Journal of Financial Economics 131 (2019) 404–432

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Mark Twain's Cat: Investment experience, categorical thinking, and stock selection☆,☆☆

Xing Huang

Department of Finance, Washington University in St. Louis, Simon Hall 211, Olin Business School, One Brookings Dr, St. Louis, MO 63130-4899, USA

ARTICLE INFO

Article history:
Received 13 August 2015
Revised 19 October 2017
Accepted 20 December 2017
Available online 16 August 2018

JEL classification:
G11
G40
G41

Keywords:
Investment experience
Categorical thinking
Household finance

ABSTRACT

This paper examines the effect of prior investment experience in specific industries on subsequent investment decisions. Using households' trading records from a large discount broker between 1991 and 1996, I find that prior success in a given industry increases the likelihood of subsequent purchases in the same industry. The effect is stronger for more recent experiences and for less sophisticated or diversified investors, and it is not wealth enhancing. The results suggest investors categorize industries at a highly resolved level, finer than the Fama–French ten-industry classification. Similar effects are also apparent for size- and value-based categories but at smaller magnitudes.

1. Introduction

How do individuals select stocks? Do they incorporate all available historical information and update their beliefs in a Bayesian fashion? Or do they weight their personal investment experiences and other statistical information differently? A growing body of evidence shows that past experience affects investors' choices in many financial decisions, including household risk-taking and corporate financing decisions.[1] Past experience can influence households' purchase decisions on common stocks as well. Barber, Odean, and Strahilevitz (2011) show that the effect of positive past investment experiences with a particular stock increases the likelihood of repurchasing that stock. But do past investment experiences also influence investors' propensity to buy other stocks?

As Barber and Odean (2000) point out, investors face a substantial search problem when choosing stocks to buy. To reduce the dimensionality of an investor's purchase decisions, I hypothesize that, to the extent that past investment experiences have impact, the impact will be stronger for stocks in the same category as (or more similar to) previously purchased stocks. Through the lens of how previously experienced outcomes affect choices within a category, I further study the granularity of in-

☆ The cat, having sat upon a hot stove lid, will not sit upon a hot stove lid again. But he won't sit upon a cold stove lid, either. – Mark Twain

☆☆ I am grateful for the very helpful suggestions and comments of Nicholas Barberis, Stefano DellaVigna, David Hirshleifer, David Laibson, Ulrike Malmendier, Gregor Matvos, Terrance Odean, Andrei Shleifer, Adam Szeidl, Nancy Wallace, my colleagues at MSU and seminar participants at UC Berkeley (Finance, Financial Economics Lunch), and the London Business School Trans-Atlantic Doctoral Conference. I would also like to thank an anonymous discount broker for providing data on individual investors' trades and Terrance Odean for his help in obtaining and understanding the dataset. All errors are my own.

E-mail address: xing.huang@wustl.edu

[1] For example, Barber, Odean, and Strahilevitz (2011); Campbell, Ramadorai, and Ranish (2014); Choi et al. (2009); Kaustia and Knupfer (2008); Malmendier and Nagel (2011); Malmendier, Tate, and Yan (2011), among others.

https://doi.org/10.1016/j.jfineco.2018.08.003
0304-405X/

vestors' categorization schemes. For example, how finely do investors categorize stocks? Do investors categorize stocks by industry, size, or value? Empirical evidence on the granularity of investors' categorization schemes could enhance our understanding about categorical thinking/learning in investment decisions that is the central feature considered in some recent papers (for example, Barberis and Shleifer, 2003; Peng and Xiong, 2006).

In this paper, I begin with an exploration of one relevant categorization scheme—industry. Do investors put more weight on their idiosyncratic personal experience of investment in an industry when they make decisions about purchasing new stocks? Consider, for example, two individual investors, Mr. Fortune and Miss Fortune ("Misfortune"), who both form portfolios in 1991. They both invest in the insurance industry but in different companies—Mr. Fortune picks Conseco, Inc., while Miss Fortune picks Pioneer Financial Services, Inc. One year later, Mr. Fortune's investment doubles, whereas Miss Fortune suffers a paper loss of 30%. Given the idiosyncratic realization of returns after enjoying a huge gain from investing in the insurance industry, will Mr. Fortune be more prone than Miss Fortune to investing in *other* stocks in that industry? The psychology literature suggests that personally experienced outcomes have a greater impact on personal decisions than does information acquired merely by reading, which lacks personal involvement (Weber et al., 1993; Hertwig et al., 2004). The experience hypothesis predicts that when investors have prior investment experience in an industry, they are more likely to buy new stocks in that industry if they have experienced good, rather than bad, returns.

This paper exploits data from households' detailed trading records at a large discount broker between 1991 and 1996 (Barber and Odean, 2000) as a measure of investors' personal investment experiences, and it explores whether the past experiences of these investors affect their subsequent purchase decisions at the industry level. The results indicate that the likelihood of investors purchasing new stocks in an industry they have previously experienced increases with their experience of excess returns. Furthermore, the effect of experienced outcomes is weaker for purchase decisions concerning stocks in industries with fewer similarities. Specifically, experience in one industry has less influence on purchases in a similar industry and almost no effect on purchases in a dissimilar industry.

While these results are consistent with the view that past experience influences stock selection, there are four confounding effects that can also drive correlations between past experiences and future purchases in the same industry. First, momentum traders (Hong and Stein, 1999; DeLong et al., 1990; Barberis and Shleifer, 2003) may be more likely to purchase stocks in an industry that performed well in the past, regardless of their personal investment experience in that industry. To control for this momentum effect, I include past industry average returns or industry-year fixed effects in the regressions; my results are robust to this inclusion.

A second confounding effect is a wealth effect: previous high returns experienced by investors can increase their wealth, generating new purchases in all industries. The results described above—that the experience effect spills over only slightly to other industry purchases—do not seem to support this story, but in addition I show that my results are robust to controls measuring changes in investors' portfolio value.

Third, one possible omitted variable is the investment skill of investors, which provides an alternative story that skilled investors are more likely to gain high returns and they can also be more active, which could drive the positive relation between experienced outcomes and future purchases. To control for the investment ability, I use three variables for trading characteristics and the results remain the same. In addition, I also test how the relation between experienced outcomes and future purchases varies with different degrees of investor sophistication. This alternative story does not predict that the relation between experienced outcomes and future purchases would vary with investor sophistication, but the results show a stronger relation among less sophisticated individuals, which is more consistent with the experience hypothesis. Although I could not completely rule out this alternative story, the evidence suggests that it is unlikely to drive the results.

A fourth potential confounding effect is portfolio rebalancing. But because households should decrease their holdings in an industry that earned them relatively higher past returns to rebalance their portfolio, this story would result in an outcome that is the opposite of the prediction of the experience hypothesis.

After controlling for industry momentum and time-varying investor heterogeneity, the results show that investors with a history of positive market-adjusted returns in one industry have a 1.86 percentage points (pp) higher propensity to purchase new stocks in the same industry, as opposed to those who earned negative market-adjusted returns. This magnitude corresponds to 16.60% of the average probability of 11.18 pp for purchasing new stocks in an industry. The difference between the impact of the top and bottom decile experienced outcomes is even larger, 3.91 pp (= 1.26 pp + 1.46 pp – (– 1.19 pp)), which is a 34.97% difference (= 3.91 pp/11.18 pp). However, this effect, which results from an investor's experience of positive excess returns, drops significantly with regard to the purchase of new stocks outside of the experienced industry. The magnitude of this effect on the purchase of new stocks in the industry that is the most similar to the experienced industry is only 40% of that for the purchase of new stocks in the original industry, and the effect on purchasing new stocks in an industry that differs significantly from the experienced industry is negligible.

I also test for the dynamic effects of experienced outcomes. The results indicate that more recent experience has a stronger influence on an investor's purchase of new stocks. The effect of experienced outcomes drops dramatically if the experience happens more than one year before the month of purchase. Gallagher (2014) finds a similar effect in the context of flood insurance; after a flood, the take-up rate steadily decreases in the flooded communities.

If the experience hypothesis drives the relation between experienced outcomes and future purchases, the magnitude and significance of the results should vary with investor sophistication and portfolio diversification. As investors become more sophisticated, their past experiences can cancel each other out so that more recent experiences have less influence. I divide the households in the sample into four subgroups according to their self-reported investment sophistication as provided in the dataset and test the effects of experienced outcomes separately for each subgroup. The results indicate that the effect is most pronounced in the group with no experience or knowledge of investing and insignificant for the group with extensive investment experience; in terms of magnitude, the gap between the top and bottom decile is a 39.40% difference (relative to the average probability of buying new stocks in an industry) for the least sophisticated group, which is about four times as large as that for the most sophisticated group (10.37%). This is further evidence in support of the experience hypothesis.

Furthermore, investors with less concentrated, nondiscretionary stock portfolios may care less about the individual performance of each component; consequently, the experienced outcome on the industry level would have less effect. I explore the variation in the concentration of investor nondiscretionary portfolios by adding an interaction term between the experience variables and measures for portfolio concentration in the regression. As predicted above, the results indicate that the influence of past experiences declines as the concentration of a household's portfolio decreases.

Though I motivate the hypothesis with a natural explanation of naive extrapolation (or reinforcement learning), there are several potential explanations for this experience effect. The effect can also be explained by "learning about ability" (i.e., investors learning about their industry specific information acquisition skill, Nieuwerburgh and Veldkamp, 2010), or "learning by trading" (Seru et al., 2010). I find supportive evidence for the "learning about ability" channel in that investors will trade (including both buying and selling) more frequently in an industry after they have had a good experience in that industry. This evidence could not be explained by naive extrapolation/reinforcement learning alone, because naive extrapolation/reinforcement alone would not encourage investors to sell more frequently after good experience.

I further compare the performance between investors' actual portfolios and a set of counterfactual portfolios. No matter which counterfactual we consider, investors' actual purchases deliver a reliably lower alpha. The results suggest that individual investors do not systematically pick the wrong *industries* for investment; rather, they appear to systematically pick the wrong *stocks*. Given the investor's inferior stock-picking ability, to the extent that the "learning about ability" explanation is relevant, the learning is more likely to be incorrect. As a result, categorization can make investors' inaccurate estimation of their ability last longer and lead to even more losses before they decide to switch to a more diversified portfolio or exit the market.

To gain additional insight into categorical thinking in conjunction with the experience effect, I investigate how the experience effect varies with the refinement of categories. I create a proxy to measure the experienced stocks' ability to represent other stocks in the same industry. The results indicate that the effect is stronger when the experienced stocks are more representative. I also test whether investors divide stocks into subcategories beyond the Fama-French ten-industry (FF10) classification. The results suggest that individuals do categorize on a finer level (i.e., the Fama-French 48 subindustry classification)—the impact of good investment experience on another subindustry from the same FF10 industry is in general smaller than the investment's effect on the experienced subindustry, and the magnitude of the effect decreases if the subindustry is economically more distant from the experienced subindustry. I also provide suggestive evidence that this granularity of investors' categorization schemes may be related to comentions in news articles.

I also consider the possibility that investors use alternative schemes to categorize investments. In particular, given their prominent role in empirical asset pricing models, I consider categorizations based on size and value. Past experience in categories based on these characteristics also has an effect on future purchases within the same category, though it is of a smaller magnitude.

This paper is related to several strands of literature. First, it contributes to the growing literature on investor behavior[2]; specifically, I address how investors choose which stocks to purchase. Some papers focus on investors' cross-sectional preferences for stocks,[3] while others discuss investors' stock purchase in a time series that connects buying decisions with past investment experiences.[4] This paper takes the latter approach and emphasizes the effects of past experience on future purchases of new stocks at different levels of categorization. Related works include papers by Campbell, Ramadorai, and Ranish (2014), who study the influence of investors' past experiences on their investment behaviors, including trading frequency, disposition bias, and diversification, and Anagol, Balasubramaniam, and Ramadorai (2015), who exploit the setting of India's initial public offering (IPO) lotteries to identify the causal effect of investment experiences on future investment behavior.

The literature also indicates that experience affects a number of other financial decisions, including IPO subscriptions (Kaustia and Knupfer, 2008), 401(k) account portfolio choices (Choi et al., 2009), stock market participation (Malmendier and Nagel, 2011), corporate external financing (Malmendier and Tate, 2005; Malmendier, Tate, and Yan, 2011), etc. Even though the tests employed in this study are based on the idiosyncratic personal experiences of investors, an examination of the variation in experiences across cohorts (as in Malmendier and Nagel, 2011) can lead to similar results. Specifically, consider industry

[2] For example, Calvet, Campbell, and Sodini (2007, 2009); Grinblatt and Keloharju (2000, 2001).

[3] For example, Barber and Odean (2007) study attention-grabbing stocks and find that individual investors are net buyers of attention-grabbing stocks. French and Poterba (1991) and many other papers show the home bias puzzle; individuals and institutions in most countries hold modest amounts of foreign equity, even though observed returns on national equity portfolios suggest substantial benefits from international diversification.

[4] For example, Barber, Odean, and Strahilevitz (2011) establish trading patterns in which investors' repurchasing of stocks they previously sold is affected by their experiences with those stocks.

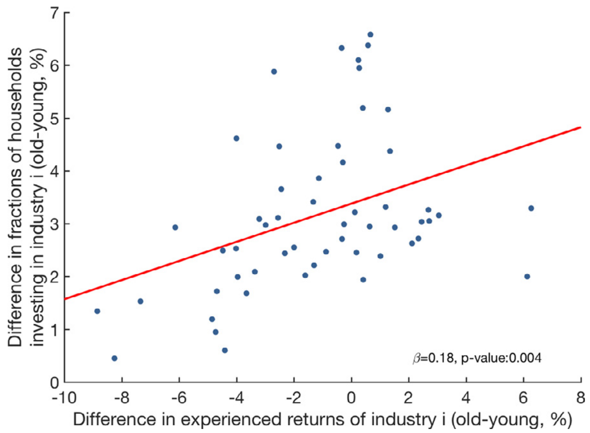

Fig. 1. Motivating example: difference in stock holdings by industry of young and old groups versus difference in their experienced industry returns. The stock holdings of an age group in industry *i* are measured by the fraction of households within the age group investing in industry *i* in a certain year. The young group is defined as households with ages between 30 and 40, while the old group is defined as households with ages between 50 and 60. The vertical axis denotes the difference in stock holdings by industry of these two groups. The horizontal denotes the difference in their experienced industry returns. The experience of the young group in industry *i* is measured as the average of industry *i* returns over the prior ten years, while the experience of the old group is measured over the prior 30 years. Each observation corresponds to a year-industry pair. Industry is classified according to the Fama-French ten-industry classification. The high tech industry is excluded in this analysis.

as the categorization scheme and assume investors begin investing in the stock market in their mid-twenties. For households whose members are between the ages of 30 and 40, I measure experiences using the average of industry returns during the preceding ten years. Likewise, for households whose members are between the ages of 50 and 60, experiences are measured by the average of industry returns during the preceding 30 years. Fig. 1 plots the difference in the fraction of households investing in an industry between the two age cohorts listed above against the difference in experience between the same two groups. The figure suggests that more investors hold stocks in an industry when their experience in that industry is better, which is consistent with the experience hypothesis.

This study also relates to the literature on investors' categorical thinking, which has implications for asset-price dynamics. Barberis and Shleifer (2003) assume that investors categorize risky assets into different styles and trade among these styles depending on their relative performance (not necessarily their experienced performance), which drives excessive comovement of assets in the same style but little comovement of assets in different styles. Peng and Xiong (2006) provide justification for category-learning behavior when attention is a scarce cognitive resource; they also generate features in return comovement and other predictions. My paper helps to understand the role of personal investment experiences in investor's categorical thinking and provides insights for studies with categorical thinking/learning as an important feature considered in investors' decision-making.

Section 2 describes the datasets and methodology and presents summary statistics. Section 3 details the results of my examination of the influence of experience on the future purchase of new stocks. Section 4 explores the underlying mechanisms through an examination of how the experience effect varies with investor sophistication and portfolio diversification. Section 5 investigates alternative categorization schemes and implications. Section 6 concludes.

2. Datasets and methodology

2.1. Data description

The dataset used in this paper includes the trading records of 78,000 households at a large discount brokerage house from 1991 to 1996. This dataset is used by Barber and Odean (2000) and others. Each household has at least one account, but some have many. I combine the trades of accounts within the same household and build observations at the household level. This paper focuses on investors' direct investments in common stock, so I exclude their investments in mutual funds, American depository receipts (ADRs), warrants, and options. The sample is further refined by removing observations with errors in trading records, short selling trades, etc. The number of households in the final sample is 47,993. More details about the restrictions I impose to select the sample for analysis are listed in the appendix. I use the Center for Research in Security Prices (CRSP) database to obtain information on stock prices for calculating investors' experienced returns and portfolio-related variables.

One feature of this paper is to discuss investors' stock selection choices among a manageable number of categories. I start with an exploration of one relevant categorization scheme—industry. The Standard Industrial Classifications (SIC) codes are obtained from two sources: CRSP and Compustat.[5] For most of the tests, stocks are classified into ten industry groupings based on their SIC code according to an algorithm devised by Fama and French (1997). The ten industry groupings are (1) consumer nondurables; (2) consumer durables; (3) manufacturing; (4) oil, gas, and coal extraction and products; (5) high technology; (6) telephone and television transmission; (7) wholesale, retail, and some services; (8) healthcare, medical equipment, and drugs; (9) utilities; and (10) others. I also exploit the Fama-French 48-industry classification in robustness tests and to define subindustries for further analysis. Besides industry, in Section 5.3, I also consider alternative schemes, such as size and value, that investors can use to categorize investments.

2.2. Investors' investment experiences and purchase decisions

I construct investors' experienced returns in each industry from their trading records and select a fixed window to

[5] The first step is to match the CUSIP numbers of the stocks in which the households invest with the corresponding SIC code in CRSP. If a corresponding SIC code cannot be matched in CRSP, the second step is to search for a match in Compustat.

measure experienced returns. For example, the experience window spans from the beginning to the end of each year. Note that the returns in a fixed window could be either realized or not. I do not use realized returns to measure experiences, because I want to avoid introducing any potential endogeneity.

The measures of experienced outcomes build on market-adjusted experienced returns. For every household, h, I denote by r_{hjt} the annualized return (either realized or not) of the stock j during experience window t. The experienced returns of industry I, r_{hIt}, is defined as the value-weighted average returns of stocks belonging to the industry ($j \in I$):

$$r_{hIt} = \frac{\sum_{j \in I} x_{hjt} r_{hjt}}{\sum_{j \in I} x_{hjt}}, \tag{1}$$

where x_{hjt} is the dollar value allocated on stock j at the date of purchase or at the beginning of the window if the date of purchase is before the window starts. The market-adjusted experienced return in industry I, er^m_{hIt}, is the difference between r_{hIt} and the market return of period t, R_{mt}[6]:

$$er^m_{hIt} = r_{hIt} - R_{mt}. \tag{2}$$

I consider three measures of investors' previously experienced outcomes, which are three indicator variables denoting good, top, and bottom experiences. The dummy for good experience, $Goodexp_{hIt}$, equals one if household h earns a positive market-adjusted return in industry I during period t (i.e., $er^m_{hIt} > 0$). The dummy variable for a top (bottom) experience, $Topexp_{hIt}$ ($Bottomexp_{hIt}$), equals one if er^m_{hIt} is above the 90th (below the 10th) percentile of market-adjusted experienced returns by all households during period t. While these indicator variables are all based on market-adjusted experienced returns, which are relative measures of experience, the results are robust to other measures, such as those based on the raw level of experienced returns (as shown in appendix, Section A.2).

Investors' purchase decisions are measured by an indicator variable, $B^{new}_{hI,t+1}$, that takes a value of one if household h purchases new stocks (those not previously owned in the experience window) in industry I during decision period $t+1$ following experience period t. Note that the purchase decisions focus on new stocks, not previously owned stocks. If I do not exclude previously owned stocks, the effect of past experiences on the industry level may be confounded by the effect of the experienced stocks on themselves. Barber, Odean, and Strahilevitz (2011) find that investors are more likely to repurchase stocks that have been previously sold for a gain; they also find that investors prefer to purchase additional shares of stocks that have lost value since being purchased. Since the experienced return (previously defined) could be either realized or not, the effect on previously owned stocks can push the results either way. Therefore, to separately identify the effect on the industry level, I only consider purchases of new stocks in the following tests.

[6] I use lowercase letters to denote the experienced returns of households, such as r_{hIt}, er^m_{hIt}, and uppercase letters to denote market and industry average returns, such as R_{mt}, R_{It}.

2.3. Summary statistics

Table 1 presents summary statistics. Panel A reports the frequencies of the purchase of new stocks and the repurchase of previously owned stocks over the years of the decision period (1992–1996). Purchases of new stocks account for a large portion (about 85%) of the investors' overall purchase decisions.

Panels B and C summarize statistics related to investors' past experiences. Panel B reports the distribution of households' experiences across industries for each year of the experience window. There are relatively more households trading in some of the industries, such as (3) manufacturing, (5) high technology, (7) wholesale, and (8) health care, which each attracts more than 10% of the households. But overall, households' participation in each industry is roughly balanced, which can help rule out the possibility that the results are driven by a concentration of trades in a particular industry. Second, this table could also show that the distribution of households' experience among the industries is stable across the years.

Panel C provides a first look at the distribution of experience outcomes within each industry for each year of the experience window. We can observe both the cross-sectional and time series variations. For example, during 1991, a large portion of the households (81.2%) had bad experiences in the energy industry, while over half of the households had good experiences in other industries, such as wholesale, health care, and utilities. However, households investing in the energy industry do not always have bad experiences. In 1993, over half of these households had good experiences.

3. Industry investment experience and stock selection

In this section, I will study the effects of returns experienced in an industry on the decision to purchase new stocks in the same industry. I present graphical evidence and estimate a baseline specification using a probit model, followed by several robustness tests, such as using different measures of experienced outcomes or different industry classifications, addressing alternative explanations, etc. Finally, I examine the long-term effect of previously experienced outcomes on stock selection.

3.1. Graphical evidence

Fig. 2 illustrates the relationship between previously experienced outcomes and future purchases. I divide all of an industry's experienced outcomes into five bins, ordered by market-adjusted returns, for each industry during each period. The right most bars (group 5) correspond to the top 20% of experienced returns, while the left most bars (group 1) correspond to the bottom 20% of experienced returns. The black bars in both figures represent the probability of buying new stocks in the experienced industry. Fig. 2(a) employs the original data. It shows a roughly monotonically increasing relationship: as experienced returns increase from the bottom to the top quintile, investors become more likely to buy new stocks in the experienced industry, especially in the upper tail of the experienced

Table 1
Summary statistics.

Panel A reports the number and percentage of trades purchasing stocks not previously owned and trades repurchasing stocks previously owned. The sample period includes all the years used as decision windows in the baseline model (1992–1996). Panel B reports the percentage of households investing in each industry grouping from 1991 to 1995 for the full sample. Panel C reports the percentage of a household's good (bad) experience in each industry grouping from 1991 to 1995 for the full sample. Good (bad) experience is when the experienced market-adjusted return in the industry is greater than (smaller than or equal to) zero. The sample period includes all the years used as experience windows in the baseline model.

Panel A: Distribution of household stock purchases, 1992–1996

	1992	1993	1994	1995	1996	Total
Purchase other stocks	77,057	70,912	57,626	68,519	72,727	346,841
Repurchase previously owned stocks	10,455	12,013	12,200	10,946	13,839	59,453
Total	87,512	82,925	69,826	79,465	86,566	406,294
Purchase other stocks	88.1%	85.5%	82.5%	86.2%	84.0%	85.4%
Repurchase previously owned stocks	11.9%	14.5%	17.5%	13.8%	16.0%	14.6%
Total	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Panel B: Distribution of household stock investment experience across industries, 1991–1995

	1991	1992	1993	1994	1995	Total
(1) Consumer nondurables (%)	8.7	8.7	9.5	9.3	8.8	9.0
(2) Consumer durables (%)	4.6	5.0	4.4	4.6	4.9	4.7
(3) Manufacturing (%)	11.9	11.0	10.5	10.1	10.6	10.7
(4) Oil, gas, and coal extraction and products (%)	4.5	4.8	4.5	4.2	4.2	4.4
(5) High technology (%)	17.4	17.1	16.6	16.3	17.0	16.8
(6) Telephone and television transmission (%)	4.8	5.0	5.6	6.5	6.9	5.9
(7) Wholesale, retail, and some services (%)	10.3	10.7	11.3	11.7	12.0	11.3
(8) Health care, medical equipment, and drugs (%)	12.5	14.7	15.3	14.4	13.3	14.1
(9) Utilities (%)	5.1	5.8	5.8	6.4	6.1	5.9
(10) Others (%)	20.2	17.2	16.6	16.5	16.2	17.0
N	58,392	86,513	101,525	107,534	116,928	470,892

(Continued on next page)

Table 1
Continued.

Panel C: Household stock investment outcomes, 1991–1995

	(1) Consumer nondurables	(2) Consumer durables	(3) Manufacturing	(4) Oil, gas, and coal extraction and products	(5) High technology	(6) Telephone and television transmission	(7) Wholesale, retail, and some services	(8) Health care, medical equipment, and drugs	(9) Utilities	(10) Others	Total
Year = 1991											
Bad (< 0) (%)	50.8	65.3	57.6	81.2	60.6	52.9	44.8	40.2	44.0	61.2	55.3
Good (> 0) (%)	49.2	34.7	42.4	18.8	39.4	47.1	55.2	59.8	56.0	38.8	44.7
N	5,058	2,715	6,927	2,633	10,177	2,793	6,012	7,297	2,959	11,821	58,392
Year = 1992											
Bad (< 0) (%)	55.1	34.9	53.3	63.9	53.6	33.7	45.2	74.4	57.2	43.4	52.9
Good (> 0) (%)	44.9	65.1	46.7	36.1	46.4	66.3	54.8	25.6	42.8	56.6	47.1
N	7,567	4,334	9,510	4,188	14,773	4,312	9,214	12,725	5,035	14,855	86,513
Year = 1993											
Bad (< 0) (%)	68.8	28.7	44.2	41.4	45.6	47.1	60.1	74.8	58.3	38.4	52.5
Good (> 0) (%)	31.2	71.3	55.8	58.6	54.4	52.9	39.9	25.2	41.7	61.6	47.5
N	9,633	4,429	10,621	4,579	16,871	5,691	11,447	15,575	5,870	16,809	101,525
Year = 1994											
Bad (< 0) (%)	53.1	67.3	45.9	49.3	37.9	69.2	64.2	39.7	78.6	62.4	54.0
Good (> 0) (%)	46.9	32.7	54.1	50.7	62.1	30.8	35.8	60.3	21.4	37.6	46.0
N	9,981	4,910	10,908	4,537	17,558	6,960	12,600	15,438	6,893	17,749	107,534
Year = 1995											
Bad (< 0) (%)	50.4	80.4	61.9	64.9	62.6	49.7	70.2	34.0	80.5	53.3	58.2
Good (> 0) (%)	49.6	19.6	38.1	35.1	37.4	50.3	29.8	66.0	19.5	46.7	41.8
N	10,302	5,706	12,415	4,961	19,928	8,019	14,038	15,547	7,111	18,901	116,928

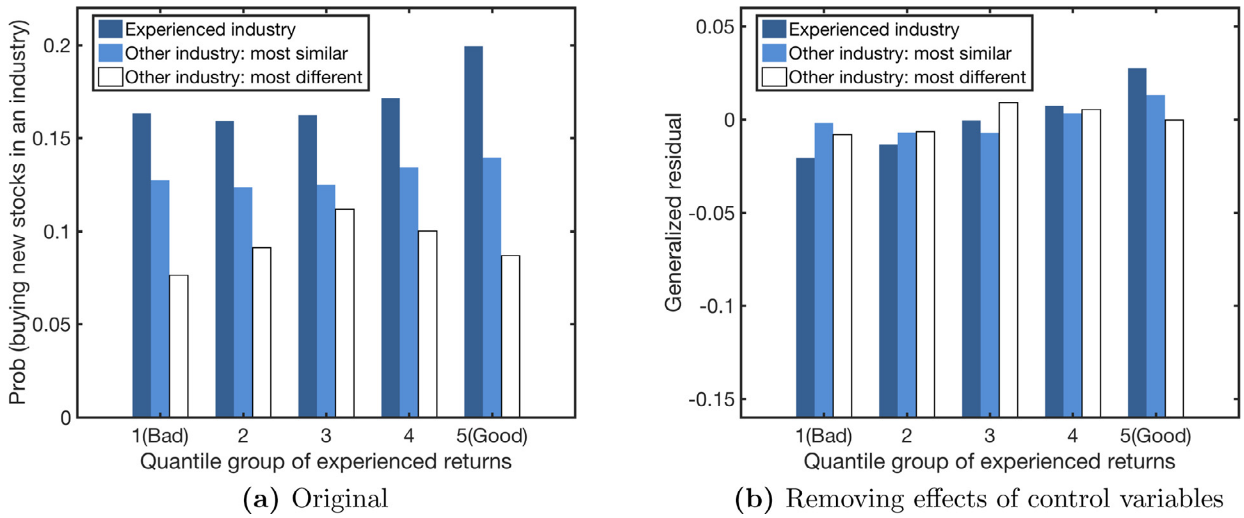

Fig. 2. The impact of personal investment experience in an industry on subsequent investment in the experienced industry versus other industries. The observations are sorted by the value-weighted average annualized excess return on the investment in an industry and divided into five groups. Group 1 has the lowest experienced return, while group 5 has the highest. Figure (a) plots the original probability of buying new stocks, that is, the percent of households buying new stocks in an industry. Figure (b) plots the generalized residuals from a probit model of regressing purchasing new stocks in an industry on controls of the baseline model.[7] The dark blue bars correspond to the experienced industry. The light blue and white bars correspond, respectively, to the industry that is the most similar to or different from the experienced industry. The most similar and the most different industries are selected by measuring the distance between the stock returns in that industry and those in the experienced industry. The distance between two industries is measured by averaging the 10-K text-based similarity scores (constructed by Hoberg and Phillips, 2010[8]) across all pairs of firms in these two industries. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

returns. Fig. 2(b) plots the average of generalized residuals within each quintile. The residuals are obtained from a probit model by regressing purchases of new stocks in one industry on control variables. (The control variables will be discussed in detail later.) After removing the effects of the controls (industry average, wealth effect, etc.), the monotonically increasing relationship becomes more striking.

3.2. Baseline model

I start by modeling the probability of purchasing new stocks in a single industry with a probit model. The dependent variable ($B^{new}_{hI,t+1}$) indicates the purchases of new stocks (those not previously owned during period t) in industry I in period $t+1$. The specification is written as follows:

$$P\left(B^{New}_{hI,t+1} = 1\right) = \Phi(\beta_0 + \beta_1 Exp_{hIt} + \beta_2 Goodexp_{hIt} + \beta_3 Topexp_{hIt} + \beta_4 Bottomexp_{hIt} + \Gamma' X_{hIt}), \quad (3)$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution function. Exp_{hIt} is a dummy variable equal to one if household h has an investment in industry I in period t. In the model, each observation corresponds to a household/industry/year set. Because I use the Fama-French ten-industry classification system, each household/year pair corresponds to ten observations. The primary coefficients of interest are on the variables of experienced outcomes. The outcomes are measured by three dummy variables, $Goodexp_{hIt}$, $Topexp_{hIt}$, and $Bottomexp_{hIt}$, which equal one if household h has positive, above the 90th percentile, or below the 10th percentile market-adjusted returns, respectively, in industry I in period t. The experience hypothesis—that investors overweight their experienced outcomes—predicts a monotonic positive relationship between experienced outcomes and future purchases (i.e., $\beta_2 > 0$, $\beta_3 > 0$, $\beta_4 < 0$).

3.2.1. Control variables

These predictions can be consistent with some confounding effects, such as (1) industry momentum trading, (2) wealth effects, or (3) investors' heterogeneity. I include a vector of controls (X_{hIt}) to address these issues.

Industry Momentum Trading. Moskowitz and Grinblatt (1999) find a strong and prevalent momentum effect in the industry component of stock returns, which provides a way for investors to conduct momentum trading at the industry level. Industry-level momentum trading could also lead to a positive relationship between previously experienced outcome and subsequent purchases. To control for this confounding effect, I include three industry average variables, all of which are based on the market-adjusted industry average return, $ER_{It} = R_{It} - R_{mt}$, where R_{It} denotes the value-weighted average return of industry I during period t. Corresponding to the dummy variables measuring

[7] The generalized residual of the probit model $Pr(Y_i = 1) = \Phi(X_i'\beta)$ is computed as:

$$\frac{Y_i - \Phi(X_i'\beta)}{\Phi(X_i'\beta)(1 - \Phi(X_i'\beta))}\phi(X_i'\beta),$$

where $\phi(\cdot)$ is the density function of the normal distribution, and $\Phi(\cdot)$ is the cumulative distribution.

[8] Hoberg and Phillips (2010) design these firm pairwise similarity scores from text analysis of firm 10-K product descriptions.

experienced outcomes, the three industry average variables are: (1) $Goodind_{It}$, an indicator of industries with positive market-adjusted industry average returns during the experience window (i.e., $ER_{It} > 0$); (2) $Topind_{It}$, an indicator of the industry with the highest market-adjusted industry average return; and (3) $Bottomind_{It}$, an indicator of the industry with the lowest market-adjusted industry average return.

Wealth Effect. Investors with good experiences in some industries are more likely to have increases in their overall stock portfolios. If investors tend to purchase new stocks in all industries when their overall portfolios earn profits, that can lead to a positive correlation between good experiences and future purchases in a specific industry. To address this explanation, I include a dummy variable indicating if the overall value of the household's portfolio of common stocks increases during the experience window.

Investors' Heterogeneity. Investors differ in their investment ability and level of expertise. Some investors with superior ability may gain high returns and have good experiences because they are better at picking misvalued securities or predicting economic prospects. These investors are also more likely to make new purchases. Even if new purchases are made randomly across industries, we can observe a positive correlation between good past experiences and future purchases. To ensure that investment ability is not driving the relation, I create three variables measuring trading characteristics as proxies of an investor's investment ability. Investors with greater ability can trade more frequently, own larger portfolios, or hold greater numbers of stocks in their portfolios. These three variables are calculated using the beginning-of-month position data from the experience window, and they are: (1) average number of stocks in the beginning-of-month portfolios; (2) the logarithm of the average size of the beginning-of-month portfolios; and (3) the logarithm of the average of the monthly turnover rate calculated following Barber and Odean (2000).[9]

3.2.2. Basic results

Table 2 presents the results. In addition to the controls mentioned above, all specifications include year effects and industry effects. The standard errors are clustered at the industry-year level. I report the marginal effect of each variable in the table.

The results for experience-related variables are consistent across all specifications. I find a significant and monotonically increasing relationship between households' experienced outcomes and their future purchases within the same industry. Columns (1)–(4) include all three experience-related variables and divide investors' experienced outcomes into four categories: (1) bottom, (2) bad but not bottom (base category), (3) good but not top, and (4) top. As indicated in Column (1), a household's propensity to buy new stocks in an industry is significantly (1.15 pp) higher when the investor enjoys a good, but not top, experience relative to a bad, but not bottom, experience. The propensity increases by an additional 1.93 pp when the investor has a top experience. In addition, a bottom experience makes the investor even more reluctant to make further purchases in the same industry: the propensity drops 1.29 pp relative to when the investor has a bad, but not bottom, experience. These results are consistent with the experience hypothesis, which states that investors have a higher propensity to purchase new stocks in an industry if they have experienced higher returns from their past investments in that industry.

To better understand the economic magnitude, I include only the dummy indicating a good experience in Column (5). Relative to a bad experience, an investor with a good experience has a 1.86 pp higher probability of buying new stocks in that industry. This magnitude corresponds to a 16.60% increase (= 1.86 pp/11.18 pp) if we normalize it by the average probability of buying new stocks in an industry (11.18 pp).[10] Another way to interpret the magnitude is to compare the top and bottom decile of experienced outcomes. According to Column (4), the likelihood difference between the top and bottom decile is 3.91 pp (= 1.26 pp + 1.46 pp – (– 1.19 pp)), which accounts for a 34.97% difference (= 3.91 pp/11.18 pp), normalized by the average probability of buying new stocks in an industry.

In comparison with the effect of personal experience, the effect of industry average returns is quite different and exhibits a U-shaped relationship. The nonextreme industry average variable does not have a significant influence, but industries with the highest and lowest market-adjusted returns both have positive impacts on households' future purchases. This evidence is consistent with the "attention-grabbing" effect found in Barber and Odean (2007). Individual investors tend to purchase stocks in industries that exhibit large price changes, because the stocks in those industries catch their attention.

The results for the wealth effect and trading characteristics controls are reported in Column (4). Consistent with our expectations, investors have a higher probability

[9] In each month during the sample period, I identify the common stocks held by each household at the beginning of month t from their position statements. To calculate the monthly sales turnover, I match these positions to sales during month t. The monthly sales turnover is calculated as the shares sold times the beginning-of-month price per share divided by the total beginning-of-month market value of the household's portfolio. To calculate the monthly purchase turnover, I match these positions to purchases during month $t-1$. The monthly purchase turnover is calculated as the shares purchased times the beginning-of-month price per share divided by the total beginning-of-month market value of the portfolio. Finally, monthly turnover is calculated by averaging monthly sales turnover and monthly purchase turnover.

[10] In addition, I want to point out that the setting of the baseline model is actually estimating a lower bound of the effect. If the investors are sorted into an industry in which they think they have an information advantage or with which they are more familiar, their experienced outcomes can have less influence. And as the investors get more and more experience in an industry, the effect of new experience tends to decrease. Ideally, I would like to identify the industries that investors are not sorted into (for example, the investors are attracted by an exogenous event and then start to invest in an industry) and be able to observe investors' trading records from their first entrance into that industry. Therefore, given the setting of current dataset, the estimated magnitude of the experience effect can be dampened due to sorting or diminishing influence as investors accumulate more and more experience in a given industry.

Table 2
Investment experience and the propensity to purchase new stocks in the experienced FF10 industry.

The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one industry, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is coded as one if the household owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics variables. Both personal experience variables and industry average variables are based on the value of market-adjusted (experienced or industry average) returns. "Good experience" ("Good industry") is coded as one when the market-adjusted experienced (industry average) return is greater than zero. "Top experience" ("Bottom experience") is coded as one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. "Top 1 industry" ("Bottom 1 industry") is one when an industry's average return is the highest (lowest) among the ten industries. "Increase of portfolio size" equals one if the size of a household's portfolio increased in the past year. The three individual characteristics variables are created from beginning-of-month position data. They denote the average monthly number of stocks, the average monthly size, and the average monthly turnover rate in the past year. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Experience dummy	0.07982***	0.08760***	0.09239***	0.06858***	0.06571***
	(0.00559)	(0.00640)	(0.00810)	(0.00943)	(0.00937)
Good experience (> 0)	0.01148***	0.01069***	0.01271***	0.01260***	0.01856***
	(0.00266)	(0.00246)	(0.00283)	(0.00298)	(0.00312)
Top experience (over the 90th percentile)	0.01925***	0.01879***	0.02011***	0.01459***	
	(0.00492)	(0.00488)	(0.00472)	(0.00415)	
Bottom experience (below the 10th percentile)	−0.01288***	−0.01263***	−0.01468***	−0.01185***	
	(0.00211)	(0.00204)	(0.00251)	(0.00333)	
Good industry (> 0)	0.00396	0.00575	0.00653	0.00621	0.00618
	(0.00580)	(0.00444)	(0.00525)	(0.00688)	(0.00689)
Top 1 industry		0.02282**	0.02914**	0.03682**	0.03703**
		(0.01128)	(0.01276)	(0.01433)	(0.01446)
Bottom 1 industry		0.02667**	0.02845**	0.03109**	0.03132**
		(0.01164)	(0.01312)	(0.01490)	(0.01499)
Increase of portfolio size			0.01833***	0.01489***	0.01482***
			(0.00202)	(0.00168)	(0.00168)
Average num of stocks > 5				0.05680***	0.05685***
				(0.00204)	(0.00205)
log(Average portfolio size)				0.01528***	0.01539***
				(0.00071)	(0.00072)
log(Average turnover rate)				0.00832***	0.00851***
				(0.00076)	(0.00078)
Industry effect	Yes	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes	Yes
Avg prob (buying new stocks in an industry)	0.0846	0.0846	0.0930	0.1118	0.1118
Observations	1,550,980	1,550,980	1,094,860	777,630	777,630

of buying new stocks when they earned money in the past year, trade more frequently, and have a larger portfolio invested in the stock market.

As a side note, the marginal effect of the experience dummy is also significantly positive. There are several ways to interpret this coefficient. One interpretation is that it implies a significantly positive unconditional effect of prior investment experience. In other words, regardless of outcome, personal involvement, on average, can have a positive effect on the probability of future purchases in that industry. For example, the involvement can trigger investors' attention, making the stocks in this industry more likely to enter the choice set for future purchases. However, the positive coefficient is also likely to result from sorting. Investors are sorted into certain industries because they, for example, have worked in those industries and thus have information advantages. The experience dummy is not the focus of this paper, and its sign can arise from multiple interpretations, but it is important to include it as a control in the regression. Moreover, if I drop all industry-year observations without past experience and run the same specification without the experience dummy, the results remain the same, as shown in Panel B in Table 3.

3.3. Robustness tests

These results are robust to alternative experience measures, subsample, explanations, and industry classifications. Table 3 presents a series of robustness tests.

Instead of using dummy variables to measure experienced outcomes and industry average returns, Panel A directly applies the level of market-adjusted experienced returns and industry average returns. The coefficient on the experienced return is significant and positive, confirming that investors have a higher propensity to buy new stocks in an industry if they earned higher returns in the same industry in the past. The effect of the industry average return is nonlinear and presents in a U-shape, which is different from the effect of past experience and consistent with the evidence displayed in Table 2.

In Panel B, I drop the industry-year observations if the household has not invested in the industry during the past

Table 3
Robustness tests.

The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Panel A exploits the level of market-adjusted experienced and industry average returns as the experience and industry average variables, instead of dummy variables. Panel B only includes observations with experience in the industry in the past year. Panel C includes industry-time fixed effects instead of time-varying industry variables and time fixed effects. In Panel D, the dependent variable is defined as one if the household purchases stocks not previously owned in the given industry and does not have sales in that industry within 30 days prior to the purchases. In Panel E, the experience variables are limited to experience with stale positions. In other words, the experience variables are only based on the returns from the beginning of the experienced year if the individual held stocks in the industry the previous year and contiuned to passively held them in the experienced year. Panel F creates investment experiences based on the Fama-French 48-subindustry classification system. Without specification, each observation corresponds to a household/industry/year group, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is coded as one if the household owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. Both personal experience variables and industry average variables are based on the value of market-adjusted (experienced or industry average) returns. "Good experience" ("Good industry") is coded as one when the market-adjusted experienced (industry average) return is greater than zero. "Top experience" ("Bottom experience") is coded as one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. "Top 1 industry" ("Bottom 1 industry") is one when an industry's average return is the highest (lowest) among the ten industries. The wealth effect and individual characteristics controls are defined as in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

Panel A: Measure of experienced outcomes by the level of market-adjusted returns

	(1)	(2)	(3)	(4)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry			
Expereience dummy	0.08553***	0.08912***	0.09281***	0.07058***
	(0.00356)	(0.00453)	(0.00474)	(0.00495)
Experienced excess return	0.00119***	0.00119***	0.00121***	0.00124***
	(0.00030)	(0.00029)	(0.00040)	(0.00042)
Industry average excess return	0.02065	−0.01171	−0.00949	−0.00855
	(0.02785)	(0.02553)	(0.02769)	(0.03034)
Square of industry average excess return		0.48734***	0.50777***	0.57314***
		(0.16751)	(0.18348)	(0.21884)
Wealth effect control	No	No	Yes	Yes
Individual characteristics control	No	No	No	Yes
Industry effect	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes
Observations	1,550,980	1,550,980	1,094,860	777,630

Panel B: only industry-year observations with past experience

	(1)	(2)	(3)	(4)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry			
Good experience (> 0)	0.01385***	0.01256***	0.01494***	0.01412***
	(0.00449)	(0.00437)	(0.00470)	(0.00458)
Top experience (over the 90th percentile)	0.01125*	0.01006*	0.00979*	0.00375
	(0.00613)	(0.00598)	(0.00578)	(0.00415)
Bottom experience (below the 10th percentile)	−0.01182**	−0.01156**	−0.01486***	−0.01350***
	(0.00500)	(0.00494)	(0.00425)	(0.00451)
Good industry (> 0)	0.00437	0.00207	0.00045	−0.00068
	(0.01134)	(0.00955)	(0.01019)	(0.01083)
Top 1 industry		0.05638***	0.07273***	0.08898***
		(0.02056)	(0.02073)	(0.02073)
Bottom 1 industry		0.03172	0.03165	0.03378
		(0.02477)	(0.02615)	(0.02684)
Wealth effect control	No	No	Yes	Yes
Individual characteristics control	No	No	No	Yes
Industry effect	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes
Observations	385,552	385,552	278,427	225,142

Panel C: Controlling for industry-year fixed effects

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Experience dummy	0.07826***	0.07913***	0.08222***	0.06060***	0.05757***
	(0.00376)	(0.00314)	(0.00374)	(0.00387)	(0.00367)
Good experience (> 0)	0.01191***	0.01133***	0.01342***	0.01380***	0.01952***
	(0.00293)	(0.00214)	(0.00260)	(0.00276)	(0.00307)
Top experience (over 90th percentile)	0.01908***	0.01681***	0.01818***	0.01234***	
	(0.00487)	(0.00446)	(0.00448)	(0.00384)	

(*continued on next page*)

Table 3 (*continued*)

Panel C: Controlling for industry-year fixed effects					
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Bottom experience (below 10th percentile)	−0.01293***	−0.01338***	−0.01545***	−0.01280***	
	(0.00210)	(0.00195)	(0.00232)	(0.00305)	
Increase of portfolio size			0.01882***	0.01530***	0.01525***
			(0.00200)	(0.00162)	(0.00162)
Average num of stocks > 5				0.05676***	0.05684***
				(0.00170)	(0.00171)
log(Average portfolio size)				0.01521***	0.01531***
				(0.00058)	(0.00059)
log(Average turnover rate)				0.00829***	0.00845***
				(0.00074)	(0.00077)
Industry effect	Yes	No	No	No	No
Year effect	Yes	No	No	No	No
Industry X year effect	Yes	Yes	Yes	Yes	Yes
Avg prob (buying new stocks in an industry)	0.0846	0.0846	0.0930	0.1118	0.1118
Observations	1,550,980	1,550,980	1,094,860	777,630	777,630

Panel D: Robustness Check for Mental Accounting					
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks in one industry without recent sales				
Experience dummy	0.07781***	0.08581***	0.09048***	0.06635***	0.06466***
	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)
Good experience (> 0)	0.00562**	0.00478*	0.00669**	0.00687**	0.00917***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Top experience (over the 90th percentile)	0.00642*	0.00580	0.00536	0.00306	
	(0.004)	(0.004)	(0.003)	(0.003)	
Bottom experience (below the 10th percentile)	−0.00683**	−0.00658**	−0.00834***	−0.00783***	
	(0.003)	(0.003)	(0.003)	(0.003)	
Good industry (> 0)	0.00405	0.00574	0.00654	0.00625	0.00628
	(0.006)	(0.004)	(0.005)	(0.007)	(0.007)
Top 1 industry		0.02289**	0.02922**	0.03681**	0.03691**
		(0.011)	(0.013)	(0.014)	(0.014)
Bottom 1 industry		0.02660**	0.02850**	0.03106**	0.03118**
		(0.012)	(0.013)	(0.015)	(0.015)
Wealth effect control	No	No	Yes	Yes	Yes
Individual characteristics control	No	No	No	Yes	Yes
Industry effect	Yes	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes	Yes
Avg prob (buying new stocks in an industry)	0.0831	0.0831	0.0913	0.0808	0.0808
Observations	1,550,980	1,550,980	1,094,860	777,630	777,630

Panel E: Robustness check for information story (eliminating experience returns after changes in positions)					
	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Experience dummy	0.07796***	0.08540***	0.09095***	0.06902***	0.06675***
	(0.00537)	(0.00634)	(0.00797)	(0.00961)	(0.00956)
Good experience (> 0)	0.00893***	0.00810***	0.00954***	0.00969***	0.01454***
	(0.00266)	(0.00245)	(0.00292)	(0.00311)	(0.00327)
Top experience (over 90th percentile)	0.01455***	0.01414***	0.01600***	0.01222***	
	(0.00408)	(0.00402)	(0.00433)	(0.00412)	
Bottom experience (below 10th percentile)	−0.00955***	−0.00936***	−0.01140***	−0.00935***	
	(0.00220)	(0.00214)	(0.00265)	(0.00363)	
Good industry (> 0)	0.00393	0.00544	0.00634	0.00626	0.00622
	(0.00571)	(0.00439)	(0.00518)	(0.00683)	(0.00683)
Top 1 industry		0.02298**	0.02914**	0.03677**	0.03693**
		(0.01099)	(0.01253)	(0.01432)	(0.01442)
Bottom 1 industry		0.02572**	0.02773**	0.03036**	0.03052**
		(0.01152)	(0.01302)	(0.01485)	(0.01491)
Wealth effect control	No	No	Yes	Yes	Yes
Individual characteristics control	No	No	No	Yes	Yes
Industry effect	Yes	Yes	Yes	Yes	Yes

(*continued on next page*)

Table 3 (*continued*)

Panel E: Robustness check for information story (eliminating experience returns after changes in positions)

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Year effect	Yes	Yes	Yes	Yes	Yes
Industry X year effect	No	No	No	No	No
Avg prob (buying new stocks in an industry)	0.0846	0.0846	0.0930	0.1118	0.1118
Observations	1,550,980	1,550,980	1,094,860	777,630	777,630

Panel F: Fama-French 48-industry Classification

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry				
Experience dummy	0.02205***	0.02349***	0.02404***	0.01583***	0.01550***
	(0.00217)	(0.00232)	(0.00297)	(0.00252)	(0.00251)
Good experience (> 0)	0.00163**	0.00154**	0.00190***	0.00175**	0.00276***
	(0.00072)	(0.00061)	(0.00073)	(0.00076)	(0.00077)
Top experience (over the 90th percentile)	0.00385***	0.00370***	0.00408***	0.00304***	
	(0.00080)	(0.00073)	(0.00078)	(0.00068)	
Bottom experience (below the 10th percentile)	−0.00120**	−0.00121**	−0.00142**	−0.00121	
	(0.00054)	(0.00051)	(0.00062)	(0.00075)	
Good industry (> 0)	0.00159*	0.00095	0.00143	0.00189	0.00189
	(0.00096)	(0.00116)	(0.00141)	(0.00160)	(0.00160)
Top 5 industries		0.00313**	0.00341***	0.00400***	0.00400***
		(0.00125)	(0.00132)	(0.00147)	(0.00147)
Bottom 5 industries		0.00094	0.00109	0.00115	0.00116
		(0.00181)	(0.00216)	(0.00244)	(0.00244)
Wealth effect control	No	No	Yes	Yes	Yes
Individual characteristics control	No	No	No	Yes	Yes
Industry effect	Yes	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes	Yes
Avg prob (buying new stocks in an industry)	0.0262	0.0262	0.0216	0.0196	0.0196
Observations	6,990,528	6,990,528	4,961,136	3,490,752	3,490,752

period. Then, I run specification (3) without the experience dummy. As summarized in Panel B of Table 3, the results remain the same, indicating a monotonically increasing relationship between experienced outcomes and future purchases within the same industry.

Another way to control for industry momentum is to include industry-time fixed effects, instead of industry average return variables. This is a stronger control and a simpler way to identify experience effects across individual variations within industry and time. Panel C reports the results with industry-time fixed effects, which are qualitatively and quantitatively similar to those with time-varying industry variables and time fixed effects.

Panel D considers an alternative story: mental accounting. If investors apply mental accounting (Thaler, 1999) and regard each industry as an account, they may rebalance portfolios only within an industry but not across industries. Such investors purchase stocks in an industry with the money from recent sales in that industry. Due to the disposition effect (Odean, 1998), sales are more likely to happen following good experiences. Hence, we are more likely to observe purchases following good experiences. To address this issue, I exclude purchases that happen within the 30 days that follow the household's most recent sale in the same industry. In other words, the dependent variable is a dummy variable that indicates purchases of new stocks without recent sales. The results are virtually unchanged, though the magnitude does drop slightly. According to Column (5), relative to bad experience, good experience increases an investor's propensity to buy new stocks without recent sales by 0.92 pp, which corresponds to an increase of 11.35% (= 0.92 pp/8.08 pp), if normalized by the average probability of buying new stocks in an industry.

Panel E considers another potential confounding story regarding the role of information. If the experienced outcomes occurred because of receiving a signal (e.g., a stock tip) on the industry level, individuals may have increased their confidence in the signal after experiencing high returns and so subsequently increase their investment in the industry. If we assume that "tips" tend to generate superior returns at shorter horizons, the story of superior private information is then unlikely to be relevant to experienced outcomes after we eliminate all returns experience arising after changes in positions. In Panel E, I construct the experience variables using only returns from the beginning of the experienced year if the individual still passively held stocks in the industry from the previous year. The results remain the same, suggesting that the results cannot be fully explained by the story of superior private information.

I also exploit the Fama-French 48-subindustry classification as a robustness check. To make the definitions of the top and bottom industries match with the definitions of the top and bottom experiences, I define the top (bottom) industries as those with the top five (bottom five) for industry average returns. The results are reported in Panel

F and are consistent with those derived using the Fama-French ten-industry classification. Moreover, Column (5) implies that the economic magnitude of the experience effect on the Fama-French 48-subindustry level is also close to that on the Fama-French ten-industry level, as shown in Table 2; the effect of a good experience in one of the 48 Fama-French sub-industries corresponds to an increase of 14.08% (= 0.27 pp/1.96 pp) in the probability of future purchases in that industry, if normalized by the average probability of buying new stocks in an industry.[11] I will further analyze how the experience effect varies with different levels of categorization in Section 5.

3.4. Dynamic effects

I have thus far investigated the influence of prior stock investment experience on a one-year horizon. In this section, I explore whether the effect lasts for a longer time, and how the significance and magnitude change for lagged experience at different time horizons.

To show more detail about the dynamic effects, I use a monthly window in this section. I estimate a vector autoregression (VAR) model with 20 lags using linear probability. Because some households open accounts or stop trading during the sample period, their trading records do not cover the entire sample period and may not be applicable for testing the long-term effects of experience. Therefore, for this analysis, I restrict the sample to households that opened accounts prior to 1991 and remain active traders in 1996.

Fig. 3 shows the impulse response function of the probability of purchasing new stocks in an industry to "Good experience" in the same industry. The figure shows that the effect of experience decreases with the progression of time—the more recent the experience, the stronger the effect. The impulse responses of a "Good experience" are significantly positive for the near future (specifically, the next 16 months) but statistically insignificant for future periods more than 17 months away. The impulse responses of monthly experiences for the next year are, on average, 0.30 pp. This magnitude is lower than that of the baseline model because of the change to monthly frequency. However, when normalized by the average probability of buying new stocks in an industry in the decision month (1.83 pp), the magnitude still remains at the same level; good experience increases an investor's propensity to buy new stocks in the same industry by about 16.39% (= 0.30 pp/1.83 pp).

3.5. Summary

Investors' future purchases are influenced by their previously experienced returns. If they experience a higher return in an industry, their propensity to purchase new stocks in that industry increases. This effect is robust after controlling for industry momentum trading, the wealth effect, investor heterogeneity, and mental accounting. The results remain the same using different subsamples, measures of experience, and industry classifications. The influence of experienced outcomes becomes weaker as the experiences recede further into the past; after 16 months have passed, the experience has no significant effect.

4. Underlying mechanisms and implications

The results of the baseline model show a significantly positive relationship between an investor's good experience and probability of buying new stocks in the same industry. This section further examines the underlying mechanisms. Does the relationship vary with investor sophistication and the concentration of investors' direct stock investments? What are the performance implications for investors' portfolios? What channels best explain the experience effect?

4.1. Investor sophistication

The influence of experienced outcomes can vary with investor sophistication. If the experience hypothesis is true, the influence of experienced outcomes on future stock purchase decisions is likely to be stronger for less sophisticated investors, because given the longer history of more sophisticated investors, recent good or bad experienced outcomes may be canceled by previous ones and have less influence.

I use self-reported experience/knowledge as a proxy for investor sophistication. When a household opened an account with the discount brokerage, they were asked to classify their perception of their experience in or knowledge of investment in one of four levels: extensive, high, limited, or none.[12]

I conduct the regressions from specification (3) separately for each group of investors and see how the coefficients change among the subgroups. The results are presented in Table 4. As shown, the marginal effect of good experience becomes weaker as investors become more sophisticated. For the most sophisticated subgroup with extensive knowledge about investing, the coefficient is insignificant. The gap of the marginal effect of good experience is statistically significant ($p\,value < 0.05$) between

[11] I want to point out that even though individuals apply a finer categorization scheme than the FF10 industry classification, the magnitude would not become stronger if we considered the effect of good experience within a finer categorization (e.g., the FF48 subindustries) than the FF10 industry scheme. It's important to understand how the number of companies in an industry group affects inference when analyzing this issue. This obviously affects how we think about finer industry categorization (e.g., FF48) versus coarse industry categorization (e.g., FF10), since the coarse categorization yields far more companies in each industry group. Consider an individual who bought a stock last period that resulted in a good experience. Let P(industry i) denote the probability of purchasing another stock in industry i. Suppose the individual categorizes stocks based on the FF48 sub-industry classification, then P(the same FF48 subindustry) > P(another FF48 subindustry). Therefore, P(the same FF10 industry) = P(the same 48 subindustry) + P(another FF48 subindustry within the same FF10 industry), which suggests that P(the same FF10 industry) ≥ P(the same FF48 subindustry).

[12] Because of missing self-reported data, the sample used in this part is much smaller than the whole sample, but in unreported tables, I show that the main summary statistics of the distribution of trades among industries and distribution of experience across industries and years don't change much.

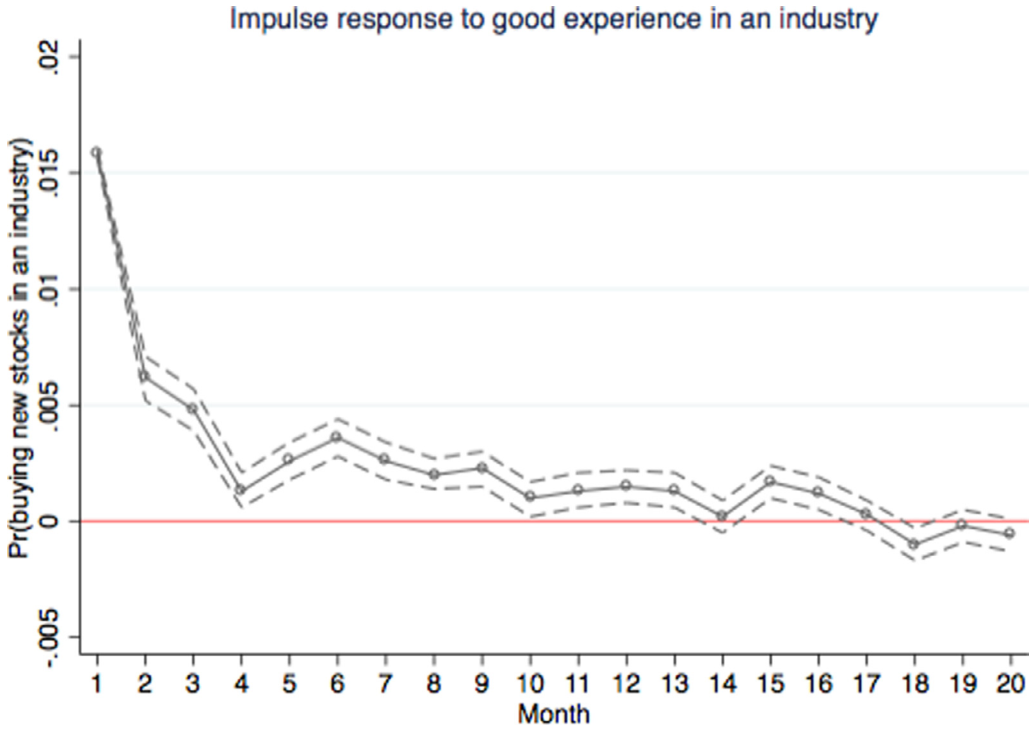

Fig. 3. Dynamic effects of monthly lagged experienced outcomes in an industry. The figure shows the dynamic effects of monthly experienced outcomes on subsequent purchases of new stocks. The sample includes all investors who already had investments in 1991 and were still trading in 1996 in the dataset. The solid line presents the impulse response function of the probability of purchasing new stocks in an industry to "Good experience" in the same industry. "Good experience" is defined as the market-adjusted experienced return is greater than zero. The impulse responses are from an estimated vector autoregression (VAR) model with 20 lags. The dashed lines indicate 90% confidence bands.

the no knowledge group and the extensive knowledge group. In terms of magnitude, for the investors with no knowledge, the difference between the top and bottom decile is 3.31 pp, which accounts for a 39.40% difference (3.31 pp/8.40 pp), relative to the average probability of buying new stocks in an industry. The difference between the top and bottom decile (1.36 pp) is much smaller for the investors with extensive knowledge and only accounts for a 10.37% difference (1.36 pp/13.12 pp), relative to the average probability of buying new stocks in an industry.

This evidence is consistent with the experience hypothesis. The most sophisticated investors have a long history with many experiences and can acquire specialized skills in investing in specific industries through their historical experienced outcomes. They are more likely to make new purchases within these specialized industries but are not affected much by their recent outcomes. In contrast, the least sophisticated investors are likely new to stock investing and do not have a long history of experience to cancel out recently experienced outcomes. Consequently, they are more likely to switch across industries depending on whether their experienced outcomes in that industry are good or bad.

4.2. Directly held portfolio concentration

I next explore how the experience effect varies with the concentration of investors' directly held portfolios. Investors with different concentrations in their directly held portfolios can take different approaches to investing. Investors with more concentrated portfolios can devote themselves to seeking stocks with extraordinary alphas, while others can attempt to diversify their portfolios to reduce idiosyncratic risk. Note that, given that the sample only covers households' directly held stock positions with the broker, we focus our discussion on the concentration of households' directly held stock portfolios or nondiscretionary accounts,[13] rather than their overall portfolios, including retirement plans, mutual funds, real estate, and human capital. The concentration of their directly held portfolio has been shown to be relevant for their trading behavior and performance in these positions (Ivkovic, Sialm, and Weisbenner, 2008; Goetzmann and Kumar, 2008).

I hypothesize that past investment experience matters more to alpha seekers than to diversification-oriented investors. People who put all their eggs in one basket have to pick the best basket, and their past experience might play a role in their decision-making. In contrast, people who put their eggs in multiple baskets would have less concern about their choice of baskets. Therefore, I expect to see that the relationship between previously experienced outcomes and future new purchases in the same industry is stronger for investors with more concentrated portfolios.

[13] Ivkovic, Sialm, and Weisbenner (2008) compare this sample with the Survey of Consumer Finance and show that the common stock positions that the households hold in their accounts with this broker likely represent the entire stock holdings for many households.

Table 4
Variation in investor sophistication.
The table reports maximum likelihood regression results for probit regressions for a subgroup sample. The sample is divided into four groups by self-reported investing experience (knowledge): Extensive, Good, Limited, and None. The results are reported as the marginal effects of independent variables. Each observation corresponds to a household/industry/year group, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The “Experience dummy” is coded one if the household owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. Both personal experience variables and industry average variables are based on the value of market-adjusted (experienced or industry average) return. “Good experience” (“Good industry”) is coded as one when the market-adjusted experienced (industry average) return is greater than zero. “Top experience” (“Bottom experience”) is one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. “Top 1 industry” (“Bottom 1 industry”) is one when an industry’s average return is the highest (lowest) among the ten industries. “Increase of portfolio size” is coded as one when the size of a household’s portfolio has increased in the past year. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)	(4)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry			
Sophistication group:	None	Limited	Good	Extensive
Experience dummy	0.07743***	0.07533***	0.08738***	0.11594***
	(0.01099)	(0.00491)	(0.00460)	(0.00830)
Good experience (> 0)	0.01935*	0.01128**	0.01249***	0.00563
	(0.01113)	(0.00469)	(0.00378)	(0.00721)
Top experience (over the 90th percentile)	−0.01203	0.00257	0.00867	0.00118
	(0.01390)	(0.00664)	(0.00626)	(0.00804)
Bottom experience (below the 10th percentile)	−0.02573**	−0.00688	−0.00763	−0.00677
	(0.01307)	(0.00463)	(0.00548)	(0.00952)
Good industry (> 0)	−0.01277	0.00522	0.00473	−0.00526
	(0.01210)	(0.00489)	(0.00634)	(0.00859)
Top 1 industry	0.02227	0.03381***	0.04296***	0.06025**
	(0.02009)	(0.01094)	(0.01500)	(0.02479)
Bottom 1 industry	−0.00019	0.01045	0.02446	0.02722
	(0.01730)	(0.01063)	(0.01557)	(0.02082)
Increase of portfolio size	0.01895***	0.01352***	0.02015***	0.00978***
	(0.00566)	(0.00310)	(0.00281)	(0.00346)
Industry effect	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes
Avg prob (buying new stocks in an industry)	0.0840	0.0824	0.1086	0.1312
Observations	43,580	143,570	97,440	10,080

I test this hypothesis by adding an interaction term between the experience-related variable and a concentration measure to the baseline model. The specification is:

$$\begin{aligned} P(B^{New}_{hI,t+1} = 1) = \Phi(&\beta_0 + \beta_1 Exp_{hIt} + \beta_2 Goodexp_{hIt} \\ &+ \beta_3 Exp_{hIt} \cdot LessConcentrated_{ht} \\ &+ \beta_4 Goodexp_{hIt} \cdot LessConcentrated_{ht} \\ &+ \beta_5 LessConcentrated_{ht} + \Gamma' X_{hIt}). \end{aligned} \quad (4)$$

I use three measures of the concentration of the directly held portfolio, following Ivkovic, Sialm, and Weisbenner (2008). The first measure is a portfolio Herfindahl Index (HI), which is defined as the sum of the squared weights of each stock k in the household stock portfolio (w_k) (i.e., $HI = \sum_k w_k^2$). The less concentrated the portfolio is, the smaller the Herfindahl index. If a household owns only one common stock, the Herfinndahl index reaches its maximum and equals one. I define the dummy variable, $LessConcentrated_{it}$, as equal to one when the Herfindahl index of investor i’s portfolio is smaller than the median in the sample. The second measure is the number of stocks held in the portfolio. A portfolio with more stocks is considered to be less concentrated. Similarly, I define the $LessConcentrated_{it}$ variable as equal to one when the number of stocks held is greater than the sample median, which is five. The third measure is based on the number of industries instead of the number of stocks. $LessConcentrated_{it}$ is equal to one when the number of industries held is greater than the sample median, which is two.

According to the results, shown in Table 5, the influence of experienced outcomes is stronger among investors who hold more concentrated portfolios, since the coefficient on the interaction term between good experiences and diversified portfolios is negative and statistically significant. For investors with concentrated portfolios, good experience in an industry increases the probability of buying new stocks in the same industry by 1.32 pp. This effect is dampened by about two-thirds when the portfolio is more diversified.

4.3. Possible explanations

In this section, I explore potential explanations for the experience effect.[14] These explanations may not be mutually exclusive. I examine some additional testable predic-

[14] Experience could also influence investors’ behavior through other channels. For example, in the context of Barber, Odean, and Strahilevitz (2011), previously experienced buying or selling prices could affect investors’ reference points and further influence investors’ actions, since investors tend to avoid anticipated regret. Because my research focuses on the decision to buy stocks not previously owned, the role of affecting the investor’s reference points may not apply.

Table 5
Variation in directly held portfolio concentration.
The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to a household/industry/year group, regardless of whether the household previously owned stocks in the industry or not. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is coded as one if the household has owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. Both personal experience variables and industry average variables are based on the value of market-adjusted (experienced or industry average) returns. "Good experience" is one when the market-adjusted experienced return is greater than zero. "Top experience" ("Bottom experience") is one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. The independent variables also include the interaction terms between experience-related variables and portfolio concentration variables. The "Less concentrated" variable equals one when either (1) the Herfindahl index[15] is smaller than the median; (2) the number of stocks in the portfolio is greater than the sample median, which is five; or (3) the number of industries in the portfolio is greater than the sample median, which is two. The industry average return variables, the wealth effect, and individual characteristics controls are also included. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry		
Diversified portfolio measured by:	Herfindahl $<$ Median	Num of stocks $>$ 5	Num of industries $>$ 2
Experience dummy	0.06365***	0.06163***	0.06554***
	(0.00972)	(0.00960)	(0.01073)
Good experience ($>$ 0)	0.01392***	0.01324***	0.01395**
	(0.00429)	(0.00360)	(0.00603)
Top Experience (over 90th percentile)	0.00240	0.00256	0.00957
	(0.00527)	(0.00387)	(0.00667)
Bottom Experience (over 10th percentile)	−0.01052***	−0.00927***	−0.01586***
	(0.00371)	(0.00281)	(0.00457)
Experience $\times$ Less concentrated	0.00997***	0.01362***	−0.00083
	(0.00226)	(0.00204)	(0.00294)
Good experience $\times$ Less concentrated	−0.00756**	−0.00853***	−0.00821*
	(0.00306)	(0.00265)	(0.00451)
Top experience $\times$ Less concentrated	0.00229	0.00254	−0.01028**
	(0.00546)	(0.00412)	(0.00500)
Bottom experience $\times$ Less concentrated	0.00458	0.00301	0.00638
	(0.00522)	(0.00363)	(0.00526)
Less concentrated	0.04277***	0.05231***	0.05078***
	(0.00181)	(0.00198)	(0.00209)
Industry return control variable	Yes	Yes	Yes
Wealth effect control	Yes	Yes	Yes
Individual characteristics control	Yes	Yes	Yes
Industry effect	Yes	Yes	Yes
Year effect	Yes	Yes	Yes
Observations	777,630	777,630	777,630

tions to determine whether they are supported by the data and discuss which explanations may be more relevant to explain most of my findings and what implications these explanations have for investors' welfare.

Naive reinforcement learning. As I motivated in the introduction, a natural explanation for why experienced outcomes affect purchase decisions is that investors may put more weight on their experienced outcomes than on other available historical information when undertaking a Bayesian updating of their beliefs about stocks in the same industry. Good experiences drive posterior expected returns upward, and consequently investors are more likely to buy new stocks in that industry.

Under this explanation, investors' past experiences may bias their beliefs and lead them to miss good investment opportunities. As one practitioner notes, "The problem is that in accumulating experience, he also acquires prejudices against industries and stocks because he has lost money in them. It is easy to . . . become an investment bigot with a closed mind on many subjects."[16] To infer whether this behavior hurts investors' performances, I will later examine investors' industry picking ability by comparing counterfactual performances between buying another stock in the chosen industry and a stock in a different industry.

Learning about ability. Another explanation is that investors can construe their experienced outcomes as indications of their ability to invest in a particular industry. Consider an investor under the framework of Nieuwerburgh and Veldkamp (2010) with specialized information acquisition ability,[17] who first chooses to acquire information on the assets that she is well informed about and then makes the investment decisions based on the acquired signals. If

[15] The Herfindahl index of the portfolio is defined as $\sum_k w_k^2$, where w_k denotes the portfolio weight allocated in stock k and $\sum_k w_k = 1$.

[16] See http://dailyreckoning.com/the-worst-possible-investing-mistake/

[17] In Nieuwerburgh and Veldkamp (2010), given different investors' preferences and information learning technology, the framework could generate either specialized information acquisition (i.e., investors specialize in assets they are well informed about) or generalized information acquisition (i.e., investors broaden their knowledge by acquiring information on assets they are most uncertain about). Specialized information acquisition will be more relevant when explaining the experience effect shown in this paper.

Table 6
Return performance: actual purchases versus counterfactual portfolios.

The table reports the monthly abnormal returns of calendar-time portfolios controlling for the Fama-French three factors (Fama and French, 1993). For each day, the calendar-time portfolio of actual purchases is formed to include all stocks bought by households within the prior 21 trading days. The counterfactual portfolios are formed as follows: (1) invested industries: to include all industries within which households have bought stocks during the prior 21 trading days; (2) not invested industries: to include all industries within which households have not bought stocks during the prior 21 trading days; (3) beginning-of-the-year portfolio: to include the beginning-of-the-year portfolio for each purchase made by households during the prior 21 trading days; (4) pure experienced-based strategy: to include all industries that gave good experiences in the previous year for each purchase made by households during the prior 21 trading days; and (5) experience-neutral industry-level momentum: to include the industry with highest past-year return for each purchase made by households during the prior 21 trading days. The returns of the portfolio are then accumulated within each month to obtain monthly returns. The difference between actual purchases and the counterfactual portfolios returns is the monthly abnormal return of going long on the portfolio of actual purchases and short on the counterfactual portfolios. Standard errors are shown in parentheses. *10%, **5%, ***1% significance.

	Portfolio 3-factor alpha		Difference (actual − counterfactual)	
Actual purchases	−0.0047	(0.0028)		
Counterfactual portfolios:				
(1) Invested industry	0.0023**	(0.0011)	−0.0070***	(0.0023)
(2) Not invested industry	0.0001	(0.0006)	−0.0048	(0.0031)
(3) Beginning-of-year portfolio	−0.0009	(0.0017)	−0.0037**	(0.0018)
(4) Pure experience-based strategy	0.0018**	(0.0007)	−0.0065**	(0.0026)
(5) Experience-neutral industry-level momentum	0.0090***	(0.0029)	−0.0137***	(0.0033)

Table 7
The impact of personal investment outcomes on subsequent trade frequencies in an industry.

The table reports the results of Tobit regression of the number of trades in an industry during the subsequent year on the investor's investment outcomes in that industry. Each observation corresponds to a household/industry/year group. The observation is only included if the household previously owned stocks in that industry. The dependent variable is the number of trades the investors execute in the experienced industry, including both buys and sells. The personal experience outcome variables are based on the value of their market-adjusted experienced returns. "Good experience" is one when the market-adjusted experienced return is greater than zero. "Top experience" ("Bottom experience") is one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. The control variables include industry average variables, portfolio size, and individual characteristics. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)
Dependent variable:	Number of trades in the same industry in the subsequent year	
Good experience (> 0)	0.62076***	0.45660***
	(0.09937)	(0.08877)
Top experience (over the 90th percentile)	1.86391***	1.25045***
	(0.21580)	(0.13850)
Bottom experience (below the 10th percentile)	−0.05677	0.10375
	(0.09374)	(0.13370)
Good industry (> 0)	−0.08504	−0.08349
	(0.19169)	(0.18384)
Top 1 industry	1.03377***	1.44166***
	(0.37715)	(0.30588)
Bottom 1 industry	0.64351	0.59849
	(0.42882)	(0.38981)
Wealth effect control	No	Yes
Individual characteristics control	No	Yes
Industry effect	Yes	Yes
Year effect	Yes	Yes
Observations	385,552	225,142

she (either correctly or incorrectly) learns that she has an advantage in collecting information about the energy industry when her energy stocks outperform the market, she allocates more capacity to learning about the energy industry and hence is more likely to purchase stocks in the energy industry.

Table 7 provides suggestive evidence for this explanation. I run a Tobit regression of the number of trades in one industry (including both purchases and sales) on the previous year's experience-related variables pertinent to that industry, controlling for the industry average and individual characteristics. The results show that investors will trade more frequently in an industry after they have had a good experience in that industry. This evidence could not be explained if naive reinforcement learning were the only channel for the experience effect, because higher posterior expected returns after good experiences only imply higher probability of future purchases but have no implications for selling decisions. The fact that investors both purchase and sell more frequently in an industry in which they have had good past experiences supports the existence of the "learning about ability" story.

There are two possible scenarios that explain how investors learn about their ability, given different priors about their information acquisition ability across industries. In the first scenario, investors have evenly distributed priors about their ability across industries. They start with a more diversified portfolio and experiment with different types of industries. After learning about their advantages in a certain industry that yielded beneficial results (good experiences), they specialize their information acquisition in that certain industry and hold a more concentrated portfolio. In the second scenario, investors initially believe they are better informed about a certain industry than others (due, e.g., to geographic proximity, past work experiences, etc.), so they start with a more concentrated portfolio. They can switch to other industries later on if they learn that their ability is inferior from a series of bad experiences.

These two scenarios can both explain the higher propensity of purchases in the industry with good past experiences, but can have different implications for investors' portfolio performances, which will be further discussed in Section 4.4. Here, to see which scenario is more relevant to the data, we can go back to the results in Table 5. In the first scenario, the experience effect would be stronger for investors who hold a more diversified (directly owned) stock portfolio; in the second scenario, the experience effect would be stronger for investors whose stock portfolios are more concentrated. The results in Column (3) of Table 5 are more consistent with the second scenario and suggest, at least to some extent, that the second scenario may be more relevant to the data.

Learning by trading. A third possibility is that investing in an industry can help investors improve their ability to process information and obtain more precise information about that industry for future investments (Seru, Shumway, and Stoffman, 2010). If good outcomes improve an investor's ability to process information about that industry more than bad outcomes do,[18] the next time they invest, investors will have a higher propensity to collect information about and invest in the industry. To identify whether investors improve their trading ability by category is beyond the scope of this paper and is not easy to do with a short sample period in the current dataset, but in light of the performance comparison later in Section 4.4, we can infer that either investors improve their ability but have a fairly low level of ability to start with, or they do not learn much by trading. Given that the stocks picked by investors in the sample are far inferior in terms of performance than the diversified portfolio strategy, chances are that investors do not learn much through past trading experience.

4.4. Implications on portfolio performances

How does the experience effect influence investors' portfolio performances? To answer this question, I compare the returns of actual purchases with the performances of several counterfactual portfolios:

(1) Invested industry: I assume investors buy another stock in the industry from which the investors have picked stocks.
(2) Not invested industry: I assume investors buy a stock in a different industry from which investors could pick but choose not to.
(3) Beginning-of-year portfolio: I assume investors buy more of their beginning-of-year portfolio. This benchmark allows investors to choose their own style (as in Barber and Odean, 2001).
(4) Pure experience based: I assume investors buy stocks in all industries in which they have had good experience in the previous year.
(5) Experience-neutral industry-level momentum: I assume investors buy stocks in the industry that earned the highest average return in the previous year, regardless of their experiences.

I form calendar-time portfolios corresponding to these strategies. For each day, the calendar-time portfolio of actual purchases (or counterfactual strategies) is constructed to include all stocks bought by households (or in counterfactual portfolios as described above) within the prior 21 trading days.[19] The weight is equally allocated to each stock-household (portfolio-household) pair.

Table 6 reports the monthly abnormal returns of these portfolios, controlling for the Fama–French three factors (Fama and French, 1993). According to the figure, past experiences do not seem matter much for each portfolio, and in general the returns of actual purchases are negative, while the returns of industry averages are positive.

As the table shows, investors hurt their performance by trading—they would have lost less from buying and holding their beginning-of-year portfolios. If we assume investors pick stocks in their preferred industries, they lose on their inferior stock-picking ability, rather than by picking the wrong industries; their invested industries, shown in portfolio (1), actually earn significantly positive alphas. Consistent with the experience effect, we also find that the alphas earned by the invested industry, shown in portfolio (1), are close to (even slightly higher than) the alphas earned by a purely experience-based strategy, as shown in portfolio (4). Moreover, we can see that both portfolios (1) and (4) earn significantly positive alphas. The reason is that the industries that have given good experiences to individual investors are also likely to have performed well in general and hence earn higher subsequent returns according to the well-known industry momentum effect (Moskowitz and Grinblatt, 1999). However, the alphas of both these experience-based portfolios are still lower than the alpha of the experience-neutral industry-level momentum portfolio, shown in portfolio (6).

The main message here is that no matter which counterfactual we consider, investors' actual purchases deliver a reliably lower alpha. Even though investors tend to

[18] According to the "ostrich effect" pointed out by Karlsson, Loewenstein, and Seppi (2009), individuals monitor and attend to information more actively given preliminary good news but "put their heads in the sand" by avoiding additional information given adverse prior news. Therefore, the ability to process information is less likely to improve during or following a bad experience.

[19] I exclude the stocks bought on the day of forming the portfolio. This is to address the concern that if investors tend to buy stocks after observing good returns of the industry average of that day, then including the stocks bought on the day of forming the portfolio can mechanically observe the returns of actual purchases lower than the industry average returns because of the manner of constructing the portfolio.

Table 8
The impact of personal investment experience on future purchases in the experienced industry versus other industries.

The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one industry, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. Column (1) corresponds to purchases in the experienced industry, while Column (2) and (3) correspond to two other industries, one of which is the most similar to the experienced industry, while the other is the most different. The most similar and different industries are selected by measuring the distance between the stock returns in that industry and those in the experienced industry. The distance between two industries is determined by averaging the 10-K text-based similarity scores (constructed by Hoberg and Phillips, 2010[2]5) across all pairs of firms in these two industries. The "Experience dummy" is coded as one if the household has owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. The personal experience variables are based on the value of market-adjusted experienced returns. "Good experience" is one when the market-adjusted experienced return is greater than zero. The industry average return variables, the wealth effect, and individual characteristics controls are also included. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)
Dependent variable:	Buy new stocks (not owned in the past year) in		
	the experienced industry	the most similar FF10 industry	the most different FF10 industry
Experience dummy	0.06571***	0.02459***	0.00781
	(0.009)	(0.00528)	(0.00621)
Good experience (> 0)	0.01856***	0.00770***	0.00323
	(0.003)	(0.00185)	(0.00232)
Increase of portfolio size	0.01482***	0.01637***	0.00930***
	(0.002)	(0.00195)	(0.00182)
Average num of stocks > 5	0.05685***	0.05503***	0.04306***
	(0.002)	(0.00234)	(0.00176)
log(Average portfolio size)	0.01539***	0.01417***	0.01133***
	(0.001)	(0.00091)	(0.00076)
log(Average turnover rate)	0.00851***	0.00766***	0.00395***
	(0.001)	(0.00083)	(0.00139)
Experience variable corresponding to the industry in the dependent variable	–	Yes	Yes
Industry average return corresponding to the industry in the dependent variable	–	Yes	Yes
Industry effect	Yes	Yes	Yes
Year effect	Yes	Yes	Yes
Observations	777,630	777,630	777,630

purchase more in industries in which they have had good experiences, they are not missing much as a result of bias toward those industries. In other words, individual investors do not systematically pick the wrong *industries* for investment. What can actually hurt their performance is that they appear to systematically pick the wrong *stocks* for investment.

Considering the evidence that, on average, individual investors appear to have inferior stock-picking ability to the extent that the "learning about ability" explanation is relevant, it is more likely that investors incorrectly, instead of correctly, assess their ability as high after positive outcomes. If they increase their trading frequency in industries where they have had good experiences, they will have higher transaction costs without being able to increase returns. As mentioned before, the data suggests that investors are more likely to start with a concentrated portfolio and experiment with the industries one by one, rather than to start with a diversified portfolio and experimenting with multiple industries all at once. This extends the period in which investors stay in the market, experiment, and (incorrectly) learn about their ability. In the long run, as described in Odean and Gervais (2001), if investors have a self-attribution bias and tend to overestimate their trading ability based on their successes while downplaying their failures, they become overconfident. Consequently, it is less likely that self-attributors will give up a stock-picking strategy before diversifying their portfolio or exiting the stock market. Categorization will only exacerbate the problem. If investors exploit a finer categorization system, they will have more opportunities to experiment in different categories and a better chance to discover that they are good at one of them. Therefore, investors who use a finer categorization system will have more inertia and are less likely to switch to a diversified portfolio or to exit the stock market.

5. Categorization

Up to this point, I have used the Fama–French ten-industry grouping system to classify industries. To shed more light on ways in which investors categorize stocks, I will explore how the impact of past experiences varies with the use of broader or finer categorization systems. I also explore alternative categorization schemes, such as size and value.

5.1. The spill-over effect across industries

I start by testing whether past experiences of investing in the ten Fama-French industries have spillover effects on an investor's purchases in other industries. The light blue and white bars in Fig. 2 display households' probability of purchasing new stocks in other industries, sorting by their experienced market-adjusted returns during the past year. The rightmost bars (group 5) correspond to the top 20% of experienced returns, while the leftmost bars (group 1) correspond to the bottom 20% of experienced returns. I create a distance measure to select, from all other industries, the one most similar to and the one most different from the experienced industry. To explore the economic connections among industries, I exploit the text-based network industry classifications (TNIC) data developed by Hoberg and Phillips (2010) to construct distance measures among the ten Fama-French industries.[20] The distance between two industries is measured by averaging the 10-K text-based similarity scores (constructed by Hoberg and Phillips, 2010) across all pairs of firms in the two industries. The matches are intuitive as well. For example, the most similar industry to the utility industry is the oil, gas, and coal extraction and products industry, while the most different is the high technology industry; the most similar industry to the consumer nondurables industry is the wholesale, retail, and other services industry, while the most different is the oil, gas, and coal extraction and products industry.

Fig. 2(a) plots the original probability of purchasing new stocks. The probability for the most similar industry increases slightly as the household's experienced returns go up, while the probability for the most different industry stays at nearly the same level across different experienced returns. The effect of previously experienced returns becomes increasingly weak the more the industry differs from the experienced industry. Using the residuals from regressing the purchase decision on the controls, Fig. 2(b) confirms this evidence.

Table 8 uses a regression framework to test whether the spillover effect on other industries is significant while controlling for the wealth effect, investor heterogeneity, year, and industry effects. As in Fig. 2, I consider the effect on two other industries: the one most similar to and the one most different from the experienced industry. As shown in Table 8, the effect of good experiences on purchases in the most similar industry is far smaller than (about 40% of) the effect of those experiences on purchases in the same industry; the effect on purchases in the most different industry is even smaller and insignificant. When we consider all the control variables, the significance and magnitude of the effects do not change much for the experienced industry or for other industries. This result is consistent with our intuition that the wealth effect and investor characteristics should have the same effect on each industry.

In summary, the evidence shown in Fig. 2 and Table 8 suggests that the impact of past experience does not spill over to other industries or only in a tiny amount if it does.

5.2. Categorical thinking within industries

In this section, I further investigate whether investors categorize stocks more finely than the Fama-French ten-industry system does. I consider finer categorization in two dimensions.

In the first dimension, the investor can have a finer categorization scheme through the representativeness of stocks with which that investor has experience. If the investor purchases stocks that are representative of the industry, is the impact of experienced outcomes larger than if the investor purchases unrepresentative stocks? I construct a measure of the representativeness of each stock. The measure for an experienced stock j in industry I is computed as the correlation between the stock return r_{jt} and the equally weighted industry return R_{It}.[22] A higher correlation indicates that the experienced stock is more representative of its industry's stocks. The representative stocks identified by this measure are intuitive. For example, the representative stocks for the industry of wholesale, retail, and some services include May Department Store, Target, and Home Depot, while the industry's nonrepresentative stocks include Perfumania, Skyline Chili, Inc., etc.

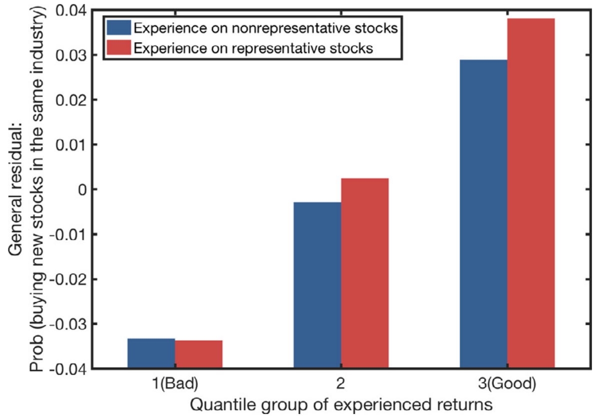

Fig. 4. The impact of investment experience: variation in the representativeness of experienced stocks. The observations are sorted by the value-weighted average annualized excess return on investment in an industry and divided into three groups. Group 1 has the lowest experienced returns, while group 3 has the highest. The figure plots the generalized residuals from a probit model of regressing the purchase of new stocks in an industry on controls of the baseline model.[21] The blue bar corresponds to observations with experience with nonrepresentative stocks, while the red bar corresponds to those with experience with representative stocks. The sample only includes observations with available representativeness measures. The representativeness measure for one experienced stock j in industry I is computed as the correlation between the stock return r_{jt} and the equal-weighted industry return R_{It}. Representative stocks are defined as stocks with representativeness measures greater than the median. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

[20] The results are robust to alternative distance measures based on return differences across industries, as shown in appendix, Section A.3.

[21] The generalized residual of the probit model $Pr(Y_i = 1) = \Phi(X_i'\beta)$ is computed as:

$$\frac{Y_i - \Phi(X_i'\beta)}{\Phi(X_i'\beta)(1 - \Phi(X_i'\beta))}\phi(X_i'\beta),$$

where $\phi(\cdot)$ is the density function of the normal distribution, and $\Phi(\cdot)$ is the cumulative distribution.

[22] I use equal-weighted industry returns to address the concern that the measure would be biased toward large firms if I used value-weighted industry returns.

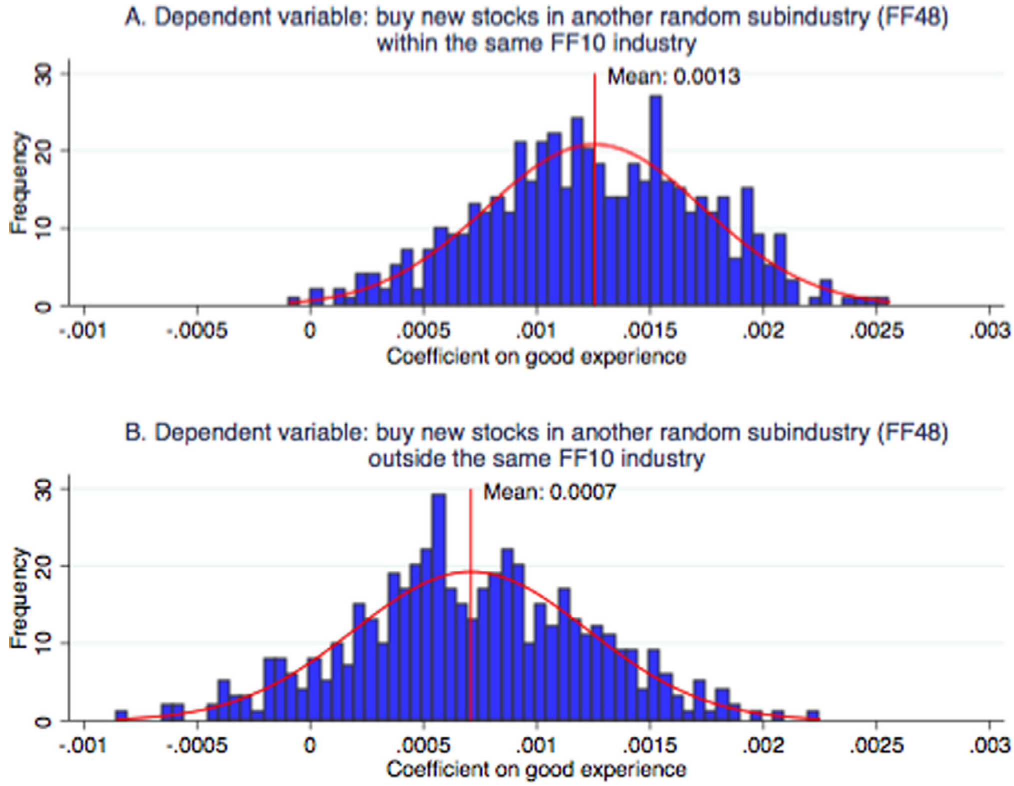

Fig. 5. The impact of personal investment experience: variation in categorization. The figure shows the impact of personal investment experience in an FF48 subindustry on subsequent purchases in another random subindustry within (Panel A) and outside (Panel B) the same FF10 industry with the random subindustry simulated 500 times. The figure plots the distribution of estimated coefficients on the "Good experience" variable from probit regressions. Each observation corresponds to one household and one subindustry, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in a simulated random other subindustry within (outside) the same FF10 industry. "Experience dummy" is coded as one if the household has owned stocks in the subindustry in the past year. Other independent variables are dummy variables related to personal experience, subindustry average, portfolio size, and individual characteristics variables. Both personal experience variables and subindustry average variables are based on the value of market-adjusted (experienced or subindustry average) returns. "Good experience" ("Good industry") is coded as one when the market-adjusted experienced (subindustry average) return is greater than zero. "Increase of portfolio size" equals to one if the size of a household's portfolio has increased in the past year. The three individual characteristics variables are created from beginning-of-month position data. They denote the average monthly number of stocks, the average monthly size, and the average monthly turnover rate in the past year.

I divide investors' experiences into two groups according to representativeness. Fig. 4 displays the probability, after removing the effects of controls, of purchasing new stocks after experiences with representative stocks and nonrepresentative stocks. According to the figure, the probability of purchasing new stocks increases with experienced returns for both groups, and the probability increases by a larger amount for the more representative group. This result indicates that investors are more influenced when they have experience with more representative stocks. If they profit from investing in Target, there is a higher probability that they will buy other stocks in the wholesale, retail, and some services industry again. In contrast, a good experience investing in Perfumania may not have as much impact.

Representativeness captures one dimension of finer categorization based on the statistical features of stock returns, and individual investors can perceive finer categorizations in a more intuitive way. They can further divide stocks into subcategories beyond the Fama-French ten-industry (FF10) classification, using something more like the Fama-French 48-subindustry (FF48) system. If investors group stocks using the FF48 classification, it should be true that the influence of past experience in the same subindustry (an industry in the FF48 classification) is stronger than experience in another subindustry, even if these two subindustries belong to the same industry in the FF10 classification.

I start by simulating 500 times of random subindustries within and outside of the same FF10 industry as the experienced FF48 subindustry.[23] Fig. 5 shows the distribution of estimated coefficients on good experiences. As the figure shows, the mean of the estimated coefficients for the impact on another FF48 subindustry within the same FF10 industry (0.0013) is about half of the impact on the experienced FF48 subindustry (0.0028); and the impact on

[23] One complication arises from the fact that in the FF48 classification, a subindustry can belong to more than one super industry. To address this multiple mapping issue, I assign each FF48 subindustry to the FF10 industry that includes the largest number of firms from the given FF48 subindustry.

Table 9
The impact of personal investment experience: variation in categorization.
The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one FF48 subindustry, regardless of whether the household previously owned stocks in the subindustry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the subindustry. Column (1) corresponds to purchases in the experienced sub-industry. Column (2) corresponds to the closest subindustry in the same FF10 industry as the experienced one. Column (3) corresponds to the furthest subindustry in the same FF10 industry. Column (4) corresponds to the furthest sub-industry not in the same FF10 industry. The distance between two subindustries is measured by averaging the 10-K text-based similarity scores (constructed by Hoberg and Phillips, 2010[2]6) across all pairs of firms in these two subindustries. The "Experience dummy" is one if the household has owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. The personal experience variables are based on the value of market-adjusted experienced returns. "Good experience" is one when the market-adjusted experienced return is greater than zero. The industry average return variables, the wealth effect, and individual characteristics controls are also included. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)	(4)
Dependent variable:	Buy new stocks (not owned in the past year) in			
	the experienced subindustry	the most similar subindustry in the same FF10	the most different subindustry in the same FF10	the most different subindustry outside the same FF10
Experience dummy	0.01554***	0.00525***	0.00616***	0.00716***
	(0.003)	(0.00108)	(0.00137)	(0.00223)
Good experience (> 0)	0.00276***	0.00193***	0.00117**	0.00054
	(0.001)	(0.00048)	(0.00052)	(0.00062)
Increase of portfolio size	0.00317***	0.00284***	0.00325***	0.00473***
	(0.000)	(0.00021)	(0.00017)	(0.00037)
Average num of stocks > 5	0.01163***	0.01026***	0.01047***	0.01694***
	(0.000)	(0.00035)	(0.00031)	(0.00065)
log(Average portfolio size)	0.00302***	0.00296***	0.00296***	0.00473***
	(0.000)	(0.00013)	(0.00013)	(0.00028)
log(Average turnover rate)	0.00189***	0.00156***	0.00170***	0.00231***
	(0.000)	(0.00015)	(0.00009)	(0.00021)
Experience variable corresponding to the subindustry in the dep. var.	–	Yes	Yes	Yes
Industry average return corresponding to the subindustry in the dep. var.	Yes	Yes	Yes	Yes
Industry effect	Yes	Yes	Yes	Yes
Year effect	Yes	Yes	Yes	Yes
Observations	3,490,752	3,490,752	3,490,752	3,490,752

another FF48 subindustry outside the same FF10 industry is even lower, 0.0007. This suggests that individuals categorize industries using a more refined scheme than FF10; otherwise, we would expect the impact on the experienced subindustry to be the same as the impact on another subindustry within the same FF10 industry.

To further establish that individuals categorize on a finer level (the FF48 subindustry classification), I directly test the hypothesis that the likelihood of purchasing a new stock in another FF48 subindustry decreases if the subindustry is economically less similar to or connected with the experienced subindustry. This hypothesis will not hold if individuals hold a broader categorization scheme, such as the FF10 industry classification; in that case, the likelihood of purchasing a new stock in any FF48 subindustry within the same FF10 industry should be the same. Using the distance measure constructed by 10-K text-based similarity scores (Hoberg and Phillips, 2010), I show in Table 9 that a good investment experience in a subindustry has less influence on subsequent purchases in another subindustry when the two industries are less similar or more distant. The impact of experience on the most similar sub-industry in the same FF10 industry is 69.93% (=0.00193/0.00276) of its impact on the experienced subindustry; the impact on the most different sub-industry in the same FF10 industry is 42.39% (=0.00117/0.00276); the impact on the most different subindustry outside the same FF10 is the lowest, 18.12% (=0.00054/0.00276).

Overall, I find that, within the FF10 classification, experiences with more representative stocks have a stronger impact on future purchases in the same industry. Moreover, when I segment the FF10 classification into the more highly resolved categories of the FF48 subindustry classification, we find the effect of a good investment experience on another subindustry is in general smaller than its effect on the experienced subindustry, and the magnitude decreases if the sub-industry is economically more distant from the experienced subindustry.

This result suggests that individuals group investments into finer categories than the FF10 industry classification. This granularity of investor categorization can be related to comentions in news article. If media exposure influences investor categorization, then we would expect that firms in industries with more firms comentioned in news articles would be more likely to be recognized by investors and

Table 10
News article comention of stocks within industries.

Panel A reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to a household/industry/year group, regardless of whether the household previously owned stocks in the industry or not. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is coded as one if the household has owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. Both personal experience variables and industry average variables are based on the value of market-adjusted (experienced or industry average) returns. "Good experience" is one when the market-adjusted experienced return is greater than 0. The independent variables also include the interaction terms between experience-related variables and a high media exposure indicator. The "High media exposure" indicator equals one when the probability that firms from the same industry are comentioned in a news article is above the median, based on the news data (2003–2011) provided by Reuters. The industry average return variables, the wealth effect, and individual characteristics controls are also included. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance. Using news data (2003–2011) provided by Reuters, Panel B records the probability of firms from the same or different industries being comentioned in a news article. For each ticker mentioned in the news, we calculate the sample probability of its being comentioned with other firms, (Column (1), and the simulated probability of being comentioned with random other firms, Column (2), in the same or different industries. The probability reported in the table is computed as the average probability across all tickers in the news.

Panel A: Variation in media exposure		
	(1)	(2)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry	
Industry classification:	FF10	FF48
Experience dummy	0.06967***	0.02734***
	(0.010)	(0.008)
Good experience (> 0)	0.01234***	0.00188***
	(0.004)	(0.001)
Experience × Higher media exposure	−0.00654	−0.00154
	(0.006)	(0.003)
Good experience × Higher media exposure	0.01282**	0.00141*
	(0.006)	(0.001)
Higher media exposure	0.08188***	0.02089***
	(0.016)	(0.004)
Industry return control variable	Yes	Yes
Wealth effect control	Yes	Yes
Individual characteristics control	Yes	Yes
Industry effect	Yes	Yes
Year effect	Yes	Yes
Observations	777,630	3,490,752

Panel B: Comention of stocks from the same versus different industries		
	(1)	(2)
	Comentioned pairs in news articles	Random sampled pairs
Prob(different FF10 industries)	43.16%	86.20%
Prob(same FF10 industry)	56.84%	13.80%
Prob(different FF48 subindustry|same FF10 industry)	20.65%	63.34%
Prob(same FF48 subindustry|same FF10 industry)	79.35%	36.66%

hence have a stronger experience effect. Using news data from the Thomson-Reuters News Analytics (TRNA) dataset (2003–2011), I divide the industries into two groups. The industries with High media exposure are those with an above-median probability that firms from the same industry are comentioned in a news article. I then interact the High media exposure indicator with the experience outcome variables; as the results displayed in Panel A of Table 10 show, the experience effect is indeed stronger in the industries with higher media exposure.

I further compute the probabilities that comentioned pairs of firms in news articles are from the same FF10 industry/FF48 subindustry. As Panel B of Table 10 shows, firms that are comentioned in news articles are more likely to belong to the same FF10 industry than randomly paired firms (56.84% versus 13.80%). Moreover, given that firms are from the same FF10 industry, firms that are comentioned in news articles are more likely to belong to the same FF48 subindustry than randomly paired firms (79.35% versus 36.66%). This tendency for similar firms to be grouped together in news articles can contribute to finer investor categorization of the investment universe.

5.3. Categorization by size or value

The preceding results are consistent with the hypothesis that investors naturally categorize stocks by (broadly defined) industry membership, and that their experience in a category has an effect on their beliefs about and subsequent investment behavior within that category. This suggests the possibility that other firm characteristics, in addition to their industry, are employed by individuals in developing and updating their categorical thinking. I thus consider groupings based on size and value, as these characteristics play a prominent role in empirical models of returns. I exploit the same empirical strategy to test the

Table 11
The impact of personal investment experience: categorizing by size and value.
The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one category, regardless of whether the household previously owned stocks in the category. Column (1) groups stocks into ten categories by size; Column (2) groups stocks into ten categories by value. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is coded as one if the household owned stocks in the category in the past year. Other independent variables are dummy variables related to personal experience, category average, portfolio size, and individual characteristics. Both personal experience and category average variables are based on the value of market-adjusted (experienced or category average) returns. "Good experience" ("Good category") is one when the market-adjusted experienced (category average) return is greater than zero. "Top experience" ("Bottom experience") is one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. "Top 1 category" ("Bottom 1 category") is one when a category's average return is the highest (lowest) among the ten categories. "Increase of portfolio size" equals one if the size of a household's portfolio increased in the past year. The three individual characteristics variables are created from beginning-of-month position data. They denote the average monthly number of stocks, the average monthly size, and the average monthly turnover rate in the past year. Standard errors, shown in parentheses, are clustered at the category-year level. *10%, **5%, ***1% significance.

	(1)	(2)
Dependent variable:	Buy new stocks (not owned in the past year) in the same category	
Categorize by:	Size	Value
Experience dummy	0.06644***	0.05227***
	(0.00454)	(0.01113)
Good experience (> 0)	0.00015	0.00346
	(0.00348)	(0.00336)
Top experience (over 90th percentile)	0.00276	0.00019
	(0.00286)	(0.00465)
Bottom experience (below 10th percentile)	−0.01129**	−0.02679***
	(0.00480)	(0.00458)
Good category (> 0)	0.00699	−0.00950
	(0.00454)	(0.01735)
Top 1 category	0.00436	−0.03729*
	(0.00535)	(0.02058)
Bottom 1 category	0.01259	−0.00873
	(0.00854)	(0.02941)
Wealth effect control	Yes	Yes
Individual characteristics control	Yes	Yes
Industry effect	Yes	Yes
Year effect	Yes	Yes
Avg prob (buying new stocks in a category)	0.1257	0.2029
Observations	787,390	786,150

experience effect. The only difference here is that stocks are grouped by market capitalization (size) and book-to-market ratio (value), rather than by industry. Similarly, I control for category momentum trading, the wealth effect, and individual characteristics.

The results are shown in Table 11. Column (1) shows the results for size-based categories. The coefficients of good experience and top experience are statistically insignificant for both columns. Only extremely bad (bottom) outcomes significantly reduce an investor's probability of purchasing new stocks in a particular category. To compare the effects between top and bottom experiences, the probability of buying a new stock in the same category is 1.43 pp (= 0.02 pp + 0.28 pp – (– 1.13 pp)) higher for top experiences, which accounts for an 11.38% (= 1.43 pp/12.57 pp) increase, normalized by the average probability of buying new stocks in a size category. The results of value-based categories are very similar. The difference between top and bottom experiences is 3.05 pp, which accounts for a 14.59% increase. Though the experience effect is still apparent with these two alternative categorization schemes, the magnitudes are much smaller than when I use industry categories. The small magnitudes echo the evidence shown by Campbell, Ramadorai, and Ranish (2014), based on the data of Indian retail equity investors. They find a long-term tendency (with only moderate magnitudes) for investors to accumulate styles (size, value, and momentum) in which they have experienced positive returns.

The stronger effects found with industry categories, as opposed to size or value categories, can relate to how investors shape their categorization schemes. For example, information sources (e.g., news articles, as shown above) can matter. From observational evidence, comentions in news articles tend to be firms within the same industries rather than firms with similar size or growth prospects. To build upon the theoretical model of Barberis and Shleifer (2003), the heterogeneous experience effects across categorization schemes can indicate different patterns of asset prices across schemes.

6. Conclusion

This paper investigates the influence of investors' personal investment experience on their subsequent stock selection decisions. Using households' trading record data from a large discount broker collected between 1991 and 1996, I consider industry as a categorization scheme potentially used by investors and demonstrate that investors have a greater propensity to purchase new stocks in an industry if they previously earned positive excess returns in that same industry. I provide evidence that the significance and magnitude of the influence of previously experienced outcomes varies over different time horizons, degrees of investors' sophistication, and diversification. Through the lens of the experience effect, we also learn about the granularity of categorization in an investor's mind; I find that investors use highly resolved categorization schemes that are as fine as the Fama-French 48-subindustry classification. The impact of good investment experience on another subindustry decreases if the subindustry is economically more distant from the experienced subindustry. The experience effect is also apparent when size and value are used as categorization schemes, but the magnitudes are smaller.

This behavior can have an effect on investor welfare. The influence of past investment experience can result from investors overweighting their own experience relative to other available historical information when updating their beliefs about stocks in a category (e.g., the industries examined in this paper). The results can also be explained by investors learning about their ability to pick stocks in a certain category from their experience. I provide evidence that investors are not likely to systematically miss good opportunities as a consequence of this biased belief updating. Nevertheless, inept investors' positive experiences in a given category can lead them to assume that they possess insight regarding decisions in this category. In this case, finer categorization will delay the exit of such inferior investors and cause welfare loss.

Furthermore, the aggregation of the influence of investors' personal investment experiences can have a systematic effect on asset pricing. For example, this experience effect can be a source of the industry momentum effect that differs from the existing explanations in the literature. The evidence about how finely investors categorize industries can imply how strong momentum effects are for different levels of industry classifications. The heterogeneous experience effects across categorization schemes can shed light on asset pricing patterns related to different schemes. This points to a promising direction for future research: why do individual investors favor particular categorization schemes over the other options? What factors matter for the categorization process? What other categorization schemes are (or are not) implemented by investors? Better understanding of the underlying mechanisms of categorization can provide richer predictions about asset prices in the market.

Appendix A

A.1. Details on processing data

I implement some restrictions to select the trading records related to the empirical analysis in this paper, and during the process of combining trading records data with the SIC code or price data from CRSP or Compustat, I have to eliminate some observations with missing data. This appendix details the steps I took to get the final subsample used in this paper.

Before combining the trading data with data from other datasets, I select the trading records, using three steps. First, among the trades of various investments, I only retain those related to investors' direct investments on common stock. Second, I remove households if there are inconsistent buy/sell records and quantity records for their trades. For example, trading activity is recorded as "B(uy)" (or "S(ell)"), while the quantity of that trade is recorded as a negative number (or a positive number). Third, I eliminate households if they have trades including short sales—more specifically, if they once had negative cumulative shares on some stocks.

Table A.1
Number of households retained after each refinement.

	Number of households
With direct investments on common stocks	66,465
Without inconsistent buy/sell records and quantity records	62,554
Without short-selling trades	54,210
Without missing SIC code	47,793

The next step is to combine the trading records with the SIC code from CRSP and Compustat. As I described before, I get the CUSIP codes for the stocks invested and match them first with CRSP and then with Compustat if I can't find a match in CRSP. Only if all the investments of the household could match with a SIC code either from CRSP or Compustat are the observations of this household selected for the final subsample used for the baseline regression.

In other analysis (variation in sophistication, time horizon, or categorization), due to missing data for specific related variables, I have to further restrict the sample. Robustness checks, basic summary statistics, and the main results remain the same with different samples.

A.2. Alternative measures for experienced outcomes

The main results shown in Table 2 are also robust to experience measures other than market-adjusted experienced returns. Table A.2 reports the results using alternative experience indicator variables based on the raw level of experience returns. Accordingly, the industry-level control variables are based on the raw (rather than the market-adjusted) industry average returns. Column (6) and (7) control for industry momentum trading by including industry-year fixed effects. Our main results remain the same.

A.3. Alternative measures for distance between FF10 industries

The results regarding the spillover effect across FF10 industries in Table 8 are robust to distance measures based on return differences across industries. For each year, the distance is calculated by averaging the daily absolute difference between the stock returns of these two industries. Suppose there are N trading days in one year; the distance between industry I and industry J (both defined by the Fama-French ten-industry classification)

Table A.2
Alternative investment experience measure and the propensity to purchase new stocks in the same industry (Fama-French ten-industry classification).

The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one industry, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. The "Experience dummy" is one if the household owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. Both personal experience variables and industry average variables are based on the value of raw experienced or industry average returns. "Good experience" ("Good industry") is coded as one when the raw experienced (industry average) return is greater than zero. "Top experience" ("Bottom experience") is coded as one when the raw experienced return is above (below) the 90th percentile (10th percentile) of the sample. "Top 1 industry" ("Bottom 1 industry") is one when the industry average return is the highest (lowest) among the ten industries. "Increase of portfolio size" equals one if the size of a household's portfolio increased in the past year. The three individual characteristics variables are created from beginning-of-month position data. They denote the average monthly number of stocks, the average monthly size, and the average monthly turnover rate in the past year. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	Buy new stocks (not owned in the past year) in one industry						
Experience dummy	0.07123***	0.08035***	0.08479***	0.05863***	0.05593***	0.05698***	0.05299***
	(0.00662)	(0.00620)	(0.00813)	(0.00921)	(0.00923)	(0.00399)	(0.00359)
Good experience (> 0)	0.01348***	0.01283***	0.01554***	0.01536***	0.02165***	0.01682***	0.02296***
	(0.00250)	(0.00254)	(0.00301)	(0.00333)	(0.00334)	(0.00334)	(0.00332)
Top experience (over 90th percentile)	0.01997***	0.01929***	0.02026***	0.01449***		0.01240***	
	(0.00498)	(0.00491)	(0.00466)	(0.00411)		(0.00389)	
Bottom experience (below 10th percentile)	−0.01289***	−0.01263***	−0.01467***	−0.01242***		−0.01332***	
	(0.00233)	(0.00229)	(0.00244)	(0.00289)		(0.00262)	
Good industry (> 0)	−0.00045	0.00027	0.00140	0.00120	0.00143		
	(0.00910)	(0.00642)	(0.00733)	(0.00915)	(0.00917)		
Top 1 industry		0.02464**	0.03133**	0.03907***	0.03916***		
		(0.01132)	(0.01288)	(0.01466)	(0.01485)		
Bottom 1 industry		0.02167**	0.02344**	0.02633**	0.02664**		
		(0.01041)	(0.01099)	(0.01188)	(0.01194)		
Increase of portfolio size			0.01841***	0.01496***	0.01491***	0.01537***	0.01535***
			(0.00202)	(0.00166)	(0.00165)	(0.00161)	(0.00161)
Average num of stocks > 5				0.05675***	0.05680***	0.05668***	0.05676***
				(0.00203)	(0.00205)	(0.00170)	(0.00171)
log(Average portfolio size)				0.01522***	0.01533***	0.01514***	0.01525***
				(0.00071)	(0.00072)	(0.00059)	(0.00060)
log(Average turnover rate)				0.00832***	0.00853***	0.00830***	0.00848***
				(0.00076)	(0.00079)	(0.00074)	(0.00078)
Industry effect	Yes	Yes	Yes	Yes	Yes	No	No
Year effect	Yes	Yes	Yes	Yes	Yes	No	No
Industry X year effect	No	No	No	No	No	Yes	Yes
Avg prob (buying new stocks in an Industry)	0.0846	0.0846	0.0930	0.1118	0.1118	0.1118	0.1118
Observations	1,550,980	1,550,980	1,094,860	777,630	777,630	777,630	777,630

Table A.3
The impact of personal investment experience on future purchases in the experienced industry versus other industries (alternative distance measure by return differences).

The table reports maximum likelihood regression results for probit regressions. The results are reported as the marginal effects of independent variables. Each observation corresponds to one household and one industry, regardless of whether the household previously owned stocks in the industry. The dependent variable is based on a dummy variable coded as one when a household purchases stocks not previously owned in the industry. Column (1) corresponds to purchases in the experienced industry, while Columns (2) and (3) correspond to two other industries, one of which is the most similar to the experienced industry, while the other is the most different. The most similar and different industries are selected by measuring the distance between the stock returns in that industry and those in the experienced industry. The distance is measured by averaging the daily absolute difference between the stock returns of the two industries. The "Experience dummy" is coded as one if the household owned stocks in the industry in the past year. Other independent variables are dummy variables related to personal experience, industry average, portfolio size, and individual characteristics. The personal experience variables are based on the value of market-adjusted experienced returns. "Good experience" is coded as one when the market-adjusted experienced return is greater than zero. "Top experience" ("Bottom experience") is coded as one when the market-adjusted experienced return is above (below) the 90th percentile (10th percentile) of the sample. The industry average return variables, the wealth effect, and individual characteristics controls are also included. The definitions of these variables are noted in Table 2. Standard errors, shown in parentheses, are clustered at the industry-year level. *10%, **5%, ***1% significance.

	(1)	(2)	(3)
Dependent variable:	Buy new stocks (not owned in the past year) in		
	the experienced industry	the most similar FF10 industry	the most different FF10 industry
Experience dummy	0.06571***	0.03103***	0.01257***
	(0.00937)	(0.00190)	(0.00105)
Good experience (> 0)	0.01856***	0.00249*	0.00188
	(0.00312)	(0.00129)	(0.00137)

(continued on next page)

Table A.3 (*continued*)

	(1)	(2)	(3)
Dependent variable:	Buy new stocks (not owned in the past year) in		
	the experienced industry	the most similar FF10 industry	the most different FF10 industry
Increase of portfolio size	0.01482***	0.01535***	0.01226***
	(0.00168)	(0.00214)	(0.00254)
Average num of stocks >5	0.05685***	0.06603***	0.03713***
	(0.00205)	(0.00280)	(0.00157)
log(Average portfolio size)	0.01539***	0.01560***	0.01166***
	(0.00072)	(0.00108)	(0.00146)
log(Average turnover rate)	0.00851***	0.00752***	0.00721***
	(0.00078)	(0.00157)	(0.00144)
Experience variable corresponding to the industry in the dependent variable	–	Yes	Yes
Industry average return corresponding to the industry in the dependent variable	–	Yes	Yes
Industry effect	Yes	Yes	Yes
Year effect	Yes	Yes	Yes
Observations	777,630	777,630	777,630

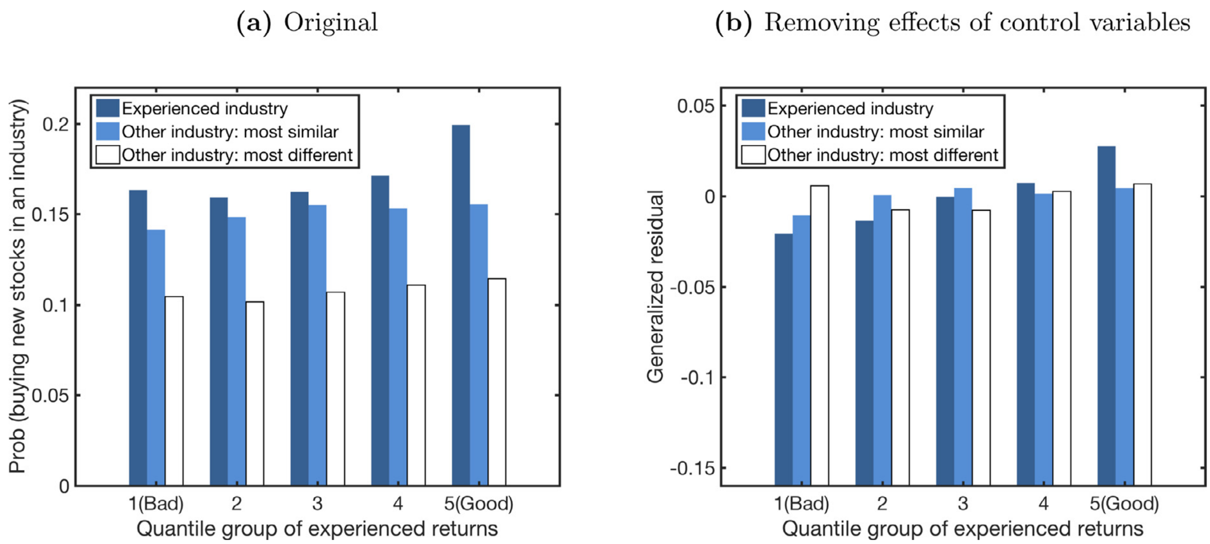

Fig. A.1.. The impact of personal investment experience in an industry on subsequent investment in that industry versus other industries (alternative distance measured by return differences) The observations are sorted by value-weighted average annualized excess returns on the investment in an industry and divided into five groups. Group 1 has the lowest experienced return, while group 5 has the highest. Figure (a) plots the original probability of buying new stocks, that is, the percent of households that buy new stocks in an industry. Figure (b) plots the generalized residuals from a probit model of regressing purchasing new stocks in an industry on controls of the baseline model.[25] The dark blue bars correspond to the experienced industry. The light blue and white bars correspond, respectively, to the industry that is the most similar to or different from the experienced industry. The most similar and the most different industries are selected by measuring the distance between the stock returns in that industry and those in the experienced industry. The distance is measured by averaging the daily absolute difference between the stock returns of the two industries.

is described as $D_{IJ} = \frac{1}{N}\sum_{t=1}^{N}|R_{It} - R_{Jt}|$.[24] The matches are intuitive as well. For example, for the majority of the years, the utility industry is the most similar to the oil, gas, and coal extraction and products industry, while the high technology industry is the most different. As Fig. A.1. and Table A.3 show, the results remain the same—the impact of past experience, if any, only spills over in a tiny amount to other FF10 industries.

[24] I also experiment with another distance measure $D_{IJ} = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(R_{It} - R_{Jt})^2}$, and the results remain the same.

[25] The generalized residual of the probit model $Pr(Y_i = 1) = \Phi(X_i'\beta)$ is computed as:

$$\frac{Y_i - \Phi(X_i'\beta)}{\Phi(X_i'\beta)(1 - \Phi(X_i'\beta))}\phi(X_i'\beta),$$

References

Anagol, S., Balasubramaniam, V., Ramadorai, T., 2015. The effects of experience on investor behavior: evidence from India's IPO lotteries. Working Paper.

Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. J. Finance 55, 773–806.

Barber, B.M., Odean, T., 2001. Boys will be boys: gender, overconfidence, and common stock investment. Q. J. Econ. 116, 261–292.

Barber, B.M., Odean, T., 2007. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financial Stud. 21, 785–818.

Barber, B.M., Odean, T., Strahilevitz, M., 2011. Once burned, twice shy: how pride and regret affect the repurchase of stocks previously sold. J. Market. Res. 48, 102–120.

Barberis, N., Shleifer, A., 2003. Style investing. J. Financial Econ. 68, 161–199.

Calvet, L.E., Campbell, J.Y., Sodini, P., 2007. Down or out: assessing the welfare costs of household investment mistakes. J. Polit. Econ. 115, 707–747.

Calvet, L.E., Campbell, J.Y., Sodini, P., 2009. Fight or flight? Portfolio rebalancing by individual investors. Q. J. Econ. 124, 301–348.

Campbell, J. Y., Ramadorai, T., Ranish, B., 2014. Getting better or feeling better: how equity investors respond to investment experiences. Working Paper.

Choi, J.J., Laibson, D., Madrian, B.C., Metrick, A., 2009. Reinforcement learning and savings behavior. J. Finance 64, 2515–2534.

DeLong, J.B., Shleifer, A., Summers, L.H., Waldmann, R.J., 1990. Positive feedback investment strategies and destabilising rational speculation. J. Finance 45, 375–395..

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financial Econ. 33 (1), 3–56.

Fama, E.F., French, K.R., 1997. Industry costs of equity. J. Financial Econ. 43, 153–193.

French, K.R., Poterba, J.M., 1991. International diversification and international equity markets. Am. Econ. Rev. 81, 222–226.

Gallagher, J., 2014. Learning about an infrequent event: evidence from flood insurance take-up in the United States. Am. Econ. J. Appl. Econ. 6, 206–233.

Goetzmann, W.N., Kumar, A., 2008. Equity portfolio diversification. Rev. Finance 12, 433–463.

Grinblatt, M., Keloharju, M., 2000. The investment behavior and performance of various investor types: a study of Finland's unique data set. J. Financial Econ. 55, 43–67.

Grinblatt, M., Keloharju, M., 2001. What makes investors trade? J. Finance 56 (2), 589–616.

Hertwig, R., Barron, G., Weber, E.U., Erev, I., 2004. Decisions from experience and the effect of rare events in risky choice. Psycholog. Sci. 15, 534–539.

Hoberg, G., Phillips, G., 2010. Product market synergies and competition in mergers and acquisitions: a text-based analysis. Rev. Financial Stud. 23 (10), 3773–3811.

Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. J. Finance 54, 2143–2184.

Ivkovic, Z., Sialm, C., Weisbenner, S., 2008. Portfolio concentration and the performance of individual investors. J. Financial Quant. Anal. 43, 613–656.

Karlsson, N., Loewenstein, G., Seppi, D., 2009. The ostrich effect: selective attention to information. J. Risk Uncertain. 38, 95–115.

Kaustia, M., Knupfer, S., 2008. Do investors overweight personal experience? Evidence from IPO subscriptions. J. Finance 63, 2679–2702.

Malmendier, U., Nagel, S., 2011. Depression babies: do macroeconomic experiences affect risk-taking? Q. J. Econ. 126 (1), 373–416.

Malmendier, U., Tate, G., 2005. CEO overconfidence and corporate investment. J. Finance 60, 2661–2700.

Malmendier, U., Tate, G., Yan, J., 2011. Overconfidence and early-life experinces: the impact of managerial traits on corporate financial policies. J. Finance 66, 1687–1733.

Moskowitz, T.J., Grinblatt, M., 1999. Do industries explain momentum? J. Finance 4, 1249–1290.

Nieuwerburgh, S.V., Veldkamp, L., 2010. Information acquisition and under-diversification. Rev. Econ. Stud. 77, 779–805.

Odean, T., 1998. Are investors reluctant to realize their losses? J. Finance 53, 1775–1798.

Odean, T., Gervais, S., 2001. Learning to be overconfident. Rev. Financial Stud. 14, 1–27.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. J. Financial Econ. 80, 563–602.

Seru, A., Shumway, T., Stoffman, N., 2010. Learning by trading. Rev. Financial Stud. 23, 705–739.

Thaler, R.H., 1999. Mental accounting matters. J. Behav. Decis. Making 12, 183–206.

Weber, E.U., Bockenholt, U., Hilton, D., Wallace, B., 1993. Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. J. Exper. Psychol. Learn. Memory Cognit. 19, 1151–1164.

where $\phi(\cdot)$ is the density function of the normal distribution, and $\Phi(\cdot)$ is the cumulative distribution.