Categorical Thinking in Portfolio Choice

Swasti Gupta-Mukherjee[*]

Quinlan School of Business, Loyola University Chicago

This version: January, 2013

ABSTRACT

Psychologists commonly believe that categorization, i.e. grouping many subjects into a few broad categories, is elemental to thinking. This study shows that the categorization of assets into coarse groups (e.g. industries, styles) is associated with some portfolio investors exhibiting *categorical thinking*– thinking where investors emphasize category-wide information and ignore asset-specific information during investment decisions in assets. To discern whether such categorical thinking in portfolio decisions reflects an information-processing bias that has potentially negative economic effects, I propose the following three hypotheses. First, psychological evidence predicts that if categorical thinking is an information-processing bias, it should increase with information uncertainty and complexity. Second, if a portfolio investor's categorical thinking is an information-processing bias, it should lead to forecasting errors about asset values and, thus, should be negatively related to portfolio performance. Third, if categorical thinking leads to biased forecasting, portfolio investors should display less (more) skill in valuing assets with more (less) coarse categorizations, i.e. for which the category-wide information is less (more) informative. Based on actively managed equity mutual funds, the empirical results strongly support these hypotheses. The findings link an information-processing bias arising from categorization to the quality of decision-making in financial markets where asset categorization is pervasive.

KEYWORDS: Limited attention, categorization, information-processing bias, behavioral finance, mutual funds

JEL Codes: G02, G11, G14, G23

[*] Address correspondence to Swasti Gupta-Mukherjee, Quinlan School of Business, Loyola University Chicago, 1 E. Pearson St, Chicago, IL 60611; Ph(o): 312-915-6071; e-mail: sguptamukherjee@luc.edu.

"To perceive is to categorize, to conceptualise is to categorize, to learn is to form categories, to make decisions is to categorize."   Jerome S. Bruner, *Actual minds, possible worlds*, 1987

I. INTRODUCTION

Since humans have limited cognitive resources, the tendency to simplify complex tasks pervades human behavior. One such simplification during information-processing tasks is the use of categorization, i.e. grouping many subjects into a few broad categories, with psychologists commonly believing that categorization is elemental to thinking. Thus, it has been used in numerous models explaining aspects of human behavior, like social and gender stereotyping, racial discrimination, consumer preferences, and economic decision-making.[2] The unifying premise is that the complex problem of inference under uncertainty is simplified by *categorical thinking*, where the properties typically associated with a category are used to make inferences about a category member with uncertain properties (e.g. Mullainathan [2002]).

Theoretical models in finance have recently started illustrating explicitly the role of categorization during the decision-making process of attention-constrained investors. For instance, Barberis and Shleifer (2003) model the asset pricing implications of style investing, where investors reduce the choice set of assets by allocating funds at the style (i.e. category) level rather than at the asset level.[3] Mullainathan (2002) argues that categorical thinking manifests as an information-processing bias ("categorization bias") in valuing assets when investors use coarse categorizations and overextrapolate a category's usual properties to infer about an asset in the category.[4] In Peng and Xiong's (2006) model of investors with limited attention, investors exhibit "category-learning" behavior in which they process market- and category-wide information to the exclusion of asset-specific information. Moreover, this behavior becomes more pronounced when the investor is more cognitively constrained.

---

[2] For overviews of the voluminous psychology literature see, for example, Laurence and Margolis (1999). Also see Allport (1954) and Fryer and Jackson (2008).

[3] Barberis, Shleifer, and Wurgler (2005) provide empirical evidence of style investing based on the inclusions and exclusions of stocks from the S&P 500 Index.

[4] In a related paper, Mullainathan, Schwartzstein, and Shleifer (2008) study thinking based on coarse categorizations to explain persuasion in advertising and product branding.

It is important to note that the general notion of categorical thinking does not automatically imply limited attention or a "bias" per se, and could be interpreted as a rational strategy. For instance, Peski (2011) proposes that the use of categorization may represent an efficient simplification during information-processing tasks, where it reduces cognitive effort without a serious detriment to inference. Also, given the evidence that assets could be mispriced due to categorization-driven phenomena like return comovement from style investing, informed investors could follow rational strategies emphasizing category-wide information that predicts future asset prices. This behavior can appear similar to the category-learning behavior of investors with limited attention in Peng and Xiong (2006).

The above theoretical insights coupled with the ubiquitous categorization of assets in financial markets into coarse groups (e.g. industries, styles) suggest that categorical thinking could have a nontrivial impact on investment decisions, and underscore some broader questions. For example, does categorical thinking in investment decisions reflect an information-processing bias as opposed to a rational strategy? If so, what factors magnify or diminish the bias? If categorical thinking reflects an information-processing bias, it should diminish the quality of decision-making. In sum, theory predicts the near-instinctive tendency towards categorical thinking, but it typically does not quantify the real economic effects of such behavior and provides little guidance in identifying the investors who exhibit this behavior. Using data on actively managed U.S. equity mutual funds, this study aims to fill these gaps in our understanding of the role of categorical thinking in an important decision setting in financial markets − portfolio choice.

A first step in exploring categorical thinking in portfolio decisions is to select an empirical framework of categorical thinking. To this end, motivated by the insights in Peng and Xiong (2006), this paper introduces a measure capturing a portfolio manager's categorical thinking called the *Categorical Thinking Index* (*CTI*).[5] *CTI* captures the portfolio manager's propensity to emphasize and respond to category-wide information, and to deemphasize and be unresponsive to asset-specific information during investment decisions (e.g. trades that change their active positions in specific stocks). In other words,

---

[5] In this study, I do not distinguish between the broad notion of categorical thinking and "category-learning" as described by Peng and Xiong (2006), since the main goal is to refer to the inference process of investors who mainly use information about an asset's category when they forecast asset returns− a feature of inference implied by both Mullainathan's (2002) model of categorical thinking and Peng and Xiong's category-learning.

categorical thinking is detected when the variation in the portfolio manager's investment decisions in assets is explained to a large (small) extent by the information set capturing category-wide (asset-specific) signals about asset prices.[6] To my knowledge, this is the first empirical study that attempts to explicitly measure categorical thinking of a group of participants in financial markets.

To discern the limited attention versus rational interpretations of categorical thinking, this study proposes the following three hypotheses to test if categorical thinking affects portfolio choice as an information-processing bias. First, if a portfolio investor's categorical thinking reflects an information-processing bias rooted in limited attention, it should increase with information uncertainty and complexity. This prediction is founded in the behavioral finance and psychology literatures which posit that cognitive biases increase with information uncertainty (e.g. Hirshleifer [2001]), and that the limits on attention are more binding when the complexity of the information-processing tasks is greater (Kahneman [1973]).

Second, if a portfolio investor's categorical thinking is an information-processing bias, it should lead to forecasting errors about asset values and, thus, should be negatively related to portfolio performance. This prediction also aligns with the premise in Peng and Xiong (2006) that the more an investor processes category-wide information and ignores asset-specific information, the more constrained is the investor's cognitive capacity and, therefore, the poorer should be the quality of decision-making. This hypothesis clearly delineates the limited attention versus rational interpretations of categorical thinking, since in the latter case we should observe an insignificant or positive relation between categorical thinking and portfolio performance.

Third, since categorical thinking as an information-processing bias implies portfolio investors processing category-wide information to the exclusion of relevant asset-specific information, they should display less (more) skill in valuing assets for which the category-wide information is less (more) informative. The informativeness of the category with respect to a constituent stock is assumed to decrease with the coarseness of the stock's categorization, where coarseness refers to the uncertainty in the information contained in the category-wide signals (e.g. the returns of the category) about the asset (e.g.

---

[6] In spirit, the construction of the measure of responsiveness of investment decisions to information is analogous to the Reliance on Public Information (*RPI*) measure in Kacperczyk and Seru (2007).

returns of the asset). Collectively, I refer to the predictions relating categorical thinking to an information-processing bias as the *Categorization Bias Hypotheses*.

A key step in constructing the *CTI* measure of categorical thinking is identifying a categorization mechanism that is likely to be used by portfolio investors during information processing. For the main tests, I consider a stock's official industry captured by its 2-digit SIC code as a salient and publicly available, but often coarse, categorization.[7] Next, a portfolio manager's responsiveness (i.e. sensitivity) to *category-wide* signals is computed for each fund in a quarter as the unadjusted R^2 ($R^{2,category}$) of the cross-sectional regression of the absolute change in the active position (i.e. position relative to the fund's benchmark portfolio) in a traded stock on returns of the stock's 2-digit SIC industry category in prior quarters lagged up to four quarters. Similarly, the portfolio manager's responsiveness to *asset-specific* signals is computed as the unadjusted R^2 ($R^{2,stock}$) of the cross-sectional regression of the absolute change in the active position in a traded stock on the stock's category-adjusted abnormal returns lagged up to four quarters. In its baseline specification, the *Categorical Thinking Index* (*CTI*) is computed for each fund in a quarter as the $R^{2,category}$, scaled by one plus the $R^{2,stock}$. The interpretation of *CTI* is that it is higher for portfolio managers whose investment decisions are more sensitive to category-wide signals (i.e. have high $R^{2,category}$), and are less sensitive to asset-specific signals (i.e. have low $R^{2,stock}$). Importantly, as described later, the results hold on measuring portfolio managers' responsiveness to different information sets using alternative proxies for category-wide and asset-specific information.

It is worth noting that *CTI* could contain measurement errors due to various reasons. For example, portfolio managers could use (unobservable) categorizations other than those considered in the analyses. Also, portfolio managers may exhibit categorical thinking based on category-wide *private* information which may not be reflected in their responsiveness to information in the public domain used to construct *CTI*. In the context of the empirical tests in this study, potential measurement errors in *CTI* should overestimate the cross-sectional differences in *CTI* and generally work against finding significant results.

---

[7] Hoberg and Phillips (2010) argue that stocks are often crudely categorized based on industry codes, but could differ substantially based on their fundamentals. Kruger, Landier, and Thesmar (2012) also use the 2-digit SIC industry code of stocks in their study of how the categorization bias of investors affects pricing efficiency.

The findings in this study strongly support the Categorization Bias Hypotheses. First, consistent with the notion that categorical thinking in portfolio choice reflects an information-processing bias, *CTI* increases with information uncertainty and complexity. Information uncertainty and complexity is measured at the market level using implied volatility (VIX), and at the stock level using cash flow volatility, firm complexity (see Cohen and Lou [2011]), or the firm's intangible R&D investments. The results hold in regression settings that control for a variety of time-varying fund-specific attributes, time-invariant fund characteristics captured by fund fixed effects, benchmark and year fixed effects. Managers of small cap, growth, and contrarian funds show more categorical thinking. *CTI* decreases with fund size, turnover, and flows, and increases with the fund's herding tendency in trades, fund age, manager's tenure, expense ratio, and activeness relative to benchmarks. Team-managed funds tend to have lower *CTI*. Funds concentrated in fewer industries tend to display more *CTI*, indicating that categorical thinking may be more likely when there is a chance of increasing the economies of scale of industry-level information.

Second, the portfolio managers exhibiting higher levels of categorical thinking significantly underperform the managers exhibiting lower levels of categorical thinking. When funds are ranked into deciles at the end of each quarter based on their *CTI*, the funds in the lowest *CTI* decile significantly outperform the funds in the highest *CTI* decile in the next quarter by 6.13% per year in terms of Daniel et al.'s (1997) characteristic-adjusted before-cost returns, and 2.21% per year in terms of Carhart's (1997) four-factor alpha based on after-cost returns. The results hold in multivariate settings and are robust to alternative specifications of *CTI* which vary in the stock categorization selection, like when *CTI* is measured using the stocks' style categories based on size, book-to-market ratio, and momentum attributes (instead of SIC codes). The findings are also robust to *CTI* computed using alternative asset-specific signals to which portfolio managers' responsiveness is measured, like the returns of the stock's peer firms matched based on fundamentals (see Hoberg and Phillips [2010]).

To test the third hypothesis linking portfolio managers' investment performance to the coarseness of a stock's categorization, I propose a stock level measure called *categorization coarseness* (*COARSE*) which quantifies the ambiguity with which the category-wide information represents the stock-specific

information. *COARSE* is computed on a monthly rolling basis as the standard deviation of the monthly difference in returns between the stock and its 2-digit SIC industry category over the prior 24 month period, scaled by the standard deviation of the stock's monthly returns over the same period. So, as the return difference between the stock and its category becomes more (less) volatile, the category-wide information is less (more) informative about the stock and the coarseness of categorization increases (decreases).[8]

Based on this measure of coarseness of categorization, I find that portfolio managers on average display significantly less stock selection ability in holdings that are poorly represented by their category (i.e. have high *COARSE*) relative to holdings that are closely represented by their category (i.e. have low *COARSE*). For instance, the mean characteristic-adjusted return generated by funds in holdings in the lowest quintile of *COARSE* is a significant 3.29% per year higher than for holdings in the highest quintile of *COARSE*. The results hold across funds with different attributes, across various alternative specifications of *COARSE*, and are robust to potential omitted risk factors in performance measurement.

This study contributes to the literature in several ways. Firstly, this study contributes to the literature in behavioral finance, specifically on categorical thinking and style investing. The theoretical literature in this area is small but growing (e.g. Barberis and Shleifer [2003]; Peng and Xiong [2006]). Recent empirical studies find that a mutual fund's style and a stock's industry classification are categorizations that affect investor behavior and preferences (e.g. Cooper, Gulen, and Rau [2005]; Kruger, Landier, and Thesmar [2012]; Chen, Cohen, and Lou [2012]). To date, existing studies on the role of categorization in financial markets have focused primarily on asset pricing efficiency. The key distinction of this study from prior work is that it offers initial evidence linking categorical thinking to real economic outcomes (i.e. performance) of financial market participants, specifically, portfolio managers.

---

[8] An important feature of this measure is that as long as the difference between the category's return and the stock's return is stable, irrespective of whether it is large or small, the category is considered informative about the stock. Some industries which have relatively low categorization coarseness based on the median across all the firms are "Fishing, Hunting, and Trapping", "Forestry", "Legal Services", "Railroad Transportation", and "Metal Mining". Some industries which have relatively high categorization coarseness based on the median across all the firms are "Real Estate", "Food and Kindred Products", "Amusement and Recreational Services", "Miscellaneous Manufacturing Industries", and "Agricultural Production Crops".

Secondly, this study sheds new light on the recent evidence uncovering the cognitive biases and limitations of institutional investors (e.g. Jiang [2010]; Gupta-Mukherjee [2012]). In this strand of research, the closest studies to this paper are Gupta-Mukherjee and Pareek (2012), Chen, Cohen, and Lou (2012), and Fang, Peress, and Zheng (2011) who also posit that mutual fund managers have limited attention. The main departure of this study is the finding that institutional investors exhibit information-processing biases rooted in categorization and limited attention, and that these biases have economically meaningful effects on investor returns. Given the ongoing debate about the value of active management, and the fact that a significant fraction of the wealth of U.S. households is now tied to their actively managed mutual funds (see French [2008]), increasing (decreasing) the assets allocated to managers who display more (less) information-processing capacity could reduce the costs and increase the potential benefits of investing in actively managed portfolios.[9]

Finally, this paper extends the literature on the role of limited investor attention in financial markets. Prior empirical work has used investor inattention to explain asset pricing anomalies like the post-earnings announcement drift (e.g. Hirshleifer, Lim, and Teoh [2009]; DellaVigna and Pollet [2009]), and mispricing due to neglect of public information (Hirshleifer and Teoh [2003]; DellaVigna and Pollet [2007]; Cohen and Frazzini [2008]; Huberman and Regev [2001]). To my knowledge, this is the first paper to explore how limited attention interacts with categorization and subsequently impacts portfolio decisions, where the categorization of assets is widespread in financial markets.

The remainder of the paper is organized as follows. Section II formulates the empirical predictions proposed and investigated by this paper. Section III describes the two key measures used in the empirical analyses: a portfolio manager's *Categorical Thinking Index* (*CTI*), and a stock's categorization coarseness (*COARSE*). Section IV outlines the data and sample selection criteria. Section V reports empirical results, and Section VI presents concluding remarks.

---

[9] The extensive literature on mutual fund performance cannot be comprehensively summarized here. Some studies that find evidence of value created by fund managers' skill are Grinblatt and Titman (1992), Daniel, Grinblatt, Titman, and Wermers (1997), Wermers (2000), Cohen, Coval, and Pastor (2005), Kosowski, Timmermann, Wermers, and White (2006), Kacperczyk, Sialm, and Zheng (2005, 2008), and Cremers and Petajisto (2009). Representative studies with contrasting evidence include Brown and Goetzmann (1995), Carhart (1997), and Fama and French (2010) who conclude that fund managers create little or no net value with their skill.

II. EMPIRICAL PREDICTIONS

This section outlines several predictions associated with categorical thinking within the context of a portfolio investor's decision-making and performance.

II.A. Categorical Thinking and Information Uncertainty

The literature on behavioral finance has identified several biases in information processing that affect investors' belief formation and decision-making in financial markets. Moreover, Hirshleifer (2001) posits that investor biases are more pronounced when the uncertainty about the value of an asset is high and accurate feedback on the asset's fundamentals is lacking or deferred. In addition, Kahneman (1973) states that the cognitive capacity of humans is limited, and attention to one task necessarily reduces the availability of cognitive resources for other tasks. If categorical thinking affects decision-making as an information-processing bias, categorical thinking should be higher when investors process more uncertain and complex information which place more demands on their cognitive ability. Given that the source of uncertainty and complexity in value-relevant information can be asset-specific (i.e. firm level) or market-wide, the following statements formalize this prediction:

Prediction 1A: *A portfolio investor's categorical thinking will increase with the uncertainty in market conditions.*

Prediction 1B: *A portfolio investor's categorical thinking will increase with the uncertainty of the asset values assessed by the investor.*

To capture uncertainty at the market level, I use the widely-used measure of implied volatility of the S&P 500 index options (VIX), calculated as the expected market volatility over the next 30 days. In other words, I assume that the market uncertainty that can impact asset prices increases with VIX. To capture uncertainty at the asset level, I employ three alternative proxies of the uncertainty in the value of the stocks

assessed by a portfolio investor, which I assume to be the set of stocks held in the portfolio.[10] First, following Zhang (2006), I assume that the uncertainty about a stock's fundamentals increases as the firm's cash flow volatility increases. Second, following Cohen and Lou (2011), I assume that a firm becomes more difficult to value as the number of business segments in the firm increases. Finally, since the feedback on intangible, long-term inputs like R&D is deferred and ambiguous, and the value of R&D is hard to evaluate, I assume that firms with R&D expenses are associated with more information uncertainty than firms which do not spend on R&D (Daniel, Hirshleifer, and Subrahmanyam [2001]). Note that the proxies for information uncertainty are also correlated with information complexity, so they are used as combined measures reflecting information uncertainty and complexity in the remainder of the paper.

II.B. Categorical Thinking and Investment Performance

If a portfolio investor's categorical thinking is associated with an information-processing bias, she will overestimate the precision with which the category-wide signal reflects the asset-specific signal of future asset returns (see Mullainathan [2002]). Consequently, ceteris paribus, a portfolio investor who exhibits this bias is more likely to misvalue assets than an investor who does not exhibit the bias. A related notion that can be inferred from Peng and Xiong (2006) is that investors who process more category-wide information to the exclusion of asset-specific information have more constraints on their attention and, thus, should make poorer decisions. These ideas lead to the following prediction concerning performance

Prediction 2: *A portfolio investor's categorical thinking should be negatively associated with portfolio performance.*

Note that an underlying assumption in order to make this prediction is that the information-processing biases linked to categorical thinking are large enough to have a meaningful impact on performance.

---

[10] The results reported in the paper remain unchanged in additional tests where the set of stocks assessed by a manager is expanded to include all stocks in the fund's portfolio and the stocks in the fund's benchmark index.

Another aspect of investment performance that can be studied in the mutual fund setting is the abnormal returns portfolio managers generate from specific holdings using their stock-selection ability. A few studies look at the types of investments where portfolio managers are likely to display stock-selection ability. For example, studies find that mutual fund managers are likely to generate favorable abnormal returns in local holdings (Coval and Moskowitz [2001]), firms in which the officials are connected to the fund manager via social networks (Cohen, Frazzini, and Malloy [2008]), and firms which are difficult to value (Schultz [2010]). In this paper, a related question is how the propensity for categorical thinking is likely to affect stock selection ability in holdings which differ in the coarseness of their categorization.

If investors process category-wide information to the exclusion of relevant asset-specific information when they predict future asset returns, their stock selection ability with respect to particular stocks should increase with the precision with which the category-wide signal captures the asset-specific signal about future returns for the stock. As the relation between the category-wide signal and the asset-specific signal grows ambiguous, this precision decreases and investors are more likely to misvalue the asset. This notion is captured by the following prediction

Prediction 3: *If a portfolio investor commits information-processing errors related to categorization, she will exhibit less (more) skill in valuing stocks which have a more (less) ambiguous relation with their category.*

III. EMPIRICAL METHODOLOGY

In this section, I describe the key empirical measures used in this study.

III.A.   Measuring categorical thinking: Categorical Thinking Index (CTI)

In reality, investors could group stocks into a multitude of categories when they process information and predict future returns. But for the purposes of detecting potential categorical thinkers in the data, the goal is to select a salient and publicly available categorization that is likely to be used by a substantial portion of

portfolio investors. The 2-digit SIC code appears to fit this criterion since it is a widely-used and intuitive stock categorization. Compared to the official SIC industries, some other categorizations (e.g. value versus growth) are more subjective, vary over time, and could be used more idiosyncratically by investors.[11]   In contrast, the 2-digit SIC code is a "hard" and generally time-invariant categorization. Nevertheless, for the sake of robustness, later sections present analyses that consider alternative categorizations like style.

Peng and Xiong (2006) propose that when investors have limited attention, they process more category-wide information to the exclusion of asset-specific information. Moreover, the more constrained the investor is, the more she processes category-wide information and ignores asset-specific information. Motivated mainly by these theoretical insights, the premise based on which I develop the empirical metric of categorical thinking is that a portfolio manager who exhibits categorical thinking emphasizes category-wide signals and neglects the stock-specific signals about future returns when she makes investing decisions. In the empirical framework, if an information signal S1 explains a larger extent of the variation in a portfolio manager's investment decisions than signal S2, i.e. the manager's decisions are more sensitive to S1 than S2, it is inferred that the manager places more emphasis on processing signal S1 than on processing signal S2.

In the first step to implement the above premise, I measure how sensitive a portfolio manager's investment decision in a stock is to the stock's category-wide information. The investment decision in a stock is calculated as the absolute change in active weight after trades. The proxy for category-wide information is the lagged returns of the stock's official industry category. For this, the following cross-sectional regression is estimated for each fund f in each quarter t based on stocks $i=1$ to N in f's portfolio

(1) $\quad \left| \Delta activewt_{i,f,t} \right| = \beta_{f,t}^{0} + \beta_{1,f,t}^{cat} ret_{i,t-1}^{cat} + \beta_{2,f,t}^{cat} ret_{i,t-2}^{cat} + \beta_{3,f,t}^{cat} ret_{i,t-3}^{cat} + \beta_{4,f,t}^{cat} ret_{i,t-4}^{cat} + \varepsilon_{f,t}$

where $\left| \Delta activewt_{i,f,t} \right| = \left| activewt_{i,f,t} - activewt_{i,f,t-1} \right|$

and $\quad activewt_{i,f,t} = wt_{i,f,t} - benchwt_{i,f,t}$

---

[11] For example, a stock in the lowest 25% based on the book-to-market ratio could be categorized as a growth stock by one portfolio manager, while for another manager growth stocks could be those in the lowest 10%.

$wt_{i,f,t}$ is the portfolio weight allocated to stock i by fund f in quarter t, and $benchwt_{i,f,t}$ is the weight of stock i in fund f's benchmark portfolio in quarter t. So, $\Delta activewt_{i,f,t}$ is the directional change in active weight in fund f's traded holdings on a split-adjusted basis between quarter t-1 and t, and captures the direction and magnitude of the change in fund f's position in the stock i.[12] $ret^{cat}_{i,t-p}$ is the equal-weighted average monthly return of all valid stocks in stock i's official 2-digit SIC industry category in quarter t-p, and p=1,2,3,4 is the number of lags in the observation of returns.[13] I only consider the sample of holdings which experienced a change in the number of shares by trading to capture a portfolio manager's information-driven investment decisions, since changes in active weight can also occur inactively, like when the passive benchmark weight of a stock changes. Basing the measure of the investment decision on the active weight clearly reflects both the change in the direction (e.g. overweighting versus underweighting) and size (e.g. degree of overweighting) of a portfolio managers' bet in a stock relative to the benchmark versus which the manager is compared during performance evaluation. Although I consider alternative measures as robustness checks in later sections, it is worth noting that some alternative measures of investment decisions, like fraction of shares traded, may increase in magnitude while decreasing the size of the bet or active position (e.g. if the trade brings the stock's portfolio weight closer to the weight in the benchmark).

The unadjusted R^2 ($R^{2,category}$), of the regression in Equation (1) is then used to infer how much emphasis a portfolio manager's strategies place on category-wide information, i.e. official industry returns. A notable feature of the $R^{2,category}$ measure is that it does not differentiate between momentum and contrarian strategies where the change in active positions would have opposite signs. Also, this method of computing a fund manager's responsiveness from the R^2 is similar to the RPI (Reliance on Public Information) measure in Kacperczyk and Seru (2007) who regress changes in a fund's holdings on lagged changes in analyst stock recommendations to obtain responsiveness to public information.

---

[12] Alternative specifications of the model in unreported tests included using non-absolute changes in active weight and using absolute returns. The results remain unaltered and the regression estimates across these specifications are highly correlated (>0.8).

[13] Valid stocks are selected from the Center for Research in Security Prices (CRSP) monthly stock files as the sample of ordinary common shares (share codes 10 and 11).

In the second step, I account for the notion that categorical thinking is expected to be more pronounced when, in addition to having high values of $R^{2,category}$, the manager ignores asset-specific information. To this end, I develop a measure of how sensitive a portfolio manager's investment decisions are to the stock's category-adjusted returns, i.e. the abnormal stock returns not explained by the category. For this, the following cross-sectional regression is estimated for each fund f in each quarter t

$$(2) \qquad \left|\Delta activewt_{i,f,t}\right| = \beta_{f,t}^0 + \beta_{1,f,t}^{stk} ret_{i,t-1}^{stk} + \beta_{2,f,t}^{stk} ret_{i,t-2}^{stk} + \beta_{3,f,t}^{stk} ret_{i,t-3}^{stk} + \beta_{4,f,t}^{stk} ret_{i,t-4}^{stk} + \varepsilon_{f,t}$$

$ret_{i,t-p}^{stk}$ is the average category-adjusted monthly return of stock i computed as the stock's monthly return minus the return of the stock's official 2-digit SIC industry category ($ret_{i,t-p}^{cat}$) in quarter t-p, where p=1,2,3,or 4. The higher is the value of the R^2 ($R_{f,t}^{2,category}$) in Equation (1) and lower is the value of the R^2 ($R_{f,t}^{2,stock}$) in Equation (2), the more the portfolio manager emphasizes category-wide signals and ignores asset-specific signals. Based on this intuition, a measure called the *Categorical Thinking Index* (*CTI*) is constructed in each quarter t for each fund f to capture the portfolio manager's categorical thinking as

$$(3) \qquad CTI_{f,t} = \frac{R_{f,t}^{2,category}}{1 + R_{f,t}^{2,stock}}$$

where $R_{f,t}^{2,category}$ and $R_{f,t}^{2,stock}$ denote the unadjusted R^2 from Equations (1) and (2), respectively.[14] Note that $R_{f,t}^{2,category}$ and $R_{f,t}^{2,stock}$ are likely to also capture to some extent the portfolio manager's reliance on public information (Kacperczyk and Seru [2007]), since managers who make decisions based on differential private information are likely to have lower values of $R_{f,t}^{2,category}$ and $R^{2,stock}$. So, *CTI* can be viewed as measuring the *type* of public information to which the manager responds, with higher values representing managers who rely more on category-wide public information and rely less on asset-specific public information.

It is important to note some features and potential concerns related to the specification of *CTI* in Equation (3) as the measure of categorical thinking. First, as noted earlier, high values of *CTI* do not

---

[14] To ensure that the regression estimates are meaningful, I only use the measures obtained from funds which have at least 30 stocks in their portfolios.

automatically imply a bias or suboptimal strategy per se, and could also be high for portfolio managers who follow rational strategies that emphasize categories. In this case, we should observe a positive relation between *CTI* and investment performance. On the other hand, if *CTI* has a negative association with subsequent investment performance, the information-processing bias interpretation of *CTI* is more suitable. Second, there could be sources of measurement error in *CTI*. For example, while this study argues that funds with low *CTI* exhibit less categorical thinking, low *CTI* values could also be a result of portfolio managers using categories not captured in the empirical apparatus, using a category-wide information set other than past returns, or using private category-wide information. However, these potential measurement errors in *CTI* should work against finding a significantly negative effect of categorical thinking captured by *CTI* on portfolio performance, since more portfolio managers who use category-wide information sets absent in measuring *CTI* are likely to be classified as not exhibiting categorical thinking. Thirdly, the *CTI* measure does not differentiate between portfolio managers who deliberately respond to category-wide signals, and portfolio managers who involuntarily divert their attention to category-wide information due to stimuli (e.g. as a response to news).[15] In the context of this study, it is not particularly important whether categorical thinking arises due to deliberate as opposed to involuntary attention directed towards categories, since the goal is to evaluate the ex-post economic implications of this behavior.

III.B.   Measuring Categorization Coarseness (*COARSE*)

An important premise in this study is that if investors simplify information processing by predicting a stock's future returns based on signals about its category, information-processing errors are more likely when they value stocks for which the category is less informative, i.e. where the stock is coarsely categorized. Suppose an investor uses categorization in a simple model where the return of each stock in month *m* is made up of two components: the stock's category returns (r_m^{cat}), and an error term (ε_m) as

---

[15] The agent deliberately choosing the focus of their attention forms the basic paradigm in studies like Gabaix et al. (2006) and the psychological evidence reviewed in Yantis (1998). Alternatively, an agent involuntarily allocating attention at the onset of stimuli is the premise in studies like Odean (1999) and Barber and Odean (2008).

(4)
$$r_m = r_m^{cat} + \varepsilon_m$$

The underlying idea that agents use oversimplified and parsimonious models during forecasting echoes the general "simple paradigms" argument made by Hong, Stein, and Yu (2007), and the more specific categorical-learning thesis of Peng and Xiong's (2006) limited attention model. Using Equation (4) as the intuitive basis, we expect the category-wide signal r_m^{cat} in the investor's forecasting model to be more (less) informative about the stock's return when the variance of the error, ε_m (= $r_m - r_m^{cat}$), is small (large). To capture the coarseness of the categorization, I measure the ambiguity in the relation between the stock and its category as the stock's *categorization coarseness*. A stock's categorization coarseness is computed in each month m as the standard deviation of the monthly return difference between the stock and its category over the prior T months, scaled by the standard deviation of the stock's return over the same period as

(5)
$$COARSE_m = \frac{\sigma_T(r_m - r_m^{cat})}{\sigma_T(r_m)}$$

Here $\sigma_T(r_m - r_m^{cat})$ is the standard deviation of the monthly return difference between the return of the stock (r_m) and the return of its 2-digit industry category (r_m^{cat}) in the T months $m-T$ to $m-1$ prior to month m where, for empirical purposes, I use T=24.[16] $\sigma_T(r_m)$ is the standard deviation of the stock return over the T months $m-T$ to $m-1$. Since $\sigma_T(r_m - r_m^{cat})$ is expected to be positively correlated to the overall standard deviation of the stock's returns, $\sigma_T(r_m)$ is used to scale $\sigma_T(r_m - r_m^{cat})$ in order to separate the effect of general stock volatility from the volatility of the stock's return relative to its category. The standard deviation of the return difference represents the uncertainty in the relation between the stock's returns and its category's returns, with the category-wide returns becoming less informative about the stock returns, and the categorization becoming more coarse, as $\sigma_T(r_m - r_m^{cat})$ increases relative to $\sigma_T(r_m)$. Alternative specifications of *COARSE* are considered for robustness checks in later sections.

IV. DATA AND SAMPLE SELECTION

---

[16] The results reported in this paper are robust to alternative measuring periods like T=12 and T=36.

The primary data sources used in this study are the CRSP Survivor-Bias Free US Mutual Fund Database (MFDB) and the Thomson Financial CDA/Spectrum holdings database. The initial sample consists of all unique mutual funds that appear in the monthly net returns data provided by CRSP during 1991 to 2010.[17] This sample is matched to the CRSP data including total net assets (TNA), fees, and other fund level characteristics. The sample of funds from CRSP is then matched to the Thomson Financial fund identifiers using the MKLINKS interface provided by WRDS. While some funds report holdings semiannually as per mandatory disclosure requirements, most mutual funds report holdings on a quarterly basis since 1980 (see Wermers, 2000). A fund size threshold of $10 million is imposed to eliminate funds that have relatively high capital constraints and various screens are then employed to select actively managed diversified domestic equity funds.[18] Funds which hold less than 30 stocks are excluded to ensure a meaningful number of observations for quarterly cross-sectional analysis of the fund's portfolio required in the empirical methodology outlines in Section III.

Next, I select funds which have non-missing passive benchmark assignments in a given quarter. As benchmarks, I include almost all the indexes used by mutual funds during the sample period. This gives a total of 19 indexes from three index families: S&P/Barra, Russell, and Wilshire. Then for each fund, following the methodology in Cremers and Petajisto (2009), the index with the lowest deviation of holdings from the fund's portfolio is assigned as its benchmark. Given that the passive benchmark data ends in 2010, I choose funds with quarterly benchmark assignments over the period 1990 to 2010. Actively managed funds are selected by excluding the funds with Active Share below 30%. Since the empirical analyses are based on stock holdings that can be matched to CRSP's stock files, I only include funds for which the market value of the reported holdings represent at least 65% of the total net assets of the fund at

---

[17] CRSP MFDB often includes multiple identifiers for the same fund if it has different share classes which vary in terms of expense ratios and loads. Two steps are taken to eliminate duplicated observations of the same fund. First, the fund identifier with the longest time series history of returns is selected. If this step does not identify a fund uniquely, the identifier associated with the highest TNA in the year prior to the return observation is selected.

[18] Index, sector, bond, international, and money market funds are excluded based on stated objectives or, if missing the objectives information, using keywords in the fund's name like "Index", "Healthcare", etc. Using the objective categories from Wiesenberger, Strategic Insight, and Lipper, funds which have objectives defined as aggressive growth, growth, growth and income, equity income, growth with current income, income, long-term growth, maximum capital gains, small cap core/growth/value, large cap core/growth/value, mid cap core/growth/value, multi cap core/growth/value, unclassified, or missing are chosen.

the end of a quarter. For the median fund in the sample, the market value of the stock holdings represents over 90% of the TNA. The stock holdings are matched to COMPUSTAT annual and segments databases to obtain data on firms.

V. RESULTS

In this section, I report summary statistics for the fund and stock samples, explore the determinants of categorical thinking, and examine how categorical thinking affects investment performance.

V.A. Summary Statistics

This section presents descriptive statistics on mutual funds' categorical thinking (*CTI*), stocks' categorization coarseness (*COARSE*), as well as other fund and stock characteristics.

Table I presents the descriptive statistics on the final sample of 2,812 U.S. equity funds with benchmark assignments over 1990 to 2010. The funds map to 67,841 unique fund-quarter observations for portfolio holdings. The mean (median) fund in the sample has a TNA of $1,041 million ($172.3 million), and holds a mean (median) of 116.4 (76) stocks in their portfolio. Validating that the sample includes actively managed funds, the funds have a mean turnover of 87.2% and Active Share of 79.4%. The correlation coefficients of various fund and portfolio attributes with *CTI* show that *CTI* has low correlations with fund and portfolio attributes (less than 0.3).

Table II reports descriptive statistics on stock characteristics for the full sample and by *COARSE* quintiles for the stocks held by mutual funds. The stock characteristics reported include *COARSE*, number of firms in the stock's 2-digit SIC category, stock level proxies for information uncertainty, and style attributes. The proxies for information uncertainty include Cash Flow Volatility defined as the standard deviation of cash flow from operations in the past five years (with a minimum of three years of available data), Firm Complexity defined as the number of business segments in the most recent year reported in the COMPUSTAT segments file, and R&D Dummy which is an indicator variable equaling one if the firm

reports non-zero R&D expenses (Compustat XRD) in the most recent year in the COMPUSTAT annual file and zero otherwise.[19] The style attributes are the stock's Size, B/M, and Momentum quintiles as defined in Daniel et al. (1997) (hereafter DGTW).

The first column which presents the correlation coefficients of *COARSE* with other stock characteristics shows that *COARSE* has very low correlations with the other stock characteristics (less than 10%). The median industry category includes 278 firms, but a standard deviation of about 360 firms indicates a large degree of variation in the category sizes. The Q5–Q1 values represent the differences in the mean stock characteristic in the lowest quintile Q1 and the highest quintile Q5 of *COARSE*. In general, stocks with lower *COARSE* tend to belong to categories with more firms, perhaps because larger categories include more stocks that are representative of the stock during the calculation of *COARSE*. In theory, as noted by Peski (2011), using fewer and larger categories alleviates the problem of "overfitting" because of the presence of many objects to compare an object with, i.e. if comparisons are based on a larger number of observations. So, the negative relation between a stock's *COARSE* and the number of firms in its category provides some validation of *COARSE* as an empirical measure capturing the likelihood that a stock is miscategorized. There are some other stock attributes that also differ significantly between the stocks in Q1 and Q5 of *COARSE*. Stocks with higher *COARSE* tend to be smaller in size and slightly more growth-oriented (i.e. have lower book-to-market (B/M) rankings) than those with lower *COARSE*, potentially reflecting the more idiosyncratic nature of these stocks. Interestingly, *COARSE* does not appear to have a clear and consistent relation with the proxies for information uncertainty, since the Cash Flow Volatility decreases and the likelihood of the firm having R&D inputs increases with *COARSE* quintiles. Additionally, Firm Complexity is not significantly different between Q1 and Q5 of *COARSE*.

---

[19] The cash flow from operations is computed as net income before extraordinary items and discontinued operations minus total accruals. Total accruals (*TACC*) in period t is calculated as
$$TACC_t = (\Delta CA_t - \Delta CL_t - \Delta Cash_t + \Delta STDEBT_t - DEPTN_t)$$

Where ΔCA_t = change in current assets in period t (Compustat ACT); ΔCL_t = change in current liabilities in period t (Compustat LCT); $\Delta Cash_t$ = change in cash and cash equivalents in period t (Compustat CHE); $\Delta STDEBT_t$ = current maturities of long-term debt and other short-term debt included in current liabilities in period t (Compustat DLC); $DEPTN_t$ = depreciation and amortization expense in period t (Compustat DP).

V.B.   *Determinants of CTI: Information Uncertainty and Fund Attributes*

In this section, I explore the determinants of categorical thinking in a multivariate setting. Table III reports the results of panel regressions explaining *CTI*, with all specifications including objective fixed effects and year fixed effects. The *p*-values for significance tests are reported based on Newey-West corrected standard errors that adjust for the heteroskedasticity and autocorrelation in observations, and are clustered by fund. Columns (1)-(5) employ four explanatory variables proxying for information uncertainty in the portfolio manager's information environment: log (VIX), log (Cash Flow Volatility), log (Firm Complexity), and % R&D Stocks. Here log (VIX), log (Cash Flow Volatility), and log (Firm Complexity) are the natural logarithms of VIX (as defined earlier), the equal-weighted Cash Flow Volatility across all stocks held by the fund, and the equal-weighted Firm Complexity across all stocks held by the fund, respectively. % R&D Stocks is the fraction of the portfolio invested in stocks of firms that incurred R&D expenses in the most recent year.

The positive and significant coefficients in columns (1)-(5) on log (VIX), log (Cash Flow Volatility), log (Firm Complexity), and % R&D Stocks show that as the macroeconomic uncertainty and uncertainty about asset values in the portfolio increase, portfolio managers' categorical thinking increases. Moreover, the proxies for information uncertainty have economically significant effects on *CTI*. For instance, in column (1), a one standard deviation increase in VIX increases *CTI* by 2.7% relative to the median value of *CTI* (=0.064). Similarly, in column (2)-(4), a one standard deviation increase in Cash Flow Volatility, Firm Complexity, and % R&D Stocks increases *CTI* by 7.0%, 12.9%, and 2.5% relative to the median, respectively.

These results support Prediction 1A and Prediction 1B stated earlier in Section II. If a portfolio manager's tendency to emphasize category-wide information to the exclusion of asset-specific signals captured by *CTI* is an information-processing bias, the findings support prior psychological evidence that behavioral biases increase with the uncertainty of the information. The variables used to proxy for information uncertainty can also be viewed as proxies for information complexity, so the interpretation of results is similar whether they are used as proxies for information uncertainty or complexity, where

information complexity is also known to make information-processing biases and limits on attention more pronounced. It should be noted that the evidence so far is not sufficient to rule out categorical thinking reflected in *CTI* as a rational strategy. For instance, if the profitability of category-driven strategies (e.g. a strategy exploiting return reversals arising from style investing) is higher in more uncertain market environments, portfolio managers could follow rational strategies where the *CTI* is higher when market uncertainty is higher. However, it is not clear whether a similar reasoning can rationally explain the positive relation between *CTI* and the portfolio-level information uncertainty of stocks held by the fund.

In column (6) of Table III, I add to the regression a fund's herding behavior in trades in a specification including the four information uncertainty proxies, where herding behavior is captured by the Herding Measure computed following Lakonishok, Shleifer, and Vishny (1992).[20] Bikhchandani, Hirshleifer, and Welch (1992) and Scharfstein and Stein (1990) propose that herding behavior increases with information uncertainty because investors have incentives to ignore private information signals and reduce reputational risk of following strategies that differ from the herd when the uncertainty about asset values is high. It is possible that investors' categorical thinking is correlated to some aspects of herding behavior (e.g. industry momentum-chasing) and, since information uncertainty increases herding tendencies, the positive relation between *CTI* and information uncertainty could be a representation of the relation between *CTI* and herding. While the estimated regression in column (6) shows that portfolio managers who herd indeed tend to exhibit more categorical thinking, the significant impact of information uncertainty on *CTI* remains on controlling for herding behavior. Another interesting observation from the specification in column (6) is that the Herding Measure only explains an incremental 0.8% of the variation

---

[20] In Lakonishok, Shleifer, and Vishny (1992), a stock i's herding measure ($HM_{i,t-1}$) is computed in quarter t-1 as

$$HM_{i,t-1} = |p_{i,t-1} - \overline{p_{t-1}}| - E(|p_{i,t-1} - \overline{p_{t-1}}|)$$

where $p_{i,t-1}$ is the proportion of mutual funds that buy stock i during the quarter t-1, out of all the funds that trade stock i in the quarter t-1; $\overline{p_{t-1}}$ is a proxy for the expected value of $p_{i,t-1}$, computed as the mean value of $p_{i,t-1}$ across all the stocks traded by mutual funds in the quarter t-1. $E(|p_{i,t-1} - \overline{p_{t-1}}|)$ is an adjustment factor, which equals the expected value of $|p_{i,t-1} - \overline{p_{t-1}}|$ under the null of no herding. Following Wermers (1999), to ensure that the measure is based on meaningful herding in trades, only stocks traded by at least five funds are included in the computation of *HM*. Additionally, I exclude stocks which started trading following an initial public offering any time during the four quarters prior to t-1.

in *CTI*, as evident from comparing the R^2 of 6.7% in column (5) to the R^2 of 7.5% in column (6). So, *CTI* contains information on managerial behavior distinct from that captured by herding in trades.

In column (7) of Table III, the regression includes as controls a fund's style attributes− Size Style, B/M Style, and Momentum Style− calculated for a fund as the value-weighted Size, B/M, and Momentum DGTW quintiles across all holdings, respectively. The information uncertainty proxies remain positive and almost always significant at the 1% level in column (7). Interestingly, since a fund's Size Style and B/M Style can also be interpreted as proxies for the difficulty in valuing the assets held by the fund, the significantly negative relation between Size Style and B/M Style with *CTI* provides further support to the prediction relating information uncertainty and categorical thinking. As Size Style and B/M Style decreases, the fund's portfolio is concentrated in the difficult-to-value small cap and growth stocks, and the *CTI* of the portfolio manager increases.

The results remain qualitatively unchanged in column (8) which adds several other fund-specific variables, namely, log (TNA), log (Fund Age), log (Manager Tenure), Team-managed Dummy, and Past Flows. Past Flows is the average monthly flow in the quarter.[21] Team-managed Dummy is an indicator variable that equals one if the fund is managed by more than one manager, and zero otherwise. Among notable results, the portfolio manager's categorical thinking decreases with fund size, and increases with the length of manager tenure and fund flows. It is possible that since larger funds presumably have more human and technological resources based on which they can pay more attention to firm-specific information, they exhibit less categorical thinking than smaller funds on average. Fund flows could be increasing a portfolio manager's categorical thinking if flows impose more cognitive constraints on the manager since the new funds need to be profitably allocated. To the extent that *CTI* represents more simplistic strategies, the positive relation of *CTI* with length of manager tenure is somewhat counterintuitive, since there is some evidence that less experienced managers follow more simplistic

---

[21] Following prior studies, the fund flows in month t, i.e. the growth in TNA due to new investments, is calculated as

$$Flow_t = \frac{TNA_t - TNA_{t-1}(1+R_t)}{TNA_{t-1}}$$

where, R_t is the monthly net return of the fund during month t, and TNA_t is the fund's total net asset value at the end of month t as reported in CRSP. Outliers are eliminated by winsorizing the 2.5% tails.

strategies than more experienced managers (e.g. Greenwood and Nagel [2009]). One explanation could be that manager tenure is not a true proxy for experience, whereas Greenwood and Nagel (2009) use a more direct measure of experience, i.e. manager age. Alternatively, if managers with longer tenure have less reputation-building incentives than newly-appointed managers, they could have a higher propensity to rely on category-wide information and simplify their cognitive tasks than newly-appointed managers. The significantly negative coefficient on Team-managed Dummy indicates that teams are less likely to display categorical thinking than single-manager funds, perhaps because the cognitive constraints become less binding when the fund is managed by multiple managers.

In column (9), several variables typically associated with a fund's activeness are included. In general, funds with higher *CTI* have higher expense ratios, industry concentration, and Active Share, and lower turnover than funds with low *CTI*. The variables representing information uncertainty remain statistically significant in column (9). In column (10), fund fixed effects are included in addition to time-varying fund attributes and proxies for information uncertainty to explain *CTI*. The significantly positive effects of the proxies for information uncertainty on *CTI* remain after controlling for the time-invariant fund characteristics. Notably, the most inclusive specification in column (10) explains 39% of the variation in *CTI*, indicating that *CTI* contains information distinct from that captured by other variables.

V.C.   CTI and Portfolio Performance

V.C.1.   Baseline Analyses

In the baseline analyses in Table IV, I use the measure of portfolio managers' categorical thinking, *CTI*, to study the cross-sectional variation in portfolio performance. The mean *CTI* of funds in each decile portfolio reported in column (1) shows a significant cross-sectional dispersion in *CTI* with a mean of 1.21% (38.4%) for funds in Decile 1 (Decile 10). The Decile 1-10 returns represent a zero-investment strategy that goes long (short) on funds which exhibit the lowest (highest) *CTI*, i.e. least (most) categorical thinking.

In columns (2)-(5), Table IV reports the relation between *CTI* and the components of fund performance based on before-cost (gross) returns. Since my primary goal here is to assess the portfolio managers' investment ability as reflected in their sensitivity to category-wide versus asset-specific information, before-cost returns are first used to capture the value added by managers using their skills. The components of fund returns are reported using the performance decomposition approach of DGTW and Wermers (2000). The following components of returns are analyzed: Gross Holdings Return (i.e. holdings' buy-and-hold stock return), Characteristic Selectivity (*CS*), Average Style (*AS*), and Characteristic Timing (*CT*). The results show that the funds with low *CTI* significantly outperform the funds with high *CTI*. In columns (2) and (3), the funds in *CTI* Decile 1 pick stocks that outperform the stocks picked by funds in *CTI* Decile 10 by an economically substantial 8.03% and 6.13% per year based on Gross Holdings Return and *CS*, respectively. The Gross Holdings Return and *CS* appear to decline with increasing *CTI* deciles nearly uniformly. Overall, the results based on before-cost returns strongly support Prediction 2 in Section II stating that if categorical thinking represents an information-processing bias, then portfolio performance should decrease with categorical thinking.

While gross returns reflect the before-cost value generated by portfolio managers' skill, net returns indicate whether any value created by managers' skill is passed on to investors after deducting fees and transaction costs. In columns (6)-(9), Table IV presents fund performance for *CTI* deciles using after-cost net returns. Consistent with before-cost returns, the net excess returns and risk-adjusted net returns (i.e. alphas) show that the funds with low *CTI* significantly outperform the funds with high *CTI*. For instance, the funds in the lowest *CTI* decile outperform the highest *CTI* decile by a statistically significant four-factor alpha of 2.21% per year.[22] In sum, the results support the notion that portfolio managers' accuracy in predicting future asset returns diminishes with their categorical thinking. Hereafter, for the sake of brevity, the results of remaining nonparametric tests are reported using gross returns, but hold for net returns.

---

[22] In unreported robustness checks, results remained unchanged on using an alternative risk adjustment method. The alphas for each decile portfolio are computed following the two-step Fama and Macbeth (1973) method, where cross-sectional regressions are run in each time period for each decile on common risk factors, followed by time series tests to determine the alphas from the intercepts. The results are available upon request.

V.C.2. Long term performance and CTI

In assessing fund performance and *CTI*, one possibility that is yet to be considered is that funds with higher *CTI* follow strategies that underperform over the short term but outperform over longer horizons. Table V explores this possibility by considering the performance of *CTI* decile portfolios based on returns that are further out in the future from the portfolio formation quarter. For *CTI* computed based on holdings in quarter *t*, the *CS* is reported for the deciles in future quarters *t+2* to *t+12* in two-quarter increments. The Decile 1-10 return spread in *CS* is positive and statistically significant at a meaningful level till quarter *t+10*, ranging from 1.57% to 5.15% per year. In sum, the results based on longer term performance of funds continue to show that funds which have lower *CTI* significantly outperform funds which have higher *CTI*. Thus, *CTI* predicts long-term fund performance.

V.C.3. Multivariate regressions

The association between portfolio performance and categorical thinking has so far been documented in nonparametric settings. However, the existing literature suggests that certain fund attributes affect performance, and could alter the relation between *CTI* and fund returns (e.g. fund size in Berk and Green [2004]). In this section, I employ the following panel regression to study performance in relation to *CTI* in a multivariate setting that simultaneously controls for various factors that may affect performance

(6) $$performance_{f,t} = \alpha_t + \gamma_i + \beta_1 CTI_{f,t-1} + \beta_2 Controls_{f,t-1} + \varepsilon_{f,t}$$

where *performance*$_{f,t}$ is the measure of fund *f*'s performance in month *t* based on either gross returns or net returns, α_t stands for year fixed effects, γ_i stands for benchmark fixed effects where fund *f* is assigned to benchmark *i* in quarter *t*−1, and *Controls*$_{f,t-1}$ is a vector of time-varying fund-specific control variables which include the fund's style, size, activeness, and other attributes measured as of the end of quarter *t*−1. The main estimate of interest is the coefficient on *CTI*$_{f,t-1}$, i.e. the measure of the manager's categorical thinking in quarter *t*−1.

Table VI summarizes the results of the panel regressions explaining mutual fund performance estimated with year and benchmark fixed effects. The relevant *p*-values are based on Newey-West robust standard errors that account for clustering by fund and have a lag length of six months. Columns (1)-(4) and columns (5)-(8) present regressions explaining dependent variables calculated from monthly gross returns and net returns in the three months in quarter *t*, respectively. Supporting prior results based on nonparametric tests, *CTI* has a significantly negative relation with future performance based both on gross returns and net returns. To put the economic significance of *CTI* in perspective, a one standard deviation increase in *CTI* has the same effect on Gross Holdings Return, *CS*, and *AS* as a 17.4%, 14.7%, and 22.2% decrease in the Active Share, respectively. As with the nonparametric tests, there is no detectable relation between *CTI* and portfolio managers' ability to time styles (*CT*), with the main negative effect of *CTI* on performance being via the detrimental impact of *CTI* on their stock and style selection ability.

V.D.   *Robustness checks for CTI*

This section presents several tests relating portfolio performance to *CTI* to examine whether the main results hold across alternative specifications of *CTI*. The alternative specifications vary in the following key elements underlying the *CTI* construct: (1) a stock's categorization when measuring a portfolio manager's responsiveness to category-wide signals, (2) the stock-specific information used to measure the portfolio managers' responsiveness to asset-specific signals, and (3) the measure of the investment decision used to estimate the portfolio managers' responsiveness to information.

V.D.1.   *CTI based on Alternative Stock Categories*

To test the robustness of the results to the definition of a stock's category, I consider a stock's style as an alternative categorization that could be used commonly by portfolio managers. Since style (i.e. categorizations containing stocks with similar market capitalization, book-to-market (B/M), and momentum), market capitalization, and value versus growth are popular stock categorizations, I consider groups of stocks with similar style, size, and book-to-market (B/M) as three categorizations that may

25

impact the information set processed by investors. To implement the style categorization, I assume the stocks ranked in the same DGTW size, B/M, and momentum quintile as belonging to the same style category. Also, I assume the stocks with the same DGTW size and DGTW B/M quintile rank as being placed in the same size-based and B/M-based category, respectively. In month m for any stock i, I compute the style, size, or B/M category returns ($r_{i,m}^{cat}$) for the stock as the equal-weighted return across all the stocks assigned to the same category as stock i in the most recent ranking period, respectively. The average monthly category returns in the prior four quarters are used to obtain a portfolio manager's responsiveness to category-wide signals based on categorizations along style, B/M, and size as $R^{2,style\ category}$, $R^{2,B/M\ category}$, or $R^{2,size\ category}$ as the unadjusted R^2 from Equation (1), respectively. Next, I obtain the stock's category-adjusted returns as the difference between the stock's monthly return and $r_{i,m}^{cat}$. The average monthly category-adjusted stock returns ($r_{i,m}^{stk}$) in the prior four quarters are used to obtain the portfolio manager's responsiveness to asset-specific signals as $R^{2,style\ stk}$, $R^{2,B/M\ stk}$, or $R^{2,size\ stk}$ obtained from the unadjusted R^2 from Equation (2) by regressing the absolute change in active weights of traded stocks on the style, B/M, or size based measure of category-adjusted stock returns, respectively. Finally, I calculate three alternative style-based $CTIs$ (CTI^{style}, $CTI^{B/M\ style}$, and $CTI^{size\ style}$) as $R^{2,style\ category}/(1+R^{2,style\ stk})$, $R^{2,B/M\ category}/(1+R^{2,B/M\ stk})$, and $R^{2,size\ category}/(1+R^{2,size\ stk})$, respectively.

Table VII reports the panel regressions relating CS to alternative specifications of CTI in settings that control for other fund attributes, objective fixed effects, and year fixed effects. The main variable of interest in columns (1) is CTI^{style}, in column (2) is $CTI^{B/M\ style}$, and in column (3) is $CTI^{size\ style}$. The results show that the negative and significant relation between CS and CTI persists across these three alternative specifications of CTI built from style-based stock categorizations. For instance, a one standard deviation increase in CTI^{style} has the same effect on CS as a 14.4% decrease in Active Share, respectively.

V.D.2. CTI based on Alternative Stock-specific Information

The results reported so far are based on CTI measured by the responsiveness of the portfolio manager's investment decisions to lagged category returns relative to the responsiveness to the stock-specific signals

represented by category-adjusted stock returns. The analyses in this section use two different specifications of *CTI* that are based on the following alternative stock-specific information signals− the ambiguity in a stock's categorization (*COARSE*), and the lagged returns of the stock's peers where the peers are selected by matching firms based on fundamentals following Hoberg and Phillips' (2010) Text-based Network Industry Classification. The portfolio manager's responsiveness to *COARSE* ($R^{2,COARSE}$) and fundamentals ($R^{2,peers}$) are measured from the cross-sectional regressions of the absolute change in active weights of traded holdings on the *COARSE* and fundamentals as follows

(7)
$$\left|\Delta activewt_{i,f,t}\right| = \beta_{f,t}^{0} + \beta_{f,t}^{COARSE} COARSE_{i,t-1} + \varepsilon_{f,t}$$

(8)
$$\left|\Delta activewt_{i,f,t}\right| = \beta_{f,t}^{0} + \beta_{1,f,t}^{peers} ret_{i,t-1}^{peers} + \beta_{2,f,t}^{peers} ret_{i,t-2}^{peers} + \beta_{3,f,t}^{peers} ret_{i,t-3}^{peers} + \beta_{4,f,t}^{peers} ret_{i,t-4}^{peers} + \varepsilon_{f,t}$$

$COARSE_{i,t-1}$ is the categorization coarseness of stock i in the last month of quarter $t−1$ (see Equation (3)). $ret_{i,t-p}^{peers}$ is the average monthly return of the equal-weighted portfolio of stock i's Hoberg and Phillips' peers in quarter $t–p$, where p=1,2,3,4. Two alternative specifications of *CTI*, CTI^{COARSE} and CTI^{peers}, are obtained as $R^{2,category}/(1+ R^{2,COARSE})$ and $R^{2,category}/(1+ R^{2,peers})$, where $R^{2,category}$ is obtained as the unadjusted R^2 from Equation (1), and $R^{2,COARSE}$ and $R^{2,peers}$ equal the unadjusted R^2 from the regressions in Equation (7) and (8), respectively.

Table VII reports the panel regressions explaining *CS* in relation to CTI^{COARSE} in column (4), and to CTI^{peers} in column (5). The estimated regressions reveal significantly negative coefficients on CTI^{COARSE} and CTI^{peers}. So, the results using specifications of *CTI* that vary in the asset-specific information used to measure portfolio managers' responsiveness are similar to those obtained using lagged category-adjusted stock returns as the proxy for asset-specific information.

V.D.3. CTI based on Alternative Investment Decision (%*Trade*)

The main results so far have been presented based on using absolute changes in active weights of traded holdings as the measure of the investment decision made by the portfolio manager. Here I repeat the main tests based on the absolute fractional change in holdings as the investment decision, i.e. the dependent

variable in Equations (1) and (2), to obtain four alternative measures of categorical thinking which vary in the stock categories used for the lagged category returns. The fractional change in holdings is the percentage change in the split-adjusted holdings. For holdings representing an addition of a new stock which was not held in the prior quarter, this percentage change is set to equal 100%. CTI^{trades}, $CTI^{style, trades}$, $CTI^{B/M\ style,\ trades}$, and $CTI^{size\ style,\ trades}$ are the CTI measures obtained with trades as the measure of investment decision and the stock's 2-digit SIC industry, style, B/M style, and size style as categories, respectively. Columns (6)-(9) in Table VII report the panel regressions of CS on CTI^{trades}, $CTI^{style,\ trades}$, $CTI^{B/M\ style,\ trades}$, and $CTI^{size\ style,\ trades}$. The results based on using the fractional change in traded holdings and various stock categorizations are qualitatively similar to those obtained from the baseline specification of CTI.

V.E.   Categorization Coarseness and Holdings-level Performance

In this section, I examine the relation between the coarseness (or ambiguity) of a stock's categorization and the portfolio managers' ability to value the stock. I compare the investment returns in stocks with less coarse categorizations to the investment returns in stocks with more coarse categorizations.

To begin the holdings-level analysis of investment returns, I use the stock's categorization coarseness (*COARSE*) described in Section III.A. to sort the set of stocks held by at least one fund in a quarter into quintiles based on *COARSE*. Each fund's portfolio in the quarter is divided into groups containing stocks associated with a *COARSE* quintile, such that quintile Q1 (Q5) comprises of stocks which are likely to be the least (most) ambiguously categorized because their categories are the most (least) informative about the stock-level returns. For each fund, I then compute Gross Holdings Returns and Characteristic Selectivity (*CS*) for each *COARSE* quintile, with the weights computed from the market value of holdings and rescaled to sum to one for each *COARSE* quintile within the fund's portfolio.

Table VIII first presents the return components on an equal-weighted basis and TNA-weighted for the *COARSE* quintiles. The returns are reported as the mean values obtained by computing the cross-sectional mean return for each quintile across all the funds in a month, and then computing the time-series average of the cross-sectional means. As evident from Table VIII, portfolio managers show significantly

more stock selection ability in stocks that have a closer relation with their industry category (i.e. Q1 of *COARSE*) compared to stocks that have an ambiguous relation with their category (i.e. Q5 of *COARSE*). For example, the equal-weighted gross returns and *CS* of stocks in quintile Q1 exceed Q5 by a statistically significant 4.28% and 3.29% per year, respectively.

These findings are consistent with Prediction 3 stated in Section II. Portfolio managers on average are more likely to misvalue stocks when they make biased forecasts by inferring a stock's signals from its category-wide signals even though the stock is not well-represented by the broad category to which it is assigned. This result rules out the alternative possibility that portfolio managers on the whole play the role of informed investors who exploit the potential arbitrage opportunities arising from the misvaluation of the stocks that are coarsely categorized (see Kruger, Landier, and Thesmar (2012)), in which case returns from Q5 should be higher than Q1 of *COARSE*. The significantly positive (negative) *CS* of 1.74% (-1.55%) of *COARSE* quintile Q1 (Q5) suggest that portfolio managers select stocks that outperform (underperform) their characteristics-based DGTW benchmark when the category is a good (bad) representation of the stock's returns. TNA-weighted returns provide similar conclusions, ruling out the possibility that the results are driven solely by small funds. Moreover, the positive spread between *COARSE* Q1−Q5 remains for subsamples of funds with low and high *CTI* values, suggesting that the results are not driven solely by portfolio managers who exhibit more categorical thinking.

To gain further insights into the disparity in portfolio managers' stock selection ability in stocks that differ in the coarseness of categorization, I also examine the performance of holdings separately for subgroups of funds which vary in other attributes. Attributes that could magnify or dampen the effect of information-processing biases include the fund's portfolio characteristics, especially those that reflect the difficulty of the valuation tasks that need to be performed by the portfolio manager. For instance, the effect of a bias could be larger and more detrimental for the returns of funds that are larger and manage more complex assets that impose higher constraints on the portfolio manager's cognitive resources.

Table IX reports the *CS* for quintiles of stocks formed based on *COARSE* for various subsamples of funds. The results are similar across funds which vary in style, TNA, flows, Expense Ratio, Turnover,

Industry Concentration, and Active Share. Also, the holdings with low *COARSE* continue to outperform those with high *COARSE* for funds which vary in the information uncertainty of assets in their portfolio measured based on Cash Flow Volatility, Firm Complexity, and % R&D Stocks. So, the cross-sectional difference in stock selection ability depending on the stock's categorization coarseness persists even for funds which hold relatively easier-to-value assets.

Another issue that may affect the interpretation of the main results on categorization coarseness is whether a stock's *COARSE* measure is correlated with other proxies for the difficulty in valuing the stock. In that case, forecasting errors may be linked to other sources of valuation problems as opposed to the assets' ambiguous categorization. To separate the effect of categorization coarseness from other sources of valuation problems, in this section I conduct tests in which I double-sort stocks based on *COARSE* and various proxies for difficulties associated with valuing the stocks. I sort the set of stocks in each quarter held by at least one fund into quintiles based on their *COARSE* and, independently, based on the following proxies for valuation difficulty: Stock Return Volatility, Cash Flow Volatility, Firm Complexity, and R&D Intensity. Next, stocks held by each fund is linked to quintile ranks in the quarter based on their *COARSE* and proxies for valuation difficulty. The returns of the double-sorted portfolios allow me to draw conclusions about how, controlling for other information problems, the returns vary with *COARSE*.

Table X reports the *CS* for the portfolios of holdings formed by double-sorting on *COARSE* and proxies for valuation difficulty. The positive and significant *CS* spread in *COARSE* Q1−Q5 holds within each quintile of stocks sorted based on Stock Return Volatility, Cash Flow Volatility, Firm Complexity, and R&D Intensity. For instance, for stocks which have low volatility (Q1 of Stock Return Volatility), the stocks in the *COARSE* Q1 outperform the stocks in the *COARSE* Q5 by 2.60% per year on a characteristic-adjusted basis. Similarly, for stocks which have high volatility (Q5 of Stock Return Volatility), the stocks in the *COARSE* Q1 outperform the stocks in the *COARSE* Q5 by 5.12% per year. In sum, even for the subsample of stocks with similar uncertainty and complexity involved in their valuation, the stocks for which the category is more informative are valued more accurately by portfolio managers than the stocks for which the category is less informative.

V.F. *Robustness checks for Categorization Coarseness*

In this subsection, I report results that test whether the findings related to categorization coarseness are robust to using different definitions of categorization coarseness and to adjusting for omitted risk factors.

V.F.1. *Categorization Coarseness based on Style Categories*

For the next set of robustness checks, I consider a stock's style as an alternative to industry-based categorization in computing alternate specifications of categorization coarseness. As before in Section V.D.1., I first assume the stocks ranked in the same DGTW size, B/M, and momentum quintile rank as belonging to the same style category. Alternatively, I assume the stocks with the same DGTW size and B/M quintile rank as having the same size-based and B/M-based category, respectively. Based on these three categorizations, in each month m, I compute the style, size, or B/M category returns (r_m^{cat}) for a stock as the equal-weighted return on the stocks assigned to the same category as the stock in the most recent ranking period. The category returns, $r_{i,m}^{cat}$, are then used in Equation (5) described earlier to compute three style-based specifications of a stock's *COARSE* in the last month of quarter $t-1$. In the portfolio quarter t, all the stocks held by at least one fund are ranked into quintiles based on *COARSE* computed from these alternative categorizations.

Table XI presents the Gross Holdings Returns and *CS* for the *COARSE* quintiles within a fund's portfolio for the style-based specifications of *COARSE*. Consistent with the results using industry-based categorization, holdings that have a closer relation with their style-based category continue to significantly outperform the holdings that have an ambiguous relation with their style-based category. For instance, the gross returns and *CS* of stocks in *COARSE* quintile Q1 exceed Q5 by a statistically significant 7.01% and 6.06% per year when *COARSE* is based on style categories, respectively.

V.F.2. *Categorization Coarseness based on Deviation of Fundamentals from Category*

Next, I consider an alternative definition of a stock's coarseness in industry categorization which is motivated by the Hoberg and Phillips' (2010) premise that a stock's fundamentals can deviate from the stock's official industry category. Hoberg and Phillips (2010) define a firm's peers based on fundamentals using a text-based classification and show that this group can have attributes that differ markedly from the firm's official industry. In the context of this study, a firm can be considered coarsely categorized when its Hoberg and Phillips peers have an uncertain relation with the firm's 2-digit SIC category. To capture this notion, I compute *COARSE* in month *m* for a stock from the volatility of the relation between the stock's peer firms and the stock's official industry category as

(9)
$$COARSE_m = \frac{\sigma_T(r_m^{peers} - r_m^{cat})}{\sigma_T(r_m^{peers})}$$

r_m^{peers} is the equal-weighted return of the stock's Hoberg and Phillips peer firms in month *m*; r_m^{cat} is the equal-weighted return of the stock's 2-digit SIC category in month *m*. $\sigma_T(r_m^{peers} - r_m^{cat})$ and $\sigma_T(r_m^{peers})$ are the standard deviation of the difference $(r_m^{peers} - r_m^{cat})$ and the return r_m^{peers} over the prior 24 months *m*−24 to *m*−1, respectively. In other words, the more volatile the relation between the peer firms' returns and the industry returns, the higher is the stock's coarseness in categorization.

Table XI presents the Gross Holdings Returns and *CS* for the *COARSE* quintiles within a fund's portfolio for *COARSE* computed based on the deviation of fundamentals from the industry category. These results provide similar conclusions as before, since the holdings with low *COARSE* outperform those with high *COARSE* by 4.83% per year on a characteristic-adjusted basis (*p*-value<0.01). The positive spread in *COARSE* Q1−Q5 is driven by both the outperformance of stocks with low *COARSE* (*CS* of 1.99%) in addition to the underperformance of stocks with high *COARSE* (*CS* of -2.84%).

V.F.3. Categorization Coarseness based on R^2

As a third robustness check, I use a different metric to measure categorization coarseness of a stock. The following regression is estimated in each month *t* for stock *i* on a rolling basis explaining the monthly returns of the stock ($r_{i,t}$) using the stock's category returns ($r_{i,t}^{cat}$) over the prior 36 months *t*−36 to *t*−1

(10)
$$r_{i,t} = a_{i,t}^0 + b_{i,t} \, r_{i,t}^{cat} + \varepsilon_{i,t}$$

The unadjusted R^2 from Equation (10) is used to obtain stock i's categorization coarseness ($R2\_COARSE_{i,t}$) in month t as $(1 - R^2)$. This measure captures the intuition that the less the variation in the stock's return is explained by the category returns, i.e. lower the R^2 of Equation (10), the more is the ambiguity of the stock's relation with its category and less is the informativeness of the category regarding the stock.

Table XI presents the Gross Holdings Returns and CS for the $R2\_COARSE$ quintiles within a fund's portfolio. The Gross Holdings Returns and CS of the Q1−Q5 portfolio based on using various stock categories are positive and significant. For example, the Q1−Q5 portfolio sorted on $R2\_COARSE$ measured by regressing stock returns on their industry and DGTW style category returns generate a CS of 2.65% and 3.19% per year, respectively. Thus, the results based on $R2\_COARSE$ are similar to the results using $COARSE$ derived from the volatility of the return difference between a stock and its category.

V.F.4. *Categorization Coarseness and Omitted Risk Factors*

An alternative interpretation of the underperformance of holdings with high categorization coarseness relative to those with low categorization coarseness is that categorization coarseness may proxy for risk factors not captured in the factor models. If this is the case, it is not straightforward to interpret the results as portfolio managers exhibiting more skill in generating higher returns from holdings that have a less uncertain relation with their category than holdings that have a more uncertain relation with their category.

To examine whether this interpretation based on omitted risk factors is borne out in the data, I augment the common factor models with a new factor that captures the cross-section of expected stock returns linked to the stocks' categorization coarseness. Each month, I compute the return on a factor-mimicking Miscategorized-Minus-Well-categorized (*MMW*) portfolio that goes long on high *COARSE* stocks and short on low *COARSE* stocks.[23] The *MMW* factor could be viewed as an omitted risk factor

---

[23] For the *MMW* portfolio, at the end of each quarter, the categorization coarseness for eligible stocks are computed where eligible stocks are selected following Pastor and Stambaugh (2003). The stocks are then sorted into 10 equal-

linked to categorization coarseness, or a mispricing factor capturing systematic misvaluation of stocks which vary in their categorization coarseness. The interpretation of *MMW* is not of particular importance here, since it is meant simply to account for systematic factors linked inherently to categorization coarseness that also predict stock returns. The goal is to incorporate *MMW* as a factor into the model generating the abnormal returns on the *COARSE* quintile portfolios of the holdings, so that the loading and premium on *MMW* captures the proportion of mean return attributable to the passive strategy of going long on high categorization coarseness stocks and short on low categorization coarseness stocks.

Table XII reports the abnormal returns obtained from the augmented four-factor model which adjusts for potential omitted factors linked to *COARSE*. Columns (1)-(5) report the factor-adjusted gross holdings returns for quintiles of *COARSE* based on five alternative specifications of *COARSE* as an equal-weighted mean across all funds in a quarter averaged across all quarters. Column (1) is based on the industry category used in the baseline results. The results hold across all specifications of *COARSE*, with the Q1−Q5 alphas remaining significantly positive and ranging from 2.95% to 7.76% per year for the augmented four-factor model. In sum, controlling for omitted factors in the pricing of stocks that vary in *COARSE* does not affect the relative underperformance of holdings with high *COARSE* compared to holdings with low *COARSE*.

VI. CONCLUSION

This paper proposes that the widespread categorization of assets has a significant impact on how portfolio investors with limited attention process information and form beliefs about future asset prices. Supporting theoretical models in which economic agents have limited attention, I show that portfolio managers− an important and relatively sophisticated class of portfolio investors− are susceptible to information-processing biases linked to categorical thinking that arises when it is common to classify stocks into categories (e.g. small cap, growth). In this study, categorical thinking is captured by the propensity to emphasize category-wide information and deemphasize asset-specific information during belief formation

sized portfolios ranging from Portfolio 1 ("Low-*COARSE* stocks") to 10 ("High- *COARSE* stocks"). The return on the *MMW* portfolio is the return on the equal-weighted Portfolio 10 minus Portfolio 1.

and portfolio decisions. Consistent with categorical thinking being a reflection of an information-processing bias, portfolio managers' tendency to emphasize (deemphasize) category-wide (asset-specific) information increases with information uncertainty and complexity, where psychological evidence predicts that information uncertainty and complexity make cognitive constraints more binding and biases more pronounced. Based on before-cost and after-cost fund level returns, portfolio managers who display more categorical thinking significantly underperform those who display less categorical thinking. This result advances the idea of viewing a portfolio manager's skill from the perspective of cognitive capacity or the reduced susceptibility to behavioral biases. Further, portfolio managers are more skilled in valuing stocks that are closely represented by their category than in valuing stocks with ambiguous relations with their category. This effect of categorization ambiguity on managers' stock selection ability is not explained by other proxies for the difficulty in valuing a stock, like the stock's return volatility. The qualitative results are robust to a wide set of robustness tests involving alternative specifications of categories (e.g. style instead of industry), alternative proxies for category-wide and asset-specific information, and multivariate regression settings controlling for fund-specific attributes, and year and benchmark fixed effects.

The findings in this study have important implications and suggest avenues for future research. Prior work links asset return dynamics and mispricing to the people's tendency to simplify information-processing tasks and reduce choice sets by assigning assets into a few categories (e.g. Barberis and Shleifer [2003]). Taken together with the evidence in this study that even sophisticated money managers are susceptible to information-processing biases associated with categorization and, crucially, that such biases lead to significant economic losses, it appears very plausible that the same underlying bias in judgment could be shared with other agents in financial markets (e.g. financial analysts, individual investors). Future empirical research can add to our understanding of how systematic such biases are across different market participants and what that implies about asset pricing, information dissemination, and financial decisions.

REFERENCES

Allport, Gordon, *The Nature of Prejudice*, eds. (Reading, MA: Addison Wesley, 1954).

Anderson, Barry, Donald Deane, Kenneth Hammond, Garry McClelland, and James Shanteau, *Concepts in judgment and decision research: Definitions, sources, interrelations, comments*, eds. (New York: Praeger, 1981).

Barber, Brad, and Terrance Odean, "All That Glitters: The effect of attention and news on the buying behavior of individual and institutional investors," *Review of Financial Studies*, 21 (2008), 785–818.

Barberis, Nicholas, and Andrei Shleifer, "Style investing," *Journal of Financial Economics*, 68 (2003), 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, "Comovement," *Journal of Financial Economics*, 75 (2005), 283–317.

Berk, Jonathan, and Richard Green, "Mutual fund flows and performance in rational markets," *Journal of Political Economy*, 112 (2004), 1269–1295.

Bikchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change in informational cascades," *Journal of Political Economy*, 100 (1992), 992–1026.

Bruner, Jerome, *Actual minds, Possible worlds*, (Cambridge, MA: Harvard University Press, 1987).

Carhart, Mark, "On persistence in mutual fund performance," *Journal of Finance*, 52 (1997), 57–82.

Cohen, Lauren, and Andrea Frazzini, "Economic Links and Predictable Returns," *Journal of Finance*, 63 (2008), 1977–2011.

Cohen, Lauren, Andrea Frazzini, and Christopher Malloy, "The small world of investing: Board connections and mutual fund returns," *Journal of Political Economy*, 116 (2008), 951-979.

Cooper, Michael, Orlin Dimitrov, and P. Raghavendra Rau, "A rose.com by any name," *Journal of Finance*, 56 (2001), 2371–2388.

Coval, Joshua, and Tobias Moskowitz, "The geography of investment: Informed trading and asset prices," *Journal of Political Economy*, 109 (2001), 811-841.

Cremers, Martijn, and Antti Petajisto, "How active is your portfolio manager? A new measure that predicts performance," *Review of Financial Studies*, 22 (2009), 3329–3365.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers, "Measuring mutual fund performance with characteristic-based benchmarks," *Journal of Finance*, 52 (1997), 1035–1058.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, "Investor psychology and security market under- and overreactions," *Journal of Finance*, 53 (1998), 1839–1886.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, "Overconfidence, arbitrage, and equilibrium asset pricing," *Journal of Finance*, 56 (2001), 921–965.

Daniel, Kent, and Sheridan Titman, "Market's reactions to tangible and intangible information," *Journal of*

Finance, 61 (2006), 1605–1643.

DellaVigna, Stefano, and Joshua Pollet, ″Demographics and industry returns,″ *American Economic Review*, 97 (2007), 1667–1702.

DellaVigna, Stefano, and Joshua Pollet, ″Investor inattention and Friday earnings announcements,″ *Journal of Finance*, 64 (2009), 709–749.

Einhorn, Hillel, ″Overconfidence in judgment,″ *New Directions for Methodology of Social and Behavioral Science,* 4 (1980), 1–16.

Fama, Eugene, and Kenneth French, ″Common risk factors in the return on bonds and stocks,″ *Journal of Financial Economics*, 33 (1993), 3–53.

Fama, Eugene, and Kenneth French, ″Luck versus skill in the cross-section of mutual fund returns,″ *Journal of Finance*, 65 (2010), 1915–1947.

Fang, Lily, Joel Peress, and Lu Zheng, ″Does media coverage of stocks affect mutual funds' trading and performance?,″ INSEAD, and University of California (Irvine), Working Paper, 2011.

French, Kenneth, ″The cost of active investing,″ *Journal of Finance*, 63 (2008), 1537–1573.

Fryer, Roland, and Matthew Jackson, ″A categorical model of cognition and biased decision-making,″ *The B.E. Journal of Theoretical Economics*, 8 (2008), 1–42.

Greenwood, Robin, and Stefan Nagel, ″Inexperienced investors and bubbles,″ *Journal of Financial Economics*, 93 (2009), 239–258.

Griffin, Dale, and Amos Tversky, ″The weighting of evidence and the determinants of overconfidence,″ Cognitive Psychology, 24 (1992), 411–435.

Gupta-Mukherjee, and Swasti, ″Investing in the 'new economy': Mutual fund performance and the nature of the firm,″ *Journal of Financial and Quantitative Analysis*, forthcoming, 2012.

Gupta-Mukherjee, and Ankur Pareek, ″Limited attention and portfolio choice: The impact of attention allocation on mutual fund performance,″ American Finance Association 2013 Annual Meeting Program, 2012.

Hirshleifer, David, ″Investor psychology and asset pricing,″ *Journal of Finance*, 56 (2001), 1533–1598.

Hirshleifer, David, Sonya Seongyeon Lim, and Siew Hong Teoh, ″Driven to distraction: Extraneous events and underreaction to earnings news,″ *Journal of Finance*, 64 (2009), 2289–2325.

Hoberg, Gerard, and Gordon Phillips, ″Product market synergies and competition in mergers and acquisitions: A text-based analysis,″ *Review of Financial Studies*, 23 (2010), 3773–3811.

Hong, Harrison, and Jeremy Stein, ″A unified theory of underreaction, momentum trading and overreaction in asset markets,″ *Journal of Finance*, 54 (1999), 2143–2184.

Hong, Harrison, Jeremy Stein, and Jialin Yu, ″Simple forecasts and paradigm shifts,″ *Journal of Finance*, 62 (2007), 1207–1242.

Huberman, Gur, and Tomer Regev, ″Contagious speculation and a cure for cancer: A non-event that made stock prices soar,″ *Journal of Finance*, 56 (2001), 387–396.

Jensen, Michael, ″The performance of mutual funds in the period 1945-1964,″ *Journal of Finance*, 23 (1968), 389–416.

Jiang, Hao, ″Institutional investors, intangible information, and the book-to-market effect,″ *Journal of Financial Economics*, 96 (2010), 98–126.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, ″On the industry concentration of actively managed equity mutual funds,″ *Journal of Finance*, 60 (2005), 1983–2011.

Kacperczyk, Marcin, and Amit Seru, ″Fund manager use of public information: New evidence on managerial skills,″ *Journal of Finance*, 62 (2007), 485–528.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng, ″Unobserved actions of equity mutual funds,″ *Review of Financial Studies*, 21 (2008), 2379–2416.

Kahneman, Daniel, ″Attention and effort,″ eds. (Englewood Cliffs, NJ: Prentice-Hall, 1973).

Kruger, Phillip, Augustin Landier, and David Thesmar, ″Categorization bias in the stock market,″ University of Geneva, Toulouse School of Economics, and HEC Paris, Working Paper, 2012.

Laurence, Stephen, and Eric Margolis, "Concepts and cognitive science," *Concepts: Core readings* (1999), 3-81.

Mullainathan, Sendhil, ″Thinking through categories,″ MIT mimeo, 2002.

Mullainathan, Sendhil, Joshua Schwartzstein, and Andrei Shleifer, ″Coarse thinking and persuasion,″ *Quarterly Journal of Economics*, 123 (2008), 577–619.

Odean, Terrance, ″Do Investors Trade Too Much?″, *American Economic Review*, 89 (1999), 1279–1298.

Peng, Lin, and Wei Xiong, ″Investor attention, overconfidence, and category learning,″ *Journal of Financial Economics*, 80 (2006), 563–602.

Peski, Marcin, ″Prior symmetry, similarity-based reasoning, and endogenous categorization,″ *Journal of Economic Theory*, 146 (2011), 111–140.

Scharfstein, David, and Jeremy Stein, ″Herd behavior and investment,″ *American Economic Review*, 80 (1990), 465–479.

Schultz, Paul, ″Rational cross-sectional differences in market efficiency: Evidence from mutual fund returns,″ *Journal of Financial and Quantitative Analysis*, 45 (2010), 847-881.

Wermers, Russ, ″Mutual fund performance: An empirical decomposition into stock-picking, style, transaction costs, and expenses,″ *Journal of Finance*, 55 (2000), 1655–1703.

Yantis, Steven, ″Control of visual attention,″ *Attention*, 1 (1998), 223–256.

Zhang, Frank X., ″Information uncertainty and stock returns,″ *Journal of Finance*, 61 (2006), 105–137.

Table I
Descriptive Statistics on Sample

This table reports descriptive statistics on the sample of 2,812 actively managed U.S. equity mutual funds over the period 1990 to 2010. The mean, median, and standard deviation are reported for the following fund and portfolio characteristics: *CTI* (*Categorical Thinking Index*), TNA (total net assets), fund age, manager tenure, # Stocks, Expense Ratio, Turnover, Industry Concentration, and Active Share. A fund's *CTI* is the unadjusted R^2 of the cross-sectional regression of the absolute change in the active weight of a stock traded by the fund on the average monthly returns of the stock's 2-digit SIC industry portfolio in prior quarters lagged up to four quarters, scaled by one plus the unadjusted R^2 of the cross-sectional regression of the absolute change in the active weight on the average monthly category-adjusted returns of the stock in prior quarters lagged up to four quarters. Fund age and manager tenure are the age (in years) of the fund computed from the first offer date, and number of years that the manager has managed the fund as of the end of the quarter in which holdings are reported, respectively. # Stocks is the number of stocks held in the portfolio. Industry Concentration is the Herfindahl index across ten industry categories, computed as the sum of the squared weights allocate to each industry. Active Share is the share of portfolio allocations that differs from the fund's passive benchmark index. . *** represents statistical significance of the correlation coefficient at the 1% level.

	Correlation with *CTI*	Mean	Median	SD
CTI (%)	1.000***	9.94	6.36	10.69
TNA ($ mill)	-0.056***	1,040.6	172.3	3,949.0
Fund Age (years)	0.034***	12.0	8.7	11.8
Manager Tenure (years)	0.127***	6.0	4.9	5.1
# Stocks	-0.251***	116.4	76.0	159.6
Expense Ratio (%)	0.019***	1.257	1.216	0.005
Turnover	-0.253***	0.872	0.660	0.770
Industry Concentration	-0.016***	0.208	0.171	0.139
Active Share	0.095***	0.794	0.820	0.141

Table II
Summary Statistics on Stocks' Categorization Coarseness

This table reports summary statistics and the variation in the categorization coarseness (*COARSE*) and stock characteristics for the stocks in the quintile portfolios formed in each quarter based on the stocks' *COARSE*. The stock characteristics include those that reflect information uncertainty (Cash Flow Volatility, Firm Complexity, and R&D Dummy), and style attributes (Size Quintile, B/M Quintile, and Momentum Quintile). Cash Flow Volatility is the standard deviation of the firm's cash flows over the past 5 years (minimum 3 years). Firm Complexity is the number of business segments reporting sales for the firm in the most recent fiscal year before the quarter. R&D Dummy is a dummy variable assuming a value of one when the firm spends on R&D, and zero otherwise. The Size Quintile, B/M Quintile, and Momentum Quintile are the DGTW size, book-to-market, and momentum quintile assignment of a stock. The first column reports the correlation coefficients of each of the variables with *COARSE*. The Mean, Median, and SD are the sample mean, median, and standard deviation of the variables. For the *COARSE* quintiles, the equal-weighted mean values of the variables are reported across all available quarters. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

	Correlation with *COARSE* (*p*-value)	Mean	Median	SD	*COARSE* Quintile					Q5-Q1 (*p*-value)
					Q1 (Low)	Q2	Q3	Q4	Q5 (High)	
Categorization Coarseness	1.000*** (0.00)	0.933	0.923	0.167	0.69	0.81	0.88	0.96	1.15	0.46*** (0.00)
No. of firms in category	-0.063*** (0.00)	388.64	278.00	360.64	490.94	437.40	388.74	361.92	365.57	-125.38** (0.03)
Cash Flow Volatility	-0.023*** (0.00)	75.60	12.96	172.15	123.41	90.69	71.17	56.04	81.96	-41.45*** (0.01)
Firm Complexity	0.008*** (0.01)	2.17	1.00	1.97	2.26	2.20	2.14	2.08	2.32	0.06 (0.50)
R&D Dummy	0.037*** (0.00)	0.48	0.00	0.50	0.39	0.45	0.47	0.48	0.49	0.10*** (0.00)
Size Quintile	-0.058*** (0.00)	1.99	1.00	1.32	2.75	2.23	1.90	1.73	2.07	-0.68*** (0.00)
B/M Quintile	-0.007*** (0.00)	2.66	2.00	1.42	2.89	2.71	2.63	2.60	2.73	-0.17* (0.10)
Momentum Quintile	-0.007*** (0.00)	3.16	3.00	1.47	3.13	3.14	3.20	3.21	3.08	-0.06 (0.45)

Table III
Determinants of *CTI*

The table reports the results for regressions relating the *CTI* of actively managed mutual funds computed for each fund in each quarter *t* to variables representing information uncertainty and other fund attributes. Log (VIX), log (Cash Flow Volatility), and log (Firm Complexity) are the natural logarithms of the expected market volatility over the next 30 days, the equal-weighted mean of the standard deviation of cash flows over the past 5 years (minimum 3 years) across all firms in the portfolio, and the equal-weighted mean of the number of business segments across all firms in the portfolios, respectively. % R&D Stocks is the fraction of the portfolio invested in stocks of firms that spend on R&D. Herding Measure is computed following Wermers (2000). Size, B/M, and Momentum Style are the natural logarithms of the value-weighted DGTW size, book-to-market, and momentum quintiles of the stocks in the portfolio, respectively. Fund Flows is the mean monthly growth in TNA due to new money over the three months in quarter *t*-1. The other explanatory variables are as defined in Table 1. *p*-values based on Newey-West standard errors that account for clustering at the fund level with a lag length of three quarters are reported in parentheses. ***, **, * represent 1%, 5%, 10% confidence levels, respectively.

Variable (*t*)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Dependent Variable: *CTI* (*t*)					
log (VIX)	0.011***				0.011***	0.012***	0.011***	0.011***	0.010***	0.010***
	(0.00)				(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log (Cash Flow Volatility)		0.012***			0.008**	0.010***	0.010***	0.007***	0.018***	0.013***
		(0.00)			(0.02)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log (Firm Complexity)			0.063***		0.061***	0.062***	0.061***	0.052***	0.065***	0.017***
			(0.00)		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
% R&D Stocks				0.010***	0.009**	0.009**	0.006	0.009	0.015***	0.002
				(0.01)	(0.04)	(0.05)	(0.32)	(0.15)	(0.00)	(0.66)
Herding Measure/ 100						0.019***	0.018***	0.018***	0.013***	0.007***
						(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Size Style							-0.030***	-0.008**	-0.001	0.002
							(0.00)	(0.02)	(0.73)	(0.73)
B/M Style							-0.053***	-0.042***	-0.050***	-0.027***
							(0.00)	(0.00)	(0.00)	(0.00)
Momentum Style							-0.073***	-0.072***	-0.032***	-0.008**
							(0.00)	(0.00)	(0.00)	(0.03)
log (TNA) /100								-0.012***	-0.011***	-0.003***
								(0.00)	(0.00)	(0.00)
log (Fund Age)								0.007***	0.008***	0.003
								(0.00)	(0.00)	(0.14)
log (Manager Tenure)								0.020***	0.010***	0.003***
								(0.00)	(0.00)	(0.01)
Team-managed Dummy								-0.004***	-0.005***	-0.002
								(0.00)	(0.00)	(0.12)
Past Flow								-0.014*	-0.025***	-0.045***
								(0.09)	(0.00)	(0.00)
Expense Ratio									0.396***	-0.152
									(0.01)	(0.62)
Turnover									-0.024***	-0.015***
									(0.00)	(0.00)
Industry Concentration									0.055***	0.041***
									(0.00)	(0.00)
Active Share									0.235***	0.137***
									(0.00)	(0.00)
Objective Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fund fixed effects	No	No	No	No	No	No	No	No	No	Yes
No. of observations	51,849	51,849	51,849	51,849	51,849	51,849	51,849	51,849	51,849	51,849
R-square	0.040	0.058	0.065	0.041	0.067	0.075	0.089	0.124	0.207	0.390

Table IV
Performance of Mutual Funds Sorted on *CTI*

At the end of each quarter *t*, funds are sorted into decile portfolios based on *CTI*. Gross Return and Net Return refer to the holdings-based buy-and-hold monthly return and the monthly net fund return over the three months in quarter *t+1*, respectively. Column (1) reports the mean *CTI* for the funds in each decile portfolio based on *CTI*. Column (2) reports the monthly buy-and-hold Gross Holdings Return. Columns (3)-(5) reports the components of the gross holdings return, namely, Characteristic Selectivity (*CS*), Characteristic Timing (*CT*), and Average Style (*AS*). *CS* is a measure of stock selection ability and is defined in month *t* as $CS = \sum w_{j,t-1}[R_{j,t} - BR_t(j,t\text{-}1)]$, where $w_{j,t-1}$ is the weight of stock *j* in the fund's portfolio as of the end of month *t−1*, $R_{j,t}$ is the month *t* return of stock *j*, and $BR_t(j,t\text{-}1)$ is the month *t* return of the benchmark portfolio to which stock *j* was allocated in month *t−1* according to its size, value, and momentum characteristics. *CT* is a measure of style timing ability and is defined as $CT = \sum[w_{j,t-1}BR_t(j,t\text{-}1) - w_{j,t-5}BR_t(j,t\text{-}5)]$. *AS* is a measure of style selection ability, and is defined as $AS = \sum[w_{j,t-5}BR_t(j,t\text{-}5)]$. Column (6) reports the time series means of each decile portfolio's net excess returns computed as the equal-weighted mean monthly excess net fund return for funds in each decile portfolio in each month over the sample period. Columns (7), (8), and (9) report the One-, Three-, and Four-factor Alphas computed as the intercept from the time-series regressions of the decile portfolio's net excess return on the excess market return (Jensen (1968)), the Fama and French (1993) three factors (market, size, and value), and Carhart's (1997) four factors that add a momentum factor to the Fama and French (1993) three factors, respectively. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

		Performance based on Gross Return (% per year)				Performance based on Net Return (% per year)			
CTI Decile Rank (*t*)	Mean *CTI* (%)	Gross Holdings Return	Characteristic Selectivity (*CS*)	Average Style (*AS*)	Characteristic Timing (*CT*)	Net Excess Return	One-factor Alpha	Three-factor Alpha	Four-factor Alpha
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Decile 1 (Low *CTI*)	1.21	15.26***	3.88***	10.88***	0.52	7.12***	2.28***	1.40**	1.60***
		(0.01)	(0.00)	(0.01)	(0.38)	(0.00)	(0.01)	(0.03)	(0.01)
Decile 2	2.62	13.11**	2.64***	10.27***	0.24	6.14**	1.74**	1.24*	2.47***
		(0.02)	(0.00)	(0.01)	(0.54)	(0.04)	(0.04)	(0.07)	(0.00)
Decile 3	3.91	14.17***	2.91***	10.67***	0.60	6.73***	0.86	0.58	1.27*
		(0.01)	(0.01)	(0.01)	(0.43)	(0.01)	(0.17)	(0.36)	(0.08)
Decile 4	5.30	12.26**	1.48*	10.65***	0.11	5.90**	0.83	0.43	0.43
		(0.03)	(0.09)	(0.01)	(0.82)	(0.04)	(0.23)	(0.47)	(0.52)
Decile 5	6.90	12.65**	1.54*	10.30***	0.78	5.74*	0.65	0.70	0.91
		(0.03)	(0.07)	(0.01)	(0.35)	(0.06)	(0.55)	(0.25)	(0.22)
Decile 6	8.88	10.70*	0.26	10.22***	0.23	6.08**	0.48	0.82	0.77
		(0.06)	(0.78)	(0.01)	(0.79)	(0.05)	(0.74)	(0.13)	(0.26)
Decile 7	11.46	10.40*	-0.30	10.23***	0.47	6.41*	0.99	0.64	0.60
		(0.07)	(0.77)	(0.01)	(0.89)	(0.06)	(0.31)	(0.26)	(0.42)
Decile 8	15.16	9.65*	-0.56	10.19***	0.02	5.80*	0.34	0.36	0.14
		(0.09)	(0.62)	(0.01)	(0.30)	(0.06)	(0.83)	(0.55)	(0.87)
Decile 9	21.35	7.90	-1.91***	10.63***	-0.79	4.84*	-0.14	0.23	-0.30
		(0.16)	(0.00)	(0.00)	(0.96)	(0.08)	(0.77)	(0.58)	(0.74)
Decile 10 (High *CTI*)	38.40	7.23	-2.25***	9.86**	-0.33	3.41*	-1.03	-0.98	-0.61
		(0.24)	(0.00)	(0.02)	(0.87)	(0.08)	(0.91)	(0.25)	(0.38)
Decile 1-10	-37.19***	8.03***	6.13***	1.02	0.85	3.71**	3.31***	2.38***	2.21***
(*p*-value)	(0.00)	(0.00)	(0.00)	(0.16)	(0.45)	(0.02)	(0.00)	(0.01)	(0.00)

Table V
Long Term Performance of Mutual Funds Sorted on *CTI*

The table report monthly fund performance for *CTI* decile portfolios formed in quarter t in quarters $(t + 3)$, $(t + 4)$, $(t + 5)$, $(t + 6)$, $(t + 7)$, and $(t + 8)$, measured as the Characteristic Selectivity (*CS*) computed from gross returns. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

CTI Decile Rank (t)	$t+2$	$t+4$	$t+6$	$t+8$	$t+10$	$t+12$
	(1)	(2)	(3)	(4)	(5)	(6)
Decile 1 (Low *CTI*)	2.23***	1.57**	1.09	0.79	0.63	0.95
	(0.01)	(0.02)	(0.38)	(0.15)	(0.35)	(0.20)
Decile 2	1.56**	0.88	1.35	0.62	0.56	0.40
	(0.04)	(0.42)	(0.25)	(0.57)	(0.39)	(0.52)
Decile 3	1.27	1.19	0.80	0.81	0.90	0.76
	(0.31)	(0.26)	(0.52)	(0.40)	(0.41)	(0.52)
Decile 4	0.85	0.70	-0.09	-0.58	0.14	0.24
	(0.45)	(0..53)	(0.93)	(0.53)	(0.88)	(0.68)
Decile 5	0.31	0.12	-0.41	0.00	0.10	0.08
	(0.83)	(0.91)	(0.69)	(1.00)	(0.92)	(0.96)
Decile 6	0.03	-0.39	-0.02	0.16	-0.33	0.11
	(0.98)	(0.70)	(0.98)	(0.88)	(0.73)	(0.87)
Decile 7	-0.20	-0.11	-0.16	-0.53	-0.51	-0.33
	(0.86)	(0.92)	(0.86)	(0.55)	(0.65)	(0.58)
Decile 8	-0.56	-1.18	-1.33	-0.97	-0.90	-0.13
	(0.34)	(0.16)	(0.22)	(0.32)	(0.13)	(0.93)
Decile 9	-1.98***	-2.27**	-3.14***	-1.45**	-1.45**	-0.61
	(0.01)	(0.03)	(0.00)	(0.04)	(0.05)	(0.29)
Decile 10 (High *CTI*)	-2.92***	-2.02**	-2.42***	-1.67**	-0.94*	-0.08
	(0.00)	(0.00)	(0.00)	(0.02)	(0.06)	(0.95)
Decile 1-10	5.15***	3.59***	3.51***	2.46***	1.57**	1.03
(*p*-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.16)

43

Table VI
Multivariate Regressions Explaining Mutual Fund Performance

The table reports the results for regressions relating mutual fund performance and *CTI*. The panel regression $performance_{f,t} = \beta_0 + \beta_1 CTI_{f,t-1} + \beta_2 Controls_{f,t-1} + \epsilon_{f,t}$ is estimated, where the dependent variable $performance_{f,t}$ is the monthly fund performance measured in the three months in quarter t as percentage per year based on Net Return (Net Excess Return, One-, Three-, or Four-factor Alpha), or Gross Return (Gross Holdings Return, *CS*, *CT*, or *AS*). The explanatory variables are measured at the end of quarter t-1 and are defined in Table III. p-values based on Newey-West standard errors that account for clustering at the fund level with a lag length of six months are reported in parentheses. All specifications include objective and year fixed effects. ***, **, * represent 1%, 5%, 10% confidence levels, respectively.

Variable (t)	Performance based on Gross Return (% per year)				Performance based on Net Return (% per year)			
	Gross Holdings Return	CS	AS	CT	Net Excess Return	One-factor Alpha	Three-factor Alpha	Four-factor Alpha
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CTI	-0.153***	-0.037***	-0.131***	-0.003	-0.072***	-0.035***	-0.029**	-0.026***
	(0.00)	(0.00)	(0.00)	(0.34)	(0.00)	(0.01)	(0.02)	(0.01)
Size Style	-0.047***	0.005	-0.067***	0.012***	-0.012*	-0.007*	0.013***	0.011***
	(0.00)	(0.14)	(0.00)	(0.00)	(0.06)	(0.07)	(0.00)	(0.00)
B/M Style	-0.034***	-0.015***	-0.022*	0.005**	0.013	0.019***	-0.012***	0.014***
	(0.01)	(0.01)	(0.06)	(0.02)	(0.12)	(0.00)	(0.01)	(0.00)
Momentum Style	-0.109***	-0.063***	-0.072***	0.022***	-0.012	0.000	-0.002	-0.009
	(0.00)	(0.00)	(0.00)	(0.00)	(0.21)	(0.94)	(0.75)	(0.16)
Herding Measure	-0.021***	0.007***	-0.031***	0.004***	-0.006**	-0.001	-0.001	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.48)	(0.67)	(0.00)
log (TNA) /100	-0.747***	-0.121***	-0.710***	-0.014	-0.511***	-0.159***	0.029	-0.045
	(0.00)	(0.01)	(0.00)	(0.54)	(0.00)	(0.00)	(0.48)	(0.27)
log (Fund Age) /100	-1.119***	-0.212	-0.495	-0.133*	1.409***	0.219	0.404***	0.490***
	(0.01)	(0.14)	(0.19)	(0.09)	(0.00)	(0.18)	(0.00)	(0.00)
log (Manager Tenure) /100	-0.174	-0.178	-0.088	0.088	0.105	-0.302**	-0.036	-0.100
	(0.67)	(0.21)	(0.81)	(0.21)	(0.71)	(0.05)	(0.78)	(0.44)
Past Flow	-0.066	0.059***	-0.148***	-0.009	-0.246***	-0.020	0.020	-0.009
	(0.14)	(0.00)	(0.00)	(0.25)	(0.00)	(0.29)	(0.22)	(0.57)
Expense Ratio	-1.653***	-0.307	-1.341***	-0.037	-1.101***	-1.183***	-0.565***	-0.784***
	(0.00)	(0.13)	(0.00)	(0.69)	(0.00)	(0.00)	(0.00)	(0.00)
Turnover /100	0.060	-0.061	0.011	0.138**	-0.170	-0.083	-0.076	-0.222**
	(0.85)	(0.67)	(0.97)	(0.04)	(0.38)	(0.48)	(0.49)	(0.04)
Industry Concentration	-0.125***	-0.028	-0.178***	0.036***	-0.068***	-0.037***	0.024**	0.015
	(0.01)	(0.33)	(0.00)	(0.00)	(0.00)	(0.00)	(0.05)	(0.25)
Active Share	0.094***	0.027***	0.063***	-0.007**	0.127***	0.125***	0.054***	0.070***
	(0.00)	(0.00)	(0.00)	(0.03)	(0.00)	(0.00)	(0.00)	(0.00)
Objective Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	136,344	136,344	125,180	124,815	154,559	143,805	143,805	143,805
R-square	0.072	0.011	0.078	0.012	0.140	0.032	0.017	0.018

Table VII
Mutual Fund Performance and Alternative Specifications of *CTI*

The table reports the results for regressions relating mutual fund performance and alternative specifications of *CTI*. The panel regression $performance_{f,t} = \beta_0 + \beta_1\,CTI_{f,t-1} + \beta_2\,Controls_{f,t-1} + \epsilon_{f,t}$ is estimated, where the dependent variable $performance_{f,t}$ is the monthly fund performance measured as percentage per year based on the Characteristic Selectivity (*CS*) computed from gross returns. CTI^{style}, $CTI^{size\ style}$, and $CTI^{B/M\ style}$ are alternative specifications of *CTI* measured using stock categories based on the stock's DGTW style quintiles, DGTW size quintiles, and DGTW B/M quintiles as described in Section E.1., respectively. CTI^{stock} and CTI^{peers} are specifications of *CTI* measured using stock returns (equation (7)) and peer firms' returns (equation (8)) as the stock-specific information to which the fund managers' responsiveness is measured, respectively. CTI^{trades} is a specification of *CTI* measured based on trades as the investment decision as described in Section E.3. *p*-values based on Newey-West standard errors that account for clustering at the fund level with a lag length of six months are reported in parentheses. All specifications include objective and year fixed effects. ***, **, * represent 1%, 5%, 10% confidence levels, respectively.

Variable	CTI^{style} (1)	$CTI^{B/M\ style}$ (2)	$CTI^{size\ style}$ (3)	CTI^{COARSE} (4)	CTI^{peers} (5)	CTI^{trades} (6)	$CTI^{style,\ trades}$ (7)	$CTI^{B/M\ style,\ trades}$ (8)	$CTI^{size\ style,\ trades}$ (9)
CTI	-0.037***	-0.029***	-0.012	-0.024***	-0.035***	-0.029***	-0.022***	-0.023***	-0.020**
	(0.00)	(0.00)	(0.19)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.02)
Size Style	0.007**	0.006*	0.004	0.005	0.005	0.004	0.006*	0.006*	0.005
	(0.04)	(0.07)	(0.20)	(0.15)	(0.19)	(0.26)	(0.06)	(0.07)	(0.11)
B/M Style	-0.011**	-0.011**	-0.011**	-0.016***	-0.022***	-0.016***	-0.011**	-0.011**	-0.010**
	(0.03)	(0.03)	(0.03)	(0.01)	(0.00)	(0.01)	(0.03)	(0.03)	(0.04)
Momentum Style	-0.046***	-0.045***	-0.043***	-0.063***	-0.074***	-0.062***	-0.039***	-0.039***	-0.038***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Herding Measure	0.006***	0.005***	0.005***	0.006***	0.009***	0.006***	0.005***	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
log (TNA) /100	-0.128***	-0.116***	-0.088**	-0.118***	-0.145***	-0.101**	-0.120***	-0.116***	-0.106***
	(0.00)	(0.01)	(0.03)	(0.01)	(0.01)	(0.03)	(0.00)	(0.01)	(0.01)
log (Fund Age) /100	-0.171	-0.184	-0.212	-0.209	-0.177	-0.221	-0.210	-0.214	-0.222
	(0.23)	(0.20)	(0.14)	(0.15)	(0.29)	(0.13)	(0.14)	(0.13)	(0.12)
log (Manager Tenure) /100	-0.084	-0.100	-0.130	-0.174	-0.207	-0.183	-0.065	-0.067	-0.082
	(0.54)	(0.47)	(0.34)	(0.22)	(0.19)	(0.20)	(0.64)	(0.63)	(0.55)
Past Flow	0.052***	0.052***	0.053***	0.058***	0.070***	0.058***	0.052***	0.052***	0.052***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Expense Ratio	-0.214	-0.221	-0.238	-0.310	-0.330	-0.330	-0.197	-0.195	-0.206
	(0.26)	(0.24)	(0.21)	(0.12)	(0.14)	(0.09)	(0.30)	(0.30)	(0.27)
Turnover /100	-0.169	-0.142	-0.080	-0.054	-0.068	-0.014	-0.168	-0.165	-0.140
	(0.23)	(0.31)	(0.56)	(0.70)	(0.68)	(0.92)	(0.23)	(0.23)	(0.30)
Industry Concentration	-0.030	-0.035	-0.042	-0.029	-0.070*	-0.028	-0.050*	-0.051**	-0.052**
	(0.26)	(0.19)	(0.12)	(0.31)	(0.06)	(0.33)	(0.06)	(0.05)	(0.05)
Active Share	0.030***	0.028***	0.023***	0.027***	0.035***	0.025***	0.028***	0.028***	0.028***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Objective Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	135,722	135,722	135,722	136,344	109,397	136,194	135,488	135,488	135,488
R-square	0.022	0.023	0.019	0.024	0.027	0.023	0.022	0.022	0.022

Table VIII
Performance of Fund Holdings Sorted by Categorization Coarseness

Each quarter, the stocks held by at least one fund are sorted into quintiles Q1 (Low) through Q5 (High) based on the stock's categorization coarseness (*COARSE*) as of the last month of the prior quarter *t*-1. Each fund's portfolio is split based on the *COARSE* quintile of the stocks. The Gross Holdings Return, Characteristic Selectivity (*CS*), and the four-factor alpha obtained from the excess Gross Holdings Return are reported for each quintile portfolio on an equal-weighted basis across funds unless specified otherwise. The returns are computed on a monthly basis and reported as percentage per year. The mean values are reported by taking the time series average of the cross-sectional averages in each month. Funds with Low (High) *CTI* are the funds ranked in the bottom (top) 33.3% in the quarter based on *CTI*. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

Categorization Coarseness Quintile	Equal-weighted returns (% per year)		TNA-weighted returns (% per year)	
	Gross Holdings Return	*CS*	Gross Holdings Return	*CS*
Q1 (Low *COARSE*)	16.95***	1.74***	12.24***	1.01
	(0.00)	(0.01)	(0.00)	(0.24)
Q2	16.77***	0.76	10.31***	0.34
	(0.00)	(0.58)	(0.01)	(0.72)
Q3	15.62**	0.34	9.62***	-0.26
	(0.02)	(0.68)	(0.01)	(0.81)
Q4	15.92***	0.42	8.83***	-0.86
	(0.00)	(0.56)	(0.00)	(0.39)
Q5 (High *COARSE*)	12.67**	-1.55**	8.57***	-1.52**
	(0.05)	(0.02)	(0.01)	(0.04)
Q1-Q5	4.28***	3.29***	3.67**	2.53***
(*p*-value)	(0.01)	(0.00)	(0.02)	(0.01)
	Funds with Low *CTI*		Funds with High *CTI*	
Categorization Coarseness Quintile	Gross Holdings Return	*CS*	Gross Holdings Return	*CS*
Q1 (Low *COARSE*)	15.20***	1.69**	13.27***	1.13*
	(0.00)	(0.04)	(0.00)	(0.07)
Q2	13.44**	0.62	12.33**	0.22
	(0.02)	(0.50)	(0.03)	(0.86)
Q3	12.72**	0.41	11.21***	-0.35
	(0.02)	(0.74)	(0.00)	(0.56)
Q4	12.03**	0.54	10.11**	-0.29
	(0.03)	(0.52)	(0.02)	(0.68)
Q5 (High *COARSE*)	10.77**	-0.83	8.03**	-1.99***
	(0.03)	(0.23)	(0.02)	(0.01)
Q1-Q5	4.43**	2.52**	5.24***	3.12***
(*p*-value)	(0.03)	(0.02)	(0.00)	(0.00)

Table IX
Performance of Fund Holdings Sorted by Categorization Coarseness (by Fund Attributes)

Each quarter, the stocks held by at least one fund are sorted into quintiles Q1 (Low) through Q5 (High) based on the stock's categorization coarseness as of the last month of the prior quarter t-1. At the end of each quarter, funds are sorted into terciles based on the following fund or portfolio attributes: Size Style, B/M Style, Momentum Style, TNA, Fund Flows, Expense Ratio, Turnover, Active Share, Industry Concentration, Cash Flow Volatility, Firm Complexity, and % R&D Stocks. The fund and portfolio attributes rankings are as defined in Table VI. The Characteristic Selectivity (CS) of the portfolios are reported in percentage per year, and are computed as the time series mean of the average CS in each month. The p-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

Categorization Coarseness	Size Style		B/M Style		Momentum Style	
Quintile	Small Cap	Large Cap	Growth	Value	Contrarian	Momentum
Q1 (Low *COARSE*)	1.41**	0.36	1.92***	1.71**	2.05***	1.33*
Q5 (High *COARSE*)	-0.31	-1.23	-0.40	-0.67	-1.08	-0.03
Q1-Q5	1.72**	1.59**	2.32***	2.38***	3.13***	1.36
(*p*-value)	(0.05)	(0.05)	(0.00)	(0.00)	(0.00)	(0.15)
Categorization Coarseness	TNA		Fund Flows		Expense Ratio	
Quintile	Small	Large	Low	High	Low	High
Q1 (Low *COARSE*)	2.03***	1.72*	1.39*	1.48**	1.84**	1.56**
Q5 (High *COARSE*)	-0.67	-0.70	-1.01	-0.11	-0.88	-0.55
Q1-Q5	2.70***	2.42***	2.40***	1.59*	2.72***	2.11**
(*p*-value)	(0.00)	(0.01)	(0.00)	(0.07)	(0.00)	(0.03)
Categorization Coarseness	Turnover		Industry Concentration		Active Share	
Quintile	Low	High	Low	High	Low	High
Q1 (Low *COARSE*)	1.85***	1.20	1.94**	1.36*	1.89**	1.29
Q5 (High *COARSE*)	-0.91	-0.58	-0.85	-0.20	-1.11*	-0.52
Q1-Q5	2.76**	1.78**	2.79***	1.56*	3.00***	1.81**
(*p*-value)	(0.02)	(0.02)	(0.00)	(0.06)	(0.00)	(0.03)
Categorization Coarseness	Cash Flow Volatility		Firm Complexity		% R&D Stocks	
Quintile	Low	High	Low	High	Low	High
Q1 (Low *COARSE*)	1.33*	2.14***	1.17	1.75**	1.95***	1.45**
Q5 (High *COARSE*)	-0.31	-1.29*	-0.28	-1.16*	-0.60	-0.52
Q1-Q5	1.64**	3.43***	1.45*	2.91***	2.55***	1.97***
(*p*-value)	(0.05)	(0.00)	(0.06)	(0.00)	(0.00)	(0.01)

Table X
Fund Holdings Sorted by Categorization Coarseness and Stock Characteristics

Each quarter, the stocks held by at least one fund are sorted into quintiles Q1 through Q5 based on the stock's categorization coarseness (*COARSE*) and, independently, into quartiles Q1 through Q5 based on one of the following stock attributes: Stock Return Volatility (Panel A), Cash Flow Volatility (Panel B), Firm Complexity (Panel C), and R&D Intensity (Panel D) as of the last month of the prior quarter *t*-1. R&D Intensity is the ratio of R&D to Property, Plant, and Equipment expenses. The other fund and portfolio attributes rankings are as defined in Table VI. The quintiles based on Firm Complexity and R&D Intensity are formed by grouping all stocks with one business segment and zero R&D expenses, respectively, into Q1 and sorting the remaining stocks into four equal-sized groups Q2 through Q5. Each fund's portfolio is split based on the categorization coarseness quintile and stock attribute quartile of the stocks. The Characteristic Selectivity (*CS*) of the portfolios formed by the double-sorting are reported in percentage per year, and are computed as the time series mean of the average *CS* in each month. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

Categorization Coarseness	Panel A: Quintiles based on Stock Return Volatility				
Quintile	Q1 (Low Volatility)	Q2	Q3	Q4	Q5 (High Volatility)
Q1 (Low *COARSE*)	1.37*	1.73***	3.36***	4.55***	3.87***
Q5 (High *COARSE*)	-1.23	-1.17	-0.47	-3.11	-1.25
Q1-Q5	2.60**	2.90***	3.83***	7.66***	5.12***
(*p*-value)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)

Categorization Coarseness	Panel B: Quintiles based on Cash Flow Volatility				
Quintile	Q1 (Low Volatility)	Q2	Q3	Q4	Q5 (High Volatility)
Q1 (Low *COARSE*)	1.80**	1.35*	2.05***	2.55***	2.65***
Q5 (High *COARSE*)	-2.88***	-1.73**	-0.08	0.23	-1.16*
Q1-Q5	4.68***	3.08***	2.13**	2.32***	3.81***
(*p*-value)	(0.00)	(0.00)	(0.04)	(0.00)	(0.00)

Categorization Coarseness	Panel C: Quintiles based on Firm Complexity				
Quintile	Q1 (Low Complexity)	Q2	Q3	Q4	Q5 (High Complexity)
Q1 (Low *COARSE*)	2.13**	1.79***	2.51***	1.91***	1.03
Q5 (High *COARSE*)	-0.23	-0.64	-0.14	-0.03	-2.85***
Q1-Q5	2.36***	2.43**	2.65***	1.94**	3.88***
(*p*-value)	(0.01)	(0.02)	(0.00)	(0.04)	(0.00)

Categorization Coarseness	Panel D: Quintiles based on R&D Intensity				
Quintile	Q1 (Low Uncertainty)	Q2	Q3	Q4	Q5 (High Uncertainty)
Q1 (Low *COARSE*)	2.65***	0.23	1.64**	1.97***	3.55***
Q5 (High *COARSE*)	-0.82	-3.58***	-1.48	-0.08	-0.08
Q1-Q5	3.47***	3.81***	3.12***	2.05**	3.63***
(*p*-value)	(0.00)	(0.00)	(0.01)	(0.04)	(0.00)

Table XI

Holdings Sorted by Alternative Specifications of Categorization Coarseness

Each quarter, the stocks held by at least one fund are sorted into quintiles Q1 (Low) through Q5 (High) based on the stock's categorization coarseness as of the last month of the prior quarter t-1. In Panel A, categorization coarseness is measured using stock categories based on the stock's DGTW style quintiles, DGTW size quintiles, or DGTW B/M quintiles as described in Section G.2, or based on the deviation of the stock's peer firms' returns (based on Hoberg and Phillips' (2010) peers) from the industry category returns as described in Section G.3. In Panel B, categorization coarseness is measured using the $(1$-$R^2)$ obtained from regressing the stock's monthly returns on the category's monthly returns over 36 months as described in Section G.4., where stock categories are based on the stock's 2-digit SIC industry category, DGTW style quintiles, DGTW size quintiles, or DGTW B/M quintiles. Each fund's portfolio is split based on the categorization coarseness quintile of the stocks. The Gross Holdings Return and Characteristic Selectivity (*CS*) are reported for each quintile portfolio computed on a monthly basis and reported as percentage per year. The mean values are reported by taking the time series average of the cross-sectional averages in each month. The *p*-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

	Panel A: Categorization Coarseness based on Alternative Stock Categories				
	Style Categories			Size Style Categories	
Categorization Coarseness	Gross Holdings			Gross Holdings	
Quintile	Return	*CS*		Return	*CS*
Q1 (Low *COARSE*)	14.55***	3.21***		14.79***	2.70***
	(0.00)	(0.00)		(0.00)	(0.00)
Q2	12.38***	1.22*		13.67***	2.44***
	(0.01)	(0.10)		(0.00)	(0.00)
Q3	9.48**	-0.92		9.73**	-0.94
	(0.02)	(0.22)		(0.02)	(0.16)
Q4	8.64**	-2.56**		7.92**	-2.19**
	(0.02)	(0.03)		(0.05)	(0.04)
Q5 (High *COARSE*)	7.54**	-2.85***		7.20*	-2.48**
	(0.04)	(0.01)		(0.07)	(0.05)
Q1-Q5	7.01***	6.06***		7.59***	5.18***
(*p*-value)	(0.00)	(0.01)		(0.00)	(0.00)
	B/M Style Categories			Deviation of Fundamentals	
Categorization Coarseness	Gross Holdings			Gross Holdings	
Quintile	Return	*CS*		Return	*CS*
Q1 (Low *COARSE*)	15.46***	4.14***		7.90*	1.99**
	(0.00)	(0.00)		(0.10)	(0.03)
Q2	11.61***	1.28		6.23	2.07**
	(0.01)	(0.23)		(0.19)	(0.02)
Q3	10.34**	-0.68		7.20	1.12
	(0.03)	(0.43)		(0.13)	(0.23)
Q4	10.80***	-0.53		5.63	0.41
	(0.00)	(0.50)		(0.34)	(0.70)
Q5 (High *COARSE*)	9.15***	-0.98		3.29	-2.84***
	(0.01)	(0.25)		(0.60)	(0.01)
Q1-Q5	6.31**	5.12***		4.61**	4.83***
(*p*-value)	(0.02)	(0.00)		(0.05)	(0.00)

Table XI (continued)
Holdings Sorted by Alternative Specifications of Categorization Coarseness

	Panel B: Categorization Coarseness measured as *R2_COARSE*				
	Industry Categories			Style Categories	
Categorization Coarseness	Gross Holdings			Gross Holdings	
Quintile	Return	*CS*		Return	*CS*
Q1 (Low *R2_COARSE*)	12.76***	1.45**		13.58***	1.97***
	(0.00)	(0.03)		(0.00)	(0.00)
Q2	12.29***	1.33*		11.56***	1.47**
	(0.00)	(0.06)		(0.01)	(0.03)
Q3	11.69***	0.67		10.97***	0.53
	(0.01)	(0.35)		(0.00)	(0.32)
Q4	9.51***	-1.31**		10.02**	-0.32
	(0.01)	(0.03)		(0.02)	(0.70)
Q5 (High *R2_COARSE*)	7.83**	-1.20*		7.94**	-1.22**
	(0.03)	(0.06)		(0.04)	(0.05)
Q1-Q5	4.93***	2.65***		5.64***	3.19***
(*p*-value)	(0.00)	(0.01)		(0.00)	(0.00)
	Size Style Categories			B/M Style Categories	
Categorization Coarseness	Gross Holdings			Gross Holdings	
Quintile	Return	*CS*		Return	*CS*
Q1 (Low *R2_COARSE*)	12.70***	0.50		14.28***	2.02***
	(0.00)	(0.72)		(0.00)	(0.00)
Q2	12.95***	0.91		12.14***	0.92
	(0.00)	(0.17)		(0.01)	(0.25)
Q3	10.94***	0.08		11.03***	0.56
	(0.01)	(0.92)		(0.00)	(0.54)
Q4	9.73***	0.12		10.25**	-0.38
	(0.00)	(0.95)		(0.04)	(0.60)
Q5 (High *R2_COARSE*)	9.15**	-1.41*		7.19*	-1.99***
	(0.01)	(0.07)		(0.06)	(0.00)
Q1-Q5	3.55**	1.91**		7.09***	4.01***
(*p*-value)	(0.02)	(0.05)		(0.00)	(0.00)

Table XII
Categorization Coarseness and Holdings' Returns Adjusted for Omitted Risk Factors

Each quarter, the stocks held by at least one fund are sorted into quintiles Q1 (Low) through Q5 (High) based on the stock's categorization coarseness as of the last month of the prior quarter t-1. Categorization coarseness is measured using stock categories based on the stock's DGTW style quintiles, DGTW size quintiles, and DGTW B/M quintiles as described in Section G.2 in column (2), column (3), and column (4), respectively. In column (5), categorization coarseness is measured using the deviation of fundamentals (based on Hoberg and Phillips' (2010) peers) from industry returns as described in Section G.3. Each fund's portfolio is split based on the categorization coarseness quintile of the stocks. The alpha is obtained as the intercept from the regression of the excess Gross Holdings Return for each quintile portfolio on the four common risk factors augmented with a fifth *MMW* factor. *MMW* is the Miscategorized-Minus-Well-categorized factor that is long on stocks with high categorization coarseness and short on stocks with low categorization coarseness. The alphas are reported as percentage per year. The p-values in parentheses are based on panel-corrected standard errors that adjust for heteroskedasticity and autocorrelation. ***, **, * represent statistical significance at the 1%, 5%, and 10% level.

Categorization Coarseness Quintile	Gross Holdings Return Adjusted Alpha for Categorization Coarseness Quintiles based on				Deviation of Fundamentals
	Industry	Style	Size Style	B/M Style	
	(1)	(2)	(3)	(4)	(5)
Q1 (Low *COARSE*)	11.88***	13.82**	12.91***	10.43**	3.34*
	(0.00)	(0.02)	(0.00)	(0.02)	(0.08)
Q2	12.83***	9.64**	10.19**	12.82***	2.74*
	(0.01)	(0.03)	(0.02)	(0.00)	(0.10)
Q3	10.80***	8.60**	8.92**	9.94**	2.05
	(0.01)	(0.03)	(0.03)	(0.02)	(0.26)
Q4	9.25**	7.36**	6.63*	7.88**	0.55
	(0.02)	(0.03)	(0.07)	(0.04)	(0.74)
Q5 (High *COARSE*)	7.10**	6.18**	5.15	5.89	0.39
	(0.01)	(0.04)	(0.15)	(0.15)	(0.96)
Q1-Q5	4.78***	7.64***	7.76***	4.54***	2.95**
(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)