Ratings and Asset Allocation: An Experimental Analysis*

Robert L. McDonald† Thomas A. Rietz‡

January 5, 2026

Abstract

Investment ratings provide an ordinal measure for comparing investments. Typically, ratings are within categories, within which assets share common characteristics, but all categories use the same scale (e.g., 1 to 5). Comparing such categorized ratings across categories can be misleading. In an asset allocation experiment, subjects make repeated allocation decisions under complete information. Some subjects see categorized ratings, which conflict with uncategorized ratings. Ratings convey no new information, but they affect subject investment choices and harm performance in the experiment. Knowledge and experience help with the base allocation task but do not mitigate the harmful effect of categorized ratings.

*We thank the TIAA-CREF Institute for funding this research and Ying Xu for programming. For helpful comments we also thank Bruno Biais, Craig Furfine, Simon Gervais, Luigi Guiso, Ravi Jagannathan, Lisa Kramer, Cami Kuhnen, Owen Lamont, Daniel Martin, Olivia Mitchell, Brian Sternthal, Tom Vinaimont, seminar participants at the New York Fed, Northwestern University, University of Iowa, University of Illinois, City University of Hong Kong, Hong Kong University, Hong Kong Polytechnic University, Korea University Business School, and conference participants at the Economic Science Association meetings, UBC, and the Wharton Conference on Household Finance.

†Erwin P. Nemmers Professor of Finance, Kellogg School of Management, Northwestern University; TIAA-CREF Institute Fellow; and NBER. Contact: 847-491-8344, mailto:r-mcdonald@northwestern.edur-mcdonald@northwestern.edu

‡Soumyo Sarkar Professor of Finance, Tippie College of Business, The University of Iowa; and TIAA-CREF Institute Fellow. Contact: 319-335-0856, mailto:thomas-rietz@uiowa.eduthomas-rietz@uiowa.edu

Conflict of Interest Disclosures

Robert McDonald

Robert McDonald is a member of the FINRA National Adjudicatory Council and a director of Eris Exchange. He receives compensation for both. Neither is a conflict of interest for this research.

Thomas Rietz

As part of TIAA's funding for this research, which included subject payments and development costs, Thomas Rietz received partial summer support for one summer. This is not a conflict of interest for this research.

1 Introduction

Personal financial decisions are complicated and have ramifications that span decades. Investors face a variety of long-term risks and are confronted with numerous assets and investment strategies. Fully optimal portfolio decisions that take into account realistic extensions to the basic model (factoring in uncertain labor income, for example) are complicated in all but the simplest cases (e.g., see Merton, 1971, 1973).

Many advisory firms provide asset and fund ratings to assist investors in making investment decisions.[1] These ratings are often assigned within investment categories, with the implication that ratings are comparable within a category but not across categories. In 2023, for example, Morningstar assigned funds to 129 distinct categories, with funds in each category rated using one to five stars (worst to best), and with stars assigned using a mandatory curve based on risk-adjusted historical returns.[2] Because of the mandatory curve, the star rating is not absolute: a fund's star rating can change when other funds enter or leave the category. More importantly, the star rating is not investment advice, as an investor's optimal allocation decisions should depend on a fund's contribution to the return and risk for the investor's entire portfolio. In particular, comparing funds across investment categories makes little sense: a 5-star fund in "Equity Precious Metals" cannot be directly compared to a 3-star fund in "Corporate Bonds." Categorized ratings are thus double-edged: they compare performance to other investments of a similar style, but the use of a common scale across categories may distort choices by making non-comparable items appear comparable.

We report on an experiment designed to assess the effects of such categorized ratings. In a setting with complete information about asset characteristics, subjects allocate money across six investment alternatives, in multiple trials. Each of the investments has a binomial payoff that is perfectly correlated across the assets. Subjects have complete information about the assets at all times: the presentation in each trial includes the high and low returns for each investment, along

[1] Advisory firms supplying ratings include CFRA, Lipper, Morningstar, TheStreet, and Zacks. US News averages these five ratings to supply its own rating.

[2] See Morningstar (2023). Ten percent of funds in each category receive 5 stars and 1 star each, 22.5% receive 4 and 2 stars each, and 35% receive three stars. The original Morningstar ranking system was introduced in 1985, with all stock funds ranked in a common pool (Blume, 1998). In 1996, category rankings were introduced. In 2002, smaller category groups were introduced (Morningstar, 2008). Morningstar justified categorization as eliminating a "tail-wind" effect, in which specific industries or investment styles would by chance generate high returns and receive investor attention. By eliminating one problem, however, categorization could create a different problem, creating the appearance of comparability where there is none.

with the mean return, range, and the return/range ratio. In comparing assets, therefore, subjects do not have to perform computations, think about diversification, or remember asset characteristics from one trial to the next. This design eliminates many computational and cognitive hurdles for subjects, and was motivated by the existing literature showing that individuals have limitations and behavioral biases in perception, attention, and cognitive ability. Payoffs are revealed at the conclusion of the experiment, so there is no learning about returns, statistical inference, or income effect. Risk-averse investors should have an unambiguous preference for two of the assets; of the remaining four, two would only be optimal for risk seeking subjects and two have payoffs that are strictly dominated for any risk preference. Diversification across the six assets is suboptimal.

The main treatment is categorization. In the first trial, the assets have no ratings. Subjects are then shown star ratings, with half of subjects shown uncategorized ratings and the other half shown categorized ratings. Categorization alters the rating of four of the six assets, and reduces the rating of one of the assets that should be preferred by risk-averse subjects. At the end of the experiment, subjects take a financial knowledge quiz. We also collect demographic information. This allows us to asses the effects of demographics, knowledge and experience on choices and susceptibility to ratings and categorization effect.

We find several interesting results. In contrast to much published research on choice under uncertainty (e.g., see Kahneman et al. (1982)), many of the choices we observe seem quite rational. In particular, we find no evidence supporting naive diversification (i.e., the "'$1/n$"' rule of Benartzi and Thaler (1999)). Subjects seem to understand that diversification is not valuable in context. They focus their investment in undominated assets and avoid dominated ones. They invest the most in the highest expected value asset, then one that should be preferred by sufficiently risk averse subjects, followed by one that should be preferred by risk seeking subjects. More knowledgeable and experienced subjects focus their investment even more and invest most heavily in the highest expected value asset. While an initial task to determine risk preferences indicate that women are more risk seeking in context, we find limited evidence supporting gender effects after controlling for this.

Nevertheless, we observe a ratings effect and this effect seems robust to knowledge and experience. We find that subjects exposed to categorized ratings hold less of the lower-rated optimal asset, and as a result perform more poorly. While knowledgeable subjects perform better overall, they appear equally affected by categorization.

Subjects could be influenced by ratings for different reasons. One possibility is the experimenter-demand effect (Zizzo, 2010), in which subjects respond to a cue such as star ratings because they believe they are supposed to do so. An alternative is that subjects are cognitively challenged, have difficulty with the experimental task, and use the ratings to assist them in decision-making. A third possibility is that subjects face cognitive dissonance when a rating is inconsistent with their prior belief about the asset and they respond to this dissonance by investing less, thereby harmonizing the rating and their behavior.

While all three explanations seem plausible, subject behavior does not align with the first two. Subjects perform reasonably well in the allocation task without ratings, suggesting cognitive ability with the task. When shown ratings, subjects do not change investment behavior on average except in the case where ratings are altered by categorization. In this case, ratings affect choices when ratings conflict with subjects' own evaluation of assets: subjects reduce their investment in an asset they apparently prefer when it receives an intermediate rating, but they do not increase their investment in a preferred asset that receives a high rating. This seems to rule out the experimenter demand effect, which in its strongest form would predict that subjects going from no stars to stars would invest more heavily in the highest rated assets and less in the lower-rated. The likeliest explanation seems to be dissonance stemming from the inconsistency of categorized ratings with the belief about asset quality.

While we defer a detailed discussion of the related literature until after we have presented our results, we note that the experiment touches on four distinct but related questions that have been discussed in the literature:

- How does the presentation of summary information and investment alternatives affect investment choices? This has been studied in the context of Morningstar ratings (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2021), bond ratings (Chen et al., 2014), and in the study of menu effects (Bateman et al., 2016; Benartzi and Thaler, 2001; Huberman and Jiang, 2006; Massa et al., 2015).[3]

- How do financial knowledge and experience affect investment behavior? Bernheim et al.

[3] In the framework of Campbell (2006), ratings may be viewed as an example of "equilibrium household finance."

(2001), Bernheim and Garrett (2003), Lusardi and Mitchell (2007), and Anderson and Settle (1996) all demonstrate effects of financial literacy.

- How do cognitive limitations affect decision making (Edgell et al., 1996; Thaler, 1980; Tversky and Kahneman, 1974; Camerer et al., 1989)?
- Can subjects make appropriate portfolio choice decisions (Moore et al., 1999; Kroll et al., 1988, 2003)?

In section 2 we discuss a non-exhaustive list of several financial and non-financial ratings systems in practice to illustrate that out point goes well beyond Morningstar mutual fund ratings. In Section 3, we describe the experiment. The actual experimental instructions are in Appendix B. Section 4 presents and discusses our results and treatment effects. We first examine subject behavior in the first, untreated trial. We then look at how behavior changes in response to treatments. We discuss related literature in Section 5 and throughout the paper as appropriate. Section 6 concludes.

2 Rankings in Practice

In this section, we describe several real-world examples of ordinal ranking systems. In each case there is obvious potential benefit of the system, as well as the potential for the system to cause confusion by using a similar ranking system for non-comparable items.

2.1 Morningstar Fund Rankings

Morningstar introduced their Morningstar Star Rating system for mutual funds in 1985. Investors respond to changes in rankings with investment flows (Del Guercio and Tkac (2008)).[4] As of April 2023, there were 129 U.S. Morningstar fund categories based on fund investment goals (see Morningstar (2023), accessed 4/9/2024).

[4]Fund flows respond to ranking changes even after controlling for historical return characteristics. This finding raises the possibility that investors are overly sensitive to the ratings, which is similar to our premise. However, Del Guercio and Tkac (2008) also report that investing only in five-star funds produces a positive risk-adjusted excess return in their sample, making it difficult to rule out the possibility that Morningstar rankings do convey additional information. Del Guercio and Tkac (2008) have no investor-level data and thus are unable to examine heterogeneity of investor responses as we do. Nor do they study the inter-category ranking issues raised by our paper.

For at least three reasons, Morningstar rankings cannot be compared across categories as a guide to making investment decisions:[5]

1. Star rankings take no account of investor heterogeneity. For example, 16 of the Morningstar categories pertain to tax-exempt bonds from different states and with different maturities. Value across bond categories vary dramatically across investors in ways not captured by the rankings. Taxable bonds generally dominate tax-exempt bonds for low-tax-bracket investors and vice versa for high-tax-bracket investors. Also, investors pay state tax on bonds issued in other states, so investors in high-tax states will prefer bonds issued by their own state. The intuition applies across other category differences as well. Whether an investor should invest in a commodity fund, an emerging markets fund, or a sector equity fund depends upon the specific risk exposures, risk tolerance, and horizon of that investor.

2. Some categories are mutually exclusive. An investor would presumably not simultaneously hold both an S&P 500 index fund, ranked in the "Large Blend" category, and a leveraged short index ETF, ranked in the "Trading-Inverse Equity" category. Along similar lines, some categories deal with target-date funds, designed for investors retiring at a particular horizon. Presumably investors would pick one target date, not many.

3. Because of the mandatory curve within categories, rankings never reveal dominated or near-dominated categories. Many Morningstar categories have highly-correlated returns. A category could be close to redundant, have low returns due to high management fees, and still contain five star funds. Some groups of categories span others, but the rankings take no account of this. It is possible that a 3-star fund in one category (for example) could be superior to a 5-star fund in another category where closely-related funds have much higher fees.

Our experimental setup resembles Morningstar rankings in the sense that a given subject or investor should never consider investing in all available alternatives.

2.2 Other Financial Ratings Examples

Here are just a few financial examples of rating systems that expliciy or implicitly use categorization.

[5] To be clear, Morningstar never claims that rankings can be compared across categories.

Bonds Credit rating agencies use a familiar ordinal system (AAA, AA, etc.) to describe the credit risk of bonds and structured products. Effectively, these instruments are ranked within categories. For example, credit rating agencies use the same nomenclature to describe ordinary corporate bonds and structured products such as collateralized debt obligations (CDOs), while stating that the ratings across categories are non-comparable. Anecdotally, many investors regarded AAA CDO tranches as substitutes for other AAA claims, even though the underlying assets were quite different.[6]

Insurance Companies AM Best ranks insurance companies for financial strength using ratings similar to bond ratings. The rating is based on AM Best's "opinion of an insurers financial strength and ability to meet its ongoing insurance policy and contract obligations" (see: AM Best (2024), accessed 4/9/2024). Presumably, AM Best takes into account the differences is risk and cash flows across categories of insurance companies (e.g., life, health and property/casualty), but it is not explicit. Either way, consumers are tempted by the ratings to compare across categories.

Stocks Standard and Poor's uses a proprietary "STock Appreciation Ranking System" (STARS) where analysts rank individual stock investments in their areas of expertise relative to appropriate market benchmarks creating implicit categories. While categorization is not explicit, this effectively creates different categories, depending on both the analyst and choice of benchmark for each individual stock.

2.3 Non-financial Examples

There are numerous instances of categorized ratings other than the financial examples above.

Wine Wines with very different characteristics, and desirable in different circumstances, are commonly rated on a 100 point scale that is not intended to enable comparisons across wine categories.

Grades Student grades are commonly compared across majors, and grade-point averages are compared across schools, even though the difficulty, level of competition, and general grading

[6] Moodys considered adopting an alternative ratings system for structured products; see Moody's Investors Services (2008). For detailed discussions of credit ratings during the financial crisis, see Benmelech and Dlugosz (2009, 2010).

policies may be quite different for different majors and institutions.

Tires Tires are often rated in different categories. For example Consumer Reports rates tires on a 100 point scale in 10 different categories, ranging from ultra high performance summer tires to snow tires and all terrain truck tires. What category of tire is "best" varies dramatically with the type of vehicle and driving conditions.[7]

Nutrition information NuVal was an ordinal system that shows the clear pitfalls of such a system. It conveyed information about nutrition to a single number on a 100 point scale. Foods had high or low scores for different reasons and may not be comparable for consumers with different dietary needs and wants. For a description NuVal's shortcomings and demise, see Watson (2017).

The point of these examples is not to criticize simple summary measures, but to illustrate the ubiquity of their use in circumstances where the rating could be mistakenly used across non-comparable categories.

3 Experimental Design

In this section, we describe the essential aspects of the experiment: the investments (essentially small lottery choices) that subjects face, the optimal choice among the investments, and the main experimental treatment, categorization, as well as the other experimental treatments. Our primary goal is to see how the presentation of information about investments affects subject investment decisions.

We begin in Section 3.1 by discussing the experimental portfolio allocation problem. In Section 3.2, we discuss optimal decision making in this context. Sections 3.3 and 3.4 discuss the treatments and the logistics of the experiment.

3.1 Characteristics of Investments

Subjects perform a portfolio allocation task. In each of four trials, each subject is asked to allocate \$12 among six lotteries, with negative allocations not permitted. We will refer to these allocations

[7] For example, while an ultra high performance summer tire may score higher than an all season tire, the latter would clearly be better on a mini-van in snowy weather.

as "investments." In this subsection, we analyze the investment problem that is common to each trial.

Panel A of Figure 1 presents the information about the six investments: the possible high and low returns and summary statistics for each — the average return, the range (high return minus low return), and the ratio of the return to the range, which we call the return/risk ratio.[8] Panel C provides the same information with a slightly altered presentation in which the investments are arranged into two groups of three. This is the "categorized" display that we describe below.

Throughout the experiment, subjects at a minimum see the information in Panel A of Figure 1. The ordering in the baseline trial is the same as in Figure 1, but differs in the other trials (see Table A1 in the appendix).

In determining the outcome of a trial, either all investments earn the high return (with probability 0.5) or all investments earn the low return. This perfect correlation of investment returns is emphasized in the instructions — subjects have to correctly answer a question about this. Because subjects in every trial see at least the information in Figure 1, they do not need to compute means or standard deviations. Moreover, because the returns are perfectly correlated across investments, subjects need not understand subtleties associated with diversification.[9]

3.2 Optimal Investment Allocations

Given the information in Figure 1, what should subjects do? Theory suggests that subjects in experiments with small gambles should behave in a risk-neutral fashion unless they are extraordinarily risk averse (Rabin, 2000). Nonetheless, subjects in experiments commonly do behave in a risk-averse fashion.

Risk-Neutral Subjects A risk-neutral subject would select B exclusively, as it has the highest average return.

[8] The return/risk ratio is half the Sharpe ratio. Thus, it reflects the same information as the Sharpe ratio, but is easier to explain to subjects.

[9] This design choice avoids well known difficulties that subjects have in accounting correctly for correlation and diversification in investment portfolio decisions. See, for example, Kroll et al. (1988), Kroll and Levy (1992), and numerous citations to these works.

Figure 1: Displays of Investment Alternatives Used in the Allocation Stages of the Experiment. The information in Panels A and C is seen in every trial by every subject in the uncategorized and categorized treatments, respectively. The order in which assets are displayed changes across trials, but A, B, and C are always grouped together, as are D, E, and F. Panels B and D show the star ratings appended to Panels A and C in one of the trials.

Panel A: Basic information about the investments in the uncategorized treatment.

Alternative:		A	B	C	D	E	F
High Return:		130%	185%	125%	200%	225%	190%
Low Return:		30%	15%	-25%	-20%	-75%	-90%
Average Return:		80%	100%	50%	90%	75%	50%
Range of Returns:		100%	170%	150%	220%	300%	280%
Return/Risk Ratio:		0.8	0.5882	0.3333	0.4091	0.25	0.1786

Panel B: Star ratings assigned to each investment, presented to subjects in the uncategorized treatment in Trial 2.

Uncategorized rating:		***	***	**	**	*	*

Panel C: Basic information about the investments in the categorized display.

		Category I				Category II		
Alternative:		A	B	C		D	E	F
High Return:		130%	185%	125%		200%	225%	190%
Low Return:		30%	15%	-25%		-20%	-75%	-90%
Average Return:		80%	100%	50%		90%	75%	50%
Range of Returns:		100%	170%	150%		220%	300%	280%
Return/Risk Ratio:		0.8	0.5882	0.3333		0.4091	0.25	0.1786

Panel D: Star ratings assigned to each investment, presented to subjects in the categorized treatment in Trial 2.

Categorized Rating:		***	**	*		***	**	*

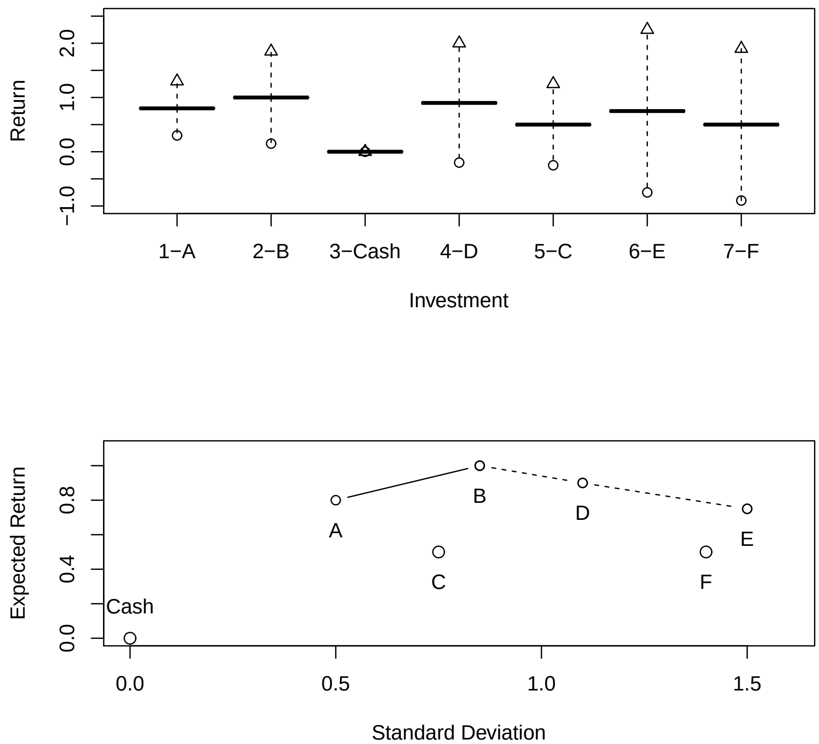

Figure 2: Top panel: Outcomes and mean return for each investment, ordered by minimum return. Investments cash, C, and F are dominated by A and B, D, and E, respectively. Bottom panel: Expected returns and standard deviations of investments along with the efficient frontier for risk averse investors (solid line) and risk seeking investors (dashed line).

Risk-Averse Subjects The optimal investment choice for a subject exhibiting risk aversion entails choosing investment A, B, or a combination of the two. To understand why, consider Figure 2, which presents two graphical depictions of the investments from Panel A in Figure 1. (Neither of these depictions was presented to the subjects.)

The top panel displays both the high and low returns from Figure 1 and ranks investments by minimum return. We can use this to consider asset choice by a risk-averse subject. It is apparent that cash, C, and F are dominated by the assets to their immediate left in the Figure. The graph also shows why a risk-averse investor strictly prefers B to both D and E. Investments D and E both have lower means than B and a greater range (greater standard deviation). Thus, a risk-averse investor would prefer B by itself to D or E, or to a combination of B with D or E. The comparison of A and B is ambiguous, however: B has a greater mean, a greater range, and the minimum return for B is below that for A.

The bottom panel of Figure 2 provides a different view of the alternatives, displaying a standard portfolio/efficient frontier graph for the six assets. We can use this to consider a subject with mean-variance utility. The frontier is simple because a single random draw determines whether all investments receive the high return or low return. If a portfolio is invested 1/3 in A and 2/3 in B, for example, both the mean and standard deviation are 1/3 that of A plus 2/3 that of B. A portfolio invested in A and D would have a lower mean than one invested in A and B or B and D. Because there is no gain from diversification, efficient portfolios for a subject with mean-variance utility consist of at most two assets.

Risk-Seeking Subjects Some evidence suggests that subjects may be risk seeking across small gambles (e.g., see Berg et al., 2010). Using the same reasoning as before, a risk seeking investor may hold B, D or E depending on the degree of preference for risk, but not A, C or F.[10]

To summarize, subjects who are not risk-seeking should invest only in some combination of A and B. It is of course possible that, whatever their risk preferences, some subjects will invest suboptimally. In all cases, we will be interested in seeing whether different experimental treatments induce different choices.

[10] Further, to achieve any given risk and return combination, a risk seeking subject never needs to hold more than two of B, D or E.

3.3 Categorization and Star Ratings

Panels A and C of Figure 1 present all of the information necessary for subjects to make investment decisions. This information is always visible to participants. The main treatment is categorization, in which the investments are split into two groups. We now discuss the role of categorization and ratings in the experiment.

3.3.1 Categorization and Display of Investments

Throughout the experiment, half of the subjects see the non-categorized display in Panel A of Figure 1, and the other half see Panel C, in which the six investments are divided into two groups, labeled "Category I" and "Category II." These categories correspond to investment risk levels: A, B, and C have a smaller range of returns than D, E, and F. Categorized information is displayed with a blank column separating the two groups of three investments. The ordering of investments changes, but the categorization grouping is always the same, based on the variability of returns, with A, B, and C always together in one category and D, E, and F in the other.[11]

Note that Category I contains the lowest-risk assets, including A and B, and Category II contains the high-risk assets. The categories thus bear a stylized resemblance to different Morningstar asset categories. Subjects were told that investments were categorized using "a commonly used financial method."

3.3.2 Star Ratings

In two of the trials, asset ratings are added to the display. Investments have ratings ranging from one star (worst) to three stars (best). The ratings use the ratio of the expected return over the risk (high minus low return), which is proportional to the Sharpe ratio. Subjects are shown star ratings in one trial and are required to assign star ratings in a different trial. In both cases we impose a uniform distribution of stars within a category (if categorized) or across the six investments (if uncategorized). Panels B and D of Figure 1 show the investment ratings based on return/risk ratios for categorized and uncategorized treatments.

[11] We use the category labels "I" and "II" to mean relatively low and high risk investments. The categorized displays in the experiment were always labeled "I" and "II" from left to right, regardless of the actual investment alternatives that appeared in each category.

The critical aspect of categorization is that it changes the way investments are ranked. Without categorization, all investments are ranked as a single group, with A and B both receiving three stars and E and F one star. With categorization, investments are ranked within a group. Investment A has the highest Sharpe ratio and has a 3-star rating whether categorized or not. Investment F has the lowest Sharpe ratio and is ranked 1-star either way. Categorization, however, drops B from 3 stars to 2 stars and C from 2 stars to 1 star. It raises D from 2 stars to 3 stars and E from 1 star to 2 stars. Our primary question is whether these rating changes affect decisions even though all of the fundamental information remains constant and is displayed throughout the trials.

We base star ratings on the return/risk ratio for several reasons. The return/risk ratio is a simple intuitive criterion that provides a ranking (in the non-categorized treatment) that is roughly consistent with optimal choices for a risk-averse investor: A and B receive the highest ratings. Conditional on categorization, the ranking is also correct for both categories: A and B are preferred to C, and D is preferred to E and F. Finally, a subject who has studied finance might believe they should perform this kind of calculation; the presentation is intended to reduce computational load for subjects by computing what they might want to compute.[12]

3.4 Description of the Experiment

The data in the paper comes 1,244 faculty and staff volunteer subjects from the University of Iowa who responded to a university wide email invitation. The email invitation asked for subjects who would participate in an on-line experimental session that would last less than an hour. Those agreeing to participate received the web address for the study, a login ID, and a random password. They could participate at their convenience. After logging in, they went through an on-line version of the instructions and exercises given in the Appendix. Subjects completing the experiment received a $5 participation fee and additional payments that depended on the investment allocation decisions described above. Because university employees are required to participate in the university's retirement program, the subjects are in a life stage where they are making important investment decisions.

[12]There are caveats to our use of the return/risk ratio. While it mirrors a common evaluation metric (the Sharpe Ratio), it is generally appropriate for comparing diversified portfolios, not individual assets. Its use assumes specific return distributions and utility functions. It does not identify dominated assets (C for example is dominated by B, but not by E, compared to which it has a higher return/risk ratio).

Figure 3: Initial Investment Alternative

Alternative:	A
High Return:	100%
Low Return:	-100%
Average Return:	0%
Range of Returns:	200%
Return/Risk Ratio:	0.0000

3.4.1 Preliminary Section

Subjects first received instructions and general information about the experiment. They were shown a sample table for one asset that mimicked the presentation in the experiment.

Second, subjects completed a three-question quiz on the determination of payments in the experiment. The three questions were intended to insure that subjects understood how to compute the return they would receive in the high and low states (two questions), and that returns on all assets were perfectly correlated (one question). Subjects could not proceed until they answered all questions correctly.

Finally, to measure their willingness to gamble, subjects chose whether to allocate $1 to a single, actuarially fair investment alternative. Figure 3 shows how this was presented to subjects.[13]

3.4.2 Trials

Each subject participates in four trials:

- Initial baseline trial: Subjects see the basic information about the investments and allocate $12.

- Rated trial: Subjects are shown star ratings for the investments together with the basic information about the investments, and allocate $12.

- Self-rated trial: Subjects see the basic information and are required to assign star ratings to the investments. They then allocate $12.

[13]We chose this procedure because elicited risk references are not necessarily stable across institutions (see for example, Berg et al. (2005)). The task here, its presentation and the choice procedure is essentially identical to tasks, presentations and procedures used in in the main experimental trials.

- Final baseline trial: Subjects see the basic information about the investments and allocate \$12.

Table 1 summarizes the structure of the different treatments. Half of the subjects experience trials in the order shown (rated trial followed by self-rated trial) while the order of these two trials is reversed for the other half. All of the subjects see a baseline display in the first and fourth trials. Subjects in the categorized treatment see a categorized display in all four trials.

Treatments include grouping the investment information in different ways (categorization), telling subjects in Trials 2 and 3 the rule used in assigning stars (the rating rule treatment), and reversing the order of the 2nd and 3rd trials (the order treatment). The three treatments—categorization, rating rule, and order—result in a $2 \times 2 \times 2$ design with 8 treatment combinations. Each treatment combination was thus presented to 1/8 of the subjects.

The ordering of investments differed across trials, but the grouping of investments did not change: in both categorized and non-categorized treatments, investments A, B, and C (not necessarily in that order) were in either the first three or the last three columns. Investments D, E and F (not necessarily in that order) were in the other three columns.[14] In uncategorized treatments, there was no mention of categorization.

3.4.3 Knowledge, Experience, and Demographics

After completing the trials, subjects filled out knowledge and demographic surveys, both of which are reproduced in the Appendix. The demographic survey, adapted from Oliven and Rietz (2004), asks about gender, age, marital status, education, etc. The knowledge survey, also adapted from other sources as noted in Appendix B, asks subjects to self-report on their own financial market knowledge and experience (four questions) and asks about simple definitions, basic concepts, applications of concepts to risk and return relationships, and asset allocation (nine questions). Obviously, pre-existing financial knowledge and experience could affect the behavior of subjects.[15]

[14] We associate the labels "A" through "F" with specific alternatives here for expositional convenience. Subjects were always shown alternatives labeled "A" through "F" in order from left to right regardless of the actual investment alternatives that appeared in each column.

[15] There is a sizable literature on knowledge and financial decision making. For example, Bernheim et al. (2001) and Bernheim and Garrett (2003) discuss how knowledge affects savings rates. Lusardi and Mitchell (2007) discuss how financial literacy affects retirement planning. Anderson and Settle (1996) show significant effects of prior knowledge on biases in financial decisions. Grinblatt et al. (2011) suggest that higher IQ is associated with greater stock market participation and a higher Sharpe ratio. Finally, Sunden and Surette (1998) suggest that gender may affect asset

Table 1: The entry in each cell reports what subjects see in each trial and treatment. In each trial, subjects allocate $12 across the six investments. In the experiment, half of the subjects perform Trial 2 before Trial 3 with the other half reversing the order. In the Rated and Self-rated trials, half of the subjects (the same half in each trial) are told the rule determining ratings.

		Treatment		
Trial		Non-categorized	Categorized	Rating Rule
1:	Initial Baseline	Figure 1, Panel A	Figure 1 Panel C	
2:	Rated	Figure 1, Panels A + B	Figure 1, Panels C + D	provided to half of subjects
3:	Self-rated	Figure 1, Panel A subject ranks investments	Figure 1 Panel C subject ranks investments	provided to same half of subjects
4:	Final Baseline	Figure 1, Panel A	Figure 1 Panel C	

An example of a question on the experience portion of the survey is "How would you classify your knowledge of financial markets?" with answers ranging from "No knowledge" to "Advanced." Treating the responses as cardinal and summing them, the maximum score is 14. We scale this to be between 0 and 1. For the knowledge portion of the survey, an example of a question is "Common stocks always provide higher returns than bonds or money market investments." The permitted answers were "True," "False," and "Don't Know." The knowledge score is the fraction of the 9 questions answered correctly. In our regressions, we normalize both by scaling them to a 0-1 range and subtracting the mean. Thus, an average participant has an experience score of 0 and a knowledge score of 0.

3.4.4 Payoffs

Finally, three random numbers determined payoffs:

1. the payoff to the initial $1 investment,
2. which trial's portfolio allocation would be used to determine payment, and
3. whether investments in the selected trial paid the high or low return.

allocation in defined contribution retirement savings plans.

After learning their total payment, the subjects answered a last question about their satisfaction with their own decisions in the experiment.

After completion of the experiment, the University of Iowa mailed checks for the total amounts to the subjects. The maximum possible payment was earned if a participant invested \$1 in the initial bet and \$12 in Alternative E in the randomly-selected trial, and then received the high payoff for both. Including the participation fee, the payoff would be $\$5+\$1\times(1+1)+\$12\times(1+2.25) = \$46$.

3.5 Interpretation and Hypotheses

Before examining the results, we discuss design and interpretative issues. We put forth several formal hypotheses and discuss other ways our design sheds light on the issues at hand.

3.5.1 Design considerations

The experimental task is not trivial, so we sought to minimize confounds with factors known to affect subject choices in investment portfolio decisions. First, we eliminated learning and income effects. Each subject's final payoff was determined at the end of the experiment, based on the results from the initial fair bet and the investment return resulting from one randomly-selected trial. Results in a given trial are unknown and cannot affect subsequent decisions, and because of random selection, subjects have the same incentive in each trial to make an optimal choice.[16]

Second, as we have discussed, in each trial the return outcomes for the six investments are perfectly correlated. This design avoids well known difficulties that subjects have in accounting correctly for correlation and diversification in investment portfolio decisions.[17]

Finally, it is important to understand that the asset allocation task here differs from one subjects would typically encounter in financial education. A subject who remembered and followed a dictum such as "diversification is beneficial" or followed a naive $1/n$ strategy (e.g., Benartzi and Thaler, 1999) would make suboptimal decisions in this experiment. Thus, if financial knowledge improves performance in the experiment, it is because the subject understands the task, not because he or she had been taught how to do it.

[16]For a discussion of income effects and ruling them out in experiments, see Kahneman et al. (1990).

[17]See, for example, Kroll et al. (1988), Kroll and Levy (1992), and numerous citations to these works.

3.5.2 General Hypotheses

Formally, across all rounds, we hypothesize:

Hypothesis 1 *With respect to overall investment levels:*

Null: Subjects naively diversify, investing in all assets in roughly the same amounts.

Alternative: Subjects concentrate their investment in some assets while avoiding others.

In short, we find strong evidence that subjects do not follow a $1/n$ rule and, instead, concentrate their investments.

If subjects are rational, they will avoid dominated Assets C, F and Cash. Given the stakes in in the experiment, Rabin (2000) argues that subjects should behave as if risk neutral while many researchers observe apparent risk preferences in experiments (e.g., Eckel and Grossman, 2008; Berg et al., 2010) leading to:

Hypothesis 2 *Null: Investments in assets follow no particular pattern.*

Risk Neutral Alternative: Subjects will invest heavily in Asset B, avoiding other assets.

Risk Averse Alternative: Subjects invest heavily in Asset A (possibly along with B), little in D and E, and avoid dominated assets C, F and Cash.

Risk Seeking Alternative: Subjects invest heavily in Assets D and E (possibly along with B), little in A, and avoid dominated assets C, F and Cash.

The risk averse and seeking alternatives assume a sufficient degree or risk preference to move subjects away from exclusive investment in B. We further hypothesize that our initial risk preference bet will indicate which alternative best explains subject behavior. We also investigate whether financial knowledge and self-reported financial experience/expertise along with demographics affect allocations. In short, we find that subjects invest most heavily in B, but also invest significant amounts in A followed by D.[18] They invest little in C, E, F and Cash. Financial knowledge and expertise along with the risk preference bet are all strongly correlated with allocations. Women invest less in B, but spread the resulting increases around so that no single other asset has a significantly higher allocation.

[18] B, A and D are all significantly positive, but clearly ordered with investment in B significantly higher than A which, in turn, is significantly higher than D.

3.5.3 Categorization

The main question we ask is whether subjects are affected by the categorization treatment and the resulting changes in star ratings for four out of the six assets. Formally:

Hypothesis 3 *Null: Allocations are unaffected by differences in star ratings across treatments.*

Ratings Effect Alternative: When a rating changes across categorization treatments, the same assets (with the same payoff/return information presented to subjects) will attract more investment under the treatment that gives it the higher rating and less under the treatment that gives it the lower rating. Thus, in categorized treatments, investment levels in B and C will fall and investment levels in D and E will rise <u>*relative*</u> *to levels in uncategorized treatments.*[19]

In addition, we ask whether apparent risk preferences, financial knowledge and self-reported financial experience/expertise along with demographics are correlated with categorization effects. Finally, we ask whether behavior is affected by explaining to subjects how ratings are determined (which shows explicitly that they contain no additional information).

In brief, we observe significant changes consistent with a ratings effect. No other variables (demographics, risk preference bet, knowledge or experience) are significantly correlated with this.

Note that differences across categorization treatments in Stages 2 and 3 drive this hypothesis. It is not a hypothesis about differences across stages. Next, we discuss several possible interpretations of the ratings effect that we will use to interpret the results.

Pure ratings effect The star ratings provide subjects with no additional information. However, as with irrelevant information (Chadd et al., 2021, e.g.,), the presence of ratings that convey no additional information may affect choices leading to a ratings effect. This may result from a "preference for simplicity" in the decision-making process that would lead subjects to rely on star rankings. We investigate whether behavior is affected by explaining the ranking rule and making it clear that ratings contain no additional information. This should mitigate any preference for simplicity effect. In addition, we study whether knowledge and experience, which ought to make it easier to solve the allocation problem, mitigate the effect. Not of these factors appear significant.

[19]Technically, there are other possible alternatives. A reactionary response may lead subjects to lower investments in highly rated assets. There may also be another or no particular pattern of changes.

Experimenter demand effect The experimenter demand effect occurs when subjects respond to cues in an experiment based on what they interpret as socially appropriate behavior (Zizzo, 2010). An interpretative problem arises when the experimenter demand perceived by subjects is correlated with the experimental predictions, in which case the cues could drive the results. In our case, a possible experimenter demand effect would be subjects responding to stars by investing more in three star investments, and less in one star investments, independently of the economic characteristics of the investment.

We argue that, if it exists, the experimenter demand effect is similar to what one might observe in financial markets, with investors feeling social pressure (from brokers, advisers, and colleagues) invest in assets with more stars. And, thus, it is simply another channel through which the ratings effect may manifest itself. However, the experimenter demand effect would primarily be an effect observed across stages and not across categorization treatments when rankings across treatments do not change. Our experiment provides us with some evidence for a categorization-based ratings effect beyond a simple demand effect.

Our design allows us to test for a demand effect and assess whether there is more to the changes observed than this alone. Subjects see no stars in Trial 1, and they are presented with stars in Trial 2. If present, the experimenter demand effect would result in subjects investing more heavily in three star investments in Trial 2 than in Trial 1 across the board. Similarly, subjects would invest less in one star investments across the board. Specifically, there would be increased investment in asset A in all treatments independent of categorization, asset B when non-categorized, and asset D when categorized. Conversely, investment in asset F should fall in all treatments, asset E when non-categorized and asset C when categorized. In contrast, the ratings effect is a prediction about the differences in investment levels across categorization treatments in Trial 2 alone. It predicts no difference in A and F with significant differences in B, C, D and E. Overall, data is more consistent with the ratings effect than the experimenter demand effect. Specifically, when the categorization and experimenter demand effects make the same prediction, investment levels move in the direction predict by both. But, when there are deviations, the data is more consistent with the ratings effect than the demand effect. [20] Thus, while the experimenter demand effect may play a role, there

[20]While there are some deviations from the demand effect predictions in E and F (e.g., investment in E falls significantly while F does not Fall in the categorized treatments after controlling for other factors), investment levels are generally small in both investments regardless of the treatment. Predictions for A, D and C are similar across

is more going on than the demand effect explains and treatment effects are consistent with the categorization effect.

Cognitive challenges Subjects could be cognitively challenged by the experiment. The experimental task is not simple and it is possible that subjects could use ratings to assist them in decision-making. Again, if this is the case, as with the experimenter demand effect we should see a systematic difference between Trial 1 and Trial 2 related to ratings. The absence of such a difference would be evidence that ratings are not being used in this fashion. Other aspects of our experimental design allow us to assess the cognitive challenges faces by subjects. The overall tendency for subject to avoid dominated assets and invest heavily in assets consistent with their measured risk preferences and their own ratings argue against this explanation. Further, financial knowledge and financial experience do little to mitigate the effect. Finally, were cognitive challenges alone responsible for the effect, explaining the ratings rule should mitigate it. Again, none of these factors appear to mitigate the effect.

Ratings create cognitive dissonance Another possibility is that subjects face cognitive dissonance when a rating is inconsistent with their prior belief about the asset. The theory of cognitive dissonance (e.g., see Akerlof and Dickens, 1982) implies that subjects will try to harmonize their behavior and belief. In our context, categorized ratings could create dissonance by, for example, reducing the star rating for B, which (as we will see) was a popular investment in Trial 1. The subject confronted with a low assigned rating can only reconcile behavior and information by reducing investment.

Interestingly, a subject assigning a rating may be less likely to experience dissonance since the subject is in control of both the rating and the investment decision, which affords the opportunity to align the two. Again, we study this in our experiment. In short, consistent with this explanation, we find that subjects invest much more heavily in assets that they self-rank highly.

both effects and results are consistent with them. So, the primary differentiator is what happens to B in Trial 2. The demand effect predicts an increase in stage 2 for uncategorized treatments and no change in categorized treatments. The opposite occurs. In contrast, the difference across treatments in this case is significant and consistent with the ratings effect.

Permanence Our design allows us to check to see whether the ratings/category effects are transitory or more permanent. In our design, subjects make four allocation decisions in stages. The fourth stage is exactly the same as the first. No results are determined until after this stage, so there are no income effects and subjects cannot learn by observing outcomes. If the effects are driven purely by observed ratings, then the should disappear in Stage 4, leading to:

Hypothesis 4 *Null: Allocations changes are transitory and do not differ significantly from each other in the two stages where there are no rankings.*

Alternative: Ratings have a residual effect after they are observed, then removed.

While at first glance the overall average investment levels appear to have significant residual effects, the regression analysis shows only one meaningful and significant difference that remains after controlling for other factors: subjects invest significantly more in D in the categorized treatment than in the uncategorized treatment in Trial 4. Thus, most of the effects appear due to the transitory impact of ratings, not from the subjects permanently changing behavior in response to them (either because of learning or sustained experimenter demand effects).

4 Results

In this section, we present the results. We first summarize the characteristics of subjects and provide basic data about the experiment. Subjects on average behave reasonably in the first, untreated trial, investing primarily in A and B, avoiding the dominated assets, and not diversifying (largely rejecting the $1/n$ null of Hypothesis 1 and generally supporting the risk averse alternative of Hypothesis 2). We then look at performance across trials, examining the effects of treatment, asking whether knowledge mitigates these effects, and exploring different economic explanations for the results. We find that categorized ratings do affect investment and worsen performance (rejecting the null of Hypothesis 3 in favor of the categorization effect alternative). We conclude that the results are more consistent with the ratings effect than an experimenter demand effect and they are unlikely to be explained by subjects finding the task too difficult, although there is clearly evidence that some subjects are at times confused. Looking at overall performance across trials, we show that categorized subjects in Trial 2 have a lower overall Sharpe ratio.

4.1 Subjects

Table 2 provides summary statistics by treatment, including the number of subjects, the gender composition, the Experience Index, Knowledge Score, how much time the subjects spend in the asset allocation portions of the experiment (the average subject spent about 6 minutes, 44 seconds in allocation rounds). Between 154 and 157 subjects participated in each treatment; 62% were women; 63% were married. The average total payoff was $27.31 with a standard deviation of $10.30, a maximum of $46.00, and a minimum of $6.20. According to Kruskal-Wallis rank sum tests, the only significant differences in summary statistics across treatments are: (1) the average knowledge score which still fell within a narrow range of 6.29 to 6.98 questions right out of 9 on average; and time spent on the allocation decisions which we expect would vary with the task.[21]

Table 3 provides the same summary statistics as Table 2 by gender. Women were significantly likelier to take the initial bet (60% vs 45% for men)[22] and had a significantly lower average knowledge scores (6.34 vs 7.24) and self-reported experience (0.15 vs 0.31). However, there were no significant difference in the time spent in allocation stages nor in total payoffs. Gender will be a control variable in regression we report later[23]

4.2 Investment in Trial 1

The first question is how subjects performed on the basic allocation task in Trial 1, without treatments. The answer is that subjects invested most heavily in A and B, mostly avoided the dominated assets (C, F and Cash), and did not naively diversify. We clearly reject the null in Hypothesis 1 in favor of the alternative that subjects concentrate their allocations. Further most subjects make allocations consistent with the risk neutral and risk averse alternatives in Hypothesis 2. While some allocations are consistent with the risk seeking alternative, few allocate to the dominated assets.

[21] While the time spent is highly variable and highly skewed, there are small, but significant differences. The median subject spent 1 minute 49 seconds in the Trial 1 allocation stage; 1 minute 8 seconds in Trial 2 (with rankings given), significantly less than in Trial 1 (rank sum z-stat=-23.886, p-value=0.0000); 2 minutes 24 seconds in Trial 3 (with self rankings required), significantly more than in Trial 1 (rank sum z-stat=13.983, p-value=0.0000); and 50 seconds in Trial 4, again again significantly less than in Trial 1 (rank sum z-stat=-27.285, p-value=0.0000). The median subject in categorized treatments spent 40.5 seconds more overall in decision trials than the median subject in uncategorized treatments (rank sum test z-stat. = 3.010, p-value = 0.0026). While interesting, it is expected. The subjects are processing different information in each stage and taking on slightly different tasks. Explaining categorization requires more text on the decision page. Ranking according to a given rule, especially within categories, may take more time than ranking according to one's own preferences. Etc.

[22] This remains significant after controlling for marital status, knowledge and experience.

[23] Table A2 in the Appendix presents correlations for gender, taking the initial bet, experience, and knowledge.

Table 2: Summary information by treatment. The Experience Index is the scaled average response on questions 1-4 on the Knowledge Survey in the Appendix, where the response values for each question are 1 through 4 in the order of the answers. The Knowledge Score is the mean score for answering questions 5-13 in the Knowledge Survey. Seconds Spent in Allocation Trials is total time spent selecting allocations for all four Trials.†

			Treatment								Kruskal-Wallis Stat.
		Overall	1	2	3	4	5	6	7	8	(p-value)
Categorized?		N/A	Yes	Yes	Yes	Yes	No	No	No	No	N/A
Rating Rule Given?		N/A	No	Yes	No	Yes	No	Yes	No	Yes	N/A
Stage II or III First?		N/A	II	II	III	III	II	II	III	III	N/A
Obs.		1244	157	155	155	157	155	156	155	154	N/A
% Female		62%	63%	61%	60%	69%	61%	57%	62%	63%	3.83
											(0.80)
% Married		63%	66%	65%	58%	62%	62%	66%	63%	62%	3.62
											(0.82)
% Making Preliminary Bet		54%	50%	55%	55%	57%	59%	53%	49%	55%	3.93
											(0.79)
Knowledge	Mean	6.68	6.98	6.92	6.39	6.44	6.76	6.80	6.29	6.84	22.99**
Score	Std. Dev.	1.71	1.59	1.60	1.73	1.85	1.69	1.72	1.83	1.50	(0.00)
Experience	Mean	0.21	0.22	0.21	0.22	0.20	0.19	0.23	0.20	0.22	3.75
Index	Std. Dev.	0.19	0.18	0.19	0.20	0.17	0.17	0.21	0.19	0.19	(0.81)
Seconds Spent in	Median	404	448	379	423	449	327	396	415	387	32.40**
Allocation Trials†	Std. Dev.	1966	245	252	673	5282	524	332	888	1071	(0.00)
Total	Mean.	$ 27.31	$ 27.50	$ 27.65	$ 26.38	$ 26.70	$ 27.56	$ 29.33	$ 26.77	$ 26.56	9.50
Payment	Std. Dev.	$ 10.30	$ 10.67	$ 10.19	$ 10.23	$ 10.30	$ 10.28	$ 10.51	$ 9.87	$ 10.28	(0.22)

†Because this variable is highly skewed, we report medians and, in the last column report the ranksum z-statistic and p-value.
*p<0.05, **p<0.01

Table 3: Summary information by gender. The Experience Index is the scaled average response on questions 1-4 on the Knowledge Survey in the Appendix, where the response values for each question are 1 through 4 in the order of the answers. The Knowledge Score is the mean score for answering questions 5-13 in the Knowledge Survey. Seconds Spent in Allocation Trials is total time spent selecting allocations for all four Trials.†

			Gender		t-test
		Overall	Male	Famale	(p-value)
Obs.		1,233	470	763	N/A
% Married		63%	69%	59%	3.63**
					(0.00)
% Making Preliminary Bet		54%	45%	60%	-5.06**
					(0.00)
Knowledge	Mean	6.68	7.24	6.34	9.38**
Score	Std. Dev.	1.71	1.56	1.70	(0.00)
Experience	Mean	0.21	0.31	0.15	15.99**
Index	Std. Dev.	0.19	0.20	0.15	(0.00)
Seconds Spent in	Median	404.5	385	413	1.18
Allocation Trials†	Std. Dev.	1975	540	2475	(0.28)
Total	Mean.	$ 27.38	$ 27.62	$ 27.22	0.66
Payment	Std. Dev.	$ 10.30	$ 10.26	$ 10.34	(0.51)

†Because this variable is highly skewed, we report medians and, in the last column, report the ranksum z-statistic and p-value.
*p<0.05, **p<0.01

Thus, overall subjects generally make allocations consistent with rational choices.

Figure 4 is a violin plot depicting the density function across subjects for investment in Trial 1 in each asset and cash. The median level of investment is represented in each plot by a white dot, the inter-quartile range by heavy line (box), and the contours of the enveloping curve depict a kernel density, effectively a rotated histogram. A and B were the only assets for which a significant percentage of subjects invested $2.50 or more, and the mean investment in the two assets combined was $7.32. For each of the dominated assets C, F, and Cash, 73% or more of subjects invested 0. Investments in Asset D constituted the bulk of investment in the high-risk group of assets.

We can also ask to what extent subjects diversify, even though there is no value to investing in more than two assets. Figure 5 depicts a violin plot for the cumulative investment across 6 of the top 7 assets (including cash) in Trial 1, with investments for each subject sorted in order of the amount invested. The figure shows that the median investor has the bulk of their cash — $9 — invested in two or fewer assets and all of their $12 invested in three or fewer assets. 75% of investors have at least $10 in three or fewer assets. Only 17.0% of investors (212) invest in 6 assets

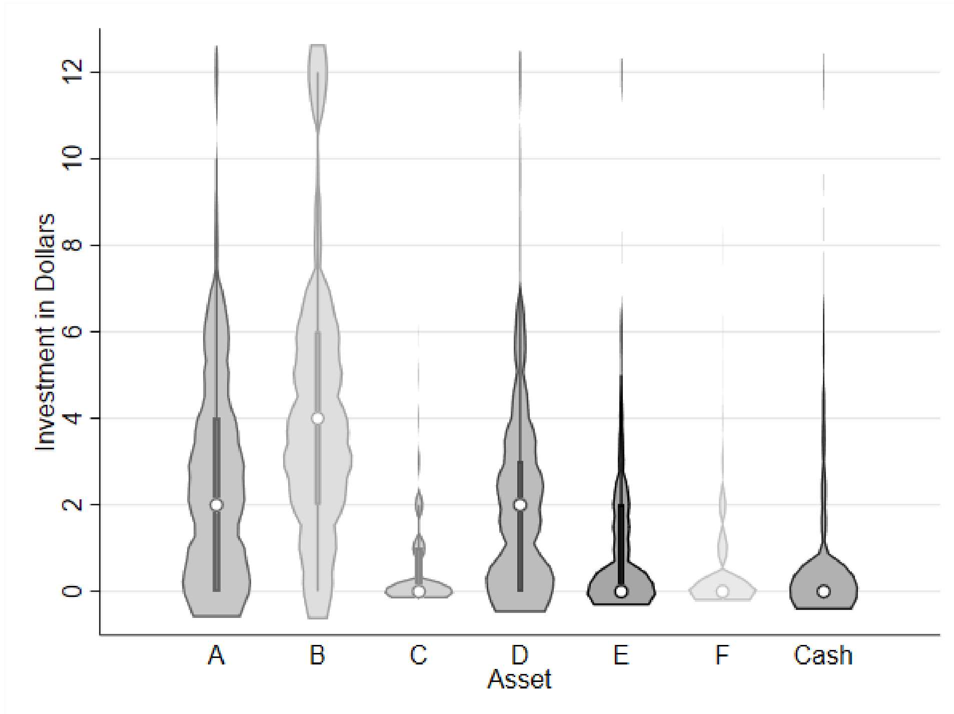

Figure 4: Violin plot depicting investment by asset in Trial 1. The box in the center shows the interquartile range, with the median a white dot. The overall shape is a kernel density plot.

at any trial and 6.2% (77) invested in all 7 assets at some point.[24] The figure demonstrates that there is heterogeneity across subjects, but also reinforces the conclusion that the bulk of investors do not naively diversify.

Finally, we checked for spurious correlation between subject behavior and treatments by computing the significance of the correlation between Trial 1 investments and each of the three treatments, none of which are in effect in Trial 1. One of the 21 correlations none were significant at the 5% level.

4.3 Investment in Trials 2 through 4

Here, we present evidence that the allocations across stages and between categorization treatments accord with the categorization effect alternative in Hypothesis 3. We also show where the evidence is more consistent with a categorization-driven ratings effect than a simple experimenter demand effect. Finally, we show in regressions that the effects are largely transitory, consistent with the transitory null in Hypothesis 4.

[24]Of these, only 4 (0.3%) subjects invested in 7 assets in every trial.

Table 4: Investment across trials and across the categorization treatments. Initial baseline trial allocations to investments are in Panel A. Panels B and C display deviations from initial allocations in later trials, in the non-categorized and categorized treatments. Panel D is the difference between Panel C and Panel B. The first line in each row in Panels B-D is the average deviation and the second is the p-value for the Wilcoxon statistic for differences in the distributions across the stages.

Panel A: Average Allocation in Trial 1								
Trial 1		**A**	**B**	**C**	**D**	**E**	**F**	**Cash**
Avg. Allocation		$ 2.809	$ 4.511	$ 0.462	$ 0.462	$ 2.043	$ 0.431	$ 0.788
Std. Dev.		$ 2.757	$ 3.390	$ 0.909	$ 0.909	$ 2.225	$ 0.998	$ 1.947
Panel B: Changes from Trial 1 in Uncategorized Treatments								
Change from Trial 1 to:		**A**	**B**	**C**	**D**	**E**	**F**	**Cash**
Trial 2	Avg. Change	$ 0.360	$ 0.326	$ 0.058	$ (0.281)	$ (0.384)	$ (0.089)	$ 0.010
	Wilcoxon P-Value	0.000**	0.013*	0.104	0.008**	0.000**	0.013*	0.053*
Trial 3	Avg. Change	$ 0.092	$ (0.442)	$ (0.048)	$ (0.290)	$ (0.329)	$ (0.052)	$ 1.069
	Wilcoxon P-Value	0.228	0.000**	0.179	0.003**	0.000**	0.100	0.000**
Trial 4	Avg. Change	$ 0.460	$ 0.055	$ (0.031)	$ (0.203)	$ (0.244)	$ (0.074)	$ 0.037
	Wilcoxon P-Value	0.000**	0.358	0.569	0.011*	0.000**	0.015*	0.366
Panel C: Changes from Trial 1 in Categorized Treatments								
Change from Trial 1 to:		**A**	**B**	**C**	**D**	**E**	**F**	**Cash**
Trial 2	Avg. Change	$ 0.785	$ (1.088)	$ (0.082)	$ 0.538	$ (0.117)	$ (0.032)	$ (0.005)
	Wilcoxon P-Value	0.000**	0.000**	0.008**	0.000**	0.054	0.015*	0.068
Trial 3	Avg. Change	$ 0.176	$ (0.827)	$ (0.051)	$ (0.056)	$ (0.229)	$ (0.112)	$ 1.099
	Wilcoxon P-Value	0.508	0.000**	0.019*	0.913	0.000**	0.002**	0.000**
Trial 4	Avg. Change	$ 0.473	$ (0.447)	$ (0.104)	$ (0.006)	$ (0.027)	$ (0.034)	$ 0.146
	Wilcoxon P-Value	0.000**	0.000**	0.001**	0.605	0.696	0.088	0.426
Panel D: Differences in Changes from Trial 1 for Categorized vs Uncategorized Treatments								
Change from Trial 1 to:		**A**	**B**	**C**	**D**	**E**	**F**	**Cash**
Trial 2	Avg. Change	$ 0.426	$ (1.414)	$ (0.140)	$ 0.819	$ 0.267	$ 0.057	$ (0.014)
	Ranksum P-Value	0.028*	0.000**	0.003**	0.000**	0.000**	0.991	0.909
Trial 3	Avg. Change	$ 0.084	$ (0.385)	$ (0.003)	$ 0.234	$ 0.100	$ (0.061)	$ 0.030
	Ranksum P-Value	0.731	0.046*	0.535	0.051	0.301	0.335	0.568
Trial 4	Avg. Change	$ 0.013	$ (0.502)	$ (0.074)	$ 0.197	$ 0.216	$ 0.041	$ 0.109
	Ranksum P-Value	0.263	0.001**	0.063	0.031*	0.003**	0.508	0.222

**p < 0:01, *p < 0:05

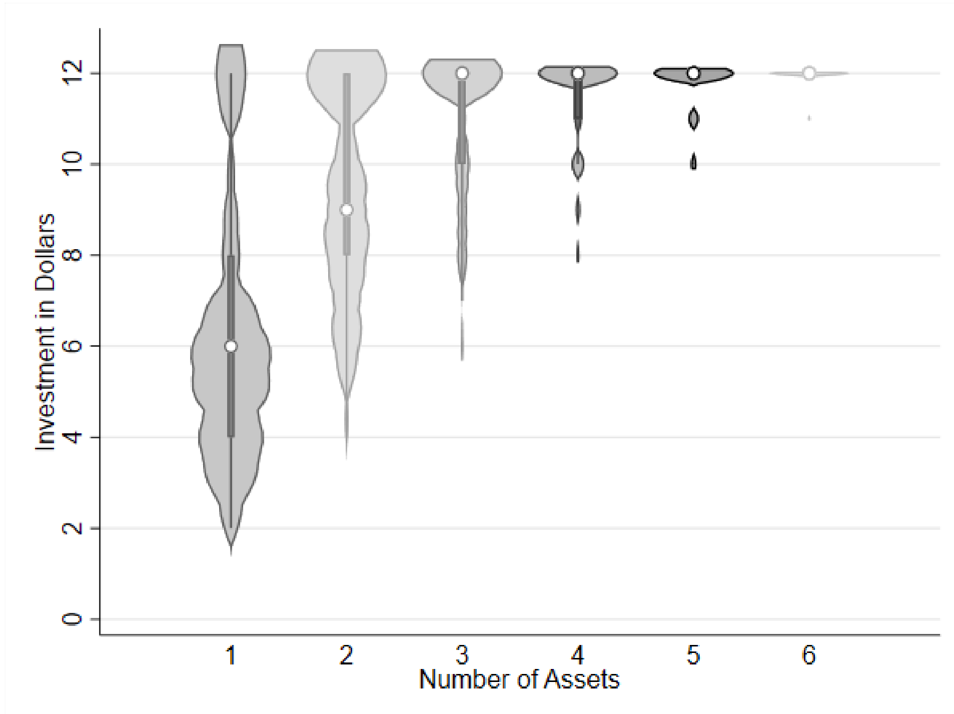

Figure 5: Violin plot depicting the distribution of the number and quantity of assets held by investors in Trial 1. For a given subject, holdings are sorted from greatest to least and cumulated. The plot shows the number of dollars invested in the largest holding, the largest two holdings, etc.

Table 4 summarizes investment levels in the first and subsequent trials with respect to the categorization treatment. Panels B and C show how investment levels for each asset change for each trial relative to the Baseline, for subjects in the non-categorized (Panel B) and categorized (Panel C) treatments. Panel D shows the difference in these changes between the two treatments (i.e., it is a difference in differences).

These difference give the reader a general idea of the impacts of treatments. However, we caution the reader not to over-interpret the results for two reasons. First, not all assets have significant investment levels. Across all stages, the median investment in A is \$3, with \$4 in B and \$2 in D. The rest of the assets have \$0 median investment overall. The data here also do not control for other variables and interactions that affect allocations. In section 4.5, we present a regression analysis that controls for these factors. A number of differences that appear significant here, do not survive the regression analysis. We focus our discussion here on assets with significant investment levels and those with significant differences also show up in the regression analysis.

Panel A shows that average investment is greatest in assets A and B (\$7.32 combined in Trial

1), somewhat less in D and E (\$2.51 combined) and lowest in the dominated assets, C and F (\$0.88). Average investment in Cash in Trial 1 was \$0.79. Thus, the lion's share of investment is in dominated assets with a lean toward the risk neutral and risk averse investments.

Panel B shows changes across trials for uncategorized treatments. In stage 2, investment increases in A and B, both popular investments in Stage 1 and both rated 3 stars. Investment also falls in D, the other popular investment in Stage 1, but ranked 2 stars. There are statistically significant investment drops in C, E and F, but investment levels in both trials are small. Without controlling for other factors, drops appear in B and D in Trial 3. However, controlling for the subject's self-rankings, the drop in B no longer remains significant. While there is a statistically significant drop in E, investment is very low in both trials. Finally, in Trial 4, there appear to be several significant differences. However, none survive the regression analysis that controls for other factors.

Panel C shows changes across trials for categorized treatments. In stage 2, investment increases in A and D, both popular investments in Stage 1 and both rated 3 stars. Investment falls in B, the other popular investment in Stage 1, but ranked 2 stars. There are statistically significant investment drops in C and F, but investment levels in both trials are small. In Trial 3, B drops relative to Trial 1. The regression analysis shows this is driven entirely by categorized treatments where subjects are told how to rank the assets. Again, there are statistically significant drops in C, E and F, but each is very low in both trials. Again, in Trial 4, there appear to be several significant differences, but only a decrease in C survives controlling for other factors.

The main result, in Panel D, is that the change in investment is consistent with the rating change due to categorization. Subjects invest more when the rating is higher (D and E) and less when it is lower (B and C). There is a increase in A at the 95% level of confidence, which is consistently rated 3-star. But, there is no change in F, which is consistently rated 1-star. Again, the apparent changes in Trial 4 do not survive the multivariate analysis except a difference in D between the uncategorized and categorized treatments.

The overall results in Table 4 are always consistent with the ratings effect, but not always consistent with the experimenter demand effect. The ratings effect predicts that, in uncategorized treatments, subjects will allocate more to B and C and less to D and E than in categorized treatments. The Panel D, Trial 2 data shows significant differences in the predicted direction in all cases. The experimenter demand effect predicts increases in A in both treatments, B in uncategorized treatments and D in categorized treatments. The data supports all three predictions. It also predicts decreases in F in both treatments, E in uncategorized treatments and C in categorized treatments. All hold except for F categorized treatments. Finally, the experimenter demand effect predicts no change in C and D in uncategorized treatments and B and E in categorized. Here, there are significant unpredicted changes. In uncategorized treatments, C increases and D falls significantly. In categorized treatments, B falls significantly.

4.4 Cash Holdings

As a consistency check, we examine investment in cash, for which positive holdings are never optimal. Across the 4 trials, 776 subjects (62.4%) never hold cash. An additional 382 subjects (30.7%) have average cash holdings of less than \$1 per trial.[25] Thus, 93.1% of the subjects essentially hold no cash. At the other extreme, 7 subjects (0.6%) only hold cash throughout the experiment and 3 more subjects (0.2%) exclusively holds cash in three trials and 13 more (1.1%) in two trials. The remaining subjects who hold cash at some point do so only once.

Cash-holding peaks in Trial 3, when subjects are asked to provide ratings. 142 subjects (11.4%) in this case invest only in cash, with 78 of those in the treatment that did not explain the rating rule. We have no definitive explanation for this behavior, but it seems likely that asking subjects to actively engage in analysis created poorer outcomes for a significant subset.[26]

We conclude that the vast majority of subjects understand the suboptimality of cash. It is most extensively used when subjects have to assign ratings, which presumably creates a significant cognitive load.

4.5 Regression Analysis of Investment in All Trials

We now use regression to examine investment behavior in more detail. We want to see how subject characteristics affect performance and whether subject knowledge mediates treatment effects. For

[25] Table A3 shows cash holdings broken down by trial and whether the subject was in the group that was told the rating rule.

[26] Subjects responded to a satisfaction survey after each trial and were asked if they wished they could change their decision. Of the 138 subjects investing only in cash in the Self-Rated trial, 7 (5.1%) said they wished they could change their decision. This compares to 2.1% of subjects who were not invested entirely in cash. We also note that 94.9% of the subjects who invested entirely in cash and said they wished they could change their allocation were in the treatments where the ranking rule was not given to them.

each asset, we run a regression across all trials controlling for subject characteristics and treatments:

$$I_{ijk} = \alpha_i + \beta_1 X_j + \beta_2 T_{k=2} X_{2,j} + \beta_3 T_{k=3} X_{3,j} + \beta_4 T_{k=4} X_{4,j} + \epsilon_{i,j} \tag{1}$$

where X_j is a vector of subject characteristics and $X_{k,j}$ is a vector of controls specific to each trial. $T_{k=n}$ is a dummy variable that takes the value 1 in Trial n. We estimate a censored regression to account for investment amounts being between 0 and 12.[27] We calculate robust standard errors, clustered by subject. Because there is effectively no treatment in Trial 1, we control for subject characteristics and, as a placebo, categorization. In trials 2-4, results depend both upon characteristics and treatments. Equation (1) should be viewed as an elaboration upon Table 4.

To summarize the main results, we find that more knowledgeable and experienced subjects behave more rationally in the baseline trial and that taking the initial bet is associated with holding riskier investments. In later trials, subjects are affected by the categorization treatment, they invest in accord with the ratings they assign, and knowledge and experience do not significantly mitigate the effect of treatments. Hence, the ratings effect alternative in Hypothesis 3 appears to hold regardless of knowledge and experience. Subjects who take the risk preference gamble invest significantly less is low risk assets A and B and more in high risk assets D and E, supporting the idea that the risk preference bet effectively measures risk preferences in context. There are surprisingly few significant interaction effects. Finally, the categorization effects largely disappear when the ratings are removed in Stage 4, giving some support to the transitory null in Hypothesis 4.

Explanatory variables, for which summary statistics are provided in Table 5, include:

Gender Dummy variable, female equals one.

Knowledge Total correct answers for the nine questions on the knowledge test, normalized to be between 0 and 1, less the mean across all subjects.

Experience Experience index, normalized to be between 0 and 1, less the mean across all subjects.

RiskBet Dummy variable for taking the initial bet.

[27] Investments range from $0 to $12 and are discrete, occurring in one dollar increments. As a robustness check, we also estimated ordered probit models and got essentially identical results. We present the censored regression models because they are easier to interpret.

Table 5: Summary statistics for regression variables. "Female" is a dummy variable where men are 0, women are 1; "Experience" and "Knowledge" are both based on a survey response scaled from 0 to 1 and normalized to have mean 0; "RiskBet" is a dummy variable for taking the initial risky bet; "Cat" and "Rule" are dummy variables for being in the treatments that are categorized and told the rating rule; and "Selfrank A - F" are the stars assigned by subjects minus the number of stars those assets have in the non-categorized treatment.

	Mean	Std. Dev.	Max	Min
Female	0.619	0.486	1.000	0.000
Experience	0.000	0.189	0.718	-0.211
Knowledge	0.000	0.189	0.258	-0.742
RiskBet	0.542	0.498	1.000	0.000
Cat	0.502	0.500	1.000	0.000
Rule	0.500	0.500	1.000	0.000
Selfrank A	-0.560	0.725	0.000	-2.000
Selfrank B	-0.525	0.677	0.000	-2.000
Selfrank C	-0.433	0.651	1.000	-1.000
Selfrank D	0.387	0.669	1.000	-1.000
Selfrank E	0.782	0.639	2.000	0.000
Selfrank F	0.349	0.712	2.000	0.000

Cat Dummy variable for the treatment with categorization.

Rule Dummy variable for the treatment in which the subject is told the rating rule.

SelfRank Rating assigned by the subject in the third trial for a given asset, less the uncategorized rating (Panel B in Figure 1).

Note that we have a choice in equation (1) about specifying subject characteristics. We can either allow their effects to vary individually in each trial or include them in the regression without conditioning on the individual trials. We include them without conditioning, so they are interpreted as providing the baseline across all trials. In some cases we interact characteristics with treatment dummies to see if knowledge (for example) alters the effect of the treatment. We discuss trial-specific effects of characteristics where relevant.

We present results separately pertaining to each trial rather than presenting the entire regression in one table. Two points relating to statistical significance are worth noting. First, because of the large number of coefficients in the regressions, we expect to see some significant coefficients at lesser significance levels as a result of random variation. We try not to over-interpret these cases. Second, across trials, more than 74% of subjects invest zero in assets C and F. Asset E also has significant zero investment. This is rational, but it means that the regression coefficients in those cases load on the small number of subjects who do invest. Idiosyncratic behavior may be over weighted. We report results for all assets, but focus our attention on A, B, and D.

4.5.1 Effects of Subject Characteristics in Trial 1

Table 6 presents those regression coefficients in equation (1) which are not multiplied by a dummy for Trials 2 through 4.[28] These coefficients are thus approximately the unconditional effect of characteristics, with one placebo variable, a dummy for the effect of categorization in Trial 1. The coefficients in all cases are computed conditional on the observation not being censored.

Table 6 shows that subjects responded to the economic characteristics of the investments, and departed in ways consistent with the personal characteristics including knowledge and experience. Assets A, B, and D have significant positive intercepts. The other assets have significant negative intercepts. The censored regression allows the independent variables to explain the non-censored variation in subject holdings.[29] Uncensored observations for C, E and F are relatively few, so the reader should use caution interpreting the results for these assets. Subjects who took the initial bet invested \$0.45, \$1.06 and \$1.43 more in the high risk assets, D, E,and F, respectively and \$0.35, \$0.85 and .06 less in the low risk assets A, B and C, respectively. A higher knowledge score is associated with increased investment in A and B and reduced investment in C, E, and F. Specifically, one additional correct question on the knowledge test was associated with a \$0.46 (= 4.10/9) increased investment in B. For all but asset D, knowledge and the risky bet had offsetting signs.Finally, subjects with more self-reported investment experience invested more in B and less in C, D, E and F. The coefficients can be interpreted as the effect of going from no experience to maximum experience.

Women invested less in B, but otherwise, gender effects were not evident (the investment was distributed across all the other assets, but no individual coefficient was significant). As a test of a

[28]Subject characteristics are effectively repeated for each trial, so there are 4,904 observations in each regression from 1,226 subjects (for whom we have all the information needed to run the regression. Clustered standard errors correct for this repetition. The alternative of a fixed effect model cannot be used because each subject only participates in one treatment.

[29]Because the regressions are censored, coefficients need not sum to zero across the assets, and the intercepts do not sum to 12 as they would with simple OLS.

Table 6: Unconditional censored normal regression coefficients for Equation (1) for subject characteristics. The explanatory variables are a gender dummy, where 1 denotes female (Gender); a dummy for having taken the initial bet (RiskBet); a dummy for the subject being in the categorized treatment (Cat); the knowledge score normalized to have a mean of 0 (Knowledge); and the Experience score, normalized to 0 (Experience). Responses were left-censored at 0 and right-censored at 12. Standard errors (in parentheses) are clustered by subject.

	A	B	C	D	E	F
Intercept	2.39**	5.16**	-2.28**	0.83**	-1.31**	-3.02**
	(0.22)	(0.24)	(0.22)	(0.19)	(0.24)	(0.27)
T1*Cat	-0.26	0.16	-0.10	0.37	-0.43*	-0.05
	(0.22)	(0.23)	(0.18)	(0.19)	(0.21)	(0.21)
Female	0.20	-0.58**	0.14	0.04	0.19	0.12
	(0.21)	(0.21)	(0.18)	(0.17)	(0.20)	(0.21)
Experience	0.14	2.09**	-1.89**	-1.87**	-1.71**	-2.25**
	(0.61)	(0.64)	(0.51)	(0.50)	(0.62)	(0.63)
Knowledge	1.79**	4.10**	-2.47**	0.47	-1.97**	-3.12**
	(0.51)	(0.54)	(0.43)	(0.46)	(0.50)	(0.51)
Risk Bet	-0.35*	-0.85**	1.06**	0.45**	1.06**	1.43**
	(0.18)	(0.18)	(0.16)	(0.16)	(0.18)	(0.19)
Num. Obs.	4,904	4,904	4,904	4,904	4,904	4,904
Trial 1:						
Left-censored	349	150	898	433	720	930
Uncensored	848	956	328	781	498	295
Right-censored	29	120	0	12	8	1
All Trials						
Left-censored	1,317	765	3,697	1,837	3,207	3,874
Uncensored	3,460	3,755	1,205	3,022	1,673	1,026
Right-censored	127	384	2	45	24	4

**$p < 0.01$, *$p < 0.05$

pure presentation effect, we included a dummy for the subject being in the categorized treatment. This dummy has significance at the 5% level for E.

4.6 The Effect of Star Ratings in Trials 2 and 3

The second and third trials exposed subjects to star ratings. Two aspects of these trials are especially important. First, in Trial 2, where subjects are shown ratings, the question is whether knowledge mediates the effect shown in Table 4. Second, in Trial 3, in which subjects assign star ratings to the investments, we impose the constraint that they assign equal numbers of one, two, and three star ratings, either within categories or not, depending on the treatment. The question is whether this active participation affects investment decisions.

4.6.1 Trial 2

Table 7 shows the coefficients from regression equation 1 for the trial in which subjects are given ratings, with half of the subjects told how the assets are rated. There are three important results in the table.

First, categorization affects investment decisions. Subjects exposed to categorization reduce significantly investment in assets B and C, both of which have one less star when categorized. In contrast, investment in D, which has one more star when categorized, increases significantly. To interpret this finding, recall from Table 4 that initial investment was greatest for B, A, and D, in that order, with average investment less than \$1 for C, E, and F. We would expect effects in B and D if anywhere. These are economically significant. The decrease in B is more than 30% of the baseline intercept while the increase in D is 175% of the baseline. In contrast, A is rated the same in both treatments and there is no significant investment difference between the two. The same holds for F, the other asset rated the same across treatments.

Second, Trial 2 allows us to differentiate between the experimenter demand effect and the ratings effect after controlling for other factors. The ratings effect is a prediction across caegoriz treatments. For assets with a maturity of uncensored observations (A, B and D), results are consistent with all predictions (no change in A, negative categorization effect in B and positive in D). The ratings effect is about investment levels relative to Trial 1 without rankings. Here, the main differentiator is asset B. The demand effect forecasts an increase increase in uncategorized treatments and no

Table 7: Regression coefficients for Equation (1) pertaining to Trial 2, denoted by the dummy variable T2. The intercept from Table 6 is duplicated for reference. The explanatory variables for trial 2 are a dummy for being in the categorization treatment (Cat); a dummy for being in the treatment where the rating rule is provided (Rule); and the knowledge score normalized to have a mean of 0 (Knowledge). Responses were left-censored at 0 and right-censored at 12. Standard errors (in parentheses) are clustered by subject.

	A	B	C	D	E	F
Intercept	2.39**	5.16**	-2.28**	0.83**	-1.31**	-3.02**
	(0.22)	(0.24)	(0.22)	(0.19)	(0.24)	(0.27)
T2	0.47*	0.36	0.11	-0.24	-1.36**	-0.64**
	(0.19)	(0.19)	(0.16)	(0.17)	(0.22)	(0.20)
T2*Cat	0.23	-1.60**	-0.46*	1.45**	0.36	0.16
	(0.21)	(0.23)	(0.19)	(0.19)	(0.23)	(0.24)
T2*Rule	0.07	0.01	-0.02	-0.34	0.40	0.31
	(0.21)	(0.23)	(0.19)	(0.19)	(0.23)	(0.24)
T2*Cat*Knowledge	0.84	-0.98	0.37	0.40	-0.80	-0.54
	(0.88)	(1.02)	(0.87)	(0.78)	(0.87)	(0.97)
T2*Rule*Knowledge	-1.39	0.75	-0.73	-0.66	-0.39	-0.39
	(0.89)	(1.03)	(0.90)	(0.79)	(0.90)	(0.96)
Num. Obs. (trial)	1,226	1,226	1,226	1,226	1,226	1,226
Left-censored	279	199	914	422	830	978
Uncensored	920	932	311	791	389	247
Right-censored	27	95	1	13	7	1

**p < 0.01, *p < 0.05

change in categorized. The opposite holds.

Third, financial knowledge does not alter the effect of the categorization treatment. Knowledge is interacted with the Trial 2 dummy individually and interacted with the category and rating rule treatment dummies. While more knowledgeable investors in the first trial invest more in B, knowledge does not mediate the effect of categorization. This remains true when performing F-tests on sums of coefficients.[30]

fourth, there are no significant effect associated with subjects being told the rating rule.

4.6.2 Trial 3

Subjects are asked to provide their own ratings while choosing investment amounts. This raises two questions:

[30] A chi-squared test showed no additional explanatory power from adding a dummy for the rating rule interacted with categorization and knowledge.

1. Did subjects assign reasonable ratings?

2. Did subjects invest in accord with the ratings they assigned?

If subjects invest in accord with their ratings, then we should generally see higher ratings correspond with more investment. Subjects in the categorized treatment, however, may be forced to assign a rating which conflicts with their investment (e.g., rating B at 2 stars). We find that self-assigned ratings are strongly associated with investment behavior, except when we force a conflict.

Because most investment is in A and B, we initially focus the discussion on those assets.

Subject ratings Subjects assign ratings that generally correspond with those in Table 1. However, they are also affected by treatments in unsurprising ways.[31] For example, subjects consistently rated A highly except in the categorized treatment when not told the rule used to construct the ratings. In that case, they tended to give the top rating to B.[32] As expected, categorization reduces the correspondence between subject-assigned ratings and the objective ratings. Overall, when given the rating rule, more than 2/3 of subjects ranked the assets in accord with the rule based on the ratio of return to risk.

Table 8 reports the results of a Probit regression analysis of the self-ratings. The table shows that categorization has a large effect on ratings, reducing the ratings assigned to A, B and C and increasing the rating assigned to D and E. Knowing the rating rule also has an effect, pushing down the ratings on B and increasing them for C and F, with the effect for F mitigated in the categorized treatment. Knowing the rule pushes up A in the categorized treatment only, as one would expect. Knowledge and/or experience increased ratings on A and B, reduced ratings on E and F, while leaving C an D unchanged. There is also some interaction between knowledge and categorization, with A and B falling and D and E rising relative to the uncategorized treatment. Subjects who take the risk bet rank low risk Assets A and B lower and the high risk assets E and F higher as we would expect. On the whole, subjects rate investments reasonably, given the restrictions we impose.

[31] Table A4 in the Appendix shows the fraction of subjects assigning each rating in each of four treatments (categorization by the rank rule being given).

[32] Panel A of Table A4 also suggests that some subjects were confused about ranks, and gave A a 1 and C a 3.

Table 8: Ordered Probit regressions explaining subject-assigned rank in Trial 3. The dependent variable is the subject-assigned rank minus 2. The explanatory variables are a gender dummy, where 1 denotes female (Female); a dummy for having taken the initial bet (RiskBet); a dummy for the subject being in the categorized treatment (Cat); a dummy for the subject receiving the rating rule (Rule), and variables denoting the knowledge score normalized to zero (Knowledge) and the experience score normalized to zero (Experience). Standard errors are in parentheses.

	A	B	C	D	E	F
Average Rank	2.44	2.48	1.57	2.39	1.78	1.35
Female	-0.11	-0.16*	0.17*	-0.11	0.06	0.20*
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)
Experience	0.57**	0.15	-0.22	0.26	-0.44*	-0.46
	(0.22)	(0.23)	(0.22)	(0.22)	(0.21)	(0.26)
Knowledge	0.81*	1.04**	-0.18	0.02	-0.42	-1.26**
	(0.37)	(0.40)	(0.35)	(0.35)	(0.37)	(0.45)
RiskBet	-0.28**	-0.16*	0.10	-0.13	0.25**	0.31**
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
Cat	-0.55**	-0.72**	-0.95**	1.05**	0.97**	0.08
	(0.10)	(0.11)	(0.10)	(0.10)	(0.10)	(0.12)
Rule	-0.12	-0.48**	0.29**	-0.17	0.06	0.34**
	(0.10)	(0.11)	(0.09)	(0.09)	(0.10)	(0.12)
Cat*Rule	0.61**	-0.14	0.01	-0.16	0.01	-0.19
	(0.14)	(0.14)	(0.14)	(0.14)	(0.13)	(0.16)
Cat*Knowledge	-1.07**	-0.96*	-0.69	1.29**	1.07**	0.20
	(0.36)	(0.38)	(0.37)	(0.36)	(0.36)	(0.43)
RiskBet*Knowledge	-0.03	0.15	0.17	-0.08	-0.70	0.61
	(0.38)	(0.38)	(0.38)	(0.37)	(0.37)	(0.45)
Num. Obs.	1,226	1,226	1,226	1,226	1,226	1,226
Psuedo-R2	0.04	0.10	0.10	0.10	0.11	0.04
Cut 1	-1.59	-2.26	-0.07	-1.13	0.24	1.31
Cut 2	-0.69	-1.08	1.31	0.23	2.03	1.60

**p < 0.01, *p < 0.05

Table 9: Regression coefficients for Equation (1) pertaining to Trial 3. The intercept from Table 6 is duplicated for reference. The explanatory variables are a Trial Dummy (T3), the rank the subject assigned the asset in the trial, a dummy for the subject being in the categorized treatment; and a dummy for the subject being in the treatment where the rank rule is given. Responses were left-censored at 0 and right-censored at 12. Standard errors (in parentheses) are clustered by subject.

	A	B	C	D	E	F
Intercept	2.39**	5.16**	-2.28**	0.83**	-1.31**	-3.02**
	(0.22)	(0.24)	(0.22)	(0.19)	(0.24)	(0.27)
T3	1.01**	-0.27	-0.33	-0.84**	-2.43**	-1.61**
	(0.27)	(0.25)	(0.22)	(0.21)	(0.35)	(0.31)
T3*SelfRank	3.11**	1.31**	0.83*	2.15**	1.87**	1.78**
	(0.37)	(0.47)	(0.35)	(0.34)	(0.43)	(0.34)
T3*Cat	0.11	0.70	0.27	0.90*	0.06	0.69
	(0.41)	(0.39)	(0.46)	(0.42)	(0.81)	(0.39)
T3*Rule	-0.10	-0.17	0.38	0.09	1.03*	1.29**
	(0.38)	(0.34)	(0.30)	(0.29)	(0.44)	(0.40)
T3*Cat*Rule	0.12	-1.43*	-0.50	-0.39	-0.79	-1.04
	(0.57)	(0.63)	(0.55)	(0.52)	(1.09)	(0.54)
T3*SelfRank*Cat	-0.55	0.31	-0.53	-1.69**	-0.80	-1.33**
	(0.56)	(0.62)	(0.55)	(0.50)	(0.81)	(0.46)
T3*SelfRank*Rule	-2.00**	-0.48	-0.67	-1.02	-0.80	-0.88*
	(0.47)	(0.55)	(0.53)	(0.59)	(0.50)	(0.42)
T3*SelfRank*Cat*Rule	0.95	-0.73	0.81	1.04	0.78	0.64
	(0.73)	(0.79)	(0.71)	(0.74)	(1.03)	(0.60)
Num. Obs. (trial)	1,226	1,226	1,226	1,226	1,226	1,226
Left-censored	386	247	945	498	849	991
Uncensored	805	902	280	720	374	235
Right-censored	35	77	1	8	3	0

**p < 0.01, *p < 0.05

Effect of subject ratings on investment Table 9 reports the Trial 3 coefficients from regression equation 1. The new explanatory variable is SelfRank, which is defined as the difference between the rating assigned by the subject and the non-categorized rating. For example, Asset B has a non-categorized rating of 3, so the self-rating is 0 if the subject assigns B the non-categorized rating of 3 stars, and -1 if a subject assigns it two stars. SelfRank is included as a standalone variable, and interacted with the categorization and rule dummies.

We will focus on Assets A, B, and D, which all have 720 or more uncensored observations in this Trial. The Trial 3 baseline investment levels are for uncategorized treatments in which subjects' self ratings equal the Trial 2 ratings without being given the rule. Investment levels are higher than in Trial 1 for A and lower for the high risk assets (D, E and F). A major factor driving investment levels are when subjects deviate from the Trial 2 ratings with positive deviations driving up investment and negative deviations driving it down. For subjects not in the categorized treatment (i.e., for whom Cat = 0) the magnitude of this effect ranges between \$0.83 and \$3.11 per assigned star. In the categorized treatments, the effects range from \$0.30 to \$2.56.

Next, we ask what happens when categorized subjects are asked to give 2 stars to B and 3 stars to D. Categorization forces subjects to give 2 stars to either A or B in the low-risk group. In itself, this makes little difference to either A or B investment levels (coefficient on T3*Cat). However, when subjects are specifically asked to rank B lower, investment in B falls significantly (coefficient on T3*Cat*Rule). When subjects are forced by categorization to give a 3-star ranking to either D, E or F, they usually give the 3 star ranking to D (81.4% of the time) and investment in D increases significantly (coefficient on T3*Cat). Asking them to rank D 3-stars through telling them the rule has little additional effect (coefficient on T3*Cat*Rule).

There are several other apparently significant effects. Several are for Assets E and F. With few investors making any positive investment, these may be over-fitted.[33] For A, there is an apparent negative relationship between investment and the subject's self ranking when asked to rate according to the ranking rule. But, this just attenuates the overall self ranking effect: adding coefficients on T3*Selfrank, T3*Rule, and T3*Selfrank*Rule gives +1.01, p-value=0.0693. For D, there is an apparent negative relationship between investment and the subject's self ranking when in categorized treatments. Again, this attenuates the overall self ranking effect: adding coefficients

[33] This is particularly apparent for F with only 235 uncensored observations.

Table 10: Regression coefficients for Equation (1) pertaining to Trial 4. The intercept from Table 6 is duplicated for reference. The explanatory variables are a Trial dummy (T4) and a dummy for the subject being in the categorized treatment. Responses were left-censored at 0 and right-censored at 12. Standard errors (in parentheses) are clustered by subject.

	A	B	C	D	E	F
Intercept	2.39**	5.16**	-2.28**	0.83**	-1.31**	-3.02**
	(0.22)	(0.24)	(0.22)	(0.19)	(0.24)	(0.27)
T4	0.36	0.08	-0.49	-0.53	-0.57	-0.47
	(0.27)	(0.30)	(0.27)	(0.26)	(0.30)	(0.32)
T4*Cat	-0.31	-0.39	-0.34	0.68**	0.14	0.03
	(0.22)	(0.22)	(0.19)	(0.20)	(0.23)	(0.23)
T4*Female	0.36	-0.39	0.40*	0.09	0.02	0.22
	(0.20)	(0.21)	(0.19)	(0.21)	(0.23)	(0.23)
T4*Experience	-0.32	0.20	1.34	0.68	-0.44	0.34
	(0.58)	(0.59)	(0.70)	(0.58)	(0.66)	(0.66)
T4*Knowledge	1.28*	0.16	-1.04	-0.82	-1.57**	-1.06
	(0.57)	(0.57)	(0.58)	(0.54)	(0.60)	(0.61)
T4*Risk Bet	0.20	0.20	-0.34*	-0.08	-0.31	-0.23
	(0.18)	(0.19)	(0.16)	(0.18)	(0.20)	(0.22)
Num. Obs. (trial)	1,226	1,226	1,226	1,226	1,226	1,226
Left-censored	303	169	940	484	808	975
Uncensored	887	965	286	730	412	249
Right-censored	36	92	0	12	6	2

**p < 0.01, *p < 0.05

on T3*Selfrank, T3*Cat, and T3*Selfrank*Cat gives +1.36, p-value=0.0000.

4.6.3 Trial 4

Table 10 reports the coefficients from regression equation 1 for the final trial, which is an untreated repeat of the first. As such, the Trial 4 regression variables measure the effects from subjects having participated in the previous trials. The results are generally not significant. There is an increase in investment in D in categorized treatments. More knowledgeable investors reduce investment in E and increase it in A. Subjects who took the risk preference bet invest a little less in C. But, in contrast to Trials 2 and 3, there are no consistent patterns of differences between Trial 1 and Trial 4.

4.7 Overall Investment Performance

The primary focus of the preceding analysis has been on investment in individual assets. Table 11 examines overall portfolio performance as measured by the portfolio Sharpe ratio. The analysis confirms that categorization harms performance. The table presents regressions explaining the Sharpe ratio and its components in Trial 1, and the difference between the Sharpe ratio in later trials and that in Trial 1.

In Trial 1, more experienced and knowledgeable subjects have a higher expected return. Women have a lower expected return. Subjects who take the initial risk preference bet take on more risk and more knowledgeable investors take on somewhat more risk as well. The net effect is that more experience and knowledgeable subjects have higher Sharpe ratios, while those who take the initial bet have lower Sharpe ratios. There is no significant gender effect on the Sharpe ratio.

Categorization creates the only significant incremental effect in Trial 2: reducing Sharpe ratios. Knowledge and experience do not attenuate this effect. In Trial three, the effect disappears. So, there is only an impact when the rankings are given to subjects, not when they rank assets themselves (even if they are told to give the same rankings as in Trial 2). Finally, more knowledgeable subjects and those who took the initial bet increase their Sharpe ratios somewhat in Trial 4 relative to Trial 1. This may be due to the secular decline in holdings of Asset E evident in Table 4. Interestingly, this only occurs in Trial 4 and not 3 when subjects are presumably spending more time considering the tradeoffs inherent in the assets.

5 Related Literature

This paper is related to several distinct areas of the behavioral literature. Our goal in this section is to provide some overarching context and relate results in the literature to our findings. As discussed in the introduction, we discuss the relation of our findings to work on presentation effects, the effects of financial knowledge, cognitive limitations, and difficulties making portfolio decisions.

5.1 Presentation Effects

Subjects have full information about investments. Why would they respond to ratings? There are at least two obvious possibilities. One is that subjects believe the ratings convey additional

Table 11: Censored normal regressions, the dependent variable is the expected return, return standard deviation (one-half the range), and the Sharpe ratio (the expected return divided by the standard deviation), for Trial 1 and across trials 2 through 4. An average Sharpe ratio could not be computed for subjects who in at least one trial invested $12 in cash.

	Trial 1			Sharpe Ratio, Trials 2 - 4		
	Exp Ret	Std Dev	Sharpe	T2 - T1	T3 - T1	T4 - T1
Intercept	0.852**	0.792**	1.092**	0.042**	0.007	0.030
	(0.010)	(0.012)	(0.013)	(0.016)	(0.017)	(0.018)
Female	-0.034**	-0.019	-0.010	0.004	-0.002	0.003
	(0.011)	(0.014)	(0.015)	(0.014)	(0.015)	(0.016)
Experience	0.089**	-0.033	0.133**	-0.038	-0.012	-0.044
	(0.031)	(0.038)	(0.041)	(0.038)	(0.040)	(0.044)
Knowledge	0.272**	0.078*	0.204**	0.057	0.021	0.120*
	(0.029)	(0.036)	(0.040)	(0.049)	(0.054)	(0.057)
RiskBet	-0.016	0.077**	-0.103**	0.014	0.036*	0.050*
	(0.010)	(0.012)	(0.013)	(0.017)	(0.018)	(0.020)
Cat				-0.049**	-0.015	-0.039
				(0.019)	(0.020)	(0.022)
Rule				0.007	0.014	-0.003
				(0.012)	(0.013)	(0.014)
RiskBet*Cat				0.017	-0.009	0.003
				(0.024)	(0.026)	(0.028)
Knowledge*Rule				0.024	0.046	-0.003
				(0.065)	(0.072)	(0.077)
Knowledge*Cat				-0.038	0.126	0.219
				(0.190)	(0.214)	(0.236)
num. Obs.	1,226	1,226	1,216	1,200	1,089	1,197
Left-censored	10	10	4	1	-	2
Uncensored	1,096	1,208	1,178	1,167	1,050	1,152
Right-censored	120	8	34	32	39	43
All cash in one of the periods in regression	10	10	10	26	142	29

**p <0.01, *p <0.05

information. In this case we would expect financial knowledge to mitigate the effect of categorized ratings. There is no significant mitigation from knowledge. Further, we would expect that explaining that ratings are based on the Return/Risk Ratio would mitigate the effect. To the degree that rule interactions are significant in Trials 2 and 3, explaining that the rating rule seems to reinforce the response to ratings.

A second possibility is that experimental subjects may respond to star rating here because they wish to please the experimenter, a response known as the experimenter demand effect (Zizzo, 2010). In our experiment, this behavior has real consequences, as blind adherence to the rating lowers the expected payoff in some treatments. The question of whether subjects follow guidance against their financial interest, however, is precisely the point of this study. This mirrors the possibility that, in practice, investors invest in accord with ratings (for example) simply because those ratings are supplied by an advisor or a ratings firm. We conclude that ratings affect performance.

The response to ratings relates to the general question of how subjects respond to presentation of information. An investor making an investment decision will examine advertisements, literature, reports, and possibly data about investment products and strategies. There is evidence that presentation of this information, including ratings, affects investment decisions for reasons unrelated to the economic characteristics of the investment.

Several papers look explicitly at the effects of ratings:

- Chen et al. (2014) examine split-rated bonds (i.e., those where rating agencies disagree about the bond rating) for which a bond index redefinition changed their status as investment grade. They show that bond yields changed in response to this purely mechanical redefinition. Because there was no new information about the bond, this real-world experiment resembles our assignment of stars, and, as with our experiment, finds real effects.

- In a related vein, Del Guercio and Tkac (2008) and Reuter and Zitzewitz (2021) examine whether equity mutual fund flows respond to Morningstar rating changes. Del Guercio and Tkac (2008) find that an increase in the star rating of a fund is accompanied by an inflow of funds, even after controlling for historical return characteristics. Reuter and Zitzewitz (2021), in a study of size and performance, also document a relationship between the rating and flows. Our findings are broadly consistent with theirs, but the settings are different.

Because they use aggregate investment data, neither study can distinguish between existing investors changing their holdings in response to a ratings change (the intensive margin) and new investors attracted to a newly higher rating (the extensive margin). For the same reason, neither can examine investor heterogeneity. Our study examines the intensive margin, examining subjects in a given treatment who either see star ratings or not, and then comparing subjects across treatments. With both papers there is also the possibility that investors believe ratings convey real information about expected returns.[34] In this regard, the findings in Chen et al. (2014) are more purely about rating changes. Finally, responses to ratings changes for a given fund does not address whether investors are misled by cross-category comparisons of ratings.

- Our experiment is complementary to that in Bateman et al. (2016), who vary the presentation of investment characteristics when asking subjects to rank portfolios. We leave the presentation unchanged, but find effects nevertheless.

Other papers consider the number of funds offered as an element of presentation.

- Benartzi and Thaler (2001) posit that investors choosing funds in a retirement account naively diversify using a $1/N$ rule (investing equal amounts in offered assets). Huberman and Jiang (2006), however, find that investors are less inclined to diversify naively as the number of investments increases.
- More recently, Massa et al. (2015) examine a variant of the 1/n rule and find that investors invest more in a style (e.g., Asian funds) when there are more funds represented in that style.

We do not alter offerings and have the same number of offerings in each category. However, even in our simple context with a relatively small number of funds, we find no evidence for a 1/n rule. However, subjects do invest across more assets than is optimal.[35]

[34] Del Guercio and Tkac (2008) report that investing only in five-star funds produces a positive risk-adjusted excess return in their sample, making it difficult to rule out the possibility that Morningstar rankings do convey information. See also Blake and Morey (2000).

[35] Other papers that look at presentation effects include (Weber et al., 2005; Ibrekk and Morgan, 1987); see also discussions in Tversky and Kahneman (1981, 1986). Anderson and Settle (1996) show that both the form and amount of information affect choices.

5.2 Financial Knowledge

We find that subjects who are more financially literate preform better in the baseline trial, but are still affected by categorization. The effect of financial literacy on performance is consistent with results in numerous papers. Among papers demonstrating positive effects of financial literacy are Bernheim et al. (2001), Bernheim and Garrett (2003), and Lusardi and Mitchell (2007). In an experiment, Anderson and Settle (1996) also show significant effects of prior financial knowledge. The most direct evidence relating intelligence and financial performance is from Grinblatt et al. (2011), who find using Finnish data that a higher IQ is associated with greater stock market participation and a higher Sharpe ratio. Our results are consistent with this: subjects with higher knowledge invest in portfolios with significantly higher Sharpe ratios. However, higher knowledge does not appear to mitigate the ratings effect we document here.

5.3 Cognitive Limitations and Decision-making Biases

Numerous studies have shown that subjects have computational limitations and exhibit biases. Most relevant are studies finding that subjects can be affected by irrelevant information. Edgell et al. (1996) list a range of areas in psychology where irrelevant information affects behavior. Examples include sunk costs (Thaler, 1980), random information signals (Tversky and Kahneman, 1974) and extra information not pertinent to the decision (Camerer et al., 1989). Subjects also have trouble ignoring irrelevant alternatives. The failure of the independence axiom is well documented in general contexts (e.g., see Davis and Holt (1993)). In an investment environment, Herne (1999) shows how a "decoy" gamble can affect choices across gambles. Our results are similar, but distinct, from the prior research. Subjects are affected by something that should not affect behavior. However, in our case, ratings do not convey irrelevant information nor do they make the decision environment more complicated. Ratings are merely redundant, yet still effect choices.

5.4 Portfolio Choices

We find that subjects on average tilt towards optimal portfolio choices, but with considerable variation. In general, research mimicking real-world portfolio decision-making has found that subjects inappropriately extrapolate from historical data (Moore et al., 1999), have difficulty forming efficient portfolios, (Kroll et al., 2003), and do not properly account for asset correlation (Kroll et al., 1988). Our design removes many of these factors to focus entirely on ratings effects. However, our Trial 1 data does shed some light on these issues: in the absence of historical data when subjects do not need to account for correlation, subject choices tend toward relatively efficient investments, with the heaviest investments in A and B.

6 Conclusion

We study the effect of categorization and ratings on investment performance in an asset allocation experiment. Ratings are ubiquitous, and often most useful for comparisons across investments with similar characteristics. This gives rise to categorized ratings, used to compare within investment classes.[36] However, there is a tension between overall (i.e., uncategorized) ratings and categorized ratings. Uncategorized ratings explicitly create comparisons across investment classes. Categorized ratings, by using a common scale across categories, implicitly create comparisons across investment classes.

In our experiment, subjects perform well in a basic allocation task, but performance deteriorates for those facing categorized ratings. Knowledge and experience help with the basic task but do not mitigate the effects of categorized ratings. We observe that rating-induced shifts are most pronounced when the categorized rating conflicts with the uncategorized ranking of the investment (which we observe to be correlated with investor choices in the uncategorized treatment). Based on this, we think the likeliest explanation is that subjects face cognitive dissonance when star rankings disagree with the subjects' own initial evaluations. In our experiment, investment is reduced in a high quality asset receiving only two stars, but is unchanged in a high quality asset receiving three stars. The results are consistent with cognitive dissonance (subjects are only swayed when stars disagree with their own evaluation) but not with the experimenter demand effect (subjects do not invest more in a high quality asset that receives a high rating).

In the end it seems that both categorized and uncategorized ratings have drawbacks. The question is then how to present expert advice without requiring the decision maker to become an

[36] As noted in the introduction, ratings are common in realms other than finance. Wines with very different characteristics, for example, are rated on the same 100-point scale, creating implicit categorized ratings. One market response has been prepackaged wine sales in the form of mail-order wine clubs.

expert in order to make use of the advice. An alternative to explicit advice is implicit advice in the form of prepackaged solutions. In the last 15 years, for example, life-cycle (or target-date) funds have become increasingly popular as default investments in defined contribution pension plans (Mitchell and Utkus, 2012; McDonald et al., 2021). These are managed investments that provide diversification and a level of risk that automatically declines with age. Since life-cycle funds from different providers are in the same category, ratings could be of help in choosing among such funds.[37] However, a given life-cycle fund may not be optimal for a given individual, who might be better off selecting from a menu of highly-rated funds in different categories.

Ratings, life-cycle funds, and the retirement investment decision illustrate an ongoing tension in financial advice: specific solutions, which may be good for many but not necessarily optimal for any, versus detailed information, offered in the belief that investors can make their own decisions if appropriately informed. Our research shows that even in a simple, full-information environment, decision aids such as ratings can sway investors. Additional research is required to understand whether it is possible to design such aids so as to avoid adverse effects.

References

Akerlof, G. and W. Dickens (1982). The economic consequences of cognitive dissonance. American Economic Review 72(3), 307–319.

AM Best (2024). Guide to best's financial strength ratings - (fsr). https://www.ambest.com/ratings/guide.pdf.

Anderson, B. F. and J. W. Settle (1996). The influence of portfolio characteristics and investment period on investment choice. Journal of Economic Psychology 17(3), 343–358.

Bateman, H., L. I. Dobrescu, B. R. Newell, A. Ortmann, and S. Thorp (2016). As easy as pie: How retirement savers use prescribed investment disclosures. Journal of Economic Behavior & Organization 121, 60–76.

[37]McDonald et al. (2021) provide evidence that pension plan participants were likelier to accept a life-cycle fund default than a money-market fund default, with the latter inducing participants to change their contributions to acquire equity exposure.

Benartzi, S. and R. H. Thaler (1999). Risk aversion or myopia? choices in repeated gambles and retirement investments. Management Science 45(3), 364–381.

Benartzi, S. and R. H. Thaler (2001). Naive diversification strategies in defined contribution saving plans. American Economic Review 91(1), 79–98.

Benmelech, E. and J. Dlugosz (2009, July). The alchemy of CDO credit ratings. Journal of Monetary Economics 56(5), 617–634.

Benmelech, E. and J. Dlugosz (2010). The credit ratings crisis. NBER Macroeconomics Annual 24(1), 161–208.

Berg, J., J. Dickhaut, and K. McCabe (2005). Risk preference instability across institutions: A dilemma. Proceedings of the National Academy of Sciences of the United States of America 102(11), 4209–4214.

Berg, J. E., J. W. Dickhaut, and T. A. Rietz (2010). Preference reversals: The impact of truth-revealing monetary incentives. Games and Economic Behavior 68(2), 443–468.

Bernheim, B. D. and D. M. Garrett (2003, August). The effects of financial education in the workplace: Evidence from a survey of households. Journal of Public Economics 87(7-8), 1487–1519.

Bernheim, B. D., D. M. Garrett, and D. M. Maki (2001, June). Education and saving: The long-term effects of high school financial curriculum mandates. Journal of Public Economics 80(3), 435–465.

Blake, C. R. and M. R. Morey (2000). Morningstar ratings and mutual fund performance. Journal of Financial and Quantitative Analysis 35(03), 451–483.

Blume, M. E. (1998). An anatomy of morningstar ratings. Financial Analysts Journal 54(2), 19–27.

Camerer, C., G. Loewenstein, and M. Weber (1989). The curse of knowledge in economic settings: An experimental analysis. Journal of political Economy 97(5), 1232–1254.

Campbell, J. Y. (2006, August). Household finance. Journal of Finance LXI(4), 1553–1604.

Chadd, I., E. Filiz-Ozbay, and E. Y. Ozbay (2021). The relevance of irrelevant information. Experimental Economics 24(3), 985–1018.

Chen, Z., A. A. Lookman, N. Schürhoff, and D. J. Seppi (2014). Rating-based investment practices and bond market segmentation. The Review of Asset Pricing Studies 4(2), 162–205.

Davis, D. D. and C. A. Holt (1993). Experimental economics. Princeton Univ Pr.

Del Guercio, D. and P. A. Tkac (2008, December). The effect of Morningstar ratings on mutual fund flow. Journal of Financial and Quantitative Analysis 43(4), 907–936.

Eckel, C. C. and P. J. Grossman (2008). Men, women and risk aversion: Experimental evidence. Handbook of experimental economics results 1, 1061–1073.

Edgell, S., N. Castellan Jr, R. Roe, J. Barnes, P. Ng, R. Bright, and L. Ford (1996). Irrelevant information in probabilistic categorization. Journal of Experimental Psychology: Learning, Memory, and Cognition 22(6), 1463.

Grinblatt, M., M. Keloharju, and J. Linnainmaa (2011, December). Iq and stock market participation. Journal of Finance 66(6), 2121–2164.

Herne, K. (1999). The effects of decoy gambles on individual choice. Experimental Economics 2(1), 31–40.

Huberman, G. and W. Jiang (2006). Offering vs. choice in 401 (k) plans: Equity exposure and number of funds. Journal of Finance 61(2), 763–801.

Ibrekk, H. and M. G. Morgan (1987). Graphical communication of uncertain quantities to non-technical people. Risk Analysis 7(4), 519–529.

Kahneman, D., J. L. Knetsch, and R. H. Thaler (1990). Experimental tests of the endowment effect and the coase theorem. Journal of political Economy 98(6), 1325–1348.

Kahneman, D., S. P. Slovic, P. Slovic, and A. Tversky (1982). Judgment under uncertainty: Heuristics and biases. Cambridge university press.

Kroll, Y. and H. Levy (1992). Further tests of the separation theorem and the capital asset pricing model. American Economic Review 82(3), 664–670.

Kroll, Y., H. Levy, and N. Alisof (2003). “homemade leverage:” theory versus experimental evidence. Journal of Portfolio Management 31(1), 84–93.

Kroll, Y., H. Levy, and A. Rapoport (1988, June). Experimental tests of the separation theorem and the capital asset pricing model. American Economic Review 78(3), 500–519.

Lusardi, A. and O. S. Mitchell (2007, January). Baby boomer retirement security: The roles of planning, financial literacy, and housing wealth. Journal of Monetary Economics 54(1), 205–224.

Massa, M., A. Simonov, and A. Stenkrona (2015, March). Style representation and portfolio choice. Journal of Futures Markets 23, 1–25.

McDonald, R. L., D. P. Richardson, and T. A. Rietz (2021). Investment defaults and retirement savings allocations. Technical report, TIAA Institute/University of Iowa Working Paper.

Merton, R. C. (1971, December). Optimum consumption and portfolio rules in a continuous-time model. Journal of Economic Theory 3(4), 373–413.

Merton, R. C. (1973, September). An intertemporal capital asset pricing model. Econometrica 41(5), 867–887.

Mitchell, O. S. and S. Utkus (2012). Target-date funds in 401 (k) retirement plans. Technical report, National Bureau of Economic Research.

Moody’s Investors Services (2008, February). Moody’s credit policy – request for comment: Should moody’s consider differentiating structured finance and corporate ratings. Technical report, Moody’s Investors Services.

Moore, D. A., T. R. Kurtzberg, C. R. Fox, and M. H. Bazerman (1999). Positive illusions and forecasting errors in mutual fund investment decisions. Organizational Behavior and Human Decision Processes 79(2), 95–114.

Morningstar (2008). The morningstar rating for funds. Technical report, Morningstar, Inc.

Morningstar (2023). The morningstar category classifications (for portfolios available for sale in the united states). `https://www.morningstar.`

com/api-corporate/midway/v1/research/download/1180969?timestamp=1694716260000050&token=eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJkYXRhIjp7ImRvY3VtZW50SWQiOjExODA5Njl9LCJpYXQiOjE3MTI2OTUxMDd9.0J6uayx0uUhg880FJz5_YqvUO3k9FlJJbJfoZwjo9jE.

Oliven, K. and T. A. Rietz (2004). Suckers are born but markets are made: Individual rationality, arbitrage, and market efficiency on an electronic futures market. Management Science 50(3), 336–351.

Rabin, M. (2000, September). Risk aversion and expected-utility theory: A calibration theorem. Econometrica 68(5), 1281–1292.

Reuter, J. and E. Zitzewitz (2021). How much does size erode mutual fund performance? a regression discontinuity approach. Review of Finance 25(5), 1395–1432.

Sunden, A. E. and B. J. Surette (1998). Gender differences in the allocation of assets in retirement savings plans. American Economic Review 88(2), 207–211.

Thaler, R. (1980). Toward a positive theory of consumer choice. Journal of Economic Behavior & Organization 1(1), 39–60.

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. Science 185(4157), 1124–1131.

Tversky, A. and D. Kahneman (1981). The framing of decisions and the psychology of choice. Science 211(4481), 453–458.

Tversky, A. and D. Kahneman (1986). Rational choice and the framing of decisions. Journal of Business 59(4), 251–278.

Watson, E. (2017). Goodbye nuval… and good riddance? https://www.foodnavigator-usa.com/Article/2017/11/13/Goodbye-NuVal-and-good-riddance#.

Weber, E. U., N. Siebenmorgen, and M. Weber (2005). Communicating asset risk: How name recognition and the format of historic volatility information affect risk perception and investment decisions. Risk Analysis 25(3), 597–609.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics 13(1), 75–98.

Internet Appendices

A Additional Tables

Table A1 shows the order, by trial, in which investments are displayed.

Table A1: Order of display of assets in each trial

Trial	Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Assset 6
Initial Baseline	A	B	C	D	E	F
Rated	D	F	E	C	A	B
Self-rated	B	C	A	E	D	F
Final Baseline	F	E	D	C	B	A

Table A2 shows correlations for gender, taking the initial bet, knowledge, and experience.

Table A2: Correlations for subject characteristics. All correlations are significant.

	Female	Bet	Experience	Knowledge
Female	1			
(p-value)	#N/A			
Bet	0.1429	1		
	(0.000)	#N/A		
Knowledge	-0.2589	-0.1711	1	
	(0.000)	(0.000)	#N/A	
Experience	-0.4156	-0.0963	0.4497	1
	(0.000)	(0.001)	(0.000)	#N/A

Table A3 shows cash holdings across trials, broken down by whether subjects were in the treatment in which they were told the ranking rule.

Table A4 shows how subjects rated the assets in the different treatments.

Table A3: Cash holdings in each trial, split by whether subjects are told the rating rule in the self-rated trial. Trial numbers indicate the baseline (1), the rated trial (2), the self-rated trial (3), and the final baseline (4).

	Trial							
	Rating Rule Not Given				Rating Rule Given			
Cash Holding	1	2	3	4	1	2	3	4
0	492	496	448	489	482	500	452	986
1	24	26	19	19	24	18	21	35
2	35	33	30	36	38	34	33	69
3	20	15	15	21	19	20	13	42
4	17	18	16	17	26	17	18	32
5	7	6	4	6	9	7	4	15
6	12	11	9	10	11	9	10	19
7	2	2	2	7	2	1	1	8
8	3	1	0	2	1	2	2	5
9	1	0	1	1	3	0	2	1
10	2	0	0	2	1	2	1	5
11	1	0	0	0	2	2	1	1
12	6	14	78	12	4	10	64	26

Table A4: Fraction of subjects assigning a given rating in the self-ranked trial, by treatment. The ratings shown to subjects in the Ranked trial are in bold. Ratings of subjects in the categorized treatment are in Panel A, and ratings of subjects in the non-categorized treatment are in Panel B. There were 758 subjects in each treatment in Panel A, and 752 subjects in each treatment in Panel B.

Panel A: Categorized Treatment

	Rank Rule Not Given			Rank Rule Given		
Asset	1	2	3	1	2	3
A	0.100	0.552	**0.348**	0.140	0.187	**0.673**
B	0.074	**0.372**	0.554	0.077	**0.760**	0.164
C	**0.826**	0.077	0.098	**0.784**	0.053	0.164
D	0.114	0.058	**0.829**	0.172	0.074	**0.755**
E	0.058	**0.879**	0.063	0.045	**0.865**	0.090
F	**0.829**	0.063	0.108	**0.784**	0.061	0.156

Panel B: Non-Categorized Treatment

	Rank Rule Not Given			Rank Rule Given		
Asset	1	2	3	1	2	3
A	0.106	0.215	**0.678**	0.176	0.101	**0.723**
B	0.069	0.082	**0.848**	0.149	0.075	**0.777**
C	0.340	**0.609**	0.051	0.136	**0.830**	0.035
D	0.075	**0.673**	0.253	0.037	**0.846**	0.117
E	**0.577**	0.335	0.088	**0.713**	0.109	0.178
F	**0.832**	0.085	0.082	**0.790**	0.040	0.170

B Experimental Instructions

Here, we give the text of the Web pages used for this experiment for the baseline treatment. We footnote the differences between treatments. Separate instruction sets for each treatment are available by request. Web formatting differed slightly from the formatting below and the Web pages were interactive. Subjects used drop-down boxes to make allocation decisions and radio buttons to answer survey questions.

General Instructions

This is an experiment in the economics of decision making. The instructions are simple. If you follow them carefully and make good decisions, you might earn a considerable amount of money which will be mailed to you at the end of the experiment.

In stages of this experiment, you will be given cash that you can keep or you can invest all of it or portions of it in one or more investment alternatives. The amounts you invest in each alternative and some random draws will determine your payoffs from participating in this experiment. We will also ask you a series of questions about your preferences, knowledge and demographics.

——— Next Web Page ———

Descriptions of Investment Alternatives

At different stages in this experiment you will be given cash that you can (a) keep or (b) can invest all of it or portions of it in one or more investment alternatives. This is called an allocation decision. The amounts you invest will be in \$1 increments. Each investment alternative will result in either a high return or a low return. A random draw will determine which occurs. Half of the time, the high return will occur and half of the time the low return will occur. The cash you keep always earns a zero return.

For each \$1 invested in an alternative, you will be paid \$1 times 1 plus the return that occurs. Returns will be expressed in percentage terms. So, for example, if the return that occurs is 80%, you will be paid \$1x(1+0.80) = \$1.80 for each dollar invested in that alternative. That is, you get the original \$1 back plus 80% of a dollar, or 80 cents in additional return. If the return that occurs is -30%, then you will be paid \$1x(1+(-0.30)) = \$0.70. That is, you lose 30%, or 30 cents, of the original dollar invested. If you had invested \$2, you would have been paid \$2x(1+0.80) = \$3.60 or \$2x(1+(0.30)) = \$1.40 if the high or low return had occurred, respectively.

Each investment alternative will be described to you by giving the high return, the low return, the average return and the range of returns in a table. Finally, you will be given a “Return/Risk” ratio which is the expected return divided by the range. For example, an investment alternative, say, “Alternative X” that has a high return of high return of 150% and a low return of -50% would be described as:

Alternative:	X
High Return:	150%
Low Return:	-50%
Average Return:	50%
Range of Returns:	200%
Return/Risk Ratio:	0.2500

If you received the high return in this case, you would be paid \$1x(1+1.5) = \$2.50 for each \$1 invested. If you received the low return in this case, you would be paid \$1x(1+(-0.50)) = \$0.50 for each \$1 invested. Recall, that the high return will occur half of the time and the low return will occur half of the time. Which actually occurs is determined by a random draw. In any given stage, you may be asked to allocate cash across several investments.

In a stage, if the high return occurs for one investment, the high return will occur for all investments in that stage. If the low return occurs for one investment, the low return will occur for all investments in that stage.

——— Next Web Page ———

Descriptions of Investment Alternatives (continued)

To see if you understand an investment alternative, please answer the following three questions about the investment alternative “Y” described by:

Alternative:	Y
High Return:	210%
Low Return:	-80%
Average Return:	65%
Range of Returns:	290%
Return/Risk Ratio:	0.2241

1. If the high return occurs for "Y", what return will occur for other investments in that stage?

   (a) All other investments in that stage will return their respective "low" returns. *Feedback if selected: "Recall that, in a given stage, either ALL investments return their respective 'high' returns or ALL investments return their respective 'low' returns. Please try a different answer."*

   (b) All other investments will return their respective "high" returns. *Feedback if selected: "Correct. Proceed to question 2."*

   (c) The returns to other investments at that stage may be mixed with some "high" and some "low" returns. *Feedback if selected: "Recall that, in a given stage, either ALL investments return their respective 'high' returns or ALL investments return their respective 'low' returns. Please try a different answer."*

2. If you invest \$3 in this alternative and the high return occurs, how much will you be paid?

   (a) \$2.10 *Feedback if selected: "This is the gain on a \$1 investment, not the total amount paid to you. Please try a different answer."*

   (b) \$3.10 *Feedback if selected: "This is the payment for a \$1 investment. Recall you are investing \$3. Please try a different answer."*

   (c) \$6.30 *Feedback if selected: "This is the gain on the \$3 investment, not the total amount paid to you. Please try a different answer."*

   (d) \$9.30 *Feedback if selected: "Correct. Please continue to Question 3."*

3. If you invest \$2 in this alternative, and the low return occurs, how much will you be paid?

(a) -\$0.80 *Feedback if selected: "This is the loss on a \$1 investment, not the total amount paid to you. Please try a different answer."*

(b) \$0.20 *Feedback if selected: "This is the payment for a \$1 investment. Recall you are investing \$2. Please try a different answer."*

(c) -\$1.60 *Feedback if selected: "This is the loss on the \$2 investment, not the total amount paid. Please try a different answer."*

(d) \$0.40 *Feedback if selected: "Correct. Please continue to the next part of the experiment."*

The continue button only becomes active after the answers are correct.

——— Next Web Page ———

Preliminary Stage

In this stage you have \$1 to allocate. You can either keep it or invest it in the following investment alternative:

Alternative:	A		
High Return:	100%		
Low Return:	-100%		
Average Return:	0%		
Range of Returns:	200%		
Return/Risk Ratio:	0.0000		Amount Kept
Amount you Invest:			*=\$1-Invested*

When you are finished with your choice, click the "continue" button below. At the end of the experiment, a random draw will determine your payoff to this stage. This will become part of your earnings for the experiment.

——— Next Web Page ———

Additional Instructions

In the rest of the experiment, you will make several investment decisions. Each decision is a "stage" in the experiment.

In each stage, you will have $12 to keep or invest. You can keep the $12 or invest some or all of it in any of 6 investment alternatives in $1 increments. You can invest the $12 in whatever combination of investments and in whatever amounts you wish so long as the total investment is less than or equal to $12. Any amount you do not invest will be kept for that stage and earn a zero return. The outcomes of any given stage will not affect the amount you have to invest in the next stage. You will always start each stage with a new $12.

At the end of the experiment, ***we will randomly select one stage***. Then, we will make a SINGLE random draw that will determine whether the gambles in that stage pay the high return or the low return. That is, either ALL investments in that stage will give you a high return or ALL of the investments will give the low return.

Earnings from this stage and random draw, together with earnings from the preliminary stage and the $5 participation fee, will contribute to your earnings for the experiment.

Between stages and at the end of the experiment, you will be asked questions that will help us categorize your responses.

——— Next Web Page ———

Stage I

In this stage you have $12 to allocate. You can either keep it or invest it in the following investment alternatives in $1 increments.[38]

Alternative:		A	B	C	D	E	F		
High Return:		130%	185%	125%	200%	225%	190%		
Low Return:		30%	15%	-25%	-20%	-75%	-90%		
Average Return:		80%	100%	50%	90%	75%	50%		
Range of Returns:		100%	170%	150%	220%	300%	280%		
Return/Risk Ratio:		0.8000	0.5882	0.3333	0.4091	0.2500	0.1786		Amount Kept
Amount you Invest:									*=$12-Invested*

Please confirm your allocation before pressing continue.

[38] In treatments with categories, the table of alternatives in each stage was presented with three investments labeled "Category I" on the left and three labeled with "Category II" on the right with a blank column between them. The sentence "These alternatives have been categorized using a commonly used financial method." was added to the description.

When you are finished with your choice, click the "continue" button below. At the end of the experiment, we will randomly select one of these stages to play out. If this stage is chosen, a single random draw will determine the returns to ALL of the investment in this stage and, as a result, your payoff to this stage. This will become part of your earnings for the experiment.

——— Next Web Page ———

Stage I Questions[39]

Please answer the following questions. Click the "Submit" button below to continue on to the next stage.

1. How satisfied are you with your allocation decision in this stage?

Very Dissatisfied			Neutral			Very Satisfied
1	2	3	4	5	6	7

2. Do you wish you could go back and change your decision?
   (a) Yes
   (b) No

3. How confused did you feel while making the allocation decision?

Very Confused			Neutral			Not Confused At All
1	2	3	4	5	6	7

——— Next Web Page ———

Stage II

In this stage you have $12 to allocate. You can either keep it or invest it in the following investment alternatives in $1 increments. In this stage, we have ranked the investments according to a common method of ranking investments.[40] Investments with more *'s in the ranking line are ranked higher according to this criterion.

[39]These questions were asked after each stage. We present them only once here for brevity.'

[40]In treatments where the ranking rule is given to subjects, this sentence was replaced with "In this stage, we have ranked the investments according to the Return/Risk ratio, which is a common method of ranking investments." The words " within categories" were added to the end of the appropriate sentence for treatments with categories.

Alternative:		A	B	C	D	E	F		
High Return:		200%	190%	225%	125%	130%	185%		
Low Return:		-20%	-90%	-75%	-25%	30%	15%		
Average Return:		90%	50%	75%	50%	80%	100%		
Range of Returns:		220%	280%	300%	150%	100%	170%		
Return/Risk Ratio:		0.4091	0.1786	0.2500	0.3333	0.8000	0.5882		
Ranking:		***	*	**	*	***	**		Amount Kept
Amount you Invest:									*=$12-Invested*

Please confirm your allocation before pressing continue.

When you are finished with your allocation, click the "continue" button below. As a reminder, at the end of the experiment, we will randomly select one of these stages to play out. If this stage is chosen, a single random draw will determine the returns to ALL of the investments in this stage and, as a result, your payoff to this stage.

——— Next Web Page ———

Stage III

In this stage you have $12 to allocate. You can either keep it or invest it in the following investment alternatives in $1 increments. In this stage, we ask you to first rank the investments. Use the drop down boxes to assign 1, 2 or 3 stars to investments. 3-star investments should be the investments with the highest rank, 1-star investments should be the investments with the lowest rank and 2-star investments should be in between. You are asked to assign two 3-star, two 2-star and two 1-star ratings.[41]

[41] In treatments where the ranking rule is given, this paragraph is replaced with "In this stage, we ask you to first rank the investments according to the return/risk ratio. Use the drop down boxes to assign 1, 2 or 3 stars to investments. 3-star investments should be the investments with the highest return/risk ratio, 1-star investments should be the investments with the lowest return/risk ratio and 2-star investments should be in between. You are asked to assign two 3-star, two 2-star and two 1-star ratings." In both treatments with categories, the last sentence is replaced with "You are asked to assign one 3-star, one 2-star and one 1-star rating within each category of investments."

Alternative:		A	B	C	D	E	F		
High Return:		185%	125%	130%	225%	200%	190%		
Low Return:		15%	-25%	30%	-75%	-20%	-90%		
Average Return:		100%	50%	80%	75%	90%	50%		
Range of Returns:		170%	150%	100%	300%	220%	280%		
Return/Risk Ratio:		0.5882	0.3333	0.8000	0.2500	0.4091	0.1786		
Ranking:									Amount Kept
Amount you Invest:									*=$12-Invested*

When you are finished with your allocation, click the "submit rankings" button below. Then, you will be allowed to allocate cash.

Please confirm your allocation before pressing continue.

When you are finished with your allocation, click the "continue" button below. As a reminder, at the end of the experiment, we will randomly select one of these stages to play out. If this stage is chosen, a single random draw will determine the returns to ALL of the investments in this stage and, as a result, your payoff to this stage.

——— Next Web Page ———

Stage IV

In this stage you have $12 to allocate. You can either keep it or invest it in the following investment alternatives in $1 increments.

Alternative:		A	B	C	D	E	F		
High Return:		190%	225%	200%	125%	185%	130%		
Low Return:		-90%	-75%	-20%	-25%	15%	30%		
Average Return:		50%	75%	90%	50%	100%	80%		
Range of Returns:		280%	300%	220%	150%	170%	100%		
Return/Risk Ratio:		0.1786	0.2500	0.4091	0.3333	0.5882	0.8000		Amount Kept
Amount you Invest:									*=$12-Invested*

Please confirm your allocation before pressing continue.

When you are finished with your allocation, click the "continue" button below. As a reminder, at the end of the experiment, we will randomly select one of these stages to play out. If this stage is chosen, a single random draw will determine the returns to ALL of the investments in this stage and, as a result, your payoff to this stage.

——— Next Web Page ———

Knowledge and Demographic Surveys

Next, we would like you to fill out brief demographic and knowledge surveys to help us classify the data. Your responses will be kept in a data file that does not include any identifying information. We will keep your participation in this research study confidential to the extent permitted by law.

Taking part in the surveys is completely voluntary. Leave blank any questions you prefer not to answer. You won't be penalized or lose any benefits for which you otherwise qualify.

At the end of the surveys, we will randomly select one of stages 1 through 4 that you have completed and use your allocation decision from this stage and a random draw to determine your payoffs for this experiment.

Earnings from this stage and random draw, together with earnings from the preliminary stage and the $5 participation fee, will contribute to your earnings for the experiment.

Press the "continue" button below to go on to the Knowledge Survey.

——— Next Web Page ———

Knowledge Survey

1. How would you classify your knowledge of financial markets?[42]

    (a) No knowledge whatsoever

    (b) Beginner level

    (c) Intermediate level

    (d) Advanced level

2. How much experience have you had with trading on organized financial markets?[43]

[42]Source: Variant (added the "No knowledge whatsoever" option) of IEM financial knowledge survey question from Oliven and Rietz (2004).

[43]Source: IEM financial knowledge survey question from Oliven and Rietz (2004).

(a) Novice (Have never traded.)

(b) Limited (Have, for example, made some trades on my own account.)

(c) Experienced Amateur (Have traded a good deal for myself, friends or family.)

(d) Professional (Have been paid for trading.)

3. How many hours in a typical week do you spend following or trading on organized financial markets?[44]

    (a) None

    (b) An hour or less

    (c) Between 1 and 5 hours

    (d) Between 5 and 10 hours

    (e) More than 10 hours.

4. How do you rate your knowledge of investments relative to other people?[45]

    (a) Much less knowledgeable

    (b) Somewhat less knowledgeable

    (c) About as knowledgeable

    (d) Somewhat more knowledgeable

    (e) Much more knowledgeable

5. What is compound interest?[46]

    (a) Earning money on your principal and your interest.

    (b) a complicated form of interest.

    (c) A long term investment.

    (d) A risk/return scenario.

[44]Source: Variant (made into multiple choice question) of IEM financial knowledge survey question from Oliven and Rietz (2004).

[45]Source: New.

[46]Source: InvestNative Project Investment Quiz, question 4. http://www.investnative.org/test.html, accessed 08/19/2010.

6. What is an example of a low risk investment?[47]

    (a) Small cap stock.

    (b) High yield "junk" bond.

    (c) FDIC insured savings account.

    (d) An international mutual fund focused on small Latin American airline companies.

7. If the value of your investment declines by 60%, what subsequent percentage increase is needed to return to your original investment amount?[48]

    (a) 60%.

    (b) 120%.

    (c) 150%.

    (d) 180%.

8. In general, stock mutual funds that are riskier tend to provide higher returns over time than stock mutual funds with less risk.[49]

    (a) True.

    (b) False.

    (c) Don't know/Not sure.

9. Which of these investments have a risk of losing value?[50]

    (a) Mutual funds

    (b) Blue chip stocks

    (c) High-yield bonds

[47]Source: InvestNative Project Investment Quiz, question 21. http://www.investnative.org/test.html, accessed 08/19/2010.

[48]Source: Pacific Life Investment Knowledge Quiz, question 4: http://www.pacificlife.com/Channel/Educational+Information/Calcul accessed 08/19/2010.

[49]Source: Financial Industry Regulatory Authority Investor Knowledge Quiz, question 7: http://apps.finra.org/Quiz/1/investorquiz.aspx, accessed 08/19/2010.

[50]Source: Pacific Life Investment Knowledge Quiz, question 5: http://www.pacificlife.com/Channel/Educational+Information/Calcul accessed 08/19/2010.

(d) They all have risk of losing value

10. What is diversification?[51]

    (a) A way to reduce risk.

    (b) Investing in different things, such as stocks, bonds, savings, property, etc.

    (c) Having an asset allocation that spreads your investment among different asset classes.

    (d) All of the above.

11. Common stocks always provide higher returns than bonds or money market investments.[52]

    (a) True.

    (b) False.

    (c) Don't know.

12. Asset allocation is a form of:[53]

    (a) Repayment.

    (b) Diversification.

    (c) Capital.

13. Generally, a portfolio that has 80% of its assets invested in stocks would be best suited for:[54]

    (a) An 18-year-old using the assets to pay for college expenses over the next 4 years.

    (b) A 35-year-old investing for retirement.

    (c) A 75-year-old investing for income and capital preservation.

    (d) None of the above.

    (e) Don't know.

[51] Source: InvestNative Project Investment Quiz, question 11. http://www.investnative.org/test.html, accessed 08/19/2010.

[52] Source: Vanguard Investment Knowledge Quiz, question 3: https://personal.vanguard.com/us/InvestmentKnowledge, accessed 08/19/2010.

[53] Source: InvestNative Project Investment Quiz, question 14. http://www.investnative.org/test.html, accessed 08/19/2010.

[54] Source: Vanguard Investment Knowledge Quiz, question 19: https://personal.vanguard.com/us/InvestmentKnowledge, accessed 08/19/2010.

——— Next Web Page ———

Demographic Survey

1. What is your gender?
   (a) Female
   (b) Male
2. What is your age?
   (a) Under 30 Years Old.
   (b) 30–39 Years Old.
   (c) 40–49 Years Old.
   (d) 50 Year Old and Older.
3. What is your current marital status?
   (a) Married.
   (b) Not Married.
4. Do you have children?
   (a) Yes.
   (b) No.
5. What is your current annual income?
   (a) $0–$19,999.
   (b) $20,000-$29,999.
   (c) $30,000-$39,999.
   (d) $40,000-$69,999.
   (e) Greater that $60,000.
   (f) Would rather not say.
6. What is the highest educational degree you have achieved?

(a) High School

(b) Bachelor's.

(c) Master's

(d) Doctorate.

(e) Other.

7. What is or was your college major (most recent degree if you are not currently a student)?

    (a) Business.

    (b) Social Science.

    (c) Humanities.

    (d) Natural Science.

    (e) Mathematics or Engineering.

    (f) Other.

    (g) None.

——— Next Web Page ———

Final Question

After outcomes and payments have been determined, subjects answer the following final question:

1. How satisfied are you with your allocation decision in the stage that was selected to determine your payoffs?

Very Dissatisfied			Neutral			Very Satisfied
1	2	3	4	5	6	7