustainability, are usually studied in equilibrium, where analysis is inherently indirect.

We circumvent these challenges by examining a novel natural experiment whereby the salience of the sustainability of over $8 trillion in mutual fund assets experienced a large shock. Sustainability went from being difficult to understand to being clearly displayed by one of the leading financial research websites, Morningstar. In March of 2016, Morningstar first published sustainability ratings, with more than 20,000 mutual funds 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). Prior to this publication, 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 two, three, or four globes do not significantly affect fund flows. The dashed vertical line corresponds to 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 nine months prior to the rating publication; to the right of the line, fund 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, funds received similar levels of flows. After the publication, funds rated highest in terms of sustainability experienced substantial inflows of roughly 4% of fund size over the next 11 months. In contrast, funds rated lowest in sustainability experienced outflows of about 6% of fund size. We estimate that over the 11 months after the sustainability ratings were first published, 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.

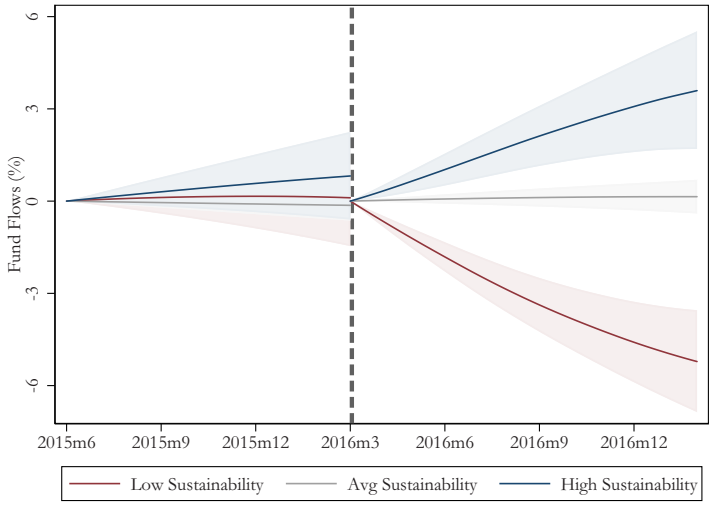

Figure 1. Cumulative fund flows by sustainability rating. Estimates are accumulated from a local linear plot of monthly flows after removing year-by-month fixed effects for nine months before and 11 months after rating publication (denoted by the dashed vertical line). Shaded areas indicate the 90% confidence interval. (Color figure can be viewed at wileyonlinelibrary.com)

Our experiment is rare in financial markets in that it examines a large quasi-exogenous shock, equivalent to approximately 40% of NYSE market cap, which does not directly impact fundamentals. The shock yields a measure of sustainability by simply repackaging publicly available information in a way that attracts attention and is easy to process. Further, 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 investment style or other general measures of sustainability.[1] The response we document is therefore a response 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 margins. If funds were systematically different before the publication of the ratings, then flows could be reflecting this difference. Figure 1 suggests that this is not the case, as do a variety of robustness checks including a matching exercise on pre-publication characteristics and a placebo test.

While the globes are a discrete rating system of five categories, Morningstar also releases each fund's sustainability score and the percentile ranks underlying the ratings. If investors responded to the five-globe rating system rather

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

than to other aspects of sustainability, we should find that the globe category itself drove the fund flows. Examining the percentile ranks that underlie the sustainability rating, we find evidence consistent with discontinuities at the extreme globe category edges, but we find a minimal impact of either the percentiles themselves or the sustainability scores. This suggests that investors focused on the simple globe rating and largely ignored the more detailed sustainability information.

We further find strong flow effects from being in the two extreme globe categories (i.e., one- or five-globe funds) relative to the three categories in the middle, but insignificant differences across funds receiving two-, three-, or four-globe ratings. This result 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, Gennaioli, and Shleifer (2012), Bordalo, Gennaioli, and Shleifer (2013a)) as well as the impact of attributes that stand out in consumer choice (Bordalo, Gennaioli, and Shleifer (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 that we observe in the short term allows us to reject the hypothesis that investors are indifferent to sustainability as well as the hypothesis that they view sustainability as a negative characteristic, but it leaves open the question of which specific aspect of sustainability drives investors to reallocate funds from one-globe funds to five-globe funds. While we are unable to definitively pinpoint the specific motive, we explore 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 those of other investors. Since noninstitutional share classes display a similar pattern, institutional constraints cannot fully account for the results.

Another possible explanation is that investors rationally view a high sustainability rating as a signal of high future returns. We examine whether funds experienced high returns after high sustainability ratings relative to a variety

[2] More broadly, our findings are consistent with the 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)). Furthermore, it is consistent with existing literature showing that people overweight extreme attributes when making judgments about people (Skowronski and Carlston (1989)) or when evaluating outcomes (Sussman and Shafir (2012), Sussman (2017)) and make choices to avoid products with attributes ranked in extreme positions when confronted with trade-offs (Simonson and Tversky (1992), Tversky and Simonson (1993)).

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 data, we do not find evidence of high returns for high sustainability funds.

If the results are not driven purely by institutions or a rational belief in higher expected returns, then it follows that some investors want to hold high sustainability funds and avoid low sustainability investments due to either an irrational belief that there is a positive correlation between future returns and sustainability or nonpecuniary motives (such as altruism, warm glow, or social norms). Because the data do not allow us to distinguish between these two possibilities, we run an experiment using MBA students and Amazon Mechanical Turk (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. This pattern of an inverse relation between expectations of risk and returns is consistent with judgments based on affect rather than reason (e.g., Slovic et al. (2004, 2005, 2007), Finucane et al. (2000)). We also find some evidence of nonpecuniary motives across both populations. Participants 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 performance and risk expectations 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 lead investors to make choices based on nonpecuniary motivations.

Our paper contributes to the literature that examines how investors value nonfinancial aspects of stocks. While other studies examine how subsets of investors value characteristics of securities, such as whether it is a "sin" (Hong and Kacperczyk (2009)), local (Huberman (2001)), or whether it offers a certain dividend yield (Harris, Hartzmark, and Solomon (2015), Hartzmark and Solomon (2019a)), our study examines a quasi-exogenous shock, which means that we evaluate how all mutual fund investors collectively value the characteristic of interest, rather than the subset of investors who hold securities with the characteristic. Perhaps, most closely related to our paper, Hong and Kacperczyk (2009) find that sin stocks yield higher returns, consistent with investors requiring a premium to hold these companies due to social norms. Our paper complements this finding by examining an exogenous shock to a significantly larger part of the market using a more direct measure of demand.

A recent literature examines the rapidly growing set of investment products with explicit social responsibility mandates (e.g., Geczy, Stambaugh, and Levin (2005), Bollen (2007), Benson and Humphrey (2008), Bialkowski and Starks (2016), Barber, Morse, and Yasuda (2017), Riedl and Smeets (2017), see Renneboog, Ter Horst, and Zhang (2008) for a review). While understanding the preferences underlying such investments represents an important area of research, it is indicative only 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 have strong

preferences for sustainability while most investors do not directly value sustainability, under standard models (e.g., Berk and Green (2004)), we would not find an effect of the ratings on net flows. If the market consists largely of Berk and Green (2004) investors who solely maximize profits, these investors who do not value sustainability would undo the effects of investors with a preference for sustainability, resulting in zero net flows. On the other hand, if most investors value sustainability, investors seeking only to maximize profit would not be numerous enough to fully offset the flows from sustainability-seeking investors. One interpretation of the flow response we observe is that the relative size of purely profit-maximizing investors in the U.S. mutual fund market as a whole is smaller than that of investors with a preference for sustainability. Our paper thus contributes to this literature by examining the preferences for sustainability among the universe of U.S. 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 the literature on investment in "doing well by doing good."[3] Some sustainable investing is clearly driven by agency issues (Cheng, Hong, and Shue (2013)), while some is consistent with efficient investment, for example, by improving morale (Edmans (2011)). As emphasized by Hart and Zingales (2017), companies should engage in investments for nonpecuniary "prosocial" reasons, such as sustainability, if such investments reflect the preferences of their shareholders. While our paper does not break sustainability down into the part due to agency versus the part due to appeasing shareholders, the general demand for sustainability from mutual fund investors suggests that a significant part of the observed investment in sustainability is not due purely to agency issues.

Finally, the evidence highlights the potential role of emotion in guiding investment decisions. 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 (e.g., Finucane et al. (2000), Slovic et al. (2004, 2005, 2007)), 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 examinations of affect have been prominent within the psychology literature on risk evaluations, its role has been studied primarily in laboratory tasks, and it has received minimal attention in the context of financial products.[4] Thus, an additional contribution of the current work is to highlight the role of affect versus analytic thought in financial decision-making and financial markets as a whole.

[3] For recent overviews, see Bénabou and Tirole (2010), Heal (2005), Kitzmueller and Shimshack (2012), Margolis, Elfenbein, and Walsh (2009), Christensen et al. (2017), and Chowdhry, Davies, and Waters (2019). For examinations of corporate social responsibility as a marketing strategy, see Du, Bhattacharya, and Sen ((2007), (2010)), Maignan and Ferrell (2004), Menon and Menon (1997), Sen and Bhattacharya (2001), and Torelli, Monga, and Kaikati (2011).

[4] For exceptions, see Hirshleifer and Shumway's (2003) study of the role of sun exposure on market movements or Birru's (2018) study of risk-taking and anomaly predictability based on shifts in mood throughout the week.

The paper is organized as follows. Section I describes the Morningstar sustainability ratings. Section II discusses data sources and presents summary statistics. Section III explores whether investors collectively value sustainability. Section IV explores the mechanism underlying the positive value placed on sustainability. Finally, in Section V, we conclude.

I. 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."[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 (ESG) issues. 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.[6] Each fund in a Morningstar category is ranked based on its sustainability score and this ranking serves as the basis of 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% and it is given two globes and rated "Below Average" if it is ranked between 67.5% and 90%. Finally, it is given one globe and rated "Low" if it is ranked in the bottom 10% of its fund category.[7] The globe ranking is prominently reported using images of one to five globes as well as the descriptive label (e.g., "High") on each fund's Morningstar page. The percentile category rank 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 that transforms holdings and ESG ratings into a globe rating, "sustainability" has become a popular term that lacks a clear and consistent definition. An investor who wishes to understand the details of Morningstar's system could easily do so, but it is likely that investors would respond based on their preconceived notion of the meaning of sustainability rather than to the specific details of the rating methodology. It is therefore useful to understand how investors interpret sustainability in general.

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

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

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

Figure 2. Example of globe rating on Morningstar website. This figure provides an example from Morningstar's website of how sustainability information is displayed on a fund's web page. (Color figure can be viewed at wileyonlinelibrary.com)

Toward this end, we recruited 482 participants from an online sample and asked them which elements of a company's business practices they believe "sustainability" refers to.[8] The results are summarized in Table I. The dominant answer was that sustainability relates to a company's environmental practices, with 79% of participants including environmental issues in their definition of sustainability. Participants included a number of other aspects as well, but none of these other aspects received more than 50% of responses. In total, participants listed 2.7 items on average, with less consistency in the selection of the additional items.[9] Importantly, while the meaning of sustainability varied across participants, there was not confusion as to what any given participant's definition was. Only 2% of participants responded that they did not know what was meant when a company's business practices became more sustainable.

II. Data Sources and Summary Statistics

All of the mutual fund data were provided to us by Morningstar. This historical data set is proprietary and not currently publicly available. The data fund-specific identifiers were anonymized by Morningstar, which means that

[8] 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 ESG issues.

[9] For example, the next-most popular item, product quality and safety, was listed by only 48% of respondents.

Table I
Survey on the Meaning of Sustainability

This table reports summary statistics from a survey on the perceived meaning of sustainability. Participants (482) on Amazon Mechanical Turk responded to the question "Recently, many companies have been trying to become 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 based on KLD definitions and were asked to select all categories that applied.

	% Selecting
Environment (e.g., pollution prevention, recycling)	79%
Products (e.g., product quality and safety, provision of products for the economically disadvantaged)	48%
Human Rights (e.g., labor rights in outsourcing, no operations in Myanmar)	34%
Community (e.g., generous giving, support for housing)	32%
Diversity (e.g., promotion of women and minorities, outstanding family benefits)	26%
Employee Relations (e.g., strong union relations, cash profit sharing)	23%
Corporate Governance (e.g., limited compensation to executives, lack of tax disputes)	22%
I Do Not Know	2%
Other	1%

we are unable to link the data to other publicly available data sources, such as holdings, and we cannot update our sample. The sample includes all U.S. domiciled open-end funds with a sustainability rating from Morningstar and is at the monthly frequency. Since Morningstar's ratings cover the entire universe of such funds, the data run the gamut of Morningstar categories.[10] The data are provided at the share class level, but we conduct our analysis at the fund level. Fund size (TNA), dollar flows, and web traffic are calculated as the sum across share classes, while expense ratios and returns are the mean. Morningstar "star" fund ratings are the rating of 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. We remove these share-class-specific attributes to form consistent categories within and across funds. We limit the sample to funds with TNA above one million dollars and winsorize continuous variables at the 1% level.

[10] These range from standard style box categories, such as large growth, to an international focus, such as Europe or China. There are also examples of more exotic styles such as long-short strategies and sector-specific investments such as energy or health.

Fund flows, our main variable of interest, are measured as monthly dollar flows divided by TNA at the end of the prior month.[11] Flows are noisy and may vary systematically based on characteristics such as size. To ensure that the results are not driven by such properties, we examine a normalized flow variable. To construct this variable, each month we split firms into deciles based on size and assign each fund to percentiles based on flows within each size decile. This normalized flow variable is inoculated from differences in flows by fund size as well as from outliers.[12]

Table II, Panel A, reports summary statistics for the funds after the publication of the sustainability ratings in March of 2016 through January of 2017. In Table II, Panel B, we report summary statistics prior to the globe publication for each globe ranking, where globe is what each fund was eventually 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 do not observe systematic differences across funds by globe rating.

In Table II, 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 flows to funds rated highest in sustainability were nearly zero. When we examine 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 II, 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 when no switching was possible. In general, for a fund ranked 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.

III. Do Investors Value Sustainability?

A. Attention to Ratings

Although Morningstar created its sustainability ratings because it believed that there would be investor interest in them, one reasonable hypothesis is that the ratings did not receive attention when published and thus had no impact. This could be because investors did not care about the ratings or did not know about the ratings. This could also be because investors were already aware of the information contained in the ratings, as the Sustainalytics score for each

[11] While it would be interesting to examine whether the effect comes from new buyers or the selling behavior of existing shareholders, our anonymized data limit the analysis to the net effects of both.

[12] We thank an anonymous referee for suggesting this methodology to construct the variable.

Table II

Summary Statistics

This table presents summary statistics for the data. Panel A examines all funds post-publication, from March 2016 through January 2017. Statistics are at the fund level. Panel B examines the data by globe for the nine 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	2,184.33	8,617.62	19.98	76.36	350.29	1,370.35	4,105.12
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	Age	Rating
All	28,713	2,112.38	0.10	50.33	229.80	−1.27	183.94	3.03
One globe	2,982	1,392.05	0.12	48.12	235.64	−1.40	178.57	2.92
Two globes	6,215	2,370.89	0.28	52.52	223.62	−1.23	184.36	3.10
Three globes	9,891	2,353.41	0.01	50.10	229.37	−1.29	183.85	3.10
Four globes	6,422	1,937.82	0.03	49.54	218.84	−1.24	194.52	3.02
Five globes	3,174	1,885.88	0.19	50.43	259.84	−1.18	167.08	2.76

(*Continued*)

Table II—*Continued*

Panel C: Post-Publication Summary Stats By Globe

	Obs	Size	Flows	Normalized Flows	Visits	Return	Age	Rating
All	34,105	2,184.32	−0.41	50.33	209.45	1.62	195.35	3.01
One globe	3,170	1,039.96	−0.90	44.69	164.34	1.72	180.58	2.74
Two globes	7,207	2,438.41	−0.32	50.36	205.35	1.71	195.89	3.05
Three globes	12,183	2,298.06	−0.41	50.70	201.52	1.62	198.11	3.10
Four globes	7,816	2,197.10	−0.45	50.69	207.97	1.56	201.51	3.00
Five globes	3,730	2,267.72	−0.10	53.13	284.43	1.53	184.94	2.83

Panel D: Transition Probability

Current Month Rating	Next Month Rating				
	One Globe	Two Globes	Three Globes	Four Globes	Five Globes
One globe	2,297	539	37	8	0
	(79.73%)	(18.71%)	(1.28%)	(0.28%)	(0.00%)
Two globes	436	4,869	1,170	29	6
	(6.70%)	(74.79%)	(17.97%)	(0.45%)	(0.09%)
Three globes	64	983	8,753	1,185	28
	(0.58%)	(8.93%)	(79.48%)	(10.76%)	(0.25%)
Four globes	18	93	1,032	5,415	512
	(0.25%)	(1.32%)	(14.60%)	(76.59%)	(7.24%)
Five globes	4	14	61	467	2,837
	(0.12%)	(0.41%)	(1.80%)	(13.80%)	(83.86%)

Figure 3. Google search for sustainability and star rating. This figure 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." Data are from a search for both terms jointly so that the magnitude of the two lines can be directly compared. The monthly measure is the average of the weekly measure, where months are defined based on the week-ending period. The data cover January 2015 through January 2017. (Color figure can be viewed at wileyonlinelibrary.com)

company was based on publicly available information, the Sustainalytics scores themselves were also publicly available (e.g., through Bloomberg), and fund holdings were publicly reported. Thus, all of the information used to construct the globe ratings was available before publication of the ratings. It is therefore possible that investors already acquired and processed the information that Morningstar aggregated into a globe rating, in which case 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" and "Morningstar sustainability rating."[13] 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" and the vertical gray line corresponds to the first publication of those ratings.

Two observations are worth noting in Figure 3. First, before their publication, there was no measurable volume of searches on the sustainability ratings.

[13] The monthly measure is the average of the weekly searches, where months are assigned based on the month in which a given week ends. Google trends normalize the results of each 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.

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 on the sustainability rating as there were on the star rating. This pattern is consistent with significant interest in the sustainability ratings, which were publicized through white papers and traditional marketing campaigns, included as a search filter option for some Morningstar clients, covered by outside media outlets, and included on every rated fund's Morningstar web page. The large search volume suggests that many investors became aware of the existence of the rating and were likely interested in issues related to sustainable investing.[14]

The paper's focus is on investors' perception of sustainability. For the ratings to provide a valid test of this mechanism, investors cannot have systematically sorted into funds based on their future rating before publication. The search frequency and subsequent findings suggest that the publication of the ratings 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 globe rating.[15] If such behavior were widespread, this would potentially affect the interpretation of the results related to the cause of return predictability (see Section IV.B) but would not change the interpretation of the paper's core results related to fund flows and investor preferences.

B. Base Results

Did the publication of the sustainability ratings affect how investors traded mutual funds? To address 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 investors' revealed preferences. If a fund is generally viewed as more desirable after its rating becomes public, money will flow in to it and it will grow. If it is viewed as less desirable, we will see money flow out of it and it will shrink. This stands in contrast to individual stocks, which have a fixed supply in the short run and therefore do not provide such a direct measure of investor responses.[16]

[14] Search volume may be elevated in the period directly after the launch of the ratings as a result of media attention surrounding the launch and the ratings system. Our results should be interpreted within this context.

[15] 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/).

[16] 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. The Internet Appendix, which is available in the online version of the *Journal of Finance* website, shows no significant flow variation 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

In addition, our setting is rare in financial markets in that we examine an event that does not change fundamentals. 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 instead examines a shock to the salience of a characteristic, so the characteristic itself is not changed and hence there is no change to the underlying business by the publication of the fund rating.

When Morningstar published its ratings, it displayed three separate measures of sustainability on a fund's page as shown in Figure 2. In particular, it displayed a fund's raw sustainability score, the percentile rank of that score within the fund's Morningstar category, and an image of how many globes the fund was rated based on percentile rank cutoffs. If investors want to invest in the most sustainable fund in the market overall, the raw sustainability score is the most informative measure, but it is difficult to interpret without exerting a significant amount of effort to understanding its scale. The percentile rank variable yields a continuous measure of within Morningstar category rank that is easier to interpret than the raw sustainability score and provides more granular detail than the globe rating. Thus, if investors want to invest in the most sustainable fund in a given Morningstar category, the percentile rank is the most informative measure. As shown in Figure 2, the globe rating is given the most space on the sustainability portion of a fund's web page. It is presented as a large image of the number of globes along with the corresponding rating label (e.g., High, Average, or Low) in a larger font than either of the two quantitative measures. However, all of the information needed to determine 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, however, investors' attention is drawn to the globe rating, they may simply be examining this salient measure and ignoring the underlying percentiles.

In Table III, we explore the reaction to each sustainability measure and find that it is the globes, rather than the other measures, which are the main driver of fund flows. Specifically, we regress fund flows on each sustainability measure and include Morningstar category-by-year-by-month fixed effects to control for time variation by category. In column (1) of Panel A, we examine the raw sustainability score and the percentile rank in category variables. We find insignificant coefficients on both. In column (2), we include dummy variables for each globe rating omitting the three globe category. One-globe funds, that is, those funds rated worst in terms of sustainability, experienced roughly −0.44% greater outflows per month than three-globe funds, with a t-statistic of −2.80 clustered by month and fund. Five-globe funds, those rated highest in terms of sustainability, experienced 0.30% greater inflows per month than three-globe funds, with a t-statistic of 1.81. These point estimates indicate that the TNA of

examining, these funds see Bialkowski and Starks (2016), Benson and Humphrey (2008), Bollen (2007), and Geczy, Stambaugh, and Levin (2005).

Table III

Fund Flows in Response to Sustainability Rating

This table shows how mutual fund flows vary with measures of sustainability. The dependent variable is fund flows, which are regressed on three proxies for sustainability, namely, the raw sustainability score, the percentile rank within category, and dummy variables for globe rankings. Columns (4) through (6) of Panel A and the even numbered columns in Panel B include as additional controls the return in the prior month, the return over the prior 12 months, the return over the prior 24 months, the log of size in the prior month, the expense ratio, the Morningstar star rating in the prior month, and the log of fund age. Panel A does not weight regressions, while Panel B weights by log of TNA in the prior month in columns (3) through (6). All columns include year × Morningstar category × month fixed effects. Data are 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 fund, 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
	(0.94)					(0.65)
Category percent rank	0.000983					0.00398
	(0.23)					(0.89)
One globe		−0.441**	−0.457***	−0.352**	−0.402***	−0.408**
		(−2.80)	(−3.21)	(−2.81)	(−3.38)	(−2.55)
Two globes		0.0964		0.134		
		(0.92)		(1.33)		
Four globes		−0.0353		0.0440		
		(−0.40)		(0.51)		
Five globes		0.297	0.281*	0.379**	0.331**	0.319*
		(1.81)	(1.87)	(2.39)	(2.35)	(1.85)
Diff: five globes − one globe		0.737	0.738	0.731	0.733	0.727
p-value: five globes = one globe		0.00384	0.00382	0.00377	0.00376	0.0296
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	No	No	No	Yes	Yes	Yes
R^2	0.0503	0.0511	0.0510	0.0909	0.0907	0.0909
Observations	34,046	34,046	34,046	32,421	32,421	32,421

(*Continued*)

Table III—*Continued*

Panel B: Normalized Flow and Size-Weighted Regressions

	Normalized Flows		Size-Weighted Flows		Size-Weighted Normalized Flows	
	(1)	(2)	(3)	(4)	(5)	(6)
One globe	−5.743***	−4.427***	−0.444***	−0.353**	−5.802***	−4.389***
	(−5.53)	(−4.62)	(−3.20)	(−3.07)	(−5.47)	(−4.59)
Five globes	2.474**	3.253***	0.302*	0.358**	2.686**	3.465***
	(2.25)	(3.19)	(2.07)	(2.64)	(2.35)	(3.29)
Diff: five globes − one globe	8.217	7.680	0.746	0.711	8.487	7.855
p-value: five globes = one globe	0.000402	0.000397	0.00299	0.00330	0.000380	0.000347
Cat by YM FE	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	No	Yes	No	Yes	No	Yes
R^2	0.0711	0.157	0.0498	0.0883	0.0725	0.160
Observations	34,046	32,421	34,046	32,421	34,046	32,421

the lowest sustainability funds decreased 5.4% per year, while the TNA of the highest rated funds increased about 3.6% per year. The difference between one- and five-globe funds is 0.74% per month, with the p-value on the test that they are equal, 0.004. 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, then the relevant test is how one- and five-globe funds compare against those rated in the middle. In column (3), we conduct such a test, two-, three- and four-globe funds comprise the omitted category. One-globe funds see −0.46% lower outflows per month than middle ranked funds, with a t-statistic of −3.21, while five-globe funds see 0.28% stronger inflows than middle ranked funds, with a t-statistic of 1.87.

The results above may be due to globe ratings systematically varying with other variables associated with flows. To address this possibility, 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 ensure that the globe ratings are not simply capturing fund flows based on size, we control for the log of fund TNA in the prior month. We also add controls for the expense ratio and the log of fund age. Because there could be a correlation between Morningstar's globe ratings and star ratings, we also control for a fund's star rating. After including these controls, we find similar effects. Specifically, one-globe funds are associated with −0.40% lower outflows with a t-statistic of −3.38, while five-globe funds have 0.33% greater inflows with a t-statistic of 2.35.

In column (6), we include all three sustainability measures simultaneously. We find that investors responded to the coarse globe ratings, not the two other measures: after including the globe rating variables, the coefficients on both the category percentile rank and the raw sustainability score are insignificant, while the coefficients on the globe ratings are materially unchanged; the one-globe variable is negative and significant, while the five-globe variable is positive and significant.[17] These results suggest 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.

In Table III, Panel B, we use the normalized flow variable to address the concern that the results are driven by systematic noise over the short sample. If the results are driven by outliers or small firms with volatile flows rather than the sustainability ratings, the results should decrease or disappear in this specification. If the normalized variable reduces noise that attenuated the estimates using raw flows, the relation should be stronger in this specification.

The first two columns of Panel B show that the results become stronger using the normalized flow variable. Examining column (2), which includes the

[17] One potential concern is that the results are due to a high correlation between the raw score and the percentile rank. In the Internet Appendix, we repeat the regression with each variable separately and find similar results.

additional controls, one-globe funds have 4.4 percentile lower outflows than middle ranked funds, with a t-statistic of -4.62, while five-globe funds have 3.3 percentile higher inflows than middle-ranked funds, with a t-statistic of 3.19. The spread of 7.7 percentiles between one- and five-globe funds has a p-value of 0.0004. Reducing the noise in flows using this normalization significantly increases the statistical significance of the results, consistent with a strong investor response to the globe ratings themselves.

Another concern is that the regressions are driven by small, economically unimportant funds. In columns (3) through (6), we repeat the analysis weighting the regressions by the log of fund size the prior month. The results are slightly stronger in terms of point estimates. For the flow measure including controls, one-globe funds underperform middle-ranked funds by -0.40%, with a t-statistic of -3.41, and five-globe funds outperform middle ranked funds by 0.33%, with a t-statistic of 2.38. The spread between the two of 0.73% has a p-value of 0.0002. Examining the normalized measures after including the additional controls in column (5) and (6), one-globe funds had outflows of -4.4 percentiles, with a t-statistic of -4.71, while five-globe funds had inflows of 3.3 percentiles, with a t-statistic of 3.25. The difference between the two of 7.7 percentiles has a p-value of 0 to four decimal places.

C. Within Globe Rating Analysis

The results above suggest that investors focus on the extreme globe ratings and largely ignore both the middle globe ratings and the underlying sustainability information that is publicly available. If so, funds within a globe rating should receive a 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 fall on different sides of an ad-hoc globe rating break point 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 categories in the middle.

In Figure 4, we explore these hypotheses by taking a more detailed look at the relation between fund flows, globe ratings, and the underlying percentile ranks. Panel A shows the average fund flow for each percentile rank from 1 through 100 after removing year-by-month fixed effects. Panel B repeats the analysis using the normalized flow measure. The dashed vertical lines indicate the globe cutoffs with the globe category listed at the top of the chart. The bars to the extreme left are five-globe 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 10 percentiles assigned to high sustainability (five-globes) funds, we see that 9 of the 10 point estimates are positive and 5 of the 10 are positive and significant at the 90% level. Examining the 11 percentiles assigned to low sustainability (one-globe) funds, we see that all 11 are negative and 5 of the 11 are negative and significant at the 90% level. Looking at the two-, three- and four-globe categories, there

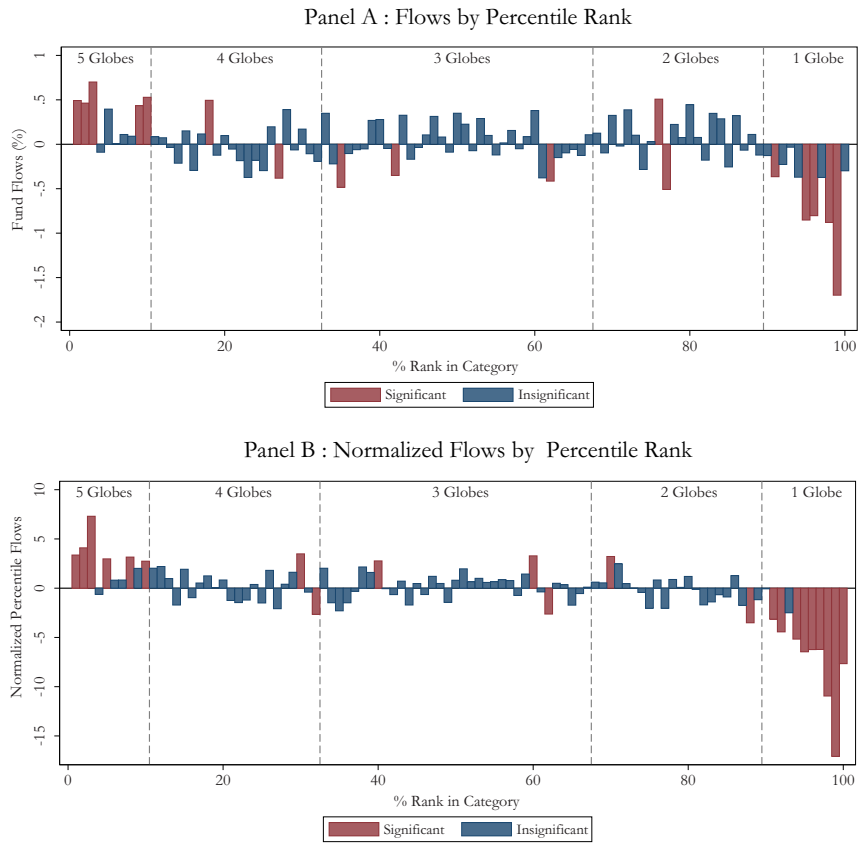

Figure 4. Flows by percentile rank of sustainability. This figure shows average percentage flows for each sustainability percentile rank after controlling for year × month fixed effects. Panel A shows the averages of this variable. Panel B examines the normalized flow measure, normalized to a mean of zero. Significant indicates the average flow is significant at the 90% level. (Color figure can be viewed at wileyonlinelibrary.com)

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 10 significant percentiles in the 21 extreme percentile categories. Panel B repeats the analysis with the percentile rank measures and if anything the results are 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, seven are significant. The evidence suggests that investors did indeed respond to the one- and five-globe categories, largely ignoring the two-, three-, and four- 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 to differ based on percentile ranks, but not at globe cutoffs, corresponds to the

extreme low sustainability funds, which observed higher outflows when ranked 98th and above. Comparing the flow in percentiles 98 and above to 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 that is −7.2 percentiles lower, with a *t*-statistic of −8.37. The effect of being in the top percentiles of high sustainability funds is more muted. The top three percentiles for five-globe funds have 0.35 greater inflows with a *t*-statistic of 3.64, while for the normalized measure, these funds receive 3.4 percentiles greater inflows 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 sustainability ratings.

If investors are responding to the globe ratings, the ad-hoc choice of cutoff will leave similar mutual funds receiving different ratings on either side of the cutoff. We examine the effect of the cutoff choice more formally in Table IV 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 Emerging Market funds category, and the top 26 received a five-globe rating. Accordingly, 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)). Panel A examines flows around one-globe break points, while Panel B examines flows around five-globe break points.

Table IV presents a series of estimates of the discontinuities surrounding the globe cutoffs. For each estimate, we provide a conventional estimate and an estimate correcting for the bias as described in Calonico, Cattaneo, and Titiunik (2014). We select bandwidths using the methods described in Calonico, Cattaneo, and Titiunik (2014) and Calonico et al. (2019) and include monthly fixed effects, using the methods described in Calonico et al. (2019).[18] We present each estimate using uniform bandwidths in odd columns and separate bandwidths in even columns. In the first two columns of Panel A, the four point estimates based on the simplest specifications (without clustering) range from −0.319 to −0.536 with *z*-statistics based on the heteroskedasticity-robust nearest-neighbor estimator from Calonico, Cattaneo, and Titiunik (2014) ranging from −2.02 to −2.85. These results suggest that moving from a two-globe rating to a one-globe rating leads to a decrease in flows of roughly 0.4% per month. Examining the same specifications for the five-globe discontinuity in Panel B, the four-point estimates range from −0.503 to −0.821 with *z*-statistics ranging from −2.56 to −3.27, which suggests that moving from a five-globe rating to a four-globe rating leads to a decrease in flows of roughly 0.6% per month. Each of these eight estimates is consistent with discontinuities around the one- and five-globe break points.

Because a variety of ad-hoc choices are involved in conducting a regression discontinuity test, we present a number of alternative estimates to show that

[18] In the Internet Appendix, we repeat the analysis on residuals controlling for month-by-category fixed effects.

Table IV

Regression Discontinuity Tests of Fund Flows around Sustainability Rating Breakpoints

This table conducts regression discontinuity tests of mutual fund flows around globe breakpoints. The first row shows the conventional RD estimate, while the second corrects for the bias described in Calonico, Cattaneo, and Titiunik (2014). Panel A examines flows at the one-globe break point, while Panel B examines flows at the five-globe break point. All specifications include monthly fixed effects. Bandwidths are selected using the procedure described in Calonico, Cattaneo, and Titiunik (2014) and Calonico et al. (2019) and are reported below. Odd columns use the same bandwidth on either side of the break point, while even columns estimate separate bandwidths. Columns (1) through (4) calculate z-statistics using a heteroskedasticity-robust nearest-neighbor estimator. Other columns calculate z-statistics using a cluster-robust plug-in residuals variance estimator as described in Calonico et al. (2019) with the clustering variable indicated below the estimates. The data are 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 at One-Globe Break Point

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	−0.463**	−0.319**	−0.466***	−0.326**	−0.412*	−0.305	−0.556***	−0.385**
	(−2.46)	(−2.02)	(−2.60)	(−2.15)	(−1.66)	(−1.42)	(−2.58)	(−2.00)
Bias corrected	−0.536***	−0.388**	−0.536***	−0.391***	−0.490**	−0.376*	−0.624***	−0.446**
	(−2.85)	(−2.46)	(−2.99)	(−2.58)	(−1.98)	(−1.75)	(−2.90)	(−2.32)
Cluster	No	No	Rank	Rank	Fund	Fund	Month	Month
Bandwidth left	42.43	98.66	42.21	98.60	47.34	102.7	34.27	69.42
Bandwidth right	42.43	42.40	42.21	41.38	47.34	43.81	34.27	35.85
Observations	31,668	31,668	31,668	31,668	31,668	31,668	31,668	31,668
Effective obs.	10,597	16,938	10,597	16,911	11,399	17,331	9,166	13,868

Panel B: Flows at Five-Globe Break Point

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Conventional	−0.747***	−0.503**	−0.777***	−0.528***	−0.712**	−0.477*	−0.846***	−0.585***
	(−2.98)	(−2.56)	(−3.19)	(−4.45)	(−2.28)	(−1.81)	(−4.59)	(−3.09)
Bias corrected	−0.821***	−0.578***	−0.852***	−0.603***	−0.807***	−0.564**	−0.883***	−0.632***
	(−3.27)	(−2.94)	(−3.50)	(−5.09)	(−2.59)	(−2.14)	(−4.79)	(−3.34)
Cluster	No	No	Rank	Rank	Fund	Fund	Month	Month
Bandwidth left	19.31	22.15	17.55	16.34	21.88	25.03	13.18	15.54
Bandwidth right	19.31	77.68	17.55	82.38	21.88	83.75	13.18	65.97
Observations	32,241	32,241	32,241	32,241	32,241	32,241	32,241	32,241
Effective obs.	7,041	15,694	6,765	15,758	7,855	16,440	5,240	14,071

no particular choice drives the results. Lee and Card (2008) show that standard errors can be misspecified with a discrete running variable and suggest clustering by the running variable to account for such an effect. Following this suggestion, in columns (3) and (4), we cluster by category rank (the running variable in our analysis) and find estimates that are similar and significant at the 95% level.[19] In columns (5) and (6), we rerun the analysis clustering by fund. For the one-globe break point, the point estimates are slightly smaller in absolute value with one insignificant (z-statistic of -1.42), two significant at the 90% level (z-statistics of -1.66 and -1.75), and one significant at the 95% level (z-statistic of -1.98). The five-globe estimates are more negative, ranging from -0.477 to -0.807, with one estimate significant at the 90% level (z-statistic of -1.81) and the other three estimates significant at the 95% level (z-statistics more extreme than -2.14). In the final two columns, estimates are clustered by month. Each of the eight estimates is slightly more negative than in the other columns with parallel bandwidth specifications and significant at the 95% level with z-statistics more extreme than -2.00 for one-globe funds and more extreme than -3.09 for five-globe funds. In the Internet Appendix, we repeat the analysis using normalized flows, flows removing category-by-month fixed effects, and normalized flows removing category-by-month fixed effects. While there is some variation in the magnitude and statistical significance of the estimates, the evidence is consistent with discontinuities at the one- and five-globe break points.

The paper presents consistent evidence suggesting that investors respond to the coarse globe ratings. In Table III, we found no evidence of a significant flow response to the raw sustainability scores, percentile ranks, or middle globe categories. In Figure 4, there is no discernible pattern in the flow response to the middle-ranked globe funds. In Table IV, we find evidence of discontinuities surrounding the one- and five-globe break points. In the Internet Appendix, we show that funds receive more extreme flows in months in which they are ranked 1 or 5 globes compared to months in which they are not. The exception noted above is that investors appear particularly averse to the absolute least sustainable funds based on percentile ranks. Taken as a whole, the results emphasize that the formation and display of information as categories can have a significant impact on investor decision-making.

D. Controlling for Preperiod

The prior section shows that there was a high correlation between globe ratings and flows. One still may be worried, however, that the prior analysis simply captures pre-period differences in funds that are not addressed by these specifications. In this section, we examine whether the globe ratings capture such preperiod effects and find that this is not likely to be the case.

[19] The choice of clustering influences bandwidth selection as well as the bias correction, and thus point estimates are not identical under different clustering assumptions in the Calonico et al. (2019) framework.

Figure 1 plots cumulative flows based on globe ratings, both before and after their publication. The globe ratings did not exist before they were published, so in the period before their publication, each 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 from a local linear plot are accumulated to construct the plot for the nine months before and 11 months after the rating's publication. Before publication, to the left of the dashed line, there are no significant differences across the groups and the trends are roughly similar. After publication, we see significant increases in flows to funds rated five globes and significant outflows from funds rated one globe.

Figure 5 examines these results further, presenting the raw averages for each month along with a version of the local linear plot without accumulating the flows. Examining the simple averages during the preperiod in Panel A, there is no clear relation. Four of the nine preperiod months have higher flows to funds that will be rated one globe than to funds that will be rated five globes, with the other five showing the opposite pattern. The smoothed local linear plots in Panel B are consistent with these patterns as there is no significant difference across globe categories in the preperiod. 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 is significantly greater. Each month post-publication, the five-globe funds have higher inflows than the one-globe funds. The results are consistent with flows being affected by the ratings and the funds being broadly similar before the ratings were published.

We examine this pattern more formally in Table V 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. We use a nearest-neighbor match based on flows, size, age, and return prior to the ratings. The results suggest 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) using the normalized measure.

While the analysis matches on observed fund characteristics, there is always a concern that we are omitting a relevant variable. In Panel B, we additionally match on the fund's preperiod loadings on orthogonal projections of Vanguard benchmarks (see Section IV.B.1 for details of their construction). To the extent that similar funds covary together on a wide variety of possible characteristics, this should control for the characteristics not explicitly included in the match. 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

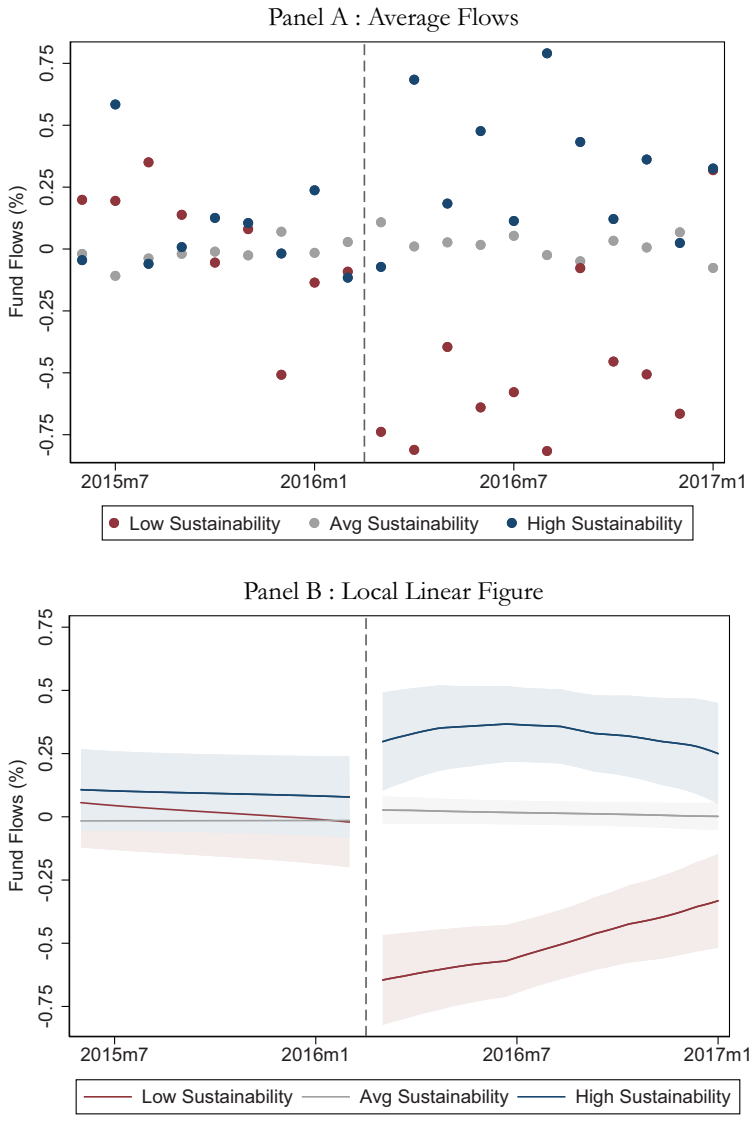

Figure 5. Flows by month. This figure shows average percentage flows by month controlling for year × month fixed effects. Panel A shows the average of the variable for each month and Panel B shows a local linear plot with 90% confidence intervals. (Color figure can be viewed at wileyonlinelibrary.com)

experience outflows of −0.19% per month. The results suggest that preperiod differences do not account for the results.

The Internet Appendix contains additional analysis further mitigating potential possible concerns. To test whether the results are due to a general trend related to sustainability, we construct pseudo-globe ratings based on

Table V
Fund Flows in Response to Sustainability Rating Matching on Preperiod 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 ratings based on flows, size, return over the prior 12 months, and fund age, each based on the value before the publication of the rating in February 2016. In Panel B, funds are matched on these characteristics as well as their loadings on Vanguard benchmark portfolios using the methodology of Berk and Van Binsbergen (2015) described in section 4.2.1. An extreme-rated fund is matched to another fund based on the initial 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)
One globe	−0.715***		−6.700***	
	(−9.07)		(−11.60)	
Five globes		0.206***		3.799***
		(2.60)		(7.44)
Observations	33,262	33,262	33,262	33,262

Panel B: Matching on Characteristics and Loadings

	Flows		Normalized Flows	
	(1)	(2)	(3)	(4)
One globe	−0.523***		−4.391***	
	(−5.91)		(−6.36)	
Five globes		0.185***		5.589***
		(2.68)		(9.75)
Observations	33,232	33,232	33,232	33,232

KLD scores in the years prior to the Morningstar rating publication. The pseudoratings do not predict fund flows. While Morningstar ratings are sticky, they are recalculated each month and funds do change categories. The results in the Internet Appendix show that funds experience more extreme flows when they possess an extreme rank compared to months in which they do not.

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 that accounts for flows. This variable would have to be related to the discrete globe ratings to account for the discontinuity analysis, but not to the underlying sustainability score or more continuous percentile ranks. Moreover, the alternate variable could not be capturing fixed fund attributes, as we find that the effect is stronger when funds are ranked high or low in sustainability than when they are not. The alternate variable must also not be captured by our explicit controls or correlations on factor loadings, and its impact must begin

only when the ratings are published as the placebo analysis shows that it was not present before. While such alternatives are not impossible, we feel that the results strongly support the parsimonious explanation that the globe ratings had a causal impact on fund flows.

E. Economic Impact

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

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

Another way to evaluate the magnitude of the ratings is to compare it to the impact of the Morningstar star ratings, which Del Guercio and Tkac (2008) argue are the "undisputed market leader" for fund ratings and "arguably the primary inputs to many investors' decisions." Reuter and Zitzewitz (2010) find that moving up one star rating results in 43 to 52 basis point higher flows per month. Thus, the impact of the sustainability rating on flows is of a similar magnitude to that of the Morningstar star rating system.

The above magnitudes are estimates of the net impact of the ratings publication and associated publicity and roll-out campaign by Morningstar and hence should be viewed in this context. The initial sorting documented here will be greater than the long-run effect that we expect to occur after investors have sorted into various funds based on their sustainability. Although we are unable to examine whether the flow effects are permanent or transitory given the data's short time series, we would not expect these effects to continue to the same extent without further ratings changes.[20] On the other hand, these are estimates of net flows, which means that they underestimate the number of investors who flowed into these funds based on sustainability ratings. On net, investors flowed into high sustainability funds, but some investors likely flowed out as well. Thus, the estimates represent a lower bar for the proportion of investors in the market as a whole that value these sustainability ratings.

[20] This is why papers examining the impact of a rating system already in equilibrium are forced to rely on estimates such as regression discontinuity to estimate their impact (e.g., Reuter and Zitzewitz 2010).

Table VI
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 used. In column (1), we show the raw counts and proportion of funds liquidating, while in the other columns, we regress the liquidation dummy variable on dummy variables for globe rankings. Columns (4) and (6) include category × year × month fixed effects. Columns (5) and (6) include as additional controls the return in the prior month, the return over the prior 12 months, the return over the prior 24 months, the log of size in the prior month, the expense ratio, the Morningstar star rating in the prior month, and the log of fund age. The data are restricted to March 2016 and after excluding the final month, the period when the globe ratings were published, and analysis is at the fund 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.

	Summary Stats	Regressions				
	(1)	(2)	(3)	(4)	(5)	(6)
One globe	13	0.00239**	0.00250***	0.00222**	0.00238***	0.00121
	[0.41%]	(2.50)	(2.78)	(2.34)	(2.58)	(1.13)
Two globes	8	−0.000613				
	[0.11%]	(−0.82)				
Three globes	21					
	[0.17%]					
Four globes	15	0.000194				
	[0.19%]	(0.45)				
Five globes	6	−0.000114				
	[0.16%]	(−0.09)				
Cat by YM FE		No	No	Yes	No	Yes
Other controls		No	No	No	Yes	Yes
R^2		0.000327	0.000285	0.0191	0.00548	0.144
Observations	34,046	34,046	34,046	34,046	32,422	32,421

Increasing size is clearly an important aspect of overall fund health and as such the effect of the flow responses should be apparent in other fund attributes. One such attribute is the probability of a fund closing down. In Table VI, we examine the probability that 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 six five-globe funds did. The one-globe rate of closure of 0.41% is more than double that of any of the other globe categories.

To shed further light on this result, we estimate linear probability models in which the left-hand-side variable is equal to one if the fund is liquidated and the right-hand side is the fund's globe rating. Here, the coefficient on a globe rating can be interpreted as the difference in the probability of a fund's liquidation in a given month relative to the omitted category.[21] Column (2) shows that a

[21] In the Internet Appendix, we conduct a similar exercise using Cox proportional hazard models and find analogous results.

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 one-globe funds are 0.25% more likely to close than all of the other funds (t-statistic of 2.78). Columns (4) and (5) add category-by-year-by-month fixed effects and the additional controls, respectively, and find similar results. Combining the fixed effects and controls together in column (6), the point estimate decreases to an insignificant 0.12%. The results suggest that being rated, one globe leads to a higher probability of closing down, but given the rarity of this event, we lack the statistical power to say for certain after including the full battery of controls and fixed effects.

IV. Why do Investors Value Sustainability?

We now explore three separate hypotheses regarding why investors place a positive value on sustainability. The first hypothesis posits that institutional investors value sustainability due to constraints imposed by their institution, the second hypothesis posits that investors (rightly or wrongly) view sustainability as a signal of higher future returns, and the third hypothesis posits that investors have a preference for sustainability for nonpecuniary reasons, such as altruism. These hypotheses are not mutually exclusive, and it is likely that each affects our results to some degree. In this section, we explore the extent to which each matters, but we are not be able to draw definitive conclusions as to the driving force underlying the demand for mutual funds with high sustainability ratings.

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 attempts to measure. This could be due to the salience of the image or to people believing that any rating Morningstar creates is a positive signal due to its reputation. While this is likely true for some investors, we believe that 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. Second, investors—especially institutional investors—presumably spend significant time and effort on their decisions and thus are likely to understand that the globe ratings were designed to capture a fund's sustainability. Finally, the Google search analysis shows that roughly as many people search for the phrase "Morningstar sustainability rating" as "Morningstar star rating." This suggests that a large number of individuals are sufficiently knowledgeable to search directly for the sustainability rating and are not simply responding to the globe image at the top of the Morningstar web page. It therefore seems reasonable to assume that the flows we observe are driven at least in part by an aspect related to sustainability.

A. 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 of institutions should be different from that of noninstitutional investors who do not share the same constraints. To test this hypothesis, we would ideally examine those institutions that are subject to such constraints. While we do not have data that identify such institutions per se, we can isolate the flows into and out of institutional share classes based on sustainability ratings.[23]

The use of institutional share classes warrants caution when interpreting the results. If institutional investors are present in the market, we assume that they are taking advantage of their size and investing primarily in institutional share classes. However, flows in these share classes may also be capturing the behavior of participants in retirement plans with access to institutional share classes (e.g., Sialm, Starks, and Zhang (2015)). If institutional share classes reflect the preferences only of retirement plan participants, this would mean that institutional investors were absent from the U.S. mutual fund market and cannot be driving the effects we document. In contrast, if institutions are the main driver of the flow patterns we observe, then as long as institutions are present in institutional share classes to some extent, the effects should be concentrated in the institutional share classes but not in the noninstitutional share classes.

Table VII repeats the analysis examining the differential impact of institutional funds based on globe ratings. The analysis is run at the share class level and standard errors are clustered by fund and month. Column (1) includes only noninstitutional share classes. If institutions were solely responsible for the results, we would not expect a large effect of the sustainability rating in this specification. Examining the regression, the spread between one- and five-globe funds was about 60 basis points per month, with a p-value of 0.008. In column (2), we examine only institutional share classes. While statistically slightly weaker, the 72 basis point estimate for the spread between one- and five-globe funds is slightly larger, with a p-value of 0.06.

We next test whether the institutional share classes behaved differently from the noninstitutional share classes. Specifically, in column (3), we use data from both institutional and noninstitutional share classes, we include globe rating dummy variables, and we also include interactions between the globe rating dummies and a dummy variable equal to one if the share class is institutional. Including the globe dummy variables and the interaction terms means that the coefficient on the institutional interaction captures how the flows into the institutional share classes with a given globe rating compare to those into the

[22] Evidence in support of this hypothesis would be consistent with prior literature, showing that institutional investors drive firms' environmental and social investments (e.g., Dyck et al. (2019)) and the 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 requires an investment of greater than $100,000.

Table VII

Institutional Share Class Flows in Response to Sustainability Rating

This table shows how mutual fund flows vary with globe ratings, comparing institutional to noninstitutional 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 between the globe ratings and the institutional dummy variable. Columns labeled "NonInst" include only noninstitutional share classes, columns labeled "Inst" include only institutional share classes, and columns labeled "All" include all share classes. Columns (1) through (3) include category × year × month fixed effects, while column (4) include fund × month fixed effects. All columns include as additional controls the return in the prior month, the return over the prior 12 months, the return over the prior 24 months, the log of size in the prior month, expense ratio, the Morningstar star rating in the prior month, and the log of fund age. 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			
	(1) NonInst	(2) Inst	(3) All	(4) All
One globe	−0.271**	−0.176	−0.219*	
	(−2.31)	(−0.71)	(−1.85)	
Two globes	0.0224	0.0255	0.0279	
	(0.27)	(0.19)	(0.34)	
Four globes	0.0527	−0.00696	0.0594	
	(0.54)	(−0.05)	(0.63)	
Five globes	0.327*	0.551*	0.347**	
	(2.14)	(2.19)	(2.28)	
One globe*Institutional			−0.0579	0.0257
			(−0.23)	(0.09)
Two globes*Institutional			0.0111	−0.0403
			(0.07)	(−0.23)
Four globes*Institutional			−0.0924	−0.180
			(−0.57)	(−1.04)
Five globes*Institutional			0.0970	0.0596
			(0.36)	(0.22)
Diff: five globes − one globe	0.598	0.727	0.565	
p-value: five globes = one globe	0.00781	0.0581	0.0104	
Cat by YM FE	Yes	Yes	Yes	No
Fund by YM FE	No	No	No	Yes
Other controls	Yes	Yes	Yes	Yes
R^2	0.0823	0.0859	0.0756	0.398
Observations	71,771	23,351	95,136	84,739

noninstitutional share classes with the same globe rating. The insignificant interaction terms, with *t*-statistics of −0.23 on the one-globe interaction and 0.36 on the five-globe interaction, suggest similar responses to the globe ratings for institutional and noninstitutional share classes. Finally, we examine whether noninstitutional and institutional share classes within the same fund received different flows. We include a fixed effect for each fund-by-month combination and the institutional dummies interacted with globe ratings. The insignificance of the coefficients, with *t*-statistics of 0.09 on the one-globe interaction

and 0.22 on the five-globe interaction, suggests that institutional and noninstitutional share classes within the same fund received similar flows based on their sustainability rating.

While the institutional share classes explain part of the effect that we observe, the effects are still present and significant in the noninstitutional 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 noninstitutional investors. This could be because institutions have similar preferences as noninstitutional investors, or because institutions face constraints forcing them to behave as if their preferences were similar. Another interpretation is that this analysis does not represent the preferences of institutional investors at all as the behavior represents individual investors trading in their retirement accounts. It could also be the case that institutions behave differently than noninstitutional investors, but they are combined together in the institutional share class in such a way that we lack the power to detect different behavior between the two sets of investors. Under any of these interpretations, including the likely combination of all of them, the results suggest that institutions are not the sole driver of the results that we document.

B. 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. If investors had a rational belief that high sustainability funds would deliver high performance, we would hope that such outperformance would manifest itself in the data. While it is difficult to draw a definitive conclusion examining 11 months of return data, we find evidence more consistent with an inverse relation or no relation between globe ratings and returns rather than the positive relation that would be necessary to account for the flow results under an explanation based on rational expectations.

B.1. Observed Performance

We examine returns relative to a variety of benchmarks in Table VIII. Column (1) examines returns in excess of the risk-free rate. Column (2) examines returns minus the value-weighted return of funds in the same Morningstar category (e.g., Pástor, Stambaugh, and Taylor (2015, 2017)). In column (3), we examine returns in excess of a fund-specific benchmark based on Vanguard loadings. To do so, we follow Berk and Van Binsbergen (2015) to construct an orthogonal basis set of Vanguard index funds using data from 2014 to January 2017.[24] We estimate fund-specific betas on these projections prior to the

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

Table VIII

Returns Based on Globe Rating

This table shows how mutual fund performance varies with globe ratings. In Panels A and B, performance is regressed on a one-globe and a five-globe dummy variable. In Panel A, regressions are value-weighted based on the prior month's TNA and in Panel B, regressions are equal-weighted. Column (1) shows 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 return in a Morningstar category. Column (3) measures returns in excess of Vanguard benchmarks using the methodology from Berk and Van Binsbergen (2015). Column (4) estimates returns in excess of a fund benchmark based on the market, SMB, HML, and momentum factors. The benchmarks in columns (3) and (4) use fund-specific beta estimates from the two years prior to the globe rating. Below each regression, the difference between five- and one-globe funds is reported along with the *p*-value for the test that they are equal. Regressions are at the fund level and returns are in percentages. Standard errors are clustered by month and fund, 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 "CAPM" columns and on the market, size, value, and momentum factors in the "4-factor" columns. Returns are before fees and data are restricted to March 2016 and after, the period when the globe ratings were published. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Value Weighted

	Excess Return (1)	Morningstar Benchmark (2)	Vanguard Benchmark (3)	Four-Factor Benchmark (4)
One globe	0.311**	0.0514	0.209**	0.159
	(2.01)	(0.44)	(2.43)	(1.24)
Five globes	−0.252**	−0.158*	−0.0995	−0.193
	(−2.57)	(−1.87)	(−0.69)	(−1.35)
Diff: five globes − one globe	−0.563	−0.209	−0.309	−0.351
p-value: five globes = one globe	0.0247	0.236	0.131	0.165
R^2	0.00144	0.00224	0.000798	0.00149
Observations	34,083	34,083	33,307	33,307

Panel B: Equal Weighted

	Excess Return (1)	Morningstar Benchmark (2)	Vanguard Benchmark (3)	Four-Factor Benchmark (4)
One globe	0.0924	0.0776	0.113	0.0782
	(0.88)	(0.94)	(1.32)	(0.87)
Five globes	−0.0961***	−0.102**	−0.0494	−0.150***
	(−2.65)	(−2.00)	(−0.45)	(−2.59)
Diff: five globe − one Globe	−0.189	−0.179	−0.163	−0.228
p-value: five globes = one globe	0.0906	0.122	0.238	0.103
R^2	0.000152	0.000620	0.000129	0.000466
Observations	34,095	34,095	33,319	33,319

(*Continued*)

Table VIII—*Continued*

Panel C: Portfolios				
	Value Weighted		Equal Weighted	
	(1) CAPM	(2) Four-Factor	(3) CAPM	(4) Four-Factor
Long five globes − Short one globe	−0.460*	−0.479*	−0.138	−0.173
	(−2.03)	(−2.14)	(−1.01)	(−1.33)
Observations	11	11	11	11

globe rating publication and then use these betas to construct a fund's Vanguard benchmark return in the post-publication period. We employ a similar methodology to construct a fund's four-factor benchmark using beta estimates on the market, size, value, and momentum factors. These measures of performance are regressed on globe ratings and are value-weighted in Panel A and equal-weighted in Panel B. For example, column (1) in Panel A shows that the value-weighted excess returns of one-globe funds were 31 basis points higher than the excess returns of middle-globe funds, and five-globe funds received 25 basis points lower excess returns than middle-globe funds received. The constant in this regression is 159 basis points, which indicates that over this period, one-globe funds earned returns in excess of the risk-free rate of 1.9% per month, while five-globe funds earned excess returns of 1.34% per month. We test whether the 56 basis points difference between the five-globe and the one-globe coefficients is equal to zero and find a p-value of 0.02.

Examining the 16 point estimates, each one-globe estimate is positive and each five-globe estimate is negative. Five of the eight five-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 for 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 0.02 to 0.24.

In Panel C, we form portfolios that are long firms rated five globes and short firms rated one globe. We regress this portfolio on the market factor only 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 return is −48 basis points (t-statistic of −2.14) and equal-weighted the alpha is −18 basis points (t-statistic of −1.33). The portfolio sorts thus yield similar estimates as the panel regressions in Panels A and B.

The short time series and the 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 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 did not manifest itself in the relatively short 11 months of data. The evidence is consistent with the hypothesis that one- and five-globe funds performed similarly as well as the hypothesis that one-globe funds outperformed five-globe funds. The point estimate on five globes is lower than that on one globe in each specification, suggesting the low sustainability funds outperformed high sustainability funds, though the weak statistical significance in some specifications is also consistent with a lack of a relation between globe ratings and performance.

B.2. Potential Explanations of Return Predictability

A variety of arguments have been made that are consistent with sustainability positively predicting performance, negatively predicting performance, or having no relation with the performance. The focus of this paper is on how investors responded to the sustainability ratings, while the question of what accounts for the return patterns is more closely related to the question of how funds responded to the ratings. Although fully answering this question is beyond the scope of this paper, we discuss various possible explanations of fund performance as a function of sustainability ratings.

We group potential explanations of return predictability into three categories. The first relates to the scale of funds with decreasing returns to scale. Berk and Green (2004) assume that funds have decreasing returns to scale. This assumption is empirically supported by the findings of Grinblatt and Titman (1989), Chen et al. (2004), and Pástor, Stambaugh, and Taylor (2015) (although Reuter and Zitzewitz (2010) do not find such an effect). If an investor believed that the sustainability rating would change flows to funds already at their optimal scale in a competitive equilibrium, the investor would expect high sustainability funds to underperform after their inflows and low sustainability funds to overperform after their outflows.[25]

The second class of theories relates to funds buying assets with high Sustainalytics ratings in order to achieve better fund ratings. Such an effect could be specific to funds competing on the Morningstar rating, or could indicate general marketwide shifts in demand for sustainable investments. If funds were aware that ratings induce flows, they may actively trade to receive a higher sustainability rating, potentially at the expense of future returns. If many funds engaged in such a strategy, this could increase the price of assets with high Sustainalytics ratings. This price increase would yield a period of high returns

[25] A fund at its optimal scale is expected to earn zero abnormal returns, so inflows to high sustainability funds already at this scale would induce negative abnormal returns. In the context of Berk and Green (2004), if most investors cared only about returns, they would undo aggregate flow effects induced by the sustainability ratings. The subset of investors who valued sustainability would shift into high sustainability funds and out of low sustainability funds while profit-maximizing investors would do the opposite. If such a pattern occurred, we would see no aggregate flow response in our data.

for funds holding such assets, but would lead to subsequent underperformance as the price pressure reversed.[26]

The third class of explanations relates to the characteristics of the underlying assets, not fund behavior. For example, Hong and Kacperczyk (2009) argue that many investors prefer not to hold "sin stocks," which leads these stocks to earn higher returns. Applying this intuition to our setting, if investors believed that there was hesitance to hold low sustainability stocks, they might expect an inverse relation between returns and globe ratings. On the other hand, Edmans (2011) finds that employee satisfaction predicts positive returns, consistent with the idea that socially responsible screens can positively predict performance if the market is incorrectly pricing such signals.[27] If an investor believed that the market was not correctly pricing attributes correlated with high sustainability, they would expect higher returns for more sustainable funds.[28]

To identify the relative importance of each of these explanations to fund performance would require examining fund trading both prior to the ratings publication and in reaction to the publication itself. We cannot perform such analysis, however, since our data lack holdings and are anonymized, and thus cannot be merged to publicly available holdings data. In the Internet Appendix, we examine returns before and after the ratings publication and find mixed support for each of the three explanations as the noise inherent in fund-level returns over a short period makes drawing definitive conclusions difficult. Holdings-level data would allow for a direct test of whether funds were systematically buying positions with high sustainability ratings and if so whether such behavior impacted the valuation of the underlying stocks. We leave addressing this question to future researchers with data that can be merged with holdings information.

C. Naive Performance Expectations and Nonpecuniary Motives

The remaining explanations are that investors either naively assumed that a high sustainability rating was predictive of high future fund returns or had a

[26] This would be similar to the return patterns observed for index inclusions (Harris and Gurel (1986), Shleifer (1986), Kaul, Mehrotra, and Morck (2000)) and dividends (Hartzmark and Solomon (2013)).

[27] Existing literature suggests that sustainability can help a firm because it induces warm-glow feelings in consumers (Becker (1974), Andreoni (1989), Cahan et al. (2015); see Sen, Du, and Bhattacharya (2016) and Trudel (2019) for reviews of additional psychological drivers of sustainable purchasing), or because corporate goodness can be used to deter harmful regulation or enforcement (Baron (2001), Hong and Liskovich (2015), Werner (2015)) or to signal good governance (Deng, Kang, and Low (2013), Dimson, Karakaş, and Li (2015), Ferrell, Liang, and Renneboog (2016)). Other papers find evidence of sustainable investments being negative for a firm (e.g., Di Giuli and Kostovetsky (2014), Dharmapala and Khanna (2016), Fernando, Sharfman, and Uysal (2017)).

[28] As the interest in sustainable investing grows, it seems less likely that the market will ignore or systematically underprice characteristics related to socially responsible investing.

nonpecuniary preference for holding more sustainable mutual funds.[29] Unfortunately, the natural experiment from Morningstar does not allow for testable predictions that distinguish between naive beliefs about future returns and 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 nonpecuniary motives hypothesis, the decision is driven by altruism, warm glow, or social motives. Thus, differentiating between these two hypotheses requires a measure of expectations of future performance.

C.1. Experiment Overview

To obtain such a measure, we ran an experiment based on the Morningstar ratings to elicit the impact of the globe rating on expected future performance.[30] We gave participants information on 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 provided 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, participants in the experiment may not understand the globe rating scale in the same way as a typical Morningstar investor, which could lead to a different motivation driving the responses of our experimental subjects compared to the Morningstar investors they are meant to represent. Therefore, we replaced the text at the bottom of the Morningstar sustainability rating with a description of the globe ratings.[31]

Each participant was asked to (1) report how well she thought the fund would perform over the next year on a seven-point Likert scale, (2) report how risky she considered an investment in the fund to be on a seven-point Likert scale,

[29] For example, investors in funds with a socially responsible mandate derive utility from the social responsible aspect of the investment and are less sensitive to negative returns (Bollen (2007), Renneboog, Ter Horst, and Zhang (2011)).

[30] Additional details and survey materials are available in the Internet Appendix.

[31] 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 saw 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.

and (3) allocate $1,000 between the fund and a savings account.[32] One group of participants was composed of MBA students at the University of Chicago Booth School of Business (269 students participated), which we take to be more likely to be representative of market participants. In addition, we ran the experiment on 576 participants on MTurk to observe how decision-making differs in a presumably less financially sophisticated subject pool.[33]

C.2. Performance Expectations

If flows to high sustainability funds are driven by increased performance expectations, then more globes should be positively correlated with these expectations. We first analyze whether people associate globe ratings with higher performance and we find supportive evidence. In Figure 6, Panel A, we graph the average performance rating for each of the three globe ratings after removing individual fixed effects. To the left, we examine the MBA students. We 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-statistic of 3.23 clustered by participant. To the right, we see a similar pattern for MTurk participants, with a difference between extreme globe ratings of about 0.8 that 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 the participants expected a fund with a higher globe rating to have higher performance because they thought five-globe funds were riskier. We plot participants' 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 five-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 exhibited slightly stronger behavior with a difference of roughly 0.8, with a t-statistic of -6.86. Thus, it is not likely that the positive correlation between globe ratings and performance is due to compensation for perceived risk. Participants reported that higher globe ratings would result in higher performance at lower risk.

Although the finding that investors believe that performance will be superior and risk will be lower for funds rated high in sustainability may appear surprising, it is consistent with existing research in psychology. The affect heuristic (Alhakami and Slovic (1994), Finucane et al. (2000), Slovic et al. (2004, 2005,

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

[33] Research examining this platform finds that participants recruited through MTurk tend to perform similarly on tasks (Casler, Bickel, and Hackett (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)).

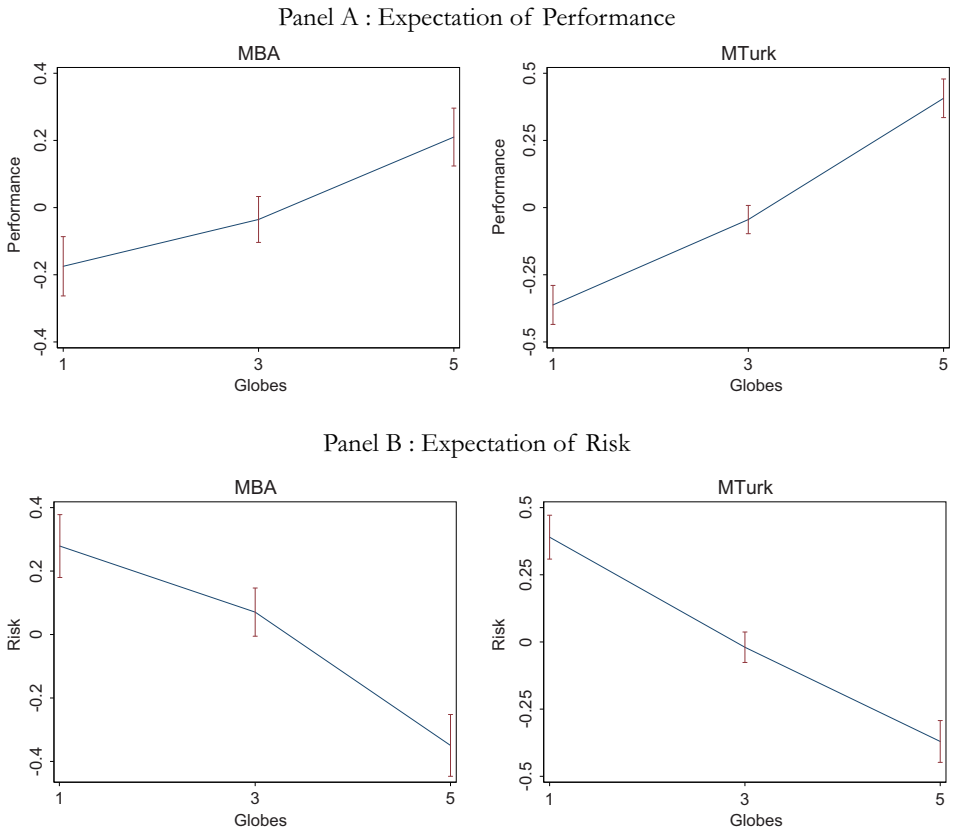

Figure 6. Experimental expectations of future performance and risk by sustainability rating. This graph shows the average performance rating in Panel A and risk rating in Panel B after taking out an individual fixed effect by globe rating. The left graphs correspond to MBA students, while the right graphs correspond to MTurk subjects. Maroon bars indicate the 90% confidence interval. (Color figure can be viewed at wileyonlinelibrary.com)

2007)) and research examining the role of affect in decision-making (Nisbett and Wilson (1977), Klauer and Stern (1992), (Loewenstein et al. (2001)) have been used to understand a range of contexts where risks and benefits are positively correlated, while people believe them to be negatively correlated (Fischoff and Lichtenstein (1978), Slovic et al. (1991), McDaniels et al. (1997)). 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.[34] To the extent that the high sustainability rating causes positive affect toward a mutual fund, the affect heuristic would suggest that people are likely to judge it to be associated with both higher returns and lower risk.

[34] 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.

C.3. Nonpecuniary Motives

While higher expected performance alone could account for the patterns we observe in Morningstar data, this does not rule out the possibility that nonpecuniary motives also play a role. That is, people may be investing in highly sustainable funds because they believe that these funds will outperform, or because they value sustainability and are willing to pay for it. This preference could derive from a number of noneconomic motivations, and would be consistent with evidence and theory that people are concerned with both self-perception (Bodner and Prelec (2003), Gneezy et al. (2012), Haws, Winterich, and Naylor (2014)) and increasing social welfare (Fehr and Schmidt (1999), Charness and Rabin (2002)). For example, investors may experience altruism or a warm glow (Andreoni (1989, 1990)), in which case they invest in sustainability because they derive value from the fact that others benefit, or because they feel good knowing that they are responsible for benefiting others. Alternatively, this preference could stem from social motives and pressures such as the desire to impress others or to avoid contempt or a social backlash (Becker (1974), Schultz et al. (2007), Olson (2009), DellaVigna, List, and Malmendier (2012), White, Simpson, and Argo (2014)).

In the context of our experiment, one potential measure of nonpecuniary motives is the extent to which an investor allocates funds toward five-globe funds or away from one-globe funds beyond what would be explained by their expectation of future performance or risk. If participants cared about the globe ratings solely as indicators of fund performance, we would expect the globes to impact expectations of future performance and risk. Under such an explanation, after controlling for these expectations, the globe ratings would have no further explanatory power. In Table IX, we examine how dollars allocated to portfolios vary with expectations of risk, performance, and globe ratings. The regressions include participant fixed effects and fund fixed effects. If there is a significant difference between the one- and five-globe dummy variables, then a participant should be 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 would be consistent with altruism. We note that interpreting the results in such a manner requires the assumption that the portfolio weights for an investor who cares only about performance and risk increase linearly in the measures based on a Likert scale. While not definitive, we believe that this analysis offers insights into a question for which little information is currently available.

The first column of Table IX 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 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 zero on the difference) and MTurk participants allocate about $130 more (again with a p-value of roughly zero).

Table IX

Experimental Results

This table shows how globe ratings impact expectations of returns, risk, 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 each 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 that 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)
One globe		−50.56**	−27.99	−13.89	−30.82
		(−2.24)	(−1.32)	(−0.43)	(−1.13)
Five globes		57.36***	20.11	−8.080	48.51*
		(2.78)	(1.00)	(−0.27)	(1.75)
Diff: five globes − one globe		107.9	48.10	5.809	79.33
p-value: fivfe globes = one 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

(*Continued*)

Table IX—*Continued*

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)
One globe		−65.69***	−39.28***	−30.29	−43.66***
		(−5.02)	(−3.15)	(−1.49)	(−2.73)
Five globes		64.43***	31.74**	11.44	42.75***
		(4.89)	(2.48)	(0.53)	(2.68)
Diff: five globes − one globe		130.1	71.03	41.73	86.42
p-value: five globes = one 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	1,728	1,728	1,728	624	1,101

Column (3) includes risk, performance, and the globe ratings to determine whether this difference in allocations is explained by performance expectations alone or whether nonpecuniary motives also play a role. After including the controls for risk and performance, the difference between funds allocated by MBA students toward one- versus five-globe funds drops, but remains meaningful at \$48 (with a 95% confidence interval of \$1 to \$95), with a p-value of 0.04. For MTurk participants, this difference is \$71 (with a 95% confidence interval of \$42 to \$100), with a p-value of roughly zero. The point estimates suggest that slightly less than half of the difference in money allocated to one- and five-globe funds can be attributed to nonpecuniary motives for the MBA students, while nonpecuniary motives account for slightly more than half of the difference for MTurk participants.

If part of the difference in allocation is driven by nonpecuniary motives related to sustainability, this effect of globe ratings should be concentrated among participants who considered sustainability when making their decisions. After making their choices, we asked participants the extent to which they considered ESG factors when making their investment decisions. Participants who said they did not consider ESG factors have no reason to exhibit nonpecuniary motives, so we would expect the globe dummy variables capturing such motives to lose their explanatory power for these investors. This is what we find when we restrict the sample to these investors in column (4). MBA students in this group exhibit only a \$5 difference in allocation between one- and five-globe funds, while MTurk subjects exhibit a marginally significant \$41 difference. Column (5) shows strong evidence consistent with nonpecuniary motives when examining participants who considered ESG factors. MBA students allocated a significant \$79 more dollars to five-globe funds (with a 95% confidence interval ranging from \$16 to \$142) and MTurk participants allocated a significant \$86 more dollars to five-globe funds (with a 95% confidence interval from \$51 to \$122). Thus, we find evidence that dollar allocations are driven by expected performance and risk, but also by altruism (or other nonpecuniary motives) above and beyond these factors.[35]

V. Conclusion

We present causal evidence that investors collectively value sustainability, and we 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 more than \$24 billion greater fund flows, while those with the lowest globe ratings face

[35] 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 present would induce the observed empirical pattern in allocations due to the picture itself, rather than the underlying context of the rating. If this were the case, the extent to which an investor considers environmental factors would not be likely to influence investment decisions. These differential responses suggest that the patterns we observe in the experimental setting were largely due to considerations related to sustainability, and not simply an attention effect unrelated to sustainability.

a reduction in fund flows of more than $12 billion. 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 and the literature showing that the choice of how information is displayed influences financial decisions (e.g., Benartzi and Thaler (1999), Hartzmark and Solomon (2019b)). Investors further respond mainly to the extreme-ranked categories, largely ignoring the others, consistent with the literature on the salience of extreme ranks. The results suggest that how categories are constructed, especially extreme categories, can have a significant effect on how decisions are made in a financial setting and impact marketwide variables such as fund flows.

Our natural experiment in which a large part 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, consistent with social constraints placed on institutions being partially responsible for the effect. However, the pattern holds among noninstitutional investors as well. We do not find evidence supporting a rational belief that more sustainable funds perform better; rather, 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 find evidence consistent with nonpecuniary motives, such as altruism or warm glow.

Taken together, our experimental findings support the importance of affect in investment decisions. Specifically, the finding that participants expect funds rated high in sustainability to 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 speak to a general phenomenon whereby attributes that are not related to performance can influence an investor' s feelings about an investment. A positive shift in affect increases investor expectations of future returns and lowers perceived risk, while a negative shift results in the opposite. This is consistent with findings on halo effects, whereby an impression formed in one area influences overall evaluations (Thorndike (1920), Asch (1946), Beckwith and Lehmann (1975), Nisbett and Wilson (1977), Chernev and Blair (2015)). In our setting, an investor who values sustainability for nonpecuniary reasons, such as moral ideals, may have positive affect influence their perception of performance.

An additional question that emerges is how investors in our data set and participants in our experiment interpret the sustainability ratings. For example, although we find that people tend to associate sustainability with the environment, they may consider 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 is being measured. We have not attempted to define sustainability in this paper, but use 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 research.

Initial submission: January 15, 2018; Accepted: January 2, 2019
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

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.

Siddiq Alhakami, Ali, 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.

Asch, Solomon E, 1946, Forming impressions of personality, *The Journal of Abnormal and Social Psychology* 41, 258–290.

Barber, Brad M., Adair Morse, and Ayako Yasuda, 2017, Impact investing, Working Paper, University of California.

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.

Beckwith, Neil E., and Donald R. Lehmann, 1975, The importance of halo effects in multi-attribute attitude models, *Journal of Marketing Research* 12, 265–275.

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.

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, Working Paper, University of Texas.

Birru, Justin, 2018, Day of the week and the cross-section of returns, *Journal of Financial Economics* 130, 182–214.

Bodner, Ronit, and Drazen Prelec, 2003, Self-signaling and diagnostic utility in everyday decision making, *The Psychology of Economic Decisions* 1, 105–26.

Bollen, Nicolas P.B., 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, *Quarterly Journal of Economics* 127, 1243–1285.
Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer, 2013a, Salience and asset prices, *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, Max H. Farrell, and Rocio Titiunik, 2019, Regression discontinuity designs using covariates, *Review of Economics and Statistics* 101, 442–451.
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, *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, *American Economic Review* 94, 1276–1302.
Cheng, Ing-Haw, Harrison Hong, and Kelly Shue, 2013, Do managers do good with other people's money?, NBER Working Paper w19432.
Chernev, Alexander, and Sean Blair, 2015, Doing well by doing good: The benevolent halo of corporate social responsibility, *Journal of Consumer Research* 41, 1412–1425.
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, 2019, Investing for impact, *Review of Financial Studies*, 32, 864–904.
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.
Diane Del, Guercio, and Paula A. Tkac, 2008, Star power: The effect of Morningstar ratings on mutual fund flow, *Journal of Financial and Quantitative Analysis* 43, 907–936.
DellaVigna, Stefano, John A. List, and Ulrike Malmendier, 2012, Testing for altruism and social pressure in charitable giving, *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* (Chicago, IL: Law School, University of Chicago).
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, *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.
Du, Shuili, Chitrabhan B. Bhattacharya, and Sankar Sen, 2007, Reaping relational rewards from corporate social responsibility: The role of competitive positioning, *International Journal of Research in Marketing* 24, 224–241.
Du, Shuili, Chitrabhan B. Bhattacharya, and Sankar Sen, 2010, Maximizing business returns to corporate social responsibility (CSR): The role of CSR communication, *International Journal of Management Reviews* 12, 8–19.

Dyck, I.J., Karl Lins, Lukas Roth, and Hannes Wagner, 2019, Do institutional investors drive corporate social responsibility? International evidence, *Journal of Financial Economics* 131, 693–714.

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, *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, 1–17.

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, Working Paper, University of Pennsylvania.

Gillan, Stuart, and Laura T. Starks, 2003, Corporate governance, corporate ownership, and the role of institutional investors: A global perspective, *Journal of Applied Finance* 13, 4–22.

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.

Gneezy, Ayelet, Uri Gneezy, Gerhard Riener, and Leif D. Nelson, 2012, Pay-what-you-want, identity, and self-signaling in markets, *Proceedings of the National Academy of Sciences* 109, 7236–7240.

Grinblatt, Mark, and Sheridan Titman, 1989, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

Harris, Lawrence, and Eitan Gurel, 1986, Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures, *Journal of Finance* 41, 815–829.

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, *Journal of Law, Finance, and Accounting* 2, 247–274.

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.

Hartzmark, Samuel M., and David H. Solomon, 2013, The dividend month premium, *Journal of Financial Economics* 109, 640–660.

Hartzmark, Samuel M., and David H. Solomon, 2019a, The dividend disconnect, *Journal of Finance* Forthcoming.

Hartzmark, Samuel M., and David H. Solomon, 2019b, Reconsidering returns, Working Paper. University of Chicago.

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.

Haws, Kelly L., Karen Page Winterich, and Rebecca Walker Naylor, 2014, Seeing the world through green-tinted glasses: Green consumption values and responses to environmentally friendly products, *Journal of Consumer Psychology* 24, 336–354.

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 Tyler Shumway, 2003, Good day sunshine: Stock returns and the weather, *Journal of Finance* 58, 1009–1032.
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, NBER working paper w21215.
Huberman, Gur, 2001, Familiarity breeds investment, *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.
Kaul, Aditya, Vikas Mehrotra, and Randall Morck, 2000, Demand curves for stocks do slope down: New evidence from an index weights adjustment, *Journal of Finance* 55, 893–912.
Kitzmueller, Markus, and Jay Shimshack, 2012, Economic perspectives on corporate social responsibility, *Journal of Economic Literature* 50, 51–84.
Christoph Klauer, Karl, and Elsbeth Stern, 1992, How attitudes guide memory-based judgments: A two-process model, *Journal of Experimental Social Psychology* 28, 186–206.
Lee, David S., and David Card, 2008, Regression discontinuity inference with specification error, *Journal of Econometrics* 142, 655–674.
Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch, 2001, Risk as feelings, *Psychological Bulletin* 127, 267–286.
Maignan, Isabelle, and O.C. Ferrell, 2004, Corporate social responsibility and marketing: An integrative framework, *Journal of the Academy of Marketing Science* 32, 3–19.
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, Working Paper, Harvard University.
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.
Menon, Ajay, and Anil Menon, 1997, Enviropreneurial marketing strategy: The emergence of corporate environmentalism as market strategy, *Journal of Marketing* 61, 51–67.
Nisbett, Richard E., and Timothy D. Wilson, 1977, The halo effect: Evidence for unconscious alteration of judgments, *Journal of Personality and Social Psychology* 35, 250–256.
Olson, Mancur, 2009, *The Logic of Collective Action* (Harvard University Press, Cambridge, MA).
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, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2015, Scale and skill in active management, *Journal of Financial Economics* 116, 23–45.
Pástor, Luboš, Robert F. Stambaugh, and Lucian A. Taylor, 2017, Do funds make more when they trade more?, *Journal of Finance* 72, 1483–1528.
Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2008, Socially responsible investments: Institutional aspects, performance, and investor behavior, *Journal of Banking & Finance* 32, 1723–1742.
Renneboog, Luc, Jenke Ter Horst, and Chendi Zhang, 2011, Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds, *Journal of Financial Intermediation* 20, 562–588.
Reuter, Jonathan, and Eric Zitzewitz, 2010, How much does size erode mutual fund performance? A regression discontinuity approach, NBER working paper w16329.
Riedl, Arno, and Paul Smeets, 2017, Why do investors hold socially responsible mutual funds?, *Journal of Finance* 72, 2505–2550.
Wesley Schultz, P., Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius, 2007, The constructive, destructive, and reconstructive power of social norms, *Psychological Science* 18, 429–434.
Sen, Sankar, and Chitra Bhanu Bhattacharya, 2001, Does doing good always lead to doing better? Consumer reactions to corporate social responsibility, *Journal of Marketing Research* 38, 225–243.
Sen, Sankar, Shuili Du, and C.B. Bhattacharya, 2016, Corporate social responsibility: A consumer psychology perspective, *Current Opinion in Psychology* 10, 70–75.

Shleifer, Andrei, 1986, Do demand curves for stocks slope down?, *Journal of Finance* 41, 579–590.
Sialm, Clemens, Laura T. Starks, and Hanjiang Zhang, 2015, Defined contribution pension plans: Sticky or discerning money?, *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, 281–295.
Skowronski, John J., and Donal E. Carlston, 1989, Negativity and extremity biases in impression formation: A review of explanations, *Psychological Bulletin* 105, 131–142.
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.
Sussman, Abigail B., 2017, Valence in context: Asymmetric reactions to realized gains and losses, *Journal of Experimental Psychology: General* 146, 376–394.
Sussman, Abigail B., and Eldar Shafir, 2012, On assets and debt in the psychology of perceived wealth, *Psychological Science* 23, 101–108.
Thistlethwaite, Donald L., and Donald T. Campbell, 1960, Regression-discontinuity analysis: An alternative to the ex post facto experiment, *Journal of Educational Psychology* 51, 309–317.
Thorndike, Edward L., 1920, A constant error in psychological ratings, *Journal of Applied Psychology* 4, 25–29.
Torelli, Carlos J., Alokparna Basu Monga, and Andrew M. Kaikati, 2011, Doing poorly by doing good: Corporate social responsibility and brand concepts, *Journal of Consumer Research* 38, 948–963.
Trudel, Remi, 2019, Sustainable consumer behavior, *Consumer Psychology Review* 2, 85–96.
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.
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.
White, Katherine, Bonnie Simpson, and Jennifer J. Argo, 2014, The motivating role of dissociative out-groups in encouraging positive consumer behaviors, *Journal of Marketing Research* 51, 433–447.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.
Replication code.