Does Aggregated Returns Disclosure Increase Portfolio Risk-Taking?

John Beshears
Harvard University and NBER

James J. Choi
Yale University and NBER

David Laibson
Harvard University and NBER

Brigitte C. Madrian
Harvard University and NBER

February 1, 2014

Abstract: Many previous experiments have found that subjects invest more in risky assets if they see their returns less frequently, see portfolio-level (rather than separate asset-by-asset) returns, or see long-horizon (rather than one-year) historical asset class return distributions. We find that aggregated returns disclosure does not increase equity allocations in an experiment where—in contrast to previous experiments—subjects invest in real mutual funds over the course of one year. Previously documented aggregation effects are highly sensitive to the distribution of the risky asset's returns and how much clock time elapses between the portfolio choice and the viewing of the returns.

This research was made possible by generous grants from the FINRA Investor Education Foundation, the National Institute on Aging (grant P01-AG005842), and the Social Security Administration through grant #10-P-98363-1-05 to the National Bureau of Economic Research as part of the SSA Retirement Research Consortium. We are grateful for the research assistance of Chris Clayton, Josh Hurwitz, Ben Hebert, Nathan Hipsman, Brendan Price, Gwendolyn Reynolds, and Eric Zwick. We have benefited from the comments of Shlomo Benartzi, Peter Bossaerts, Arie Kapteyn, Andy Lo, Jan Potters, Richard Thaler, and seminar audiences at Bentley College, NYU, UCLA, UT Dallas, University of Mannheim, Wharton, Yale, and the Annual Conference in Behavioral Economics. The findings and conclusions expressed are solely those of the authors and do not represent the views of SSA, any agency of the Federal Government, the NBER, FINRA, any component of Harvard University, or other sponsors of this workshop. Comments should be directed to the authors. The FINRA Investor Education Foundation, formerly known as the NASD Investor Education Foundation, supports innovative research and educational projects that give investors the tools and information they need to better understand the markets and the basic principles of saving and investing. For details about grant programs and other new initiatives of the Foundation, visit www.finrafoundation.org.

A remarkable series of experiments has found that subjects are more willing to invest in risky assets with positive expected returns if only aggregated returns are reported to them, rather than the individual component returns. Information aggregation along various dimensions produces this effect: reporting subjects' portfolio return over the last $n \geq 2$ periods once every n periods rather than reporting single-period returns each period (Gneezy and Potters, 1997; Thaler et al., 1997; Gneezy, Kapteyn, and Potters, 2003; Bellemare et al., 2005; Haigh and List, 2006; Sutter, 2007; Langer and Weber, 2008; Fellner and Sutter, 2009; van der Heijden et al., 2012), reporting subjects' portfolio-level returns rather than returns for each individual asset separately (Anagol and Gamble, 2013), or reporting historical long-horizon return distributions of asset classes rather than historical one-year return distributions of asset classes (Benartzi and Thaler, 1999).[1]

These results are consistent with subjects suffering from myopic loss aversion (Benartzi and Thaler, 1995), which is the combination of loss aversion (Kahneman and Tversky, 1979) and mental accounting (Kahneman and Tversky, 1984; Thaler, 1985, 1990, 1999).[2] Aggregation frames encourage subjects to integrate multiple gamble outcomes into a single mental account. If these gambles are not perfectly correlated across time or across assets and the gambles have positive expected values, the resulting integration can lower the probability that the mental account is evaluated as having borne an overall loss.[3] Thus, integration makes the gambles appear more attractive to a loss-averse subject than if each gamble occupied its own separate mental account.

The strength and consistency of the experimental results constitute compelling evidence that myopic loss aversion is a real psychological phenomenon that responds to aggregation manipulations. In this paper, we consider a related but separate question: Would a financial institution increase the portfolio risk-taking of its clients if it started disclosing returns at a more

[1] Guiso (2009) examines another aggregation manipulation that we do not test in this paper. He finds that asking subjects about their labor income risk before offering a hypothetical lottery makes them more likely to accept the lottery.

[2] Loss-averse agents derive utility and disutility directly from gains and losses, and the disutility of a loss is greater than the utility of a gain of equivalent magnitude. Agents engage in mental accounting when they evaluate outcomes within a subset of their wealth portfolio—the "mental account"—in isolation from outcomes outside the mental account.

[3] Aggregating gambles does not decrease the probability of an overall loss for all return distributions. But the loss probability does decrease, for example, when the aggregated gambles are drawn from the same normal distribution. See also Langer and Weber (2005) for examples of gambles for which prospect theory predicts that aggregation would decrease the willingness to invest due to diminishing sensitivity to magnitudes and nonlinear probability weighting.

aggregated level, and decrease risk-taking if it started disclosing returns at a less aggregated level?

Numerous authors have extrapolated from the existing experimental literature to suggest that the answer to this question is “yes.” For example, Thaler et al. (1997) write, “Decisions made by employees covered by such [defined contribution pension] plans may vary considerably depending on how their investment opportunities are described and the manner and frequency with which they receive feedback on their returns.” Gneezy and Potters (1997) observe that, “Manipulating the evaluation period of prospective clients could be a useful marketing strategy for fund managers.” Haigh and List (2005) write that “institutions may have the ability to influence asset prices through changes in their information provisioning policies.”

However, the typical investment environment has a number of features that may diminish the impact of disclosure policy on portfolio risk-taking, and previous experiments abstracted away from these features:

1. Previous experimenters have had their subjects’ undivided attention, along with complete control over information flows during the experiment, whereas aggregation manipulations by a single institution in a typical investment environment may have little power because of interference from or interactions with background information flows.
2. The experiments to date have been conducted over the course of one session, but the psychology of risk-taking over many days, months, or years may differ. For example, being shown returns more frequently may not shorten investors’ evaluation periods in settings where natural evaluation periods such as a year or a quarter already exist. Or the prospect of seeing a negative return in the few minutes may be quite psychologically aversive, but the pain of seeing a negative return next week may be heavily discounted.
3. Previous experiments have used laboratory assets whose return distributions usually differed substantially from the return distributions of typical financial assets. In addition, these assets were not given labels such as “stocks” and “bonds.”[4] When dealing with familiar assets such as stocks, bonds, and mutual funds, investors may have strong prior beliefs about what they should do. For example, they may employ context-specific heuristics such as, “Allocate 100 minus my age to stocks,” which could blunt the effect of aggregation manipulations on portfolio choices.

[4] An exception is Anagol and Gamble (2013).

We conducted an experiment that incorporates these features of typical investment environments that may be relevant for gauging the practical effect of disclosure policy. We recruited 597 subjects from the general U.S. adult population to participate in a year-long study. Each subject allocated \$325 among four real mutual funds that cover the U.S. equity, international equity, U.S. bond, and U.S. money market asset classes. Subjects were free to reallocate their portfolio throughout the year, just as if they were making real investments in these mutual funds. We paid each subject whatever the \$325 would have been worth at the end of the year if the money had been invested according to his or her choices. The large per-subject payment ensured that subjects remained interested in their experimental portfolio through the end of a one-year experiment. Haigh and List (2005) find that professional futures and options traders exhibit myopic loss aversion over individual gambles whose maximum possible gain is \$10 and maximum possible loss is \$4, so any null results that we find are probably not due to our portfolio stakes being too low.

We randomly varied the level of information aggregation along four dimensions.

The first treatment dimension varied how frequently subjects saw their returns by paying half of subjects to view their weekly returns on our study website once a week and paying the other half to view their biannual returns on our website once every six months. We paid subjects to view their returns on our website rather than simply emailing them their returns because emails are easily ignored, which would create less variation in viewing frequency across conditions and reduce our ability to detect a viewing frequency effect on portfolio risk-taking.[5]

The second treatment dimension varied the level of detail subjects saw when they viewed their weekly or biannual returns. Half of subjects saw only their overall portfolio return over the last week or six months. The other half of subjects saw the return over the last week or six months of each individual asset they were holding. Because a screen available to all subjects showed the dollar value of each asset in their portfolio, subjects in the former group could, in theory, calculate their individual asset returns if they remembered the previous value of each asset. But we did not ourselves provide convenient access to these previous values, hindering this

[5] To the extent that real-world return disclosures are more likely to be ignored than our incentivized disclosures, which were viewed 87% of the time in the weekly condition and 74% of the time in the biannual condition, varying return disclosure frequency outside our experiment will generate even less variation in return viewing frequency, resulting in even weaker potential effects on portfolio risk-taking.

calculation. Similarly, subjects in the latter group could calculate their overall portfolio return from their individual asset returns, but we did not perform this calculation for them.

The third treatment dimension varied the historical returns information shown to subjects. We showed some subjects graphs depicting the distribution of real one-year returns for U.S. equities, international equities, U.S. bonds, and U.S. money markets from 1971 to 2007. Others were shown the distributions of real annualized five-year returns for the four asset classes over the same time period. We also gave some subjects no historical returns information at all in order to see whether allocations were affected by seeing any version of the returns graphs.

The fourth treatment dimension varied whether subjects who saw the historical returns graphs could also access information about the historical performance of mixed portfolios. Some subjects could only see historical return distributions of four "pure" portfolios, each invested 100% in one of the four asset classes offered. Other subjects could, via a Web interface, see return distributions of portfolios invested in whatever mix of asset classes they wished. This latter treatment, which is the backward-looking analogue of the previously described overall portfolio return reporting treatment, might make more apparent the diversification benefits of holding multiple asset classes, thus encouraging greater investment in risky assets.

We find that none of the aggregation treatments significantly increased risk-taking, as measured by the fraction of the portfolio invested in equities. We have enough statistical power to reject increases of more than 2 to 4 percentage points, depending on the treatment. Nor are there significant aggregation effects among subjects who turned down a one-time gamble that gave them an equal chance of winning $8 or losing $5, a choice that Rabin (2000) and Barberis, Huang, and Thaler (2006) show is difficult to explain without myopic loss aversion. These results collectively suggest that information aggregation is not a tool that financial institutions can use to robustly affect risk-taking.

Our null effects are unlikely to be due to our treatments having no effect on the information subjects received. There is a significant difference between how often subjects in the weekly versus biannual return viewing treatments viewed returns on our study website, and weekly viewing treatment subjects are significantly more likely than biannual viewing treatment subjects to report in an exit questionnaire that study participation made them see market returns more often. We additionally find that showing either the one-year or five-year historical returns graphs to subjects increased their equity investment by 11 to 12 percentage points, indicating that seeing the graphs changed subjects' beliefs about future asset returns.

Our null effects instead appear to be due to the fact that the aggregation treatment effects documented in previous studies are highly sensitive to the risky asset's return distribution. Our experiment in fact replicates the Benartzi and Thaler (1999) result that showing long-horizon historical U.S. stock returns increases investment in U.S. stocks relative to showing one-year historical U.S. stock returns. But showing long-horizon historical international stock returns *decreases* investment in international stocks, so there is no change in subjects' total equity allocations.

We also find in a series of follow-up laboratory experiments that seeing ongoing returns less frequently increases risk-taking only when the risky asset has a binary return distribution of +250% with 1/3 probability and -100% with 2/3 probability, as is the case in Gneezy and Potters (1997) and the many subsequent studies that adopt their experimental design (Gneezy, Kapteyn, and Potters, 2003; Bellemare et al., 2005; Haigh and List, 2006; Sutter, 2007; Langer and Weber, 2008; Fellner and Sutter, 2009; van der Heijden et al., 2012). Such a return distribution is much more extreme than that of most common financial assets, including the mutual funds in our main experiment. The return viewing frequency effect disappears when we change the risky asset's return distribution to +25% with 1/3 probability and -10% with 2/3 probability while keeping the total potential dollars at risk the same by increasing by a factor of 10 the maximum allowable investment in the risky asset. Return viewing frequency similarly has no effect when we tell subjects that the risky asset's return distribution matches the historical U.S. stock market return distribution.

Even when the risky asset has the binary return distribution of +250% and -100%, we find that the return viewing frequency effect disappears if the amount of clock time that elapses between the portfolio choice and the return viewings lengthens. In all previous studies, subjects saw their returns immediately after making their portfolio choice. We run experimental conditions where subjects in the frequent viewing treatment see their returns once per week for three weeks, while subjects in the infrequent viewing treatment only see the sum of their three returns three weeks in the future. Return viewing frequency has no effect on risk-taking, regardless of the risky asset's return distribution, when return disclosure is delayed like this.

The remainder of the paper proceeds as follows. Section I describes our main experiment's design. Section II presents the main experiment's results. Section III presents robustness checks for the main experiment. Section IV describes our follow-on experiments' design, and Section V presents their results. Section VI concludes.

I. Main experiment design

A. Subject recruitment

We recruited subjects in the summer of 2008 for a one-year investing experiment through the market research firm MarketTools. Figure 1 shows the number of subjects active in the experiment at each calendar date (the gray bars), as well as the level of the S&P 500 normalized by its July 2, 2007 value (the thin line in the top graph) and the VIX index of expected annualized S&P 500 volatility (the thick line in the bottom graph). Even though our experiment spanned the market collapse in the fall of 2008, 94% of our subjects made their *initial* portfolio choices between June 23, 2008 and July 14, 2008. Ninety-nine percent had completed this task by July 30, 2008, and the remaining 1% had completed the task by August 30, 2008. The market's precipitous fall did not commence until after the September 15, 2008 bankruptcy of Lehman Brothers. The VIX averaged only 24.5% from June 23 to July 14, 2008, 24.1% from June 23 to July 30, 2008, and 22.6% from June 23 to August 29, 2008.[6] These averages are not far from the historical large-cap equity annualized monthly return standard deviation from 1926 to 2007 of 20.0% reported by Ibbotson Associates, and are far from the values later in 2008 that would herald the arrival of the financial crisis—the VIX rose to 31.7% on September 15, 2008 and peaked at 80.9% on November 20, 2008. Hence, our subjects made their initial portfolio allocations in a non-crisis environment, many weeks *before* we entered a bear market of historic proportions. We will discuss in Section III.C evidence that the market's 18% decline from its October 2007 peak prior to the beginning of our recruiting period does not explain our null treatment effects.

We requested that our subjects be at least 25 years old and have an annual income of at least $35,000, so that it was more likely that they had some investable assets. All interaction with the subjects occurred through the Internet; we had no direct contact with them.

[6] August 30, 2008 was a Saturday.

The initial invitation text introduced the faculty authors with our university affiliations in order to establish the credibility of the study. It then informed subjects that they would receive a \$20 up-front participation fee for allocating \$325 among four mutual funds. At the end of one year, we would pay them whatever their initial \$325 portfolio was worth at that time, plus an additional amount for periodically checking their portfolio's return on the study website. The text concluded by telling the subjects that we expected the initial portfolio allocation to take thirty minutes to an hour, and that it would take no more than thirty minutes to an hour of additional time over the course of the next year to check their portfolio's return.

People interested in participating in the study clicked a link that took them to an informed consent page that described the task, the compensation scheme, and the expected time commitment again. The informed consent document also told subjects that they would periodically receive e-mails with a link that they could click to see their portfolio's return, and that we would pay them for clicking on these links.

Giving informed consent took subjects to a registration page where they supplied their name and contact information and chose a password. In order to prevent anybody from registering for the study more than once, we blocked any attempts to register multiple times from the same IP address. Upon registration, an e-mail was sent to each subject with a link to click on in order to activate his or her account.[7] The link then took them to a login screen.

We recruited 600 subjects, but three of them did not participate after registering. Therefore, our final sample consists of 597 subjects, whom we randomly assigned to one of eighteen experimental cells. Table 1 shows the distribution of our sample among the experimental cells. All subjects who did not see a historical returns graph had their ongoing returns reported to them at the aggregated portfolio level.[8] The remaining treatment assignments are independent of each other. We will describe each experimental condition in further detail in Sections I.C and I.F. Online appendix figures show representative screenshots from the experimental website.

[7] Using an e-mailed activation link ensured that we had an active e-mail account to which we could send the returns-checking links.

[8] We allocated extra subjects to the treatment cells where no historical return graphs were shown and ongoing returns were reported at the portfolio level because these treatments were most similar to the Gneezy and Potters (1997) experiment. If we did not replicate their viewing frequency effect in the full sample, we could estimate the treatment effect using only these 120 subjects to see if interactions with the historical returns graphs and asset-by-asset return reporting were responsible. Results on this subsample are consistent with those on the full sample, so we do not report them later in the paper.

Table 2 groups the experimental cells in a different way to make clearer the comparisons we will be making in our analysis and the sample that is available for these comparisons. All of our aggregation treatment effects are estimated by comparisons between groups that contain at least 238 subjects each (40% of our sample), and the return viewing frequency effects are estimated by a comparison between groups of at least 298 subjects each (50% of our sample). The effect of seeing any version of the historical returns graph is estimated by comparing the 120 subjects who saw no graph (20% of our sample) to the 477 subjects who did see one of the two graphs (80% of our sample). Although 60% of subjects saw only aggregate portfolio-level returns and 40% of subjects saw each asset's return reported separately, 20% of the former group saw no historical returns graph while none of the latter group saw no historical returns graph. In estimating the portfolio-level returns viewing treatment effect, we will control for the historical returns graph the subject saw, so we will effectively be comparing only the 40% of subjects who saw aggregate portfolio-level returns *and* a historical returns graph to the 40% of subjects who saw individual asset returns (and a historical returns graph).

B. Opening instructions screen

After logging in, subjects received a fuller description of the study instructions. The instructions reiterated the nature of the portfolio allocation task and the compensation scheme, and informed subjects that they could reallocate their portfolio any time during the year by logging into their account on the website. Subjects were also told about the inducement to view their ongoing returns, as well as the content and frequency of the ongoing returns they would be paid to see. In some conditions, subjects were introduced to the historical returns graphing tool.

C. Historical returns graph treatments

For 80% of our subjects, the bottom of the instructions screen described above introduced a graphing tool that was intended to help them understand the historical real return distributions of four asset classes: U.S. equities, international equities, U.S. bonds, and U.S. money markets. The remaining 20% of subjects did not see the graphing tool and did not receive any alternative information on historical returns. The graphs generated by the tool are modeled after those in Benartzi and Thaler (1999). Returns for an asset class during the historical sample period are sorted from lowest to highest and displayed as a bar chart. The lowest return is the leftmost bar, and the highest return is the rightmost bar. The median return is also highlighted and labeled with its value.[9] We used the S&P 500, MSCI EAFE, Barclays Capital Aggregate Bond Index, and 30-day U.S. Treasury bill as our asset class proxies. Because the MSCI EAFE series starts in 1970, we cannot use returns prior to 1970 while maintaining identical sample periods for all asset classes. The most recent year of returns available at the start of the experiment was 2007. In order for each return series to have a unique median, we used the period from 1971 to 2007—which has an odd number of years—for all our asset classes.[10] Subjects who had the graphing tool available to them were required to click through an animation that explained how to interpret and use the graph before they could proceed to the next part of the study. This animation could also be replayed in later screens where the graphing tool was shown.

The graphs varied across treatments along two dimensions. The first dimension was whether one-year return distributions or five-year annualized return distributions were shown. We used overlapping periods for the five-year distributions, so there were 33 bars shown on the five-year graph. The second dimension was whether subjects could see only the historical return distributions of four "pure" portfolios—each of which is invested 100% in a single asset class—or could see the return distribution of any asset class mix they wanted. The graphing tool allowed subjects to compare the return distributions of two different asset classes or portfolios side-by-side. The vertical axis on all graphs was fixed to range between -40% and 70%, so that differential treatment effects across graphs could not be attributed to visual scaling effects.

D. Initial portfolio allocation

Subjects made their asset allocations by specifying portfolio percentages to be invested in each investment option. For subjects who had access to the graphing tool, this choice was made after they saw the initial instructions screen and clicked through the animated explanation of the

[9] A programming error caused the bar immediately to the left of the median return to be highlighted instead for the first six months of the experiment, even though the correct median return number was displayed in the graph's caption. The online appendix figures show the graphs with the shifted highlighting. The discrepancy was not visually apparent except in the one-year U.S. equities graph, where the median return was 10.61% but the highlighted bar corresponded to a 7.38% return.

[10] In addition, the Barclays Capital index starts in 1976. We constructed our own aggregate bond market index returns from 1971 to 1975 by weighting the returns of Ibbotson's long-term corporate bond, intermediate Treasury, and long-term Treasury indexes by the total amount of each type of issue outstanding (as reported by the U.S. Treasury) at the end of the prior year.

graphing tool. For subjects who did not see any historical returns graphs, the input boxes for the initial portfolio allocation were below the experimental instructions on the first screen.

Subjects could choose among four index funds offered by Northern Funds: the U.S. Stock Index Fund, the International Equity Index Fund, the Bond Index Fund, and the Money Market Fund.[11] We provided links to each fund's prospectus. We also informed subjects that the International Equity Index Fund charges a 2% redemption fee on the sale of shares held for less than thirty days.[12] For subjects who were shown the historical returns graphs, the graphing tool remained accessible on the same screen in which the portfolio allocation was entered in order to aid their portfolio decision. Subjects could take as long as they wanted to make their portfolio decision. We did not (and could not) prevent subjects from consulting sources of information available outside of our website.

E. Post-allocation questionnaire

After subjects submitted their initial allocation, they completed a post-allocation questionnaire that elicited information on demographics, self-assessed investment knowledge, self-assessed confidence about their portfolio allocation, and time preference. We also offered subjects a gamble with a 50% chance of winning \$8 and a 50% chance of losing \$5. The outcome of the gamble depended on whether the high temperature at San Francisco Airport on a future date, as reported on the National Weather Service website, was an odd or even integer. We applied the gains and losses from this gamble to the \$20 participation fee. Expected utility maximizers with remotely reasonable risk aversion over large-stakes gambles should always accept such a small-stakes, positive-expected-value gamble (Rabin, 2000; Barberis, Huang, and Thaler, 2006). Therefore, subjects who refuse the gamble are particularly likely to be loss averse and prone to engage in mental accounting. Fehr and Goette (2007) show that in a field experiment on labor supply, only workers who rejected a similar gamble (50% chance of winning 8 Swiss francs and 50% chance of losing 5 Swiss francs) exhibited a negative elasticity of effort per hour with respect to an exogenous increase in the piece wage rate, consistent with

[11] We chose Northern Funds because it was the largest fund family that offered U.S. equity index funds, international equity index funds, bond index funds, and a money market fund; did not charge sales loads; did not impose redemption fees on non-international funds; and did not impose frequent trading restrictions.
[12] We followed a first-in-first-out (FIFO) convention for determining which shares will incur the redemption fee, as real-life mutual funds do.

their daily labor supply being determined by loss-averse preferences that are evaluated each day with a reference point around a target daily income level.

Upon finishing the questionnaire, subjects were taken to a page that showed their current investment allocation and total balance. At this point, subjects could log out. On subsequent logins to the site that were not initiated by clicking an e-mailed link (described in Section I.F), subjects would see this portfolio status page first.

F. Ongoing returns viewing treatments

During the one-year duration of the experiment, half of subjects received e-mails once a week with a link they could click to view their previous week's return. These e-mails were sent on Saturdays, starting at the end of the subject's first full calendar week of participation. If they clicked the link within a week of receiving the e-mail, we added $1 to their final payment. Thus, if they clicked all of the e-mailed links they received during the one-year study, they would earn an additional $52. The other half of subjects received e-mails once every 26 weeks with a link they could click to view their prior six-month return. The dates these biannual e-mails were sent coincided with when these subjects would have otherwise received their 26th and 52nd e-mails if they had been assigned to receive weekly e-mails. If subjects receiving biannual e-mails clicked the link within a week of receiving the e-mail, we added $20 to their final payment.[13] We offered only $20 per viewing for this group because we anticipated that subjects receiving weekly e-mails would not click on every e-mailed link, and we wanted to equalize average return-viewing payments across treatments based on our best guess of treatment compliance.

Within each of the above two treatments, we varied the level of detail subjects saw when they clicked on the e-mailed link. Half of subjects saw a screen that showed the return of each individual asset they held. The other half of subjects saw a screen that showed only the overall return of their portfolio.[14] These return screens were only accessible via the e-mailed link (i.e., they could not be reached by following links within the study website). If a link in a given e-mail had already been clicked, clicking it again later would not lead to the return screen; this was to

[13] If an e-mail was sent on day t, the link reported returns through day t, even if the link was not clicked until day $t + n$.

[14] When ongoing returns were reported asset by asset, if the e-mail was sent on day t, only assets held on day t were included in the returns list. Returns on assets completely liquidated prior to day t were not reported. If a subject previously held no position in an asset but established a position sometime between e-mail send dates, the asset return reported was for the full period between e-mails (one week or six months) and did not adjust for the fact that the asset was held for only part of the time between e-mails.

ensure that subjects receiving biannual e-mails did not see the returns screen more frequently than once every six months.

G. Treatment of interest, dividends, and trades

Dividends and interest were automatically reinvested in the fund that paid them.[15] All subjects were free to reallocate their portfolio at any time during the year by logging into their account and clicking a button on the portfolio status page that took them to a reallocation screen. The reallocation screen showed the graphing tool relevant for the subject's experimental condition,[16] links to prospectuses, the current percentage allocations across the four mutual funds, and a note about the international fund's redemption fee. Four input boxes allowed subjects to specify what their new portfolio allocation should be. Trades were executed at the next close of the U.S. markets and could be cancelled by the subject any time up to then.

H. Exit questionnaire

At the end of the one-year investment period, we administered an exit questionnaire to subjects. We will use in our analysis the questions that elicited beliefs about stock market return autocorrelations and the effect study participation had on subjects' attention to market fluctuations. Of the 597 subjects, 569 (95%) completed the exit questionnaire.

II. Main experiment results

A. Subject characteristics

Table 3 displays demographic and financial summary statistics on our subjects, which were collected in the questionnaire administered immediately after the initial portfolio allocation. Men slightly outnumber women, and the young are slightly overrepresented in our sample—33% of subjects are 35 or younger—although all ages have substantial representation. Our subjects are

[15] We used Yahoo! Finance for our dividend and price data. On July 1, 2008, Yahoo! erroneously reported a money market fund dividend of 28.8 cents per dollar invested, which was deposited into 339 of our subjects' accounts. The mean excess windfall was 4.5% of portfolio value. After the market close on July 31, 2008, we sent an e-mail to the affected subjects informing them of the error and (if applicable) how it had affected the July 5 weekly return reported to them. We let them keep the windfall but reallocated it (at the same time the e-mail was sent) in accordance with the subjects' initially chosen asset allocation. This reallocation raised average equity allocations by 1.0 percentage points among subjects receiving weekly e-mails and 2.2 percentage points among subjects receiving biannual e-mails.

[16] The graphs shown to a given subject remained constant throughout the experimental period. That is, they were not updated in real time to reflect new returns that had been realized since the start of the experiment.

relatively well-educated, with 56% reporting holding a bachelor's degree or higher. The high average level of education is perhaps due to our request for subjects with annual incomes above $35,000; only 5% of subjects report an income less than that threshold, and the median subject reports an income between $50,001 and $75,000. The median subject reports total bank, brokerage, and retirement account assets of about $75,000, and 29% of our sample reports assets in excess of $100,000. Only 20% of our sample reports holding no stocks whatsoever in their personal portfolio.

Since the experimental setup was simple (from the perspective of an individual subject) and the assets were passively managed funds in familiar asset classes, subjects did not necessarily need a long time to make a considered decision. The median subject who received no historical returns information took 14 minutes between login and submission of the initial portfolio allocation, as did the median subject who was only given the historical returns distributions of portfolios invested 100% in a single asset class. The median subject who was able to see the historical returns distribution of any portfolio mix took 13 minutes.[17]

B. Average asset allocations

Subjects initially allocated 65.7% of their portfolio to equities—with 34.8% invested in international equities and 30.9% invested in domestic equities—18.6% to bonds, and 15.8% to money markets. The relatively high allocation to international equities may be due to this asset class's strong performance in the time immediately preceding the experiment. The most recent one-year before-tax return reported in the fund prospectus was 25.76% for the international index fund, versus only 15.56% for the domestic equity index fund. Subjects initially held positive amounts in 3.66 asset classes out of 4, on average.

C. Return viewing frequency

Table 4 shows that our periodic e-mails to subjects were successful at creating significant variation in the frequency with which they visited the study website and viewed their returns. During the one-year investment period, subjects who received weekly e-mails logged into the website 60.7 times on average, versus only 18.2 times for subjects who received biannual

[17] We report medians because of outliers for whom the time between initial login and initial portfolio submission was extremely long. These subjects likely made their allocation over the course of more than one sitting.

e-mails. Under the weekly e-mail treatment, 45.3 of those 60.7 logins occurred because subjects clicked on an e-mailed link to view the screen with their ongoing returns. Thus, compliance with the link-clicking requests was high; 87.2% of weekly links sent were clicked within a week of receipt. In the biannual e-mail treatment, subjects clicked 73.8% of links sent, so they saw the returns screen an average of 1.5 times. Subjects in both treatments logged in about 16 times on average when not prompted by an e-mail.

The extra return viewings by subjects who received weekly e-mails did not merely crowd out or coincide with return viewing that they would have engaged in anyway. In the exit questionnaire, we asked subjects, "Did participating in this study make you see the ups and downs of the market more often than you otherwise would have?" Subjects could respond that participation made them see the fluctuations "more often," "less often," or that it had "no effect." In the weekly e-mail treatment, 79% of subjects reported that participation made them see the ups and downs of the market more often, versus 57% of subjects in the biannual e-mail treatment. Because these responses do not indicate *how much* more often study participation made them see returns, this 22% gap ($p < 0.01$) likely understates the effect being in the weekly treatment had on return viewing relative to being in the biannual treatment. Only 1% of subjects in the weekly e-mail treatment and 2% of subjects in the biannual e-mail treatment reported that the study caused them to see market fluctuations less often.

The fact that a $1 payment was sufficient to induce subjects to view their portfolio returns almost every week indicates that they did not find such viewing to be particularly painful, even though this portfolio return was likely to be quite informative about the returns of their total financial portfolio. Andries and Haddad (2013) and Pagel (2013) present theoretical models where people are inattentive to their portfolios because they trade off the benefit of improved economic choices from better information against the fear of learning negative news. The high return-viewing compliance we obtain from a $1 payment places a low upper bound on how strong this fear can be.

D. Effect of returns aggregation on asset allocations

Our main dependent variable is the fraction of the experimental portfolio that is invested in equities at the *beginning* of the experiment. Any effect of the historical returns graphs would likely be most easily detected in the initial allocation, immediately after all subjects in the graph conditions were required to view the graphs. The previous literature on return feedback frequency also finds that individuals who know they will receive frequent feedback reduce their demand for risky assets starting in the very first period of the experiments, indicating that they *prospectively* anticipate the disutility from disaggregated ongoing return disclosure.[18]

The first column of Table 5 reports coefficients from OLS regressions where the dependent variable is the total fraction of the portfolio invested in equities at the beginning of the experiment and the explanatory variables are treatment dummies. We find that anticipating biannual rather than weekly e-mails had a 0.0 percentage point effect on equity shares. We can reject at the 95% confidence level the hypothesis that the increase is more than 3.6 percentage points, which would be a treatment effect that is an order of magnitude smaller than the 28.7 percentage point increase Thaler et al. (1997) find when subjects are shown yearly ongoing returns rather than monthly ongoing returns.[19] We also find that telling subjects that they would see ongoing returns consolidated at the portfolio level rather than separately by each asset *decreased* equity investment by 1.7 percentage points, with a 95% confidence interval of -5.4% to 2.0%.

Simply viewing any historical returns graph significantly raised the initial equity share by 9.2 to 11.2 percentage points relative to not viewing a historical returns graph, which suggests that many subjects were unaware how attractive stock returns have been historically. But it does not appear to matter whether the distributions of one-year returns or five-year annualized returns are presented. In fact—contrary to the increases in equity allocations ranging from 19 to 41 percentage points found by Benartzi and Thaler (1999) when their subjects were shown simulated 30-year return distributions rather than one-year return distributions—our subjects who saw the five-year returns initially allocated *less* to equities than subjects who saw the one-year graph, although the difference is not statistically significant (95% confidence interval from -5.7% to 1.6%).[20] Nor does it seem to matter whether subjects were able to see the return

[18] See Gneezy and Potters (1997), Gneezy, Kapteyn, and Potters (2003), Bellemare et al. (2005), and Haigh and List (2006).

[19] It is difficult to compare our treatment effect magnitudes with those of Gneezy and Potters (1997), Gneezy, Kapteyn, and Potters (2003), Bellemare et al. (2005), Haigh and List (2006), and van der Heijden et al. (2012), since they only offer subjects assets with binary payoffs.

[20] Although using simulated 30-year returns in the long-horizon condition, as Benartzi and Thaler (1999) did, produces a stark contrast against one-year returns, simulated returns are difficult to explain to ordinary investors and are thus less likely to be employed in a real-world educational intervention. Reasonable five-year distributions can be computed from our 37-year historical sample period without resorting to simulation.

distributions of any mix of asset classes instead of only portfolios invested entirely in a single asset class. Being able to see the mixed asset class distributions is associated with an insignificant 0.5 percentage point increase in equity allocations (95% confidence interval from -3.2% to 4.2%).

The subsequent columns of Table 5 show the treatment effect estimates for each of the four asset classes separately. Less frequent ongoing return viewing and being able to see asset class mixes in the historical returns graph have no significant effect on allocations to any asset class. Portfolio-level reporting has a significant effect only on U.S. stock allocations, but with a negative sign, which is the opposite of what Anagol and Gamble (2013) predict. The most interesting results are for the one-year versus five-year historical graph treatments. We find that seeing the five-year graph instead of the one-year graph significantly ($p < 0.01$) increases allocations to U.S. stocks by 4.2 percentage points. Thus, we qualitatively replicate the Benartzi and Thaler (1999) result, which tested the aggregation of historical U.S. stock return distributions. However, aggregation in the graph also significantly ($p < 0.01$) *decreases* allocations to international stocks by 6.3 percentage points, which is why the overall equity allocation is unchanged.

Figures 2 and 3 give some insight into why aggregation decreases allocations to international stocks. Figure 2 shows each asset class's return from 1971 to 2007, sorted from the worst one-year return to the best one-year return, while Figure 3 shows the analogous figure for five-year annualized returns. The data in these graphs match the data shown to subjects in the graph treatments, although subjects could only see information for one to two asset classes at a time.

In the one-year return graph, international stock returns look quite attractive relative to U.S. stocks when evaluated at three salient points in the distribution: the minimum, the median, and the maximum. The minimum and median international stock returns are respectively quite close to the minimum and median U.S. stock returns, but the maximum international stock return is dramatically higher than the maximum U.S. stock return. In the five-year graph, the minimum international stock return continues to be similar to the minimum U.S. stock return, but the median international stock return is now noticeably below the median U.S. stock return and is exactly equal to the median bond return. In addition, the advantage of the maximum international stock return over the maximum U.S. stock return shrinks considerably. Therefore, the overall appeal of international stocks appears lower to subjects who overweight these three salient points in the distribution.

III. Robustness checks for main experiment

A. Interactions with strength of loss aversion and mental accounting

Aggregating reported returns is thought to increase risk-taking because individuals are loss averse and engage in mental accounting. It is possible, then, that we would find positive aggregation effects on equity shares among a subset of individuals who are most prone to loss aversion and mental accounting, even though there are no positive effects on average over our entire sample. We identify such individuals as the 47% of our sample who rejected the equal chance of winning \$8 or losing \$5, a choice that is difficult to explain in the absence of loss aversion and mental accounting.

Table 6 adds interactions between the treatment dummies and a dummy variable for rejecting the small gamble to the Table 5 total equity regression. We find no significant evidence of positive aggregation effects on equity allocations among the subjects who reject the small gamble. The point estimate of the biannual e-mail treatment effect is 2.5 percentage points higher among gamble rejecters than gamble accepters, but this difference is not significant. The overall biannual e-mail effect among gamble rejecters, –1.6 + 2.5 = 0.9 percentage points, is not significantly different from zero. Reporting portfolio-level returns causes gamble rejecters' equity allocations to change by –2.9 + 2.5 = –0.4 percentage points, but this difference too is not significant. Gamble rejecters who saw the five-year graphs initially allocated a statistically insignificant 6.4 + 6.0 – (12.2 – 1.6) = 1.8 percentage points more to equities than gamble rejecters who saw the one-year graph, and gamble rejecters who were able to see graphed return distributions of asset class mixes allocated an insignificant –1.4 + 3.7 = 2.3 percentage points more to equities than gamble rejecters who could see only single asset class return distributions.

B. Do aggregation effects emerge with a delay?

The subjects in Gneezy and Potters (1997), Gneezy, Kapteyn, and Potters (2003), Bellemare et al. (2005), and Haigh and List (2006) did not need to first experience disaggregated ongoing return disclosure before reducing their portfolio risk. Instead, they reduced their demand for risky assets starting in the very first period of the experiments. It is nevertheless possible that our subjects initially did not realize how disaggregated ongoing return disclosure would affect their utility, but they gradually learned as they became exposed to these returns. This would lead to a relative decrease in the disaggregated groups' portfolio risk as the experiment progressed. Our subjects were not inactive in their experimental accounts; the median number of days on which a subject made a reallocation is 2, and the average is 4.6.

We test for the gradual emergence of a positive aggregation effect on risk-taking by using the equity share halfway into the experimental period as the dependent variable in the first column of Table 7, and equity share at the end of the experiment as the dependent variable in the second column. Equity share at the halfway point is measured eight days after subjects receiving weekly e-mails got their 26th e-mailed returns-checking link, and eight days after subjects receiving biannual e-mails got their first returns-checking link.[21]

The coefficient estimates indicate that throughout the investment period, reporting ongoing returns on an aggregated basis did not significantly increase portfolio risk-taking. The point estimate of the biannual e-mail treatment effect grows slightly from the 0.0 percentage points at the beginning of the investment period (in Table 5) to 2.4 percentage points at the halfway point, and it flips sign to –1.4 percentage points at the end of the experiment. The portfolio-level return reporting treatment effect also attenuates from –1.7 percentage points (in Table 5) to an insignificant 0.1 percentage points at the halfway point and an insignificant –1.2 percentage points at the end of the experiment. Although having seen historical returns graphs continued to raise equity share by about 7 to 9 percentage points through the remainder of the experiment, it still does not matter whether one-year or five-year return distributions were shown. Those seeing five-year graphs held 1.7 percentage points more in equities than those seeing one-year graphs at the halfway mark, and 1.5 percentage points less at the end of the experiment, but the differences are not significant.

C. Are our aggregation effects null because of negative expected equity returns?

Could our null effects be driven by some subjects believing that the expected return of equities is negative due to the market's drop prior to the start of the experiment? The same logic

[21] By measuring allocations at this point, we capture the allocations of subjects receiving both weekly and biannual e-mails right after they have been induced to see their returns on the website via an e-mailed link. It may be particularly convenient to reallocate one's experimental portfolio right after clicking on the e-mailed link. Therefore, biannual subjects have had a chance to adjust their portfolios in response to market movements and the reporting regime via the same convenient channel weekly subjects had had available each week for the prior six months.

that causes some gambles with positive expected returns to appear more attractive under aggregation causes some gambles with negative expected returns to appear less attractive under aggregation.

The fact that our subjects initially allocated an average of 65.7% of their portfolio to equities suggests that they did not in fact believe that the expected return on equities was negative. We further test this story by running regressions of initial equity share on the treatment dummies and their interactions with a dummy for a subject believing that market returns follow a random walk. We classify a subject as believing in random walk returns if he said in the exit questionnaire that neither a 10% increase nor a 10% decrease in the market last month should affect one's forecast of the market's return next month. Table 8 shows these regression results. Contrary to the hypothesis that our treatment effects are attenuated by the market's drop prior to the experiment, we find no significantly positive aggregation treatment effects among those who believe in random walk returns and thus were unlikely to forecast a negative equity premium due to the pre-experiment market decline.

IV. Viewing frequency experiment design

Our main experiment's results can be reconciled with the Benartzi and Thaler (1999) finding that seeing longer-horizon historical return distributions of the U.S. stock market increases investment in U.S. equities, since we replicate their result on U.S. equity investments but also find that it does not generalize to all other return distributions. On the other hand, our main experiment offers no reconciliation with the previous literature's findings that more aggregated disclosure of ongoing returns increases risk-taking. This inconsistency is especially surprising for the return viewing frequency treatment, since many different experimenters have replicated this effect.

Our main experiment is not well-suited to discovering what causes the return viewing frequency effect to disappear, since many aspects of its design differ from the previous literature's experiments. In this section, we describe a follow-up experiment that starts with the classic Gneezy and Potters (1997) design and modifies it step by step to move closer to our main experiment's design.[22] Our objective is to see which modifications suffice to eliminate the return viewing frequency effect.

Our follow-up experiment allows us to explore five dimensions along which our main experiment differs from the Gneezy and Potters (1997) experiment:

1. The risky asset in Gneezy and Potters (1997) has an extreme binary distribution: a 2/3 probability of a -100% net return and a 1/3 probability of a 250% net return (i.e., the investment amount plus 2.5 times the investment amount would be returned to the subject). Such return realizations are astronomically unlikely over a one-year horizon for the mutual funds offered in our main experiment.
2. The risky asset in Gneezy and Potters (1997) is an artificial laboratory asset with which subjects were unlikely to have had prior experience. Subjects are much more likely to have contextual knowledge about the real financial assets offered in our main experiment.
3. The risky asset realizations in Gneezy and Potters (1997) were shown within the course of a single 40-minute experimental session. The returns in our main experiment were revealed over the course of one year.
4. At the beginning of each round of Gneezy and Potters (1997), subjects were given an endowment, and they could invest any portion of this endowment in the risky asset. After the risky asset's outcome was seen, subjects were given a fresh endowment, and earnings from previous rounds were not available to invest in the current round. Therefore, each return-viewing event may have also been perceived as a portfolio liquidation event. In contrast, when subjects viewed returns in our main experiment, they were not forced to liquidate their portfolios. In the realization utility models of Barberis and Xiong (2009, 2012), investors receive prospect theoretic utility from returns only when they *sell* a security, not when they view its returns. Therefore, the return viewing frequency effect of Gneezy and Potters (1997) may really be a liquidation frequency effect.
5. Gneezy and Potters ran their experiment on Tilburg University students, whereas our main experiment was run on a non-student population.

[22] Like other authors in the myopic loss aversion literature, we choose to base these experiments on the Gneezy and Potters (1997) design rather than the Thaler et al. (1997) design because the former is a purer demonstration of myopic loss aversion, while the latter incorporates elements of inference and learning as well.

We recruited experimental subjects in two waves, the first in Spring 2013 and the second in Fall 2013 after we had analyzed the results of the first sample. Subjects were paid a \$5 participation fee plus any experimental earnings. The first wave consisted of students in the Harvard Decision Science Laboratory subject pool, making the sample closely comparable to the Tilburg University students used by Gneezy and Potters. The second wave included both students and non-students in this subject pool, as well as subjects recruited by distributing flyers on the Harvard campus. Only 38% of this second wave were full-time students.

Subjects in the first wave participated in both an in-lab study and a post-lab study. Subjects were randomly assigned to one of four in-lab conditions, and independently randomly assigned to one of four post-lab conditions. Within each condition, each subject was randomly assigned to a high-frequency viewing treatment or a low-frequency viewing treatment. Subjects in the second wave participated only in the in-lab study. The online appendix contains the experimental instructions.

The first in-lab condition was a direct replication of Gneezy and Potters (1997). Subjects made choices for nine rounds. Subjects in the high-frequency treatment received a \$2 endowment each round and could bet any amount from \$0 to \$2 in a gamble that had a 2/3 probability of a -100% net return and a 1/3 probability of a 250% net return. Subjects kept any amount that was not bet. Gamble earnings and amounts not bet from previous rounds could not be bet in the current round. Subjects knew that the outcome of each round's gamble would be revealed to them immediately after each round. Subjects in the low-frequency treatment also received \$2 per round, but they had to make investment decisions for three rounds at a time. For example, in round 1, they chose how much they would invest in rounds 1 through 3. Per-round investment amounts were constrained to be the same within each block of three. Subjects knew that they would learn the outcomes of the three gambles for each block simultaneously instead of one by one.

The second in-lab condition, which we call the "scaled Gneezy-Potters" condition, modified the first in-lab condition by scaling down the percentage returns of the risky asset to a 2/3 chance of a -10% return and a 1/3 chance of a 25% return, which is within the realistic range of one-year returns on a stock market investment. To keep the maximum possible dollars at risk the same, we increased the first period endowment to \$20. In order to avoid creating significant wealth effects, we did not give subjects a fresh \$20 endowment each round. Instead, the gamble earnings and amount not bet from round t constituted the maximum allowable investment amount in round $t + 1$. Subjects kept whatever balance remained at the end of round 9. Because the total balance available to bet changed from round to round, we had subjects specify what percent of their endowment they wished to bet, rather than an absolute dollar amount as in the first condition. In the low-frequency treatment, this percentage was required to be constant within each three-round block.

The third in-lab condition replaced the binary risky asset with an asset whose return distribution matched the historical U.S. stock market distribution. We explicitly labeled this asset the "stock market" in the instructions given to subjects. Subjects were told that each round corresponded to one month, and that we would randomly select a starting month between January 1923 and January 2010, with each month having an equal chance of being selected. If, for example, the starting month was March 1954, then the round 1 stock market return would be the actual March 1954 U.S. stock market return, the round 2 stock market return would be the actual April 1954 U.S. stock market return, etc. The instructions showed subjects a histogram of historical monthly stock returns from 1923 to 2012.[23] There were 18 total rounds, and subjects began with a $20 endowment. In the high-frequency treatment, subjects made an allocation choice at the beginning of each round. The instructions said, "At the end of each round, you will learn how much money you gained or lost and your resulting balance. Your stock market investment will then be completely sold, and you must decide what percent of your total balance you wish to invest in the stock market for the next round." Therefore, we retained the coincidence of return viewing and portfolio liquidation in this condition. At the end of each round, subjects were told, "Your stock market investment has been completely sold." Although this liquidation has no economic meaning, it may be psychologically meaningful.[24] A subject's experimental earnings equaled her balance at the end of round 18. In the low-frequency treatment, subjects made portfolio choices at the beginning of rounds 1, 7, and 13, and saw only their cumulated six-round returns at the end of rounds 6, 12, and 18. The six-round frequency matches the frequency of paid return viewings in our main experiment's biannual viewing treatment. Portfolios were not rebalanced within each six-round block, so the percent invested in

[23] These histograms had the return magnitude on the horizontal access and the percent of months with this return on the vertical axis. These differ from the historical return graphs shown in our main experiment.
[24] Weber and Camerer (1998) find that when shares are automatically sold at the end of each period in an experimental market, subjects' disposition effect is greatly reduced even though they can buy their previous position back costlessly, rendering the liquidation economically meaningless.

stock would move with the risky asset's return, but the instructions told subjects that their portfolio would be liquidated every six rounds.

The fourth in-lab condition was identical to the third, except that we did not force liquidation each time a return was viewed. The instructions said that after seeing returns, "you will have the option of holding onto your stock market investment from the last round or changing the percent of your balance invested in the stock market." The computer screen eliciting the subject's investment choice showed two radio buttons labeled "Keep current stock holdings" and "Change stock holdings." If the subject chose the second button, she was asked to specify what percent she wished to invest in the stock market.

The four post-lab conditions mirrored the in-lab conditions, except that the amount of real time that elapsed between rounds was one week. In the Gneezy-Potters replication condition, there were three post-lab rounds. At the beginning of the first round, which occurred at the end of the laboratory session, subjects in the high-frequency treatment chose an amount between \$0 and \$2 to bet in an asset that had a 2/3 chance of a -100% return and a 1/3 chance of a 250% return. The investment principal was taken from their experimental earnings in the in-lab session excluding their \$5 participation fee.[25] At the end of each subsequent week, subjects received an email with the outcome of the gamble and a link to an online survey. The survey asked the subject to enter the result of the gamble (to confirm that she had seen it) and how much she wished to bet in the next round. If the subject filled out the survey within five days of the email being sent, we added \$1 to her final payment. The next round's bet amount equaled the previous round's bet amount for subjects who did not fill out the survey within five days. The third and final survey did not ask for a bet amount. After the last survey, we mailed subjects a payment equal to their final balance plus any survey completion payments they earned.

Subjects in the low-frequency treatment in the Gneezy-Potters replication condition chose a per-round gamble amount that was constrained to be identical across post-laboratory rounds. The cumulative outcome was only revealed to them in an email sent three weeks in the future. The email also contained a link to a survey that asked what the result of the gambles were. If subjects filled out the survey within five days of the email being sent, we added \$3 to their final payment.

[25] For subjects who earned less than \$6 in the in-lab session, the bet amount was limited to one-third of their earnings.

The second post-lab condition modified the asset to have a 2/3 chance of a -10% return and a 1/3 chance of a 25% return. Subjects specified what percent of their ending in-lab balance (excluding their $5 participation fee) they wished to bet, and balances rolled over from round to round, as in the second in-lab condition. High-frequency treatment subjects chose their three bet percentages each week, while low-frequency treatment subjects chose one bet percentage that applied to all three rounds.

The third post-lab condition lasted for six weekly rounds, and the risky asset's returns were drawn from one historical U.S. stock market return sequence, as in the third in-lab condition. Subjects in the high-frequency condition were told that their investment would be completely liquidated at the end of each round, and they would have to choose a new allocation each week. These subjects were paid $1 for completing each of six post-lab surveys about their realized return and new portfolio choice. Subjects in the low-frequency condition were only told their six-round return at the end of six weeks, and they were paid $6 for completing one survey at that time.

The fourth post-lab condition was like the third, except it did not force liquidation at the end of each round in the high-frequency treatment. The post-lab surveys allowed subjects to click a radio button to keep their current stock holdings. The low-frequency treatment was identical to the low-frequency treatment in the third post-lab condition.

V. Viewing frequency experiment results

We recruited 320 subjects in Wave 1 and another 320 subjects in Wave 2. This allowed us to assign 80 subjects to each in-lab condition and 80 subjects to each post-lab condition. Based on the means and standard deviations reported in Gneezy and Potters (1997), a sample size of 80 gives us 79.9% power to detect an effect of their reported magnitude at 5% significance in the replication condition.

Panel A of Table 9 reports regression results from the in-lab conditions for Wave 1 subjects. The dependent variable is the percent of the available endowment that is invested in the risky asset in each round. Without controlling for any additional covariates, we do not replicate the Gneezy and Potters (1997) viewing frequency result in our replication condition. The infrequent returns reporting treatment dummy coefficient is 4.8%, which has the predicted sign, but it is insignificant. However, if we additionally control for a gender dummy, the viewing frequency treatment effect rises to 6.3% and becomes marginally significant at the 10% level (not shown in tables). We therefore weakly replicate the Gneezy and Potters finding in this sample.

Scaling the risky asset payoffs down by a factor of ten causes the positive infrequent returns reporting treatment effect to disappear altogether. The treatment dummy coefficient indicates that infrequent reporting actually *reduces* risky asset investment by 7.8 percentage points. This coefficient is not significant, but additionally controlling for a gender dummy yields a point estimate of -6.5% that is marginally significant at the 10% level. Further modifying the risky asset payoffs to match the stock market return distribution does not resurrect the Gneezy and Potters effect, whether portfolio liquidation is forced upon return viewing or not. The treatment coefficients in these last two conditions are insignificant whether or not we control for gender.

The Wave 1 in-lab results raised two questions. First, is our failure to find a robust viewing frequency effect in the replication condition the result of Type II error, or is the original Gneezy and Potters (1997) result a Type I error and the subsequent successful replications in the literature due to publication bias? Second, is the marginally significant negative effect in the scaled Gneezy-Potters condition a true effect? To answer these questions, we re-ran the in-lab conditions on Wave 2 subjects.

Panel B of Table 9 shows that in this second sample, viewing returns less frequently increased risk-taking robustly in the replication condition. The effect's magnitude is 26.9 percentage points, which is much larger than the 16.9 percentage point effect reported by Gneezy and Potters (1997), and highly significant ($p < 0.0001$). We conclude that our weak replication in Wave 1 was due to Type II error (which we expected to occur with 20% probability). In the scaled-down Gneezy-Potters condition, we find no significant effect whether or not we control for gender, and the point estimate has the opposite sign of what we found in Wave 1. Therefore, the marginally significant negative effect we estimated in Wave 1 appears to be a Type I error. Finally, return viewing frequency continues to have no significant effect with or without a gender control in the two stock conditions, and both point estimates have a sign that is the opposite of their Wave 1 counterparts. Therefore, these effects appear to be true nulls.

In Table 10, we report regressions results from the post-lab conditions, which were run only on Wave 1 subjects. These regressions are analogous to the in-lab regressions in Table 9.

We find that in none of the four conditions does infrequent return viewing significantly affect the fraction of the available endowment invested in the risky asset, whether or not we control for gender. The null effect in the post-lab replication condition is not simply due to the fact that subjects in the Wave 1 in-lab replication condition responded weakly to infrequent return viewing. Because assignment to the post-lab conditions was independent of assignment to the in-lab conditions, 75% of subjects in the post-lab replication condition were not in the in-lab replication condition.

Our follow-up experimental results indicate that the effect of return viewing frequency on risk-taking depends crucially on the risky asset having extreme percentage returns and the returns being viewed immediately after the portfolio choice is made. Scaling down the risky asset's percentage returns while keeping the total dollars subjects can put at risk constant causes the viewing frequency effect to disappear, as does creating a one-week or longer delay between when the choice is made and the first portfolio return is shown. These findings suggest that both the distribution of the asset returns and the experimental time horizon are responsible for the null effects in our main experiment. Because the return viewing effect in the in-lab replication condition is stronger in Wave 2, which contained many more non-student subjects than Wave 1, and neither wave showed significant return-viewing effects in the other three in-lab conditions, our main experiment's null results do not appear to be driven by the fact that its subjects are not students.

VI. Conclusion

Many financial behaviors are difficult to explain unless loss aversion and/or mental accounting are important determinants of economic choices. Such behaviors include aversion to small-stakes risks with positive expected values (Rabin, 2000; Rabin and Thaler, 2001), the tendency to sell stocks with paper gains and hold stocks with paper losses (Shefrin and Statman, 1985; Odean, 1998), and the failure to consider the asset allocation of non-salient accounts when making allocation decisions in a salient account (Choi, Laibson, and Madrian, 2009). Myopic loss aversion has also been proposed as a resolution of the equity premium puzzle (Benartzi and Thaler, 1995; Barberis, Huang, and Santos, 2001; Barberis, Huang, and Thaler, 2006).

The importance of loss aversion and mental accounting raises the possibility that the boundaries of investors' mental accounts could be manipulated by changes in information disclosure in order to increase risky asset demand. The experimental evidence to date has found that indeed, reporting only aggregated outcomes of multiple gambles increases subjects' willingness to take risks. However, these previous experiments abstracted away from certain features of the real-world investment environment that may moderate such effects outside the laboratory. In particular, previous experiments have all taken place within a single session and have used laboratory assets. In order to gauge the potential impact of disclosure policy, our main experiment had subjects invest in real financial assets over the course of an entire year.

We find that disclosing returns at a more aggregated level does not increase portfolio risk-taking. Further analysis and experiments indicate that the effect of aggregated return disclosure is highly sensitive to the distribution of the risky asset's returns and the amount of clock time that elapses between the portfolio choice and the return disclosure. Thus, aggregated return disclosure does not appear to be a policy lever that can be used by financial institutions to robustly affect average portfolio risk choices.

References

Anagol, Santosh, and Keith Jacks Gamble, 2013. "Does presenting investment results asset by asset lower risk taking?" *Journal of Behavioral Finance* 14, pp. 276-300.

Andries, Marianne, and Valentin Haddad, 2013. "Information aversion." Princeton University working paper.

Barberis, Nicholas, Ming Huang, and Tano Santos, 2001. "Prospect theory and asset prices." *Quarterly Journal of Economics* 116, pp. 1-53.

Barberis, Nicholas, Ming Huang, and Richard Thaler, 2006. "Individual preferences, monetary gambles, and stock market participation: A case for narrow framing." *American Economic Review* 96, pp. 1069-1090.

Barberis, Nicholas, and Wei Xiong, 2009. "What drives the disposition effect? An analysis of a long-standing preference-based explanation." *Journal of Finance* 64, pp. 751-784.

Barberis, Nicholas, and Wei Xiong, 2012. "Realization utility." *Journal of Financial Economics* 104, pp. 251-271.

Bellemare, Charles, Michaela Krause, Sabine Kröger, and Chendi Zhang, 2005. "Myopic loss aversion: Information flexibility vs. investment flexibility." *Economics Letters* 87, pp. 319-324.

Benartzi, Shlomo, and Richard H. Thaler, 1995. "Myopic loss aversion and the equity premium puzzle." *Quarterly Journal of Economics* 110, pp. 73-92.

Benartzi, Shlomo, and Richard H. Thaler, 1999. “Risk aversion or myopia? Choices in repeated gambles and retirement investments.” *Management Science* 45, pp. 364-381.

Choi, James J., David Laibson, and Brigitte C. Madrian, 2009. “Mental accounting in portfolio choice: Evidence from a flypaper effect.”*American Economic Review* 99, pp. 2085-2095.

Choi, James J., David Laibson, and Brigitte C. Madrian, 2010. “Why does the law of one price fail? An experiment on index mutual funds.”*Review of Financial Studies* 23, pp. 1405-1432.

Fehr, Ernst, and Lorenz Goette, 2007. “Do workers work more if wages are high? Evidence from a randomized field experiment.” *American Economic Review* 97, pp. 298-317.

Fellner, Gerlinde, and Matthias Sutter, 2009. “Causes, consequences, and cures of myopic loss aversion — an experimental investigation.” *Economic Journal* 119, pp. 900-916.

Gneezy, Uri, and Jan Potters, 1997. “An experiment on risk taking and evaluation periods.” *Quarterly Journal of Economics* 112, pp. 631-645.

Gneezy, Uri, Arie Kapteyn, and Jan Potters, 2003. “Evaluation periods and asset prices in a market experiment.” *Journal of Finance* 58, pp. 821-837.

Guiso, Luigi, 2009. “A test of narrow framing and its origin.” EUI Working Paper ECO 2009/02.

Haigh, Michael S., and John A. List, 2005. “Do professional traders exhibit myopic loss aversion? An experimental analysis.” *Journal of Finance* 60, pp. 523-534.

Kahneman, Daniel, and Amos Tversky, 1984. “Choice, values, and frames.” *American Psychologist* 39, pp. 341-350.

Langer, Thomas, and Martin Weber, 2005. “Myopic prospect theory vs. myopic loss aversion: how general is the phenomenon?” *Journal of Economic Behavior & Organization* 56, pp. 25-38.

Langer, Thomas, and Martin Weber, 2008. “Does commitment or feedback influence myopic loss aversion? An experimental analysis.” *Journal of Economic Behavior & Organization* 67, pp. 810-819.

Lusardi, Annamaria, and Olivia S. Mitchell, 2009. “How ordinary consumers make complex economic decisions: Financial literacy and retirement readiness.” Dartmouth College working paper.

Odean, Terrance, 1998. “Are investors reluctant to realize their losses?” *Journal of Finance* 53, pp. 1775-1798.

Pagel, Michaela, 2013. "A news-utility theory for inattention and rebalancing in portfolio choice." University of California working paper.

Rabin, Matthew, 2000. "Risk aversion and expected-utility theory: A calibration theorem." *Econometrica* 68, pp. 1281-1292.

Rabin, Matthew, and Richard H. Thaler, 2001. "Anomalies: Risk aversion." *Journal of Economic Perspectives* 15, pp. 219-232.

Shefrin, Hersh, and Meir Statman, 1985. "The disposition to sell winners too early and ride losers too long." *Journal of Finance* 40, pp. 777-790.

Sutter, Matthias, 2007. "Are teams prone to myopic loss aversion? An experimental study on individual versus team investment behavior." *Economics Letters* 97, pp. 128-132.

Thaler, Richard H., 1985. "Mental Accounting and Consumer Choice." *Marketing Science*, 4(3): 199-214.

Thaler, Richard H., 1990. "Saving, Fungibility and Mental Accounts." *Journal of Economic Perspectives*, 4(1): 193-205.

Thaler, Richard H., 1999. "Mental Accounting Matters." *Journal of Behavioral Decision Making*, 12(3): 183-206.

Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz, 1997. "The effect of myopia and loss aversion on risk taking: an experimental test." *Quarterly Journal of Economics* 112, pp. 647-661.

van der Heijden, Eline, Tobias J. Klein, Wieland Müller, and Jan Potters, 2012. "Framing effects and impatience: Evidence from a large scale experiment." *Journal of Economic Behavior & Organization* 84, pp. 701-711.

Weber, Martin, and Colin F. Camerer, 1998. "The disposition effect in securities trading: an experimental analysis." *Journal of Economic Behavior & Organization* 33, pp. 167-184.

Table 1. Sample Size in Each Experimental Cell

This table reports the number of subjects that were assigned to each experimental cell. Panel A contains cells where ongoing returns were reported only at the aggregated portfolio level. Panel B contains cells where ongoing returns were reported separately by each asset held by the subject.

Panel A: Ongoing returns reported at portfolio level		
	Return viewing inducement frequency	
Historical return graph shown	Weekly	Biannual
None	60	60
1-year returns, single asset classes	30	30
5-year returns, single asset classes	29	30
1-year returns, portfolio mixes allowed	30	30
5-year returns, portfolio mixes allowed	30	30
Panel B: Ongoing returns reported separately by asset		
	Return viewing inducement frequency	
Historical return graph shown	Weekly	Biannual
1-year returns, single asset classes	30	29
5-year returns, single asset classes	30	30
1-year returns, portfolio mixes allowed	30	30
5-year returns, portfolio mixes allowed	30	29

Table 2. Sample Divisions

	Proportion of sample
How often are subjects induced to get feedback about their own ongoing returns?	
Weekly	50%
Biannually	50%
How granular is subjects' feedback about their own ongoing returns?	
Each asset's return reported separately	40%
Only aggregated portfolio-level returns reported	60%
Are historical returns graphs available?	
No historical returns graphs	20%
Historical returns graphs available	80%
How are historical returns data aggregated across time in the graphs?	
1-year return distributions shown	40%
5-year annualized return distributions shown	40%
How are historical returns data aggregated across asset classes in the graphs?	
Return distributions of single asset classes only are shown	40%
Return distributions of mixes of asset classes are shown	40%

Table 3. Subject Characteristics

Percent male	56%	Financial assets in bank, brokerage, and retirement accounts	
Age		< \$25,000	27%
≤ 25	2%	\$25,001 - \$50,000	13%
26-35	31%	\$50,001 - \$75,000	10%
36-45	22%	\$75,001 - \$100,000	9%
46-55	19%	> \$100,000	29%
55-65	13%	Prefer not to answer	12%
≥ 66	13%		
Education		Percent of outside financial assets invested in stocks at beginning of experiment	
Some high school	1%		
High school graduate	10%		
Some college	23%	0%	20%
Associate's degree	10%	1 - 25%	32%
Bachelor's degree	28%	26 - 50%	17%
Some graduate school	7%	51 - 75%	15%
Graduate degree	21%	76 - 100%	8%
		Prefer not to answer	9%
Annual household income			
< \$35,000	5%		
\$35,000 - \$50,000	21%		
\$50,001 - \$75,000	29%		
\$75,001 - \$100,000	19%		
> \$100,000	21%		
Prefer not to answer	5%		

Table 4. Website Visits After Initial Allocation

This table shows, by return viewing inducement frequency, the average number of total visits to the study website per subject, the average total viewings of the returns screens per subject, and the average fraction of the available returns screens that were viewed by each subject. “Total visits to the website” include visits that involved viewing a returns screen. Standard errors are in parentheses.

	Return viewing inducement frequency		*p*-value of difference
	Weekly	Biannual	
Total visits to website	60.7 (2.6)	18.2 (1.9)	0.000
Viewings of returns screens from e-mail links	45.3 (0.7)	1.5 (0.0)	0.000
Fraction of possible e-mail link returns screens viewed	87.2% (1.9)	73.8% (2.5)	0.000

Table 5. Aggregation Effects on Initial Allocations

The dependent variable is the percent of the portfolio allocated to equities in total or to each asset class at the start of the experiment. *Biannual e-mail* is a dummy for whether the subject was sent an e-mail with a link to his ongoing returns biannually. *Portfolio-level return reporting* is a dummy for whether the subject's ongoing returns were reported only at the consolidated portfolio level. *1-year graph* is a dummy for whether the subject was shown graphs with one-year historical returns. *5-year graph* is a dummy for whether the subject was shown graphs with five-year historical returns. *Asset class mixes shown* is a dummy for whether the subject saw a historical returns graphing tool that could show distributions of arbitrary asset class mixes. Point estimates from an OLS regression are shown, with standard errors in parentheses.

	Total equity	U.S. stock	Intl stock	Money market	Bonds
Biannual e-mail	-0.0	-0.5	0.5	-0.5	0.5
	(1.8)	(1.5)	(1.6)	(1.1)	(1.1)
Portfolio-level return reporting	-1.7	-4.1*	2.4	1.6	0.1
	(1.9)	(1.6)	(1.8)	(1.2)	(1.2)
1-year graph	11.2**	1.8	9.4**	-6.9**	-4.4**
	(2.6)	(2.3)	(2.5)	(1.7)	(1.7)
5-year graph	9.2**	6.0**	3.1	-6.0**	-3.2
	(2.6)	(2.3)	(2.6)	(1.7)	(1.7)
Asset class mixes shown in graph	0.5	2.4	-1.9	-0.3	-0.2
	(1.9)	(1.6)	(1.8)	(1.2)	(1.2)
Constant	58.3**	29.6**	28.8**	20.3**	21.3**
	(2.8)	(2.4)	(2.6)	(1.8)	(1.8)
Sample size	597	597	597	597	597

* Significant at 5% level. ** Significant at 1% level.

Table 6. Aggregation Effects on Initial Equity Allocation Interacted with Loss Aversion

The dependent variable is the percent of the portfolio allocated to equities at the start of the experiment. *Biannual e-mail* is a dummy for whether the subject was sent an e-mail with a link to his ongoing returns biannually. *Portfolio-level return reporting* is a dummy for whether the subject's ongoing returns were reported only at the consolidated portfolio level. *1-year graph* is a dummy for whether the subject was shown graphs with one-year historical returns. *5-year graph* is a dummy for whether the subject was shown graphs with five-year historical returns. *Asset class mixes shown* is a dummy for whether the subject saw a historical returns graphing tool that could show distributions of arbitrary asset class mixes. *Loss averse* is a dummy for whether the subject turned down the win $8/lose $5 gamble we offered. Point estimates from an OLS regression are shown, with standard errors in parentheses.

Biannual e-mail	-1.6
	(2.3)
Biannual viewing ×	2.5
Loss averse	(3.4)
Portfolio-level return reporting	-2.9
	(2.6)
Portfolio-level return	2.5
reporting × Loss averse	(3.7)
1-year graph	12.2**
	(3.7)
1-year graph × Loss averse	-1.6
	(5.3)
5-year graph	6.4
	(3.7)
5-year graph × Loss averse	6.0
	(5.3)
Asset class mixes shown in graph	-1.4
	(2.6)
Asset class mixes shown	3.7
× Loss averse	(3.7)
Loss averse	-7.8
	(5.5)
Constant	62.3**
	(3.9)
Sample size	597

* Significant at 5% level. ** Significant at 1% level.

Table 7. Aggregation Effects on Later Equity Allocations

The dependent variable is the percent of the portfolio allocated to equities 27 weeks into experimental participation or at the end of the experiment. *Biannual e-mail* is a dummy for whether the subject was sent an e-mail with a link to his ongoing returns biannually. *Portfolio-level return reporting* is a dummy for whether the subject's ongoing returns were reported only at the consolidated portfolio level. *1-year graph* is a dummy for whether the subject was shown graphs with one-year historical returns. *5-year graph* is a dummy for whether the subject was shown graphs with five-year historical returns. *Asset class mixes shown* is a dummy for whether the subject saw a historical returns graphing tool that could show distributions of arbitrary asset class mixes. Point estimates from an OLS regression are shown, with standard errors in parentheses.

	27 weeks	Final
Biannual e-mail	2.4	-1.4
	(2.1)	(2.2)
Portfolio-level return reporting	0.1	-1.2
	(2.4)	(2.4)
1-year graph	7.4*	9.0**
	(3.4)	(3.4)
5-year graph	9.1**	7.5*
	(3.4)	(3.4)
Asset class mixes shown in graph	1.5	-1.0
	(2.4)	(2.4)
Constant	48.7**	54.6**
	(3.5)	(3.6)
Sample size	597	597

* Significant at 5% level. ** Significant at 1% level.

**Table 8. Aggregation Effects on Initial Equity Allocation
Interacted with Belief in Random Walk Returns**

The dependent variable is the percent of the portfolio allocated to equities at the start of the experiment. *Biannual e-mail* is a dummy for whether the subject was sent an e-mail with a link to his ongoing returns biannually. *Portfolio-level return reporting* is a dummy for whether the subject's ongoing returns were reported only at the consolidated portfolio level. *1-year graph* is a dummy for whether the subject was shown graphs with one-year historical returns. *5-year graph* is a dummy for whether the subject was shown graphs with five-year historical returns. *Asset class mixes shown* is a dummy for whether the subject saw a historical returns graphing tool that could show distributions of arbitrary asset class mixes. *Random walk* is a dummy for whether the subject believes that neither a 10% increase in the market in one month nor a 10% decrease in the market in one month should affect one's forecast of the market's return next month. Point estimates from an OLS regression are shown, with standard errors in parentheses.

Biannual e-mail	1.2 (2.4)
Biannual viewing × *Random walk*	-1.4 (3.4)
Portfolio-level return reporting	-3.9 (2.7)
Portfolio-level return reporting × *Random walk*	4.4 (3.8)
1-year graph	7.4 (3.8)
1-year graph × *Random walk*	7.0 (5.4)
5-year graph	8.5** (3.7)
5-year graph × *Random walk*	1.8 (5.4)
Asset class mixes shown in graph	-1.4 (2.7)
Asset class mixes shown × *Random walk*	4.3 (3.8)
Random walk	-2.6 (5.6)
Constant	59.2** (3.9)
Sample size	570

* Significant at 5% level. ** Significant at 1% level.

Table 9. In-Laboratory Aggregation Experiment Results

This table shows regression results where the dependent variable is the fraction of the endowment invested in the risky asset and the explanatory variable is a **dummy** for the subject having returns reported infrequently. Standard errors are clustered by subject.

	Gneezy-Potters replication	Scaled Gneezy-Potters	Stock condition, forced liquidation	Stock condition, no forced liquidation
Panel A: Wave 1				
Infrequent return viewing	4.8	-7.8	-0.8	4.0
	(7.8)	(8.0)	(7.0)	(6.9)
Constant	48.2**	49.5**	48.7**	43.1**
	(5.7)	(5.8)	(4.9)	(4.9)
N	480	480	840	840
Subjects	80	80	80	80
Panel B: Wave 2				
Infrequent returns viewing	26.9**	0.7	5.4	-5.7
	(6.0)	(6.7)	(7.0)	(5.9)
Constant	32.9**	32.9**	36.7**	34.8**
	(4.5)	(5.0)	(4.6)	(4.5)
N	480	480	840	840
Subjects	80	80	80	80

* Significant at 5% level. ** Significant at 1% level.

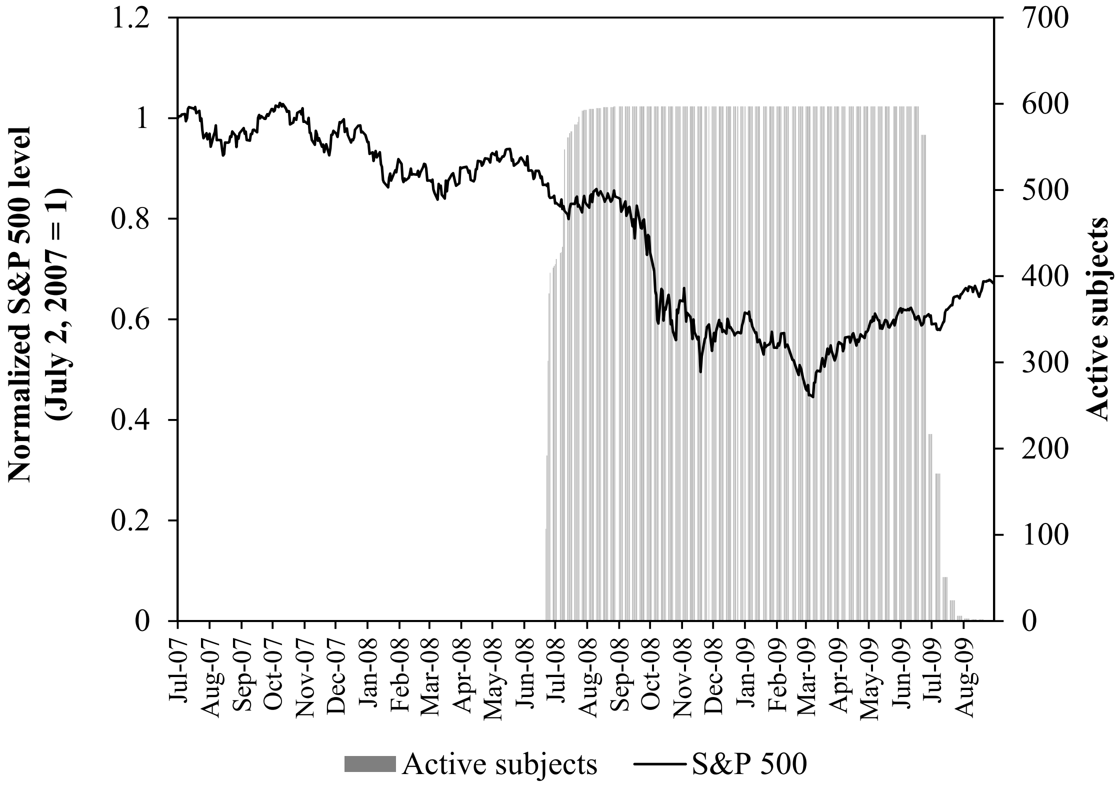
Table 10. Post-Laboratory Aggregation Experiment Results

This table shows regression results where the dependent variable is the fraction of the endowment invested in the risky asset in each round and the explanatory variable is a dummy for the subject having returns reported infrequently. Standard errors are clustered by subject.

	Gneezy-Potters replication	Scaled Gneezy-Potters	Stock condition, forced liquidation	Stock condition, no forced liquidation
Infrequent return viewing	-2.8	6.6	-4.8	-10.0
	(7.7)	(8.9)	(7.2)	(7.2)
Constant	55.7**	40.9**	46.0**	50.4**
	(5.1)	(6.0)	(4.6)	(4.9)
N	160	160	280	280
Subjects	80	80	80	80

* Significant at 5% level. ** Significant at 1% level.

Figure 1. Experimental Period



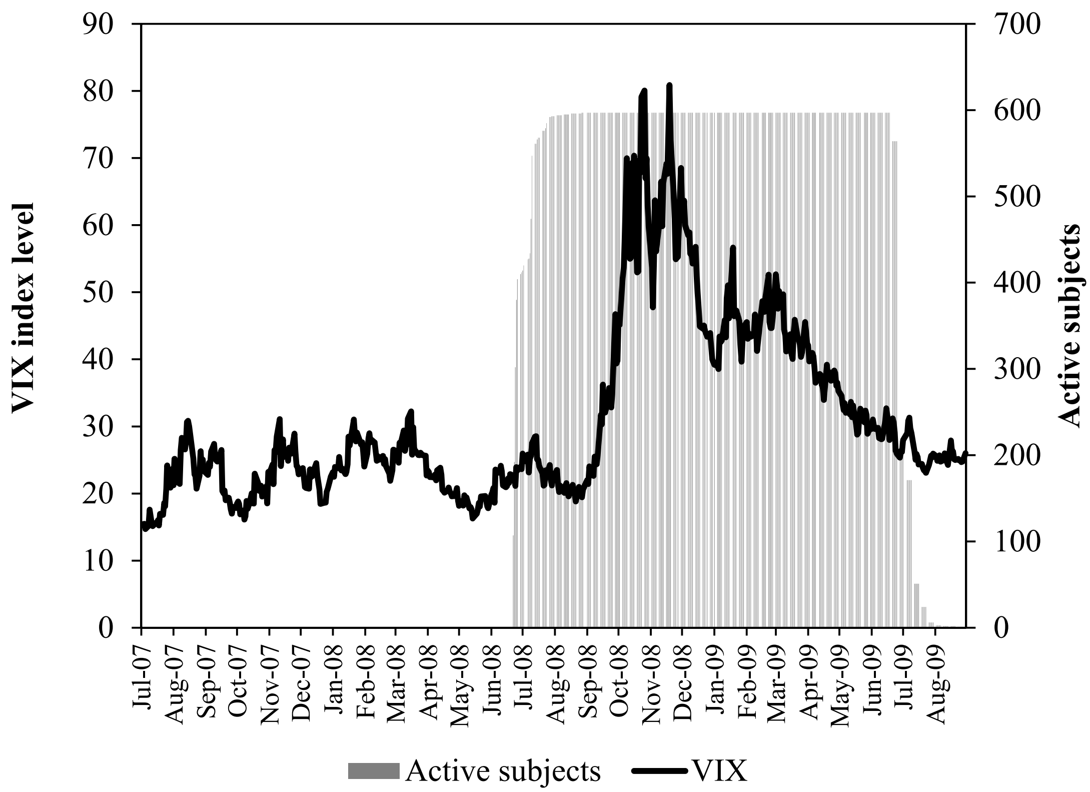


Figure 2. 1-year asset class returns, 1971-2007, sorted from worst to best

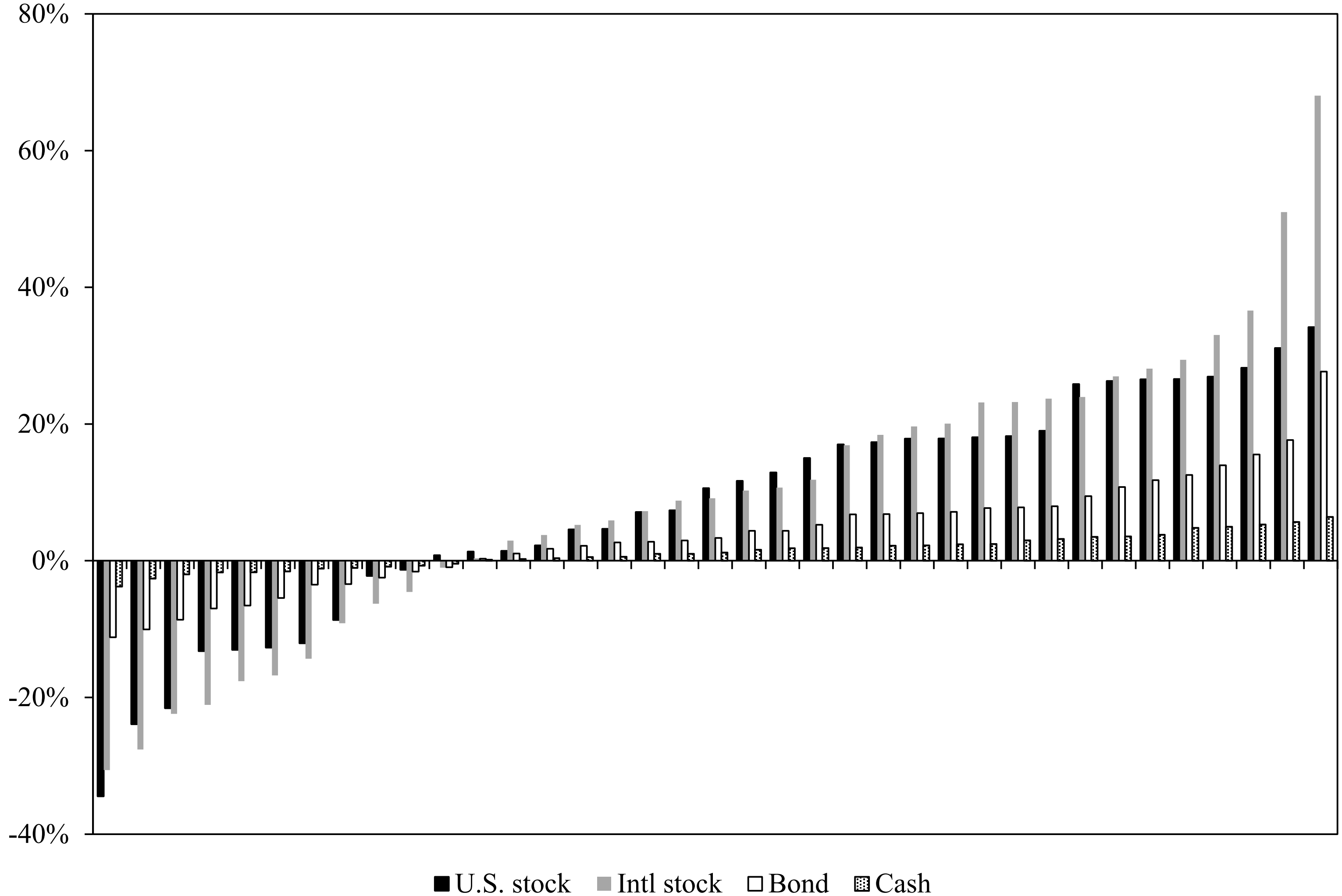


Figure 2. 5-year annualized asset class returns, 1971-2007, sorted from worst to best