Heads I Win, Tails It’s Chance: Mutual Fund Performance Self-attribution

Meng Wang
Georgia State University

Heads I Win, Tails It’s Chance:
Mutual Fund Performance Self-attribution

BY

Meng Wang

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree

Of

Doctor of Philosophy

In the Robinson College of Business

Of

Georgia State University

GEORGIA STATE UNIVERSITY
ROBINSON COLLEGE OF BUSINESS
2024

ACKNOWLEDGEMENT

I would like to express my gratitude to my dissertation committee—Baozhong Yang, Vikas Agarwal, Zhen Shi, and Sean Cao—for their invaluable feedback on my dissertation. In particular, I extend a special thanks to Baozhong Yang and Vikas Agarwal for their mentorship and friendship, and continuous support and encouragement from my first day in the PhD program. I cannot imagine having completed this journey without their guidance.

I also want to thank my parents and parents-in-law for their unconditional backing. Lastly, but most importantly, I dedicate this dissertation to my wife, Taoran, for her unwavering love and faith in me. Having her by my side is the greatest fortune of my life.

ACCEPTANCE

This dissertation was prepared under the direction of the Meng Wang Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Business Administration in the J. Mack Robinson College of Business of Georgia State University.

Richard Phillips, Dean

DISSERTATION COMMITTEE

Dr. Baozhong Yang (Co-chair)
Dr. Vikas Agarwal (Co-chair)
Dr. Zhen Shi
Dr. Sean Cao (External—University of Maryland)

ABSTRACT

Heads I Win, Tails It’s Chance:
Mutual Fund Performance Self-attribution

BY

Meng Wang

May 2, 2024

Committee Chair: Baozhong Yang & Vikas Agarwal

Major Academic Unit: Department of Finance

This paper investigates the presence of self-attribution bias among mutual fund managers and evaluates its impacts on trading outcomes. I develop a novel GPT-based Natural Language Processing (NLP) architecture designed to extract attribution information from mutual funds' self-assessments of performance in their shareholder reports. On average, mutual fund managers exhibit a significant self-attribution bias—they are 40.6% more likely to attribute performance contributors versus performance detractors to internal factors. Funds displaying stronger self-attribution bias tend to engage in excessive trading and excessive risk-taking in the subsequent reporting period, which negatively impacts their performance. In addition, funds exhibit a higher self-attribution bias following higher performance, despite the fact that biased attribution only influences fund flows when funds perform poorly. Overall, these findings suggest that biased attribution likely stems from cognitive bias rather than strategic choices.

Heads I Win, Tails It's Chance:

Mutual Fund Performance Self-attribution ¶

Meng Wang†

May. 2024

Click here for the most recent version

ABSTRACT

This paper investigates the presence of self-attribution bias among mutual fund managers and evaluates its impacts on trading outcomes. I develop a novel GPT-based Natural Language Processing (NLP) architecture to extract attribution information from mutual funds' self-assessments of performance in their shareholder reports. On average, mutual fund managers exhibit a significant self-attribution bias—they are 40.6% more likely to attribute performance contributors versus performance detractors to internal factors. Funds displaying stronger self-attribution bias tend to engage in excessive trading and excessive risk-taking in the subsequent reporting period, which negatively impacts their performance. In addition, funds exhibit a higher self-attribution bias following good performance, despite the fact that biased attribution only influences fund flows when funds perform poorly. Overall, these findings suggest that biased attribution likely stems from cognitive bias rather than strategic choices.

***JEL* classification**: C45, G11, G23, G41
Keywords: Mutual Funds, Institutional Investors, Behavioral Finance, FinTech, Investments, Deep Learning, Textual Analysis, Artificial Intelligence, Large Language Model, GPT

¶ I am deeply grateful to Vikas Agarwal and Baozhong Yang for their continuous guidance and support, and to my committee members, Sean Cao and Zhen Shi, for their invaluable feedback. I also benefited greatly from discussions with Mark Chen, Gerald Gay, Umit Gurun, Gerard Hoberg, Lixin Huang, Manish Jha, Omesh Kini, Scott Murray, Liyan Yang, and seminar participants at Georgia State University, University of South Carolina, and University of South Florida. All errors are mine alone.
† J. Mack Robinson College of Business, Georgia State University. E-mail: mwang36@gsu.edu

1. Introduction

Self-attribution bias, also known as self-serving attribution bias, refers to a cognitive bias where individuals tend to attribute their successes to internal, personal factors, and their failures to external, situational factors (Langer and Ross, 1975; Miller and Ross, 1975; Taylor and Brown, 1988). Such a bias plays an important role in a trader's trading activities and negatively affects his/her trading outcomes. In the standard learning models of behavioral economics (Daniel, Hirshleifer, and Subrahmanyam, 1998; Gervais and Odean, 2001), self-attribution bias is viewed as a learning bias that hinders a trader from objectively updating beliefs about his own ability based on past trading performance (i.e., he overweighs successes and underweights failure when learning about his own ability). This bias can lead to excessive trading, increased volatility, and lower expected profits.[1]

Given the critical role that mutual funds play in financial markets, it is important that we explore the extent to which mutual fund managers exhibit self-attribution bias and assess its potential impact on their trading outcomes.[2] Understanding whether mutual fund managers exhibit self-attribution bias may have important implications to fund investors[3] and financial markets.[4] However, it remains unclear whether fund managers are likely to exhibit such a bias. On the one

[1] In the framework of Gervais and Odean (2001), the presence of self-attribution bias is believed to result in investor overconfidence and subsequent irrational trading activities. Specifically, the expected insider profit is decreasing in the insider's learning bias parameter γ, while the expected price volatility and volume are increasing in the parameter.
[2] Note that as shown in Gervais and Odean (2001), a trader with greater attribution bias is likely to develop a higher level of overconfidence, which impacts subsequent trading activities. The main objective of this paper is to assess the impact of attribution bias on trading outcomes (through the channel described above), but not to isolate the impact of attribution bias from overconfidence. The measure developed in this paper aligns closely with the definition of attribution bias (i.e., attribute good performance internally and poor performance externally), and is therefore proposed as a measure of attribution bias.
[3] According to the Investment Company Institute (ICI), the total net assets of U.S. mutual funds stood at approximately $27.0 trillion at the end of 2022, with 52.3% of U.S. households owning US-registered fund shares.
[4] Daniel, Hirshleifer, and Subrahmanyam (1998) predict that investor self-attribution bias generates delayed overreaction to information and result in momentum. Given the significant shareholding of mutual funds, self-attribution bias of mutual fund managers may have implications for market anomalies such as momentum.

hand, one might argue that institutional investors, with their sophisticated financial knowledge, should be able to avoid biased attribution and assess performance objectively.[5] On the other hand, evidence from psychological literature (Heath and Tversky, 1991; Griffin and Tversky 1992) suggests that people who consider themselves as knowledgeable often exhibit stronger attribution bias, implying that institutional investors would actually be more likely to have such a bias.

Identifying self-attribution bias among mutual fund managers on a large scale presents a challenge due to its unobservable nature. Prior studies have relied on fund and portfolio characteristics (e.g., manager gender, past performance) as indirect measure of self-attribution bias/overconfidence, which inevitably leads to endogeneity issues.[6] In addition, measures of executive overconfidence commonly used in corporate finance literature may not be applicable in fund setting.[7] As a result, there is no direct measure or large-scale evidence either confirming the presence of self-attribution bias among mutual fund managers or evaluating its impact on their trading outcomes.[8]

This paper fills that gap by examining the narrative attribution of performance by mutual fund managers in their N-CSR filings. Under the Rule 30e-1 of Investment Company Act of 1940, mutual funds that are registered with the SEC are required to disclose performance information,

---

[5] For example, they could use sophisticated attribution methods such as the Brinson-Fachler model (Brinson and Fachler, 1985) or the risk-adjusted method (Bacon, 2008) to decompose and attribute past performance (CFA Institute Review, 2019).

[6] Choi and Lou (2010) use active share as a proxy for confidence and find that it increases after a fund's exhibits strong performance. They interpret this as evidence of the presence of self-attribution bias in mutual funds. Adebambo and Yan (2018) use a composite of manager's gender, manager's tenure, management structure, portfolio turnover, portfolio concentration, prior portfolio performance, and portfolio idiosyncratic risk to proxy for self-attribution bias and overconfidence.

[7] For example, Malmendier and Tate (2005) utilizes CEO option holdings to construct measures of CEO overconfidence.

[8] Glaser, Langer, and Weber (2005) conducted a survey of 123 professionals working at investment banks and found that expert judgement was biased, as professionals incorrectly attributed success and failure. My paper, which is based on a sample from all actively managed domestic U.S. equity funds, presents several key differences when compared to the survey study approach. First, the self-assessment in my study is based on real-world investments, not hypothetical tasks in experiments. Second, the longitudinal data allows me to track and study the time-series and cross-sectional variation in self-attribution bias among mutual fund managers.

including the management's narrative discussion of fund performance in shareholder reports (i.e., N-CSR/N-CSRS Item 1) on an annual basis.[9] In that disclosure, managers typically highlight what significantly contributed to and detracted from fund performance (SEC Investor Bulletin, 2022) as well as their views on the attributing factors behind these contributions and/or detractions. These attributions might include internal factors, such as stock selection, sector weighting, and deviation from the benchmark, as well as external factors like the economic environment, conditions in specific sectors, and common exposure with the benchmark.[10] For instance, the statement "The fund experienced a positive contribution from its overweight exposure in industrials, which we attribute to the effects of individual stock selection," implies an internal factor, as it suggests that the fund's stock selection (a fund-specific factor) was a key contributor to its performance.[11]

To accurately extract attribution information from the textual content of such disclosure, I develop a two-layer GPT-based Natural Language Processing (NLP) model. This model is capable of reading a sentence and 1) identifying performance-attribution information (i.e., perception of causality), and 2) classifying that information as i) a performance contributor vs. performance detractor, and ii) an internal factor vs. external factor.[12] Compared to traditional bag-of-words or rule-based approaches, my deep-learning-based methodology excels at understanding context, capturing semantic meaning, and dealing with unknown words and ambiguity.[13] My model is able

[9] Although performance information is required for the annual report only, most funds include performance information in their semi-annual report as well. For details, see https://www.sec.gov/investor/alerts/ib_readmfreport.

[10] Attribution theory (Fiske and Taylor, 1991) suggests that people categorize their attributions of events and behavior as either internal or external. When making an internal attribution, individuals infer that an event or someone's behavior is a result of personal factors like traits, abilities, or emotions. Conversely, when making an external attribution, individuals infer that someone's behavior is influenced by situational factors.

[11] Please refer to Appendix B for more examples.

[12] I run a horserace of eight model candidates—BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), DistillBERT (Sanh et al., 2019), XLNet (Yang et al., 2019), GPT2 (Radford et al., 2019), GPT3 (Brown et al., 2020), ERNIE (Sun et al., 2019), BLOOM (Scao et al., 2022). Based on the evaluation metrics from five-fold cross-validation results, GPT3 has the best overall performance and thus becomes the primary base model used to construct the NLP architecture. Note that ChatGPT and GPT4 are not allowed to fine-tune yet and might thus not be a proper solver for some specialized tasks in the financial context (Li, Zhu, Ma, Liu, and Shah, 2023).

[13] See Section 2.2. for a detailed illustration.

to achieve an overall out-of-sample *Accuracy* of 89.91% and an *F1 score* of 89.95%.

Using the classifications obtained from the model, I first investigate the presence of biased attribution in the narrative discussions of mutual fund managers. A sentence-level probit model with filing fixed effects reveals that, after controlling for all potential confounding textual factors (e.g., sentiments), mutual fund managers are 40.6% more likely to attribute performance contributors to internal factors than they are to attribute performance detractors to internal factors, suggesting that they indeed exhibit significant self-attribution bias.

To better understand the dynamics of self-attribution bias among mutual fund managers, I construct a filing-level self-attribution bias measure *Self-attribution Score (SAS).* This score is computed by taking the difference between the percentages of internal and external contributors in performance-attribution-related sentences (*IC* - *EC*), and subtracting the difference between the percentages of internal and external detractors (*ID* - *ED*).[14] Intuitively, *the Self-attribution Score (SAS),* which ranges from -1 to 1, is meant to capture the discrepancy in a manager's perception of causality between what contributes to, versus what detracts from, performance over the reporting period.[15]

Employing this measure, I investigate the effects of self-attribution bias on a mutual fund's trading outcomes. In line with the predictions of Gervais and Odean (2001), I find that fund managers exhibiting stronger self-attribution bias are inclined to engage in excessive trading and risk-taking in the subsequent period, which negatively impacts their performance. Specifically, the *Self-attribution Score (SAS)* positively predicts a fund's next-period turnover and idiosyncratic risk and negatively predicts its performance. When the *FFCarhart* model is used as performance

[14] Please refer to Section 2.2 for detailed construction methodology.
[15] A positive *SAS* suggests that the fund tends to attribute performance contributors to internal factors and detractors to external ones.

specification, a one-standard-deviation increase (0.37) in *SAS* results in a 0.8% decrease in cumulative alphas over the subsequent reporting period. The results are held across various factor models and free of confounding factors at both macro-level and fund-level with the inclusion of style-period and fund fixed effects.

At the same time, these findings raise the question of whether the biased attribution in shareholder reports stems from a cognitive bias (i.e., self-attribution bias) or from strategic decisions. Although the results showing a positive relationship between *SAS* and future excessive trading lend credence to the cognitive bias hypothesis, I conduct additional tests to examine the strategic choice hypothesis.

To examine whether biased attributions in shareholder reports reflect strategic signals, one must first ascertain whether and when mutual funds might be incentivized to strategically present biased attribution information to shareholders. To this end, I study the fund's incentives by exploring how mutual fund investors perceive biased attributions.

On the one hand, we might expect investors to respond positively to *SAS*. For example, they might see a fund's self-evaluation as a tool to understand its past performance and disentangle skill from luck.[16] On the other hand, there are reasons to believe that *SAS* negatively affects flows. For instance, psychology literature suggests that the perception of self-attribution bias in others can indeed elicit negative reactions, including feelings of frustration and dissatisfaction (Campbell, Sedikides, Reeder, and Elliot, 2000). I find that *SAS* mitigates outflows at poor past performance but has insignificant impact when funds perform well, suggesting that investors only respond

[16] In Mullainathan, Schwartzstein, and Shleifer's (2008) framework, a mutual fund's past performance can be decomposed into two components: skill and luck. Investors, due to their limited expertise or cognitive resources, might not be able to disentangle fund skill from luck and thus might infer the fund's skill by referring to other signals. In this scenario, a mutual fund's self-evaluation of performance might become an important source to help investors understand the fund's past performance. Consequently, mutual funds might have incentives to strategically attribute performance contributors to internal factors and detractors to external factors.

positively to *SAS* when funds perform poorly.[17]

Based on investor responses, if biased attributions in shareholder reports are the result of strategic choices, then we might expect *SAS* to be negatively correlated to the fund's past performance. Yet both the psychology and behavioral economics literatures suggest that individuals tend to have stronger self-attribution biases when they have recently experienced success (Miller and Ross 1975; Weiner, Russel, and Lerman, 1979; Anderson and Jennings, 1980; Gervais and Odean, 2001; Shepperd, Malone, and Sweeny, 2008). Thus, when biased attribution stems from self-attribution bias, we would expect *SAS* to be positively correlated to a fund's past performance. My results imply that *SAS* is significantly and positively influenced by past performance, suggesting that biased attribution is more likely to arise from cognitive biases rather than strategic decision-making. Moreover, *SAS* is insignificantly correlated to both the level of fund-specific investment during the reporting period and its interaction with past performance, indicating that *SAS* does not reflect the level of true discrepancy between what internally and externally contributes to performance relative to what detracts from it.[18]

In addition, I find that *SAS* correlates with some potential cognitive bias indicators such as manager tenure (Gervais and Odean, 2001) and portfolio concentration (Goetzmann and Kumar, 2008). Taken together, these results indicate that *SAS* indeed reflects the self-attribution biases of mutual fund managers.

---

[17] One possible explanation for investor's asymmetric response is related to the theory of loss aversion (Tversky and Kahneman, 1979). When a fund performs poorly, investors might search for reasons to justify their initial decision to invest. When a mutual fund attributes poor performance to external factors, it provides cognitive relief, supporting the initial investment decision and thereby mitigating outflows.

[18] Intuitively, if the fund claims that the good past performance was driven by internal factors, *SAS* will increase; and if the claim is true, then we should observe increases not only in past performance but also in fund-specific activities. Similarly, when the fund blames external factors for its failures, *SAS* will also increase; and if the assertion is true, then we should observe decreases not only in past performance but also in fund-specific activities. As a result, if *SAS* reflects the level of true discrepancy in what internally and externally contributes to performance relative to what detracts from it, it should positively correlate with the interaction of fund-specific investment during the reporting period and past performance.

My paper contributes to three strands of literature. First, it adds to the body of work dedicated to understanding predictors of mutual fund performance (e.g., Carhart, 1997; Kacperczyk, Sialm, and Zheng, 2008; Cremers and Petajisto, 2009; Barras, Scaillet, and Wermers, 2010; Fama and French, 2010; Amihud and Goyenko, 2013). Specifically, my research is related to studies that assess how behavioral biases of mutual fund managers impact their funds' performance (e.g., Choi and Lou, 2010; Adebambo and Yan, 2016). Unlike prior studies that rely on fund or portfolio characteristics (e.g., active share, manager gender, past performance, etc.) as indirect indicators of self-attribution bias, my methodology, which is grounded in managers' self-assessments, allows me to directly examine whether managers exhibit self-attribution bias and to assess the impact of that bias. To the best of my knowledge, this is the first paper to provide large-scale and direct evidence showing that 1) on average, mutual fund managers tend to attribute successes to internal factors and failures to external ones, and 2) fund managers who exhibit stronger self-attribution bias are more inclined to engage in excessive trading in the subsequent period, which negatively impacts their performance.

Second, my paper contributes to the literature evaluating the effects of behavioral biases on the part of various agents in financial markets (e.g., Daniel, Hirshleifer, and Subrahmanyam, 1998; Gervais and Odean, 2001; Barberis and Thaler, 2003; Chui, Titman, and Wei, 2010). For instance, existing studies have shown the influence of self-attribution bias on analyst forecast accuracy (e.g., Hilary and Menzly, 2006. Hilary and Hsu, 2010), and have demonstrated how executives' self-attribution bias and overconfidence affect corporate performance (e.g., Malmendier and Tate, 2005; Campbell et al., 2011; Galasso and Simcoe, 2011; Ahmed and Duellman, 2012; Hirshleifer, Low, and Teoh, 2012; Schrand and Zechman, 2012). Examining the impact of self-attribution in the mutual fund context offers several advantages over the corporate setting. First, mutual funds are typically managed by a small group who often exert direct control over fund portfolios. Thus, the effects of self-attribution bias on the part of managers should be more directly observable in portfolios. Second, discussions of performance in corporate disclosures (e.g., 10-K's, conference call transcripts) often encompass a variety of topics, making it harder to directly identify internal and external factors driving performance. In the mutual fund context, managerial assessments of performance drivers are typically standardized (e.g., stock selection, market condition), facilitating a cleaner classification and comparison of attribution information.

Lastly, my paper is related to the emerging literature that leverages textual analysis to uncover novel insights from financial disclosures (e.g., Li, 2010; Loughran and McDonald, 2011; Gentzkow, Kelly, and Taddy, 2019; Hassen et al, 2019; Hillert, Niessen-Ruenzi, and Ruenzi, 2021; Li, Mai, Shen, and Yan, 2021; Noh and Zhou, 2022; Cao, Yang, and Zhang, 2023; Chava, Du, Shah, and Zeng, 2022). Specifically, my study introduces a novel deep-learning-based methodology to identify self-attribution bias from textual content, an approach that is far more advanced in terms of understanding context, capturing semantic meanings, and managing unknown words and ambiguity.

The rest of this paper is organized as follows. Section 2 describes the data and measures. Section 3 presents hypotheses and empirical findings. Section 4 concludes.

2. Data and Methodology

2.1 Institutional Background

Under Rule 30e-1 of the Investment Company Act of 1940, mutual funds that are registered with the SEC must send reports to their shareholders on a semiannual basis (i.e., N-CSR/N-CSRS Item

1). In shareholder reports, mutual funds are required to disclose performance information, including a narrative discussion of fund performance produced by the fund's managers.[19] In that discussion, managers usually will point out what significantly contributed to and detracted from the fund's performance (SEC Investor Bulletin, 2022).

In most cases, instead of plainly describing which holdings contributed to and/or detracted from fund performance, managers also narrate their view of the attributing factors of these contributors/detractors. For instance, consider the following sentence: "The fund's outperformance of the benchmark was driven by security selection, with my picks in the information technology, financials and consumer discretionary sectors contributing most." In this statement, the fund manager suggests that holdings in the information technology, financial, and consumer discretionary sectors were major performance contributors and, moreover specifically attributes those contributors to the fund's skill in stock selection.

Attributing factors vary from internal (i.e., fund-specific) factors to external (i.e., non-fund-specific) factors.[20] Typical internal factors include stock selection, sector overweighting/underweighting, benchmark deviations, etc. For example, the statement "The fund experienced a positive contribution from its overweight exposure in industrials, which we attribute to the effects of individual stock selection" implies an internal factor. It suggests that the fund's stock selection (a fund-specific factor) was a key contributor to its performance. Typical external factors include economic environment, sector-wise conditions, common exposure with benchmark, etc. For instance, the statement "During the last six months, this was an impediment to the

[19] Although performance information is required for the annual report only, most funds include performance information in their semiannual report as well. For details, see https://www.sec.gov/investor/alerts/ib_readmfreport

[20] Attribution theory (Fiske and Taylor, 1991) suggests that people categorize their attributions of events and behavior as either internal or external. When making an internal attribution, individuals infer that an event or someone's behavior is a result of personal factors like traits, abilities, or emotions. Conversely, when making an external attribution, individuals infer that someone's behavior is influenced by situational factors.

performance of the funds, as value stock returns have continued to outpace growth returns" suggests an external detractor. It implies that growth funds may have underperformed relative to value funds, which is a factor beyond the management's control.[21]

2.2 Measuring Self-attribution Bias

2.2.1 Limitations of Traditional Textual Measures

Identifying, classifying, and quantifying attribution-related information from performance discussions is an essential part of exploring whether mutual fund managers make biased attributions. Doing so, however, presents a significant challenge when using traditional textual analysis methods such as bag-of-words/rule-based approaches (i.e., calculating the Tf-idf of keywords in textual content[22]), because the semantic meaning of textual information is dependent not only on lexicalized features (i.e., the meaning of each word) but also on context-level features. Consider the following two sentences, both of which contain the keywords "benchmark index," one of the most common bigrams in shareholder reports:

> *Sentence 1: "The fund's holdings in the health care sector, especially biotechnology stocks, held up considerably better than those in the benchmark index."*
>
> *Sentence 2: "The fund benefited significantly from holdings in the health care sector, as it was a strong-performing sector of the benchmark index."*

In the first sentence, the fund attributes its good performance to its benchmark-deviating behavior, which can be inferred as an internal contributor; in the second sentence, the fund suggests that its

[21] Please refer to Appendix B for more examples.

[22] For instance, Li (2010) uses the percentage of first-person pronouns relative to that of second- and third-person pronouns in the 10K MD&A to measure managers' self-serving attribution bias; Noh and Zhou (2022) first construct an "economy" dictionary to identify economy/market-related sentences from earnings call transcripts and then identify a "blame sentence" if an economy/market-related sentence has more negative words than positive ones.

good performance was not fund-specific (i.e., the benchmark index also benefited the fund because it holds similar holdings), which can thus be inferred as an external contributor. In this scenario, one cannot simply use the keywords "benchmark index" to classify textual information as denoting either an internal or external factor. More to the point, it is exceptionally difficult to explicitly define rules that would assist with classification using other textual measures (e.g., sentiment). As a result, a bag-of-words/rule-based approach might lead to non-negligible measurement errors and estimation bias in empirical analyses. To mitigate these issues, I develop a novel deep-learning-based NLP method to extract and classify target textual information. This methodology is described in the following sections.

2.2.2 Data Sources

I obtain mutual fund shareholder reports (i.e., N-CSR/N-CSRS Item 1) from 2006 to 2018 from SEC EDGAR.[23] In these reports, fund managers usually discuss performance either under section MD&A (Management's Discussion and Analysis of Fund Performance) or in shareholder letters. To extract relevant narrative contents from N-CSR/N-CSRS Item 1, I follow Hillert, Niessen-Ruenzi, and Ruenzi (2021) and Cao, Yang, and Zhang (2023) and use common phrases to locate the target textual contents, which I complement with manual work. Specifically, I use phrases such as "Managers Discussion" and "Manager's Discussion" to identify MD&A sections, and I use phrases such as "Dear shareholders," "To fellow shareholders," and "To our shareholders" to identify shareholder letters.

I retrieve fund characteristic data (e.g., *TNA, Age, Expense Ratio, Turnover*) from the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund database and fund portfolio holdings data from the Thomson Reuters Mutual Fund Holdings (formerly

[23] N-CSR fillings are available from 2003, but series and class identification information is only available after 2006. See Hillert et al. (2021) and Cao, Yang, and Zhang (2023) for more information.

CDA/Spectrum S12) database.[24] I follow Kacperczyk, Sialm and Zheng (2008) and retain actively managed domestic equity funds only.[25] To merge that database with the data from shareholder reports, I follow Cao, Yang, and Zhang (2021) and construct a link between Series ID (fund identifier in N-CSR) and the WFICN (Wharton Financial Institution Code Number, i.e., the identifier for fund portfolios in MFLINKS).[26] Over the period from 2006 to 2018, my initial sample consists of MD&A's and shareholder letters extracted from 16,270 shareholder reports associated with 1,969 unique funds.

In general, MD&A's or shareholder letters include three broad topics: 1) market recaps, in which managers discuss the economic environment over the past reporting period; 2) performance reviews, in which managers discuss what contributed to or detracted from fund performance over the past reporting period; and 3) future perspectives, in which managers give their opinions on the prospects of the fund/economy. Since the focus of this paper is managerial discussions of performance (i.e., performance reviews), it is important to ensure that a textual source is not biased by other textual information (i.e., market recaps or future perspectives). To isolate performance-related sentences, I only keep sentences containing key phrases with the closest embeddings to performance-related root words.[27] Specifically, I first train a Word2Vec model using all MD&A's and shareholder letters, constructing a dictionary of 200 performance-related key phrases (i.e., unigrams and bigrams) with the closest embeddings to root words. I then extract all sentences containing at least one performance-related key phrase from each MD&A or shareholder letter in a shareholder report. This process yields 129,185 performance-related sentences.

---

[24] I use MFlinks as provided by Wharton Research Data Services (WRDS) to merge two databases.
[25] The methodology can be found at https://wrds-www.wharton.upenn.edu/pages/wrds-research/macros/wrds-macros-return-gap/
[26] I use the Class Ticker (under each Series ID) to match with the ticker symbol in CRSP Mutual Fund database, and then drop the cases in which one Series ID is matched to multiple WFICNs.
[27] Performance-related root words include, for example, "gain," "outperform," "drive," "remain," "overhaul," "lower," "boost," "result," "upbeat," "detract," and "contribute."

2.2.3 NLP Architecture

After collecting all performance-related sentences from shareholder reports, the next step is to build a model architecture that can help identify relevant information and classify sentences into target categories. The model architecture should be able to 1) identify whether a performance-related sentence contains attribution information, and 2) classify a performance-attribution sentence (i.e., if a performance-related sentence contains attribution information) along two dimensions: contributor vs. detractor and internal vs. external.

To achieve these objectives, I develop a two-layer model architecture as illustrated in Figure 1.

[Insert Figure 1 Here]

The input to the model is a performance-related sentence extracted using the process described in the previous section. The first layer of the model has one classifier, classifier 0, which is used to identify and pass performance-attribution information (i.e., perception of causality) to the second layer. Note that not every performance-related sentence contains attribution information.[28] The second layer of model has two classifiers—classifier 1 and classifier 2, which independently classify the performance attribution sentence into a) performance contributors vs. performance detractors and b) internally attributed factors vs. externally attributed factors.

I consider a pool of eight transformer-based deep learning candidate models that are most widely used in fine-tuned text classification tasks: BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), DistillBERT (Sanh et al., 2019), XLNet (Yang et al., 2019), GPT2 (Radford et al.,

[28] For instance, the sentence could be a plain performance overview such as "XXX fund investor class has outperformed the MSCI world index by 1.20% per annum on average since its 2006 inception"; in contrast, a performance attribution sentence must be specifically related to attribution, for example, "the XXX fund experienced a positive contribution from its overweighed exposure in industrials, which we attribute to the effects of individual stock selection."

2019), GPT3 (Brown et al., 2020)[29], ERNIE (Sun et al., 2019), BLOOM (Scao et al., 2022).[30] I construct a high-quality training sample and fine-tune NLP models to achieve specialized text classification goals; I then apply a stratified 5-fold cross-validation method and use four measures to evaluate models' out-of-sample performance.[31]

[Insert Table 1 Here]

Accuracy is the ratio of correct category predictions to total number of observations (i.e., the sum of true positives and true negatives divided by the total number of observations). *Precision* is the ratio of true positives to the sum of true positives and false positives. *Recall* is the ratio of true positives to the sum of true positives and false negatives. *F1 score* = 2* *Precision***Recall* / (*Precision* + *Recall*). Among the eight candidate models, GPT3 has the best overall performance, with an average *F1 score* of 89.95% and an *Accuracy* of 89.91% and was thus chosen as the primary base model used to construct the NLP architecture.

2.2.4 Self-attribution Score (SAS)

I use the trained GPT3-based NLP architecture to obtain the final classifications of 129,185 performance-related sentences extracted from 16,270 shareholder reports. For each shareholder report, I calculate the proportion of attribution information in each category. Specifically, for each shareholder report *i*, I define:

$$IC = \frac{len(Internal\ Contributor)}{len(Performance\ Attribution)}$$

$$ID = \frac{len(Internal\ Detractor)}{len(Performance\ Attribution)}$$

[29] ChatGPT and GPT4 are not allowed to fine-tune yet and might not be a proper solver for financial text analytics (Li, Zhu, Ma, Liu, and Shah, 2023).
[30] GPT3 (Ada) is accessible using OpenAI API. Other models are open sourced and available on the Hugging Face library.
[31] See Appendix C for detailed descriptions of training and evaluation processes.

$$EC = \frac{len(External\ Contributor)}{len(Performance\ Attribution)}$$

$$ED = \frac{len(External\ Detractor)}{len(Performance\ Attribution)}$$

where *IC (Internal Contribution)* is the length (i.e., number of words) of sentences classified as internal contributors scaled by the length of sentences identified as containing attribution information; *ID (Internal Detraction), EC (External Contribution),* and *ED (External Detraction)* are similar measures capturing information on internal detractors, external contributors, and external detractors, respectively. Note that $IC + ID + EC + ED = 1$. To proxy for self-attribution bias, for each shareholder report *i,* I combine four measures and define *Self-attribution Score (SAS)* as:

$$SAS = (IC - EC) - (ID - ED)$$

Intuitively, *SAS* captures the discrepancy in a manager's perception of causality between what contributes to and detracts from fund performance. (*IC* – *EC*) amounts to the difference between the proportion of internally attributed factors versus externally attributed factors in terms of what management describes as having contributed to the fund's performance. (*ID* – *ED*) is the difference between the proportion of internally attributed factors versus externally attributed factors in terms of what management describes as having detracted from the fund's performance. *SAS* ranges from -1 to 1. A positive *SAS* suggests that a fund's management tends to attribute success to internal factors and failure to external factors.

2.3 Construction of Other Variables

2.3.1 Fund Performance and Idiosyncratic Risk

To measure fund performance, I calculate performance alphas based on various factor models using beta coefficients obtained from a rolling regression over the prior twenty-four-month period. Taking *FFCarhart* factor specification (Carhart, 1997) as an example, for fund *i* at month *t*, I first obtain factor loadings using the following regression from month *t-24* to *t-1*,

$$R_{i,t} - R_{rf,t} = \alpha + \beta_{i,mkt}MKT_t + \beta_{i,SMB}SMB_t + \beta_{i,HML}HML_t + \beta_{i,UMD}UMD_t + \varepsilon_{i,t}$$

where $R_{i,t}$ is fund *i*'s return in month *t*; $R_{rf,t}$ is the risk-free rate in month *t*; MKT_t, SMB_t, HML_t, and UMD_t are the realized excess returns on the *FFCarhart* four-factor portfolios.[32] I then calculate the monthly alpha $\alpha_{i,t}$ for fund *i* in month *t* as the difference between excess returns (i.e., $R_{i,t} - R_{rf,t}$) and risk adjustment (i.e., $\hat{\beta}_{i,mkt}MKT_t + \hat{\beta}_{i,SMB}SMB_t + \hat{\beta}_{i,HML}HML_t + \hat{\beta}_{i,UMD}UMD_t$). Finally, I calculate the cumulative alphas for fund *i* over a T-period as follows,

$$Perf_{i,[t,t+T]} = \prod_{t}^{t+T}\left(1 + \alpha_{i,t}\right) - 1$$

The idiosyncratic risk of fund *I* over a T-period is calculated as the standard deviation of residuals $\varepsilon_{i,t}$ during that period. Specifically,

$$IdioRisk_{i,[t,t+T]} = \sigma\left(\hat{\varepsilon}_{i,[t,t+T]}\right) = \sqrt{\frac{1}{T}\sum_{t}^{t+T}\left(\hat{\varepsilon}_{i,t} - \bar{\varepsilon}_{i,[t,t+T]}\right)^2}$$

2.3.2 Fund Characteristics

In addition to the above, I consider a set of variables representing fund characteristics that are commonly used in scholarship on mutual funds. *TNA* is the natural logarithm of a fund's total net assets (TNA). *Expense Ratio* is the ratio of total investment that shareholders pay for the fund's operating expenses, which include 12b-1 fees. *Age* is the logarithm of a fund's age computed from the date when a fund was first offered. *Turnover* is the fund's yearly turnover ratio as reported in

[32] Factor portfolios return data is available on Kenneth French's website, https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

CRSP: minimum (of aggregated sales or aggregated purchases of securities), divided by the average 12-month TNA of the fund. *Turnover1, Turnover2,* and *Turnover 3* are a fund's quarterly turnover ratio calculated using TR-13F data as defined in Ben-David, Franzoni, and Moussawi (2010), where *Turnover1* is the minimum of total buys and sales divided by portfolio size, *Turnover2* is the minimum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size, and *Turnover3* is the sum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size. *Active Share* is the share of portfolio holdings that differ from the benchmark index holdings as defined in Cremers and Petajisto (2009).[33] *Flow* is calculated as $(TNA_{i,t} - TNA_{i,t-1}) / TNA_{i,t-1} - r_{i,t}$ where $TNA_{i,t}$ denotes fund *i*'s total net assets in month *t* and $r_{i,t}$ denotes fund *i*'s returns in month *t*.

2.3.3 Textual Controls

FinBERT_positive, *FinBERT_negative*, and *FinBERT_neutral* are the weighted-average tone scores of all performance-related sentences in a shareholder report determined using FinBERT (Araci, 2019). *LM_Positive*, *LM_Negative*, *LM_Uncertainty*, *LM_Litigious*, *LM_Strong*, *LM_Weak*, *and LM_Constraining* measure the tf_idf (Term Frequency-Inverse Document Frequency) of relevant LM sentiment keywords (Loughran and McDonald, 2011) in all performance-related sentences in a given shareholder report.

2.4 Summary Statistics

Table 2 presents summary statistics for key variables used in the analysis. Variable constructions are described in Section 2.3 and Appendix A.

[Insert Table 2 Here]

[33] The data are available at https://www.petajisto.net/data.html.

The final sample used for regression analysis consists of 15,434 shareholder reports associated with 1,400 unique actively managed domestic equity funds over a 13-year span from 2006 to 2018. The *SAS* (Self-attribution score) measured from 15,434 shareholder reports has a mean of 0.23 and a standard deviation of 0.37.

[Insert Figure 2 Here]

The estimated distribution of *SAS* is left-skewed (as shown in Figure 2). The median SAS is 0.27, and approximately 74% of observations have a *SAS* value greater than 0.

Notably, Panel C of Table 2 indicates that variations in *SAS* arises not only from cross-section (i.e., different funds) but also time-series (i.e., within the same fund), with a cross-section standard deviation of 0.36 and a time-series standard deviation of 0.31.[34] The time-series unit-root tests further suggest that *SAS* is not a stationary time series, with an Im-Pesaran-Shin (IPS) t-bar of -24.76 and a Fisher-type Inverse Chi-squared of 1402.24.

3. Hypotheses and Empirical Findings

3.1 Do Fund Managers Exhibit Biased Performance Attribution?

I begin my empirical analysis by examining whether, on average, mutual fund managers exhibit biased performance attribution (i.e., whether they tend to attribute success to internal factors and failures to external factors). There are reasons to believe that institutional investors, given their sophisticated financial knowledge, would avoid biased attribution and evaluate performance objectively. Given their level of expertise, for example, they may decide to use sophisticated attribution methods such as the Brinson-Fachler model (Brinson and Fachler, 1985) or the risk-

[34] To calculate cross-section standard deviation of *SAS*, I first compute the standard deviation of all funds in each reporting period, and then take the average all reporting periods; similarly, to calculate time series standard deviation of *SAS*, I first compute the standard deviation of all reporting periods within each fund, and then take the average all funds.

adjusted method (Bacon, 2008) to decompose and attribute past performance (CFA Institute Review, 2019). At the same time, there are reasons to believe that institutional investors actually have a greater propensity for attribution bias than retail investors. Scholarship in the field of psychology (Heath and Tversky, 1991; Griffin and Tversky 1992) suggests that people are likely to demonstrate stronger attribution bias when they consider themselves knowledgeable. Based on these studies, I develop the following hypotheses,

H1*: Mutual fund managers tend to internalize successes and externalize failures.*

H1$_a$*: Mutual fund managers tend not to internalize successes and externalize failures.*

To test the hypotheses, I first compute the average percentage of internally and externally attributed factors to performance contributors versus detractors in shareholder reports. Specifically, for each shareholder report, I calculate the length (i.e., number of words) of performance attribution sentences classified as internal contributors (*IC*) divided by the length of internal and external contributors (*IC* + *EC*); I do the same calculation for external contributors (*EC*), internal detractors (*ID*), and external detractors (*ED*). Next, I calculate the average ratios across all shareholder reports in the sample. Figure 3 reports the results.

[Insert Figure 3 Here]

On average, 41% of the factors attributed to performance contributors are external, while 59% are internal. Conversely, 83% of the factors attributed to performance detractors are external, with 17% being internal factors. This result supports the hypothesis that mutual fund managers tend to internalize successes (i.e., performance contributors) and externalize failures (i.e., performance detractors).

To partial out potential textual confounding factors (e.g., tone), I run a sentence-level probit model that examines whether funds managers are more likely to attribute performance contributors (rather than detractors) to internal factors. Specifically, I estimate the following regression:

$$Internal_{s,i} = \beta Contributor_{s,i} + \gamma FinBERT\_positive_{s,i} + \delta LM\_sent_{s,i} + \mu_i + \varepsilon_{s,i}$$

where the dependent variable *Internal* is a dummy variable that equals one if a performance-related sentence s in filing i is internally attributed and equals zero if externally attributed. *Contributor* equals one if sentence s concerns something that contributes to the fund's performance and equals zero if it concerns something that detracts from the fund's performance. $FinBERT\_positive$ and LM_sent are textual controls as defined in Section 2.3. I include filing fixed effects μ_i to rule out the effects of potential unobserved filing-level factors such as writing style. Standard errors are clustered at the fund level. Table 3 presents the results.

[Insert Table 3 Here]

The results imply that mutual fund managers tend to internalize successes and externalize failures. In column (1), when not including textual control variables, one can observe a significant positive relationship between *Internal* and *External*, suggesting that performance contributors are more likely to be attributed to internal factors. Specifically, the likelihood of being attributed to internal factors is 48.1% higher for performance contributors than for performance detractors when not including other textual control variables (e.g., tone and other sentiments). In column (2), when including textual controls, the magnitude drops to 40.6%, suggesting that 7.5% of the explanatory power associated with the main independent variable is absorbed by other textual variables. In any case, the results imply that corpora associated with internal factors are more likely to be written in positive tones and less likely to be written with weak modality.

3.2. *Self-attribution Score* (*SAS*) and Trading Outcomes

In this section, I examine the impact of self-attribution bias on mutual funds' future trading outcomes. Gervais and Odean (2001) describe the dynamics by which self-attribution bias engenders overconfidence in traders. In this framework, self-attribution bias is defined as a learning bias, and a biased trader (with a learning bias parameter γ) who successfully forecasts a dividend weights this success too heavily when applying Baye's rule to assess their own ability and thinks their signal in the next period is more informative than it really is. Thus, a biased trader will use their information more aggressively than they should, which results in higher expected trading volume, lower expected profit, and higher expected volatility. Specifically, the expected volume and volatility in next period is increasing in the insider's learning bias parameter γ, and the expected profit is decreasing in γ.[35] Based on the theory, I have developed the following hypotheses:

H2*: Self-attribution Score* (*SAS*) *negatively predicts a fund's future performance.*

H3*: Self-attribution Score* (*SAS*) *positively predicts a fund's future turnover.*

H4*: Self-attribution Score* (*SAS*) *positively predicts a fund's future idiosyncratic risk.*

To test these hypotheses, I estimate the following specification,

$$Y_{i,\,[t+1,\,t+6]} = \beta SAS_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variables correspond to the outcome variables identified in the hypotheses. *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. *FilingControls* is a set of filing-level control variables including *FinBERT_positive, FinBERT_negative*, and *FinBERT_neutral*, which are the weighted average tone scores of performance-related sentences in the shareholder report determined by FinBERT (Araci, 2019),

[35] See Propositions 5, 6, and 7 in Gervais and Odean (2001).

and *LM_Positive*, *LM_Negative*, *LM_Uncertainty, LM_Litigious, LM_Strong, LM_Weak,* and *LM_Constraining,* which measure the *tf_idf* (Term Frequency-Inverse Document Frequency) of relevant words in an LM sentiment dictionary (Loughran and McDonald, 2011). *FundControls* is a set of fund characteristics including *Age*, *Size*, *Expense Ratio*, *Past Flow*, *Past Performance*, and *Turnover*. I include fund style * reporting period fixed effects $\mu_{style,[t-6,t-1]}$ to partial out potential style-period-level macro confounding factors and only compare funds within the same style and reporting period. I also include fund fixed effects π_i to exclude any impact of potential fund-level confounding characteristics (e.g., writing style). Standard errors are adjusted for heteroskedasticity and clustered at the fund level.

3.2.1 Future performance

[Insert Table 4 Here]

Table 4 displays results of regressing fund future performance on *Self-attribution Score* (*SAS*). Specifically, the dependent variable *Perf* is the cumulative alphas of fund *i* from one month after a shareholder report was filed, month (*t*+1), to six months after the filing, month (*t*+6), where alphas are computed from various factor models using beta coefficients obtained from a rolling regression over the prior twenty-four-month period. In columns (1) through (6), I select six candidates for fund performance proxies—*Raw Return*, *Excess Market*, *CAPM alpha*, *CAPMSP alpha*, *FF3 alpha*, and *FFCarhart alpha*—each closely aligning with the models that investors use to make their capital allocation decisions (Berk and van Binsbergen, 2016). The estimated coefficients of *SAS* are significantly negative through all columns, supporting *H2* and suggesting that a fund's performance in the subsequent reporting period decreases in the level of self-attribution bias. In particular, as shown in column (6), when using the *FFCarhart* model as a performance specification, a one-standard-deviation increase (0.37) in *SAS* results in a 0.8%

decrease in *FFCarhart* alpha over the next reporting period.

3.2.2 Future Turnover

[Insert Table 5 Here]

Table 5 reports the results on the relationship between a fund's future portfolio turnover and its *Self-attribution Score (SAS)*. I follow Ben-David, Franzoni, and Moussawi (2010) to calculate three types of turnover measures using 13F holding data. In column (1), the dependent variable *Turnover1* is calculated as the minimum of total buys and sales divided by portfolio size; in column (2), *Turnover2* is calculated as the minimum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size; in column (3), *Turnover3* is calculated as the sum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size. The estimated coefficient of *SAS* is 0.0078 (at the 1% level) in column (1) and 0.0083 (at the 5% level) in column (3), indicating that *SAS* positively predicts the fund's turnover in the next period. These results support *H3* and imply that a fund's turnover in the subsequent reporting period increases with self-attribution bias.

3.2.3 Future Idiosyncratic Risk

[Insert Table 6 Here]

Table 6 investigates the relationship between a fund's future portfolio idiosyncratic risk and its *Self-attribution Score (SAS)*, where the dependent variable *IdioRisk* is the portfolio idiosyncratic risk on fund *i*'s returns from one month after the shareholder report's filing month (t+1) to six months after the filing month (t+6). To be specific, in columns (1) and (2), *IdioRisk* is calculated as the standard deviation of a fund's raw return and excess return to market; in columns (3) to (6), *IdioRisk* is calculated as the standard deviation of model residuals using beta coefficients estimated using *CAPM*, *CAPMSP*, *FF3*, and *FFCarhart* models from a rolling regression over the prior twenty-four-month period. As shown in columns (3) through (6), when using factor model residuals as proxies for idiosyncratic risk, the estimated coefficients of *SAS* are positive and significant (at 1% level in columns (1) and (2); at 5% level in columns (3) and (4)). These results support *H4* and suggest that *Self-attribution Score (SAS)* positively predicts a fund's future idiosyncratic risk.

In summary, the results in this section are consistent with predictions made based on Gervais and Odean (2001). The extent of a mutual fund's self-attribution bias negatively predicts next-period performance through the excessive trading channel. This is evidenced by the positive relation between the *Self-attribution Score (SAS)* and both future turnover and idiosyncratic risk.

3.3. Alternative Explanations for Biased Performance Attribution

Results in Section 3.1 suggest that mutual fund managers exhibit biased performance attribution in shareholder reports, that is, they tend to attribute performance contributors to internal factors and performance detractors to external factors. A natural question arising from these results is whether this biased attribution in shareholder reports stems from a cognitive bias (i.e., a self-attribution bias) or strategic choices. These two alternatives are represented by the following hypotheses:

H5*: Self-attribution Score (SAS) reflects a fund manager's self-attribution bias.*

H5$_a$*: Self-attribution Score (SAS) reflects a fund manager's strategic choices.*

Although the results on the positive relationship between *SAS* and future excessive trading lend credence to the cognitive bias hypothesis, I conduct additional tests to examine the strategic choice hypothesis.

3.3.1. Investor Flows and Fund Incentives

To examine whether biased attributions in shareholder reports reflect strategic signals, it's important to first ascertain whether and when mutual funds might be incentivized to strategically present biased attribution information to shareholders. I study a fund's incentives by exploring how mutual fund investors perceive biased attributions.

On the one hand, there are reasons to believe that investors respond positively to *SAS*. In the framework of Mullainathan, Schwartzstein, and Shleifer (2008), a mutual fund's past performance can be decomposed into two components: skill and luck. Investors, due to their limited expertise or cognitive resources, might not be able to disentangle fund skill from luck and thus might infer the fund's skill by referring to other signals. In this scenario, a mutual fund's self-evaluation of performance might become an important source to help investors understand the fund's past performance. Consequently, mutual funds might have incentives to strategically attribute performance contributors to internal factors and detractors to external factors. On the other hand, it could likewise be reasonably supposed that *SAS* negatively affects flows. For instance, the psychology literature suggests that the perception of self-attribution bias in others can indeed elicit negative reactions, including feelings of frustration and dissatisfaction (Campbell, Sedikides, Reeder, and Elliot, 2000).

To test flow sensitivity, I employ a piecewise regression as in Sirri and Tufano (1998). This approach considers the potential impact of a convex flow-performance relationship and allows me to test the relation between flow and *SAS* across different performance intervals. Specifically, I estimate the following specification,

$$\begin{aligned} Flow_{i,\,t+1} = {} & \beta_1 SAS_{i,t} + \beta_2 LowPerf_{i,[t-6,t-1]} + \beta_3 HighPerf_{i,[t-6,t-1]} + +\beta_4 SAS_{i,m} \\ & * LowPerf_{i,[t-6,t-1]} + \beta_5 SAS_{i,m} * HighPerf_{i,[t-6,t-1]} + + \gamma FilingControls_{i,t} \\ & + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,m} \end{aligned}$$

where *Flow* is the monthly flow of fund *i* in month (*t*+1), month *t* being the month in which the shareholder report was filed. *LowPerf* and *HighPerf* are dummies that take a value of one if fund *i*'s performance over the reporting period (*t*-6 to *t*-1) is in the bottom or top quantile (within the same fund style) as defined in Sirri and Tufano (1998). I include the same set of control variables and fixed effects as specified in Section 3.2.

[Insert Table 7 Here]

As shown in Table 7, the estimated coefficients for *LowPerf* and *HighPerf* are significantly negative and positive, respectively. This coefficient captures a general impact of performance on flows, suggesting that funds generally experience outflows with poor performance and inflows with good performance. The estimated coefficients of *SAS * LowPerf* are significantly positive, indicating that *SAS* mitigates outflows at poor past performance. In particular, as shown in column (6), when using the *FFCarhart* model as a performance specification, a one-standard-deviation increase (0.37) in *SAS* results in a 0.0042 increase in flows. Compared to the impact of poor performance on flows (-0.0158), a one-standard-deviation increase in *SAS* mitigates 27% of outflows. Interestingly, flows are not sensitive to *SAS* at good performance, as the estimated coefficients of *SAS * HighPerf* are insignificantly positive. One possible explanation for this empirical finding is related to the theory of loss aversion (Tversky and Kahneman, 1979). When a fund performs poorly, investors might search for reasons to justify their initial decision to invest. When a mutual fund attributes poor performance to external factors, it provides cognitive relief, supporting the initial investment decision and thereby mitigating outflows.

3.3.2 SAS and Past Performance

The results above suggest that *SAS* mitigates outflows when a fund performs poorly but has an insignificant impact when performance is good. Thus, if the biased attribution in shareholder reports results from strategic choices, then we should expect *SAS* to be negatively correlated to the fund's past performance.

In contrast, psychological studies suggest that individuals tend to exhibit stronger self-attribution bias when they have recently experienced success (Miller and Ross 1975; Weiner, Russel, and Lerman, 1979; Anderson and Jennings, 1980; Shepperd, Malone, and Sweeny, 2008).[36] The behavioral economics literature also provides similar predictions and evidence.[37] Thus, if biased attribution stems from self-attribution bias, we should expect *SAS* to be positively correlated to a fund's past performance.

To distinguish between these two possibilities, I employ the following regression,

$$SAS_{i,\,t} = \beta_1 Perf_{i,[t-6,t-1]} + \beta_2 AS_{i,[t-6,t-1]} + \beta_3 Perf_{i,[t-6,t-1]} \times AS_{i,[t-6,t-1]} \\ + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t}$$

where the *Perf* is the cumulative alphas of fund *i* from six months before the shareholder report filing month (*t*-6) to one month before the filing month (*t*-1), and *AS* is the active share for fund *i* over past six months (*t*-6 to *t*-1), which is used to proxy for fund-specific deviation and calculated as in Cremers and Petajisto (2009). If *SAS* reflects cognitive bias, then we would expect β_1 to be negative; if *SAS* reflects strategic choices, then we would expect β_1 to be positive.

The interaction term in the specification allows me to test another alternative explanation for biased attribution in shareholder reports, namely, that the tendency of funds to attribute success to internal factors and failure to external ones might simply stem from the reality that funds indeed have more internal contributors and external detractors. Intuitively, if the fund claims that the good

[36] To be specific, recent success could bolster a positive self-perception or increase the expectation of future success, leading to a stronger self-serving bias.
[37] Gervais and Odean (2001) predict that investors, after a period of successful investing (such as one quarter or one year), are more likely to believe that their success is due to their acumen as investors rather than to factors out of their control. Li (2010) finds that managers tend to use more first-person pronouns (relative to second- and third-person pronouns) in the Management Discussions and Analysis Section of the 10-K filings when firm performance is better.

past performance was driven by internal factors, *SAS* will increase; and if the claim is true, then we should observe increases not only in past performance but also in fund-specific activities (i.e., positive β_3). Similarly, when the fund blames external factors for its failures, *SAS* will also increase; if the assertion is true, then we should observe decreases not only in past performance but also in fund-specific activities.

[Insert Table 8 Here]

Table 8 reports the results. One can observe that *SAS* is significantly and positively explained by the past performance. When using the *CAPM* model as the performance specification as in column (1), the estimated coefficient of *Perf* is 0.0068 (at the 5% level). This finding is robust to changing performance proxies, except for *Raw Return*. The positive relationship between *SAS* and past performance indicates that biased attribution likely stems from cognitive bias rather than strategic decision-making. Moreover, *SAS* is insignificantly correlated to either active shares or the interaction of two variables. The insignificant relationship suggests that *SAS* does not reflect the level of true discrepancy in what internally and externally contributes to performance relative to what detracts from performance.

3.3.3 Other tests

I next examine the relationship between *Self-attribution Score (SAS)* and a set of potential cognitive bias indicators. I include four variables that have been demonstrated to be correlated to self-attribution bias in prior literature: i) *Male*, a dummy that equals one if the fund is managed by male managers only;[38] ii) *Manager Tenure*, the average tenure of managers who manage the

[38] Previous studies have shown that males are more likely to exhibit overconfidence (Lundeberg, Fox, and Puncochar, 1994; Barber and Odean, 2001), a trait often found to be highly correlated to self-attribution bias (Daniel, Hirshleifer, and Subrahmanyam, 1998; Gervais and Odean, 2001; Glaser, Langer, and Weber, 2013). I determine the gender of the manager by applying the Python package “gender-detector” to the first names of the managers.

fund,[39] iii) *Portfolio Concentration*, the maximum weight of holdings in the portfolio,[40] and iv) *SMF*, a dummy that equals one if the fund is single-managed fund.[41]

[Insert Table 9 Here]

As shown in Table 9, when including all variables (column 5), the estimated coefficient of *Manager Tenure* is -0.0035 (at the 5% level), suggesting that inexperienced managers tend to exhibit a high *Self-attribution Score (SAS).* In addition, *SAS* is also positively correlated with *Portfolio Concentration* (at 10% level), indicating that funds with high *SAS* tend to have concentrated portfolios. These results reveal that *SAS* is correlated with some cognitive bias indicators, providing further evidence that *SAS* indeed reflects the level of self-attribution bias.

4. Conclusions

Due to its largely unobservable nature, investigating the presence of self-attribution bias among mutual fund managers and evaluating its potential implications for trading outcomes has long proved challenging. In this study, I explore the self-attribution bias of mutual fund managers by analyzing their self-assessment of performance in N-CSR filings. To accurately extract attribution information, I develop a two-layer transformer-based Natural Language Processing (NLP) architecture. This architecture is capable of reading a sentence and 1) identifying performance-attribution information (i.e., perception of causality), and 2) classifying the information as i) contributor vs. detractor and ii) internal vs. external. Using the classifications obtained from the NLP model, I discover that mutual fund managers exhibit significant self-attribution bias—they

[39] Gervais and Odean (2001) have shown that self-attribution bias is more pronounced among inexperienced managers.
[40] Goetzmann and Kumar (2008) conclude that overconfidence is related to under-diversification.
[41] Psychological studies suggest that group decision-making may counteract self-serving attribution bias, because individual group members may challenge each other's views, leading to a more balanced perspective on success and failure (Kugler, Kausel, and Kocher, 2012).

are 40.6% more likely to attribute performance contributors to internal factors than they are to attribute performance detractors to internal factors. Consistent with the predictions in Gervais and Odean (2001), funds displaying stronger self-attribution bias tend to engage in excessive trading and excessive risk-taking in the subsequent reporting period, which negatively impacts their performance. Specifically, a one-standard-deviation increase in *Self-attribution Score (SAS)* results in a 0.8% decrease in cumulative *FFCarhart* alphas over the subsequent reporting period. I further found that biased attribution information only helps mitigate outflows when funds perform poorly, whereas funds exhibit a higher self-attribution bias following successful investing outcomes. Moreover, *SAS* correlates with potential cognitive bias indicators such as manager tenure (Gervais and Odean, 2001) and portfolio concentration (Goetzmann and Kumar, 2008). Collectively, these findings suggest that *SAS* indeed reflects the self-attribution bias of mutual fund managers rather than strategic choices.

Appendix A: Variable Definitions

Panel A: filing-level variables

Variable	Definition
SAS	Self-attribution score. Measured as (the length (i.e., number of words) of sentences classified as internal contributors – the length of external contributors) – (the length of internal detractors – the length of external detractors), scaled by the length of sentences identified as containing attribution information. See Section 2.2 for details.
Contributor	The length (i.e., number of words) of sentences classified as contributors scaled by the length of sentences identified as containing attribution information.
Detractor	The length (i.e., number of words) of sentences classified as detractors scaled by the length of sentences identified as containing attribution information.
Internal	The length (i.e., number of words) of sentences classified as internal factors scaled by the length of sentences identified as containing attribution information.
External	The length (i.e., number of words) of sentences classified as external factors scaled by the length of sentences identified as containing attribution information.
FinBERT_positive	Weighted average positive tone scores of all performance-related sentences in shareholder report determined using FinBERT (Araci, 2019).
FinBERT_negative	Weighted average negative tone scores of all performance-related sentences in shareholder report determined using FinBERT (Araci, 2019).
FinBERT_neutral	Weighted average neutral tone scores of all performance-related sentences in shareholder report determined using FinBERT (Araci, 2019).
LM_uncertainty	The sum of tf-idf (Term Frequency-Inverse Document Frequency) scores for all uncertainty tone words in the LM dictionary (Loughran and McDonald, 2011)
LM_litigious	The sum of tf-idf (Term Frequency-Inverse Document Frequency) scores for all litigious tone words in the LM dictionary (Loughran and McDonald, 2011)
LM_strong	The sum of tf-idf (Term Frequency-Inverse Document Frequency) scores for all strong modal words in the LM dictionary (Loughran and McDonald, 2011)
LM_weak	The sum of tf-idf (Term Frequency-Inverse Document Frequency) scores for all weak modal words in the LM dictionary (Loughran and McDonald, 2011)
LM_constraining	The sum of tf-idf (Term Frequency-Inverse Document Frequency) scores for all constraining tone words in the LM dictionary (Loughran and McDonald, 2011)

Panel B: fund characteristics

TNA	The natural logarithm of a fund's total net assets (TNA).
Expense Ratio	Fund's expense ratio. Ratio of total investment that shareholders pay for the fund's operating expenses, which include 12b-1 fees.

Age	Logarithm of a fund's age computed from the date when a fund was first offered.
Turnover	Fund's yearly turnover ratio as reported in CRSP. Minimum (of aggregated sales or aggregated purchases of securities), divided by the average 12-month Total Net Assets of the fund.
Turnover1	Fund's quarterly turnover ratio calculated using TR-13F data as defined in Ben-David, Franzoni, and Moussawi (2010). The minimum of total buys and sales divided by portfolio size.
Turnover2	Fund's quarterly turnover ratio calculated using TR-13F data as defined in Ben-David, Franzoni, and Moussawi (2010). The minimum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size.
Turnover3	Fund's quarterly turnover ratio calculated using TR-13F data as defined in Ben-David, Franzoni, and Moussawi (2010). The sum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size.
Active Share	The share of portfolio holdings that differ from the benchmark index holdings as defined in Cremers and Petajisto (2009).
Portfolio Concentration	The maximum weight of holdings in portfolio.
Male	A dummy that equals one if the fund is managed by male managers only
Manager Tenure	The average tenure of managers who manage the fund
Closed-end	A dummy that equals one if the fund is closed-ended
SMF	A dummy that equals one if the fund is single-managed fund.
Flow (monthly)	$(TNA_{i,t} - TNA_{i,t-1}) / TNA_{i,t-1} - r_{i,t}$ where $TNA_{i,t}$ denotes fund *i*'s total net assets (TNA) in month t and $r_{i,t}$ denotes fund *i*'s return in month *t*.
Raw Return	Funds' raw returns as reported in CRSP.
Excess Market	Funds' raw returns minus market returns.
CAPM alpha	Performance alpha from a market model. The alphas are estimated using beta coefficients obtained from a rolling regression over the prior twenty-four-month period.
CAPMSP alpha	Performance alpha from a market model with market returns replaced by S&P500 returns. The alphas are estimated using beta coefficients obtained from a rolling regression over the prior twenty-four-month period
FF3 alpha	Performance alpha from an FF3 model (Fama and French, 1992). The alphas are estimated using beta coefficients obtained from a rolling regression over the prior twenty-four-month period.
Carhart alpha	Performance alpha from a Carhart model (Carhart, 1997). The alphas are estimated using beta coefficients obtained from a rolling regression over the prior twenty-four-month period.
IdioRisk	The fund's portfolio idiosyncratic risk over a T-period is calculated as the standard deviation of residuals from factors models during that period.

Appendix B: Examples of Performance Attribution Sentences

This section presents examples of performance attribution sentences classified into four categories (IC, EC, ID, ED) by the fine-tuned GPT3-based NLP model.

Internal Contributor (IC)

(Sentences in this category indicate that what contributes to fund's past performance is an internal and fund-specific factor)

Example 1: The fund's outperformance of the benchmark was driven by security selection, with my picks in the information technology, financials and consumer discretionary sectors contributing most.

Example 2: The fund experienced a positive contribution from its overweighed exposure in industrials, which we attribute to the effects of individual stock selection.

Example 3: The fund's holdings in health care sector, especially biotechnology stocks, held up considerably better than those in the benchmark index.

Example 4: Allocations to off-benchmark corporate bonds, particularly in China and Russia, were our biggest performance contributors.

Example 5: Our best contributors for the second quarter included independent refiner Sunoco, gas pipeline operator El Paso, and Newmont Mining, all of which outperformed their lagging sectors.

External Contributor (EC)

(Sentences in this category indicate that what contributes to fund's past performance is an external and non-fund-specific factor)

Example 1: An improving economic outlook, rising interest rates, and increasing trading volumes drove fund's performance.

Example 2: The fund benefited significantly from holdings in health care sector, as it was a strong-performing sector of the benchmark index.

Example 3: The declines in energy and commodity prices benefited these holdings to a greater extent than the overall market and drove significant relative outperformance during the third quarter of 2008.

Example 4: Investment grade corporate bonds and emerging markets were additional significant contributors to fund's performance as credit spreads tightened and interest rates declined materially over the final three quarters of the period.

Example 5: Information technology was also a notable outperformer, with particularly good returns generated by companies profiting from rising demand for computer hardware.

Internal Detractor (ID)
(Sentences in this category indicate that what detracts from fund's past performance is an internal and fund-specific factor)

Example 1: Our overweight in materials didn't enable us to outperform, and poor stock selection in the sector hurt us as well.

Example 2: Stock selection in Australia and off-benchmark allocations to Argentina and Canada also hindered relative results.

Example 3: Picks in information technology and industrials hurt relative performance, as did my overall positioning in consumer discretionary, where a modest overweighting in the weak-performing automobiles/components segment detracted.

Example 4: The fund's underperformance of the MSCI all country world index net was primarily due to selection of stocks in the health care sector, as well as the fund's underweights in consumer staples and utilities.

Example 5: Performance also suffered due to a stake in Constellium, a Dutch aluminum producer, as well as my decision to sell strong-performing chipmaker Skyworks solutions.

External Detractor (ED)
(Sentences in this category indicate that what detracts from fund's past performance is an external and non-fund-specific factor)

Example 1: During the last six months, this was an impediment to the performance of the funds, as value stock returns have continued to outpace growth returns.

Example 2: This quick and dramatic sector rotation caused the fund to underperform.

Example 3: The fund's performance was hampered by maintaining a significant overweight allocation to the consumer discretionary sector, as it was a weak-performing sector of the benchmark index.

Example 4: In all three cases, many of the fund's underweight stocks underperformed as the market rallied.

Example 5: The fund's performance was hurt by maintaining an overweighted allocation to the banks industry during the period, as banks were a weak-performing industry group of the S&P 500 index with a return of -2%.

Appendix C: Training and Evaluating NLP Models

This section describes the training and evaluation process of transformer-based NLP models. The model takes as its input a performance-related sentence, which is identified by key phrases that have embeddings closest to the performance-related root words, and outputs labels for the sentence.

Step 1: Training Sample Construction

A high-quality training sample must include sentences from various sub-categories within each target main category (Indurkhya and Damerau, 2010). To achieve this objective, I construct a word list that covers words associated with different sub-categories of attribution sentences, including 10 root words: "benchmark," "sector," "market," "environment," "selection," "pick," "detract," "contribute," "gain," and "outperform." For each root word, I search for 9 words with closed embeddings and add them to the word list, which gives a total of 100 words. Note that this word list contains duplicated words, as duplicates naturally occur more frequently than other words. Thus, keeping duplicates increases their "weight" in the training sample. Next, for each of the 100 words, I randomly select 20 unique sentences containing the word to form my training sample, which consists of 1,916 sentences. For each sentence in the training sample, I label it as 1) having attribution information (i.e., perception of causality) vs. lacking attribution information, 2) describing a performance contributor vs. describing a performance detractor, and 3) describing an internally attributed factor vs. describing an externally attributed factor. To minimize human errors in the labeling process, I ask two MBA students to cross-validate labels and require a consensus on the classification by three of us.

Step 2: Training and Evaluating NLP Models

I fine-tune NLP models using NVIDIA T4 GPU on Google Colab. I consider eight candidate models: BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), DistillBERT (Sanh et al., 2019), XLNet (Yang et al., 2019), GPT2 (Radford et al., 2019), GPT3 (Brown et al., 2020), ERNIE (Sun et al., 2019), and BLOOM (Dadachev et al., 2022). GPT3 is accessible using OpenAI API; the rest are open-sourced and available in Hugging Face library. To be specific, I choose "*bert-base-uncased*," "*roberta-base*," "*distilbert-base-uncased*," "*xlnet-base-uncased*," "*gpt2*," "*ada*," "*ernie-2.0-large-en*," and "*bloom-560m*." Sentences are tokenized into different kinds of inputs based on the selected model. Input sequences are padded with a special token to match the length of the longest sequence in the dataset; the maximum length is determined by finding the longest sequence among all the input data. The sequences are truncated if the length exceeds 42 characters or tokens. I apply a stratified 5-fold cross-validation method and use four measures to evaluate each model's out-of-sample performance. Specifically, 1) the dataset is randomly shuffled to avoid any inherent ordering or bias; 2) the data is divided into five folds, with each fold containing a roughly equal proportion of samples from each class (this ensures that each fold is representative of the overall class distribution); 3) the model is trained on four folds and evaluated on the remaining fold, a process that is repeated five times, with each fold being used as the evaluation set exactly once; 4) the performance metrics—accuracy, precision, recall, or F1 score—are calculated for each iteration of the cross-validation process; and 5) the final performance of the model is typically reported as the average of the performance metrics obtained from the five iterations.

References

Adebambo, B. N., & Yan, X. (2018). Investor overconfidence, firm valuation, and corporate decisions. *Management Science*, *64*(11), 5349-5369.

Ahmed, A. S., & Duellman, S. (2013). Managerial overconfidence and accounting conservatism. *Journal of accounting research*, *51*(1), 1-30.

Amihud, Y., & Goyenko, R. (2013). Mutual fund's R 2 as predictor of performance. *The Review of Financial Studies*, *26*(3), 667-694.

Anderson, C. A., & Jennings, D. L. (1980). When experiences of failure promote expectations of success: The impact of attribution failure to ineffective strategies 1. *Journal of Personality*, *48*(3), 393-407.

Bacon, C. R. (2023). *Practical portfolio performance measurement and attribution*. John Wiley & Sons.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, *116*(1), 261-292.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, *1*, 1053-1128.

Barras, L., Scaillet, O., & Wermers, R. (2010). False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, *65*(1), 179-216.

Ben-David, I., Franzoni, F., & Moussawi, R. (2010). The behavior of hedge funds during liquidity crises. *Unpublished working paper, Ohio State University*.

Berk, J. B., & Van Binsbergen, J. H. (2016). Assessing asset pricing models using revealed preference. *Journal of Financial Economics*, *119*(1), 1-23.

Brinson, G. P., & Fachler, N. (1985). Measuring non-US. equity portfolio performance. *The Journal of Portfolio Management*, *11*(3), 73-76.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901.

Campbell, T. C., Gallmeyer, M., Johnson, S. A., Rutherford, J., & Stanley, B. W. (2011). CEO optimism and forced turnover. *Journal of Financial Economics*, *101*(3), 695-712.

Campbell, W. K., Sedikides, C., Reeder, G. D., & Elliot, A. J. (2000). Among friends? An examination of friendship and the self-serving bias. *British Journal of Social Psychology*, *39*(2), 229-239.

Cao, S., Yang, B., & Zhang, A. (2023). Managerial risk assessment and fund performance: Evidence from textual disclosure. *Available at SSRN 4060307*.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, *52*(1), 57-82.

CFA Institute (2019). *CFA Institute Annual Report, Fiscal Year 2019*.

Chava, S., Du, W., Shah, A., & Zeng, L. (2022). Measuring firm-level inflation exposure: A deep learning approach. *Available at SSRN 4228332*.

Choi, D., & Lou, D. (2010, December). A test of the self-serving attribution bias: evidence from mutual funds. In *AFA 2011 Denver Meetings Paper*.

Chui, A. C., Titman, S., & Wei, K. J. (2010). Individualism and momentum around the world. *The Journal of Finance*, *65*(1), 361-392.

Cooper, M. J., Gutierrez Jr, R. C., & Hameed, A. (2004). Market states and momentum. *The journal of Finance*, *59*(3), 1345-1365.

Cremers, K. M., & Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *The review of financial studies*, *22*(9), 3329-3365.

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *the Journal of Finance*, *53*(6), 1839-1885.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Fama, E. F., & French, K. R. (2010). Luck versus skill in the cross-section of mutual fund returns. *The journal of finance*, *65*(5), 1915-1947.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Galasso, A., & Simcoe, T. S. (2011). CEO overconfidence and innovation. *Management science*, *57*(8), 1469-1484.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535-574.

Gervais, S., & Odean, T. (2001). Learning to be overconfident. *The review of financial studies*, *14*(1), 1-27.

Glaser, M., Langer, T., & Weber, M. (2005). Overconfidence of professionals and lay men: individual differences within and between tasks?. *None*.

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*(3), 433-463.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive psychology*, *24*(3), 411-435.

Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135-2202.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of risk and uncertainty*, *4*(1), 5-28.

Hilary, G., & Hsu, C. (2011). Endogenous overconfidence in managerial forecasts. *Journal of Accounting and Economics*, *51*(3), 300-313.

Hilary, G., & Menzly, L. (2006). Does past success lead analysts to become overconfident?. *Management science*, *52*(4), 489-500.

Hillert, A., Niessen-Ruenzi, A., & Ruenzi, S. (2021). Mutual fund shareholder letters: flows, performance, and managerial behavior. *Performance, and Managerial Behavior (May 17, 2021)*.

Hirshleifer, D., Low, A., & Teoh, S. H. (2012). Are overconfident CEOs better innovators?. *The journal of finance*, *67*(4), 1457-1498.

Holden, Sarah, Daniel Schrass, and Elena Barone Chism. 2023. "Defined Contribution Plan Participants' Activities, First Quarter 2023." *ICI Research Report (July)*. Available at www.ici.org/files/2023/23-rpt-recsurveyq1.pdf.

Kacperczyk, M., Sialm, C., & Zheng, L. (2008). Unobserved actions of mutual funds. *The Review of Financial Studies*, *21*(6), 2379-2416.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).

Kugler, T., Kausel, E. E., & Kocher, M. G. (2012). Are groups more rational than individuals? A review of interactive decision making in groups. *Wiley interdisciplinary reviews: Cognitive science*, *3*(4), 471-482.

Li, F. (2010). Managers' self-serving attribution bias and corporate financial policies. *Available at SSRN 1639005*.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring corporate culture using machine learning. *The Review of Financial Studies*, *34*(7), 3265-3315.

Li, X., Zhu, X., Ma, Z., Liu, X., & Shah, S. (2023). Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? An Examination on Several Typical Tasks. *arXiv preprint arXiv:2305.05862*.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of finance*, *66*(1), 35-65.

Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology*, *92*(1), 152.

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *The journal of finance*, *60*(6), 2661-2700.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction?. *Psychological bulletin*, *82*(2), 213.

Mullainathan, S., Schwartzstein, J., & Shleifer, A. (2008). Coarse thinking and persuasion. *The Quarterly journal of economics*, *123*(2), 577-619.

Noh, J., & Zhou, D. (2022). Executives' Blaming external factors and market reactions: Evidence from earnings conference calls. *Journal of Banking & Finance*, *134*, 106358.

Petajisto, A. (2013). Active share and mutual fund performance. *Financial analysts journal*, *69*(4), 73-93.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm. *Journal of personality and social psychology*, *32*(5), 880.

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., ... & Manica, M. (2022). Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

Schrand, C. M., & Zechman, S. L. (2012). Executive overconfidence and the slippery slope to financial misreporting. *Journal of Accounting and economics*, *53*(1-2), 311-329.

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, *2*(2), 895-908.

Sirri, E. R., & Tufano, P. (1998). Costly search and mutual fund flows. *The journal of finance*, *53*(5), 1589-1622.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, *103*(2), 193.

U.S. Securities and Exchange Commission (2013). *How to Read a Mutual Fund Shareholder Report.* Available at https://www.sec.gov/files/ib_readmfreport.pdf.

Weiner, B., Russell, D., & Lerman, D. (1979). The cognition–emotion process in achievement-related contexts. *Journal of Personality and Social psychology*, *37*(7), 1211.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, *32*.

Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., & Liu, Q. (2019). ERNIE: Enhanced language representation with informative entities. *arXiv preprint arXiv:1905.07129*.

Figure 1: Illustration of NLP Architecture

This figure illustrates the two-layer NLP architecture I developed to extract attribution information from mutual funds' self-assessment of performance. The input to the model is a performance-related sentence identified by key phrases with closest embeddings to the performance root words. The first layer of model has one classifier—classifier 0, which is used to identify and pass performance-attribution information (i.e., perception of causality) to the second layer; the second layer consists of two classifiers—classifier 1 and classifier 2, which independently classify the performance-attribution sentence into i) performance contributor vs. performance detractor, ii) internal factor vs. external factor. All classifiers undergo fine-tuning using GPT3 (Ada), selected from a pool of eight transformer-based model candidates based on the out-of-sample test results using a stratified 5-fold cross-validation mtehod.

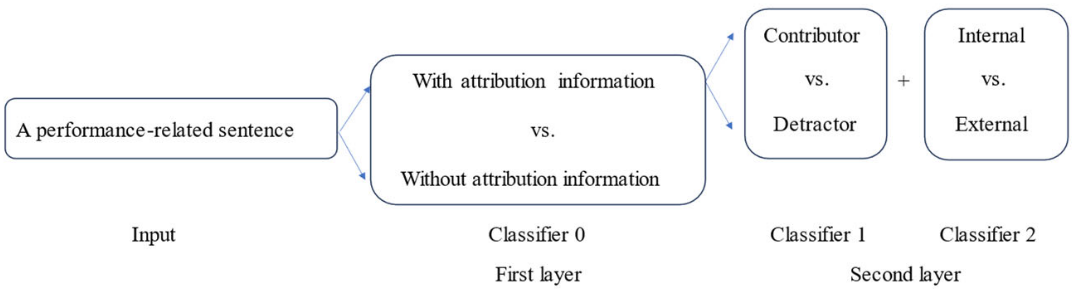

Figure 2: Distribution of *Self-attribution Score (SAS)*

This figure plots the distribution of *Self-attribution Score (SAS)* measured from 15,422 shareholder reports associated with 1,400 unique actively managed domestic equity funds over a 13-year span from 2006 to 2018. The X-axis depicts *SAS* values, while the Y-axis shows density.

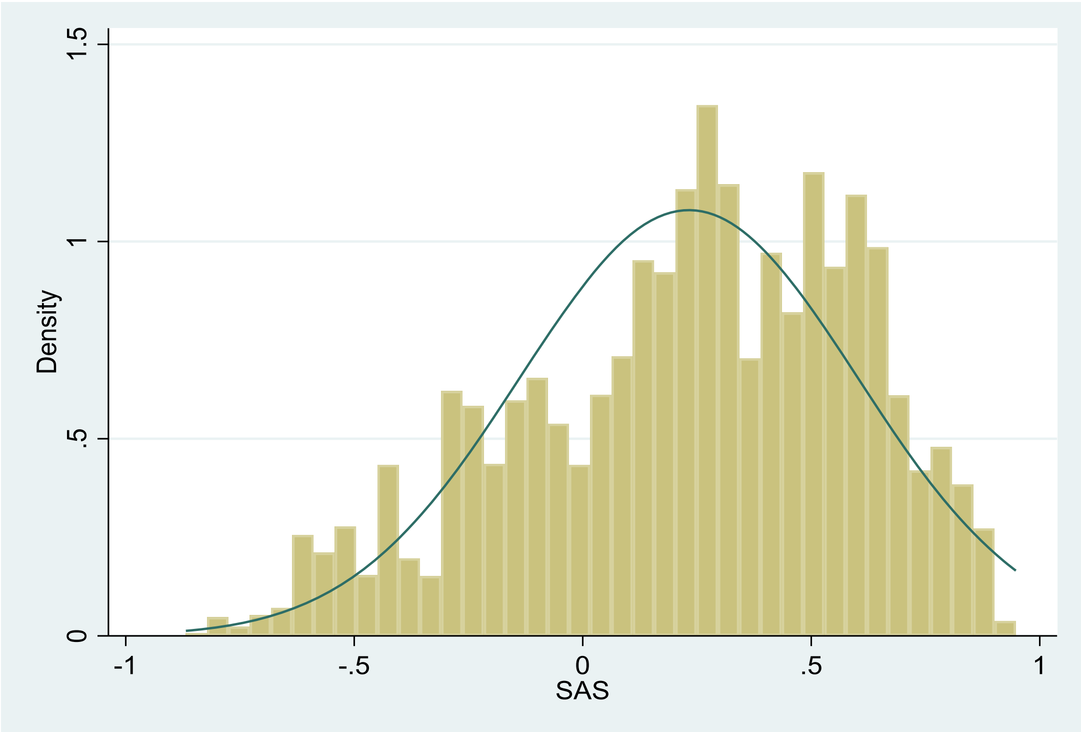

Figure 3: Attributed Factors of Performance Contributors versus Detractors

This figure displays the average percentage of internally and externally attributed factors to performance contributors versus detractors in shareholder reports. Specifically, for each shareholder report, I calculate the length (i.e., number of words) of performance attribution sentences classified as internal contributors (*IC*) divided by the length of internal and external contributors (*IC* + *EC*); I do the same calculation for external contributors (*EC*), internal detractors (*ID*), and external detractors (*ED*). Next, I calculate the average ratios across all

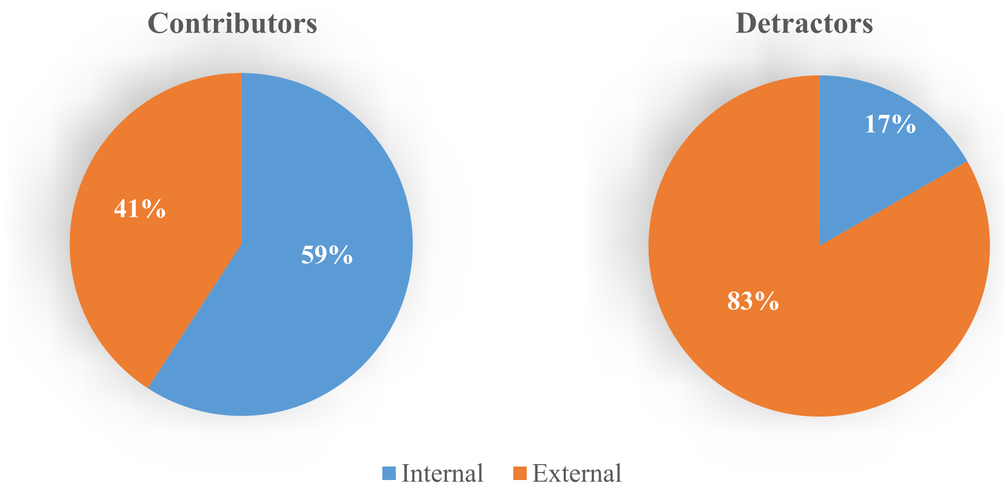

Table 1: Horserace of Model Candidates

This table reports the out-of-sample performance of eight NLP model candidates for classifiers. Models are evaluated using a stratified 5-fold cross validation method (see Appendix C for detailed descriptions). *Accuracy* is the ratio of correct category predictions to total number of observations (i.e., the sum of true positives and true negatives divided by the total number of observations). *Precision* is the ratio of true positives to the sum of true positives and false positives. *Recall* is the ratio of true positives to the sum of true positives and false negatives. *F1 score* = 2* *Precision* * *Recall* / (*Precision* + *Recall*). Classifier 0 identifies performance-attribution (i.e., the perception of causality must be inferred) sentences from performance-related sentences; Classifier 1 and 2 classify a performance-attribution sentence along two dimensions: i) performance contributor vs. performance detractor and ii) internal factor vs. external factor.

Classifier 0: Performance Attribution (i.e., perception of caulsality)

	BERT	*RoBERTa*	*DistilBert*	*XLNet*	*GPT2*	*GPT3*	*ERNIE*	*BLOOM*
Accuracy	88.45%	88.65%	87.54%	82.56%	86.73%	89.39%	84.31%	76.50%
Precision	87.55%	87.44%	86.44%	79.54%	85.42%	91.67%	81.18%	76.18%
Recall	87.79%	88.23%	86.89%	82.79%	86.12%	92.37%	85.09%	76.78%
F1 score	87.67%	87.83%	86.67%	81.13%	85.77%	92.02%	83.09%	76.48%

Classifier 1: Contributor v.s. Detractor

	BERT	*RoBERTa*	*DistilBert*	*XLNet*	*GPT2*	*GPT3*	*ERNIE*	*BLOOM*
Accuracy	93.55%	94.37%	92.22%	78.80%	87.66%	92.48%	94.63%	76.63%
Precision	93.30%	94.18%	92.11%	78.06%	88.08%	87.07%	94.53%	75.75%
Recall	93.72%	94.44%	92.16%	79.04%	87.78%	97.17%	94.62%	77.05%
F1 score	93.51%	94.31%	92.14%	78.55%	87.93%	92.73%	94.58%	76.39%

Classifier 2: Internal v.s. External

	BERT	*RoBERTa*	*DistilBert*	*XLNet*	*GPT2*	*GPT3*	*ERNIE*	*BLOOM*
Accuracy	85.39%	77.39%	84.35%	81.56%	80.52%	87.85%	81.98%	80.36%
Precision	88.79%	75.52%	88.79%	80.60%	79.84%	82.95%	80.73%	79.43%
Recall	85.59%	79.44%	84.07%	83.07%	80.49%	87.50%	82.52%	80.04%
F1 score	87.16%	77.43%	86.36%	81.82%	80.17%	85.12%	81.62%	79.74%

Table 2: Summary Statistics

This table describes the summary statistics for key variables used in the analysis. Variable constructions are described in Section 2.3 and Appendix A.

Panel A: filing-level variables	**Obs.**	**Mean**	**SD**	**P1**	**P25**	**P50**	**P75**	**P99**
SAS (Self-attribution Score)	15,434	0.23	0.37	-0.63	-0.03	0.27	0.52	0.88
Contributor	15,434	0.53	0.25	0.00	0.36	0.52	0.70	1.00
Detractor	15,434	0.47	0.25	0.00	0.30	0.48	0.64	1.00
Internal	15,434	0.37	0.29	0.00	0.10	0.34	0.58	0.90
External	15,434	0.63	0.29	0.00	0.42	0.66	0.90	1.00
FinBERT_positive	15,434	7.05	6.45	0.20	2.84	5.30	9.05	32.11
FinBERT_negative	15,434	3.52	3.54	0.00	1.10	2.45	4.85	16.14
FinBERT_neutral	15,434	7.13	9.22	0.00	1.38	4.30	9.30	41.84
LM_uncertainty	15,434	0.03	0.02	0.00	0.02	0.03	0.04	0.10
LM_litigious	15,434	0.01	0.01	0.00	0.00	0.00	0.01	0.04
LM_strong	15,434	0.01	0.01	0.00	0.00	0.00	0.01	0.02
LM_weak	15,434	0.01	0.01	0.00	0.00	0.01	0.01	0.03
LM_constraining	15,434	0.01	0.01	0.00	0.00	0.00	0.01	0.05

Panel B: fund characteristics ((N = 1,400 actively managed domestic equity funds)								
TNA ($ millions)	15,434	1570.3	5564.3	1.3	44.3	206.7	972.3	23915.2
Expense Ratio (%)	15,434	1.11	0.49	0.06	0.83	1.12	1.39	2.34
Age (years)	15,434	13.40	11.28	1.00	5.00	11.00	18.00	57.50
Turnover	15,434	0.77	0.95	0.03	0.26	0.51	0.90	5.17
Flow (monthly)	15,434	0.03	0.25	-0.42	-0.06	-0.01	0.04	1.12
Raw Return (monthly)	15,434	0.005	0.047	-0.124	-0.017	0.009	0.033	0.114
Excess Market (monthly)	15,434	-0.001	0.021	-0.065	-0.011	-0.001	0.008	0.058
CAPM Alpha (monthly)	15,434	0.001	0.024	-0.068	-0.011	0.001	0.013	0.063
CAPM SP500 Alpha (monthly)	15,434	0.001	0.048	-0.133	-0.023	0.003	0.027	0.125
FF3 Alpha (monthly)	15,434	0.001	0.264	-0.679	-0.030	0.000	0.028	0.895
Carhart Alpha (monthly)	15,434	-0.002	0.118	-0.433	-0.025	-0.002	0.021	0.435

(cont'd)

Panel C: statistics of *SAS*

Time Series Mean (average of fund*manager)	0.25
Time Series Sd (average of fund*manager)	0.31
Cross Section Mean (average of all reporting period)	0.23
Cross Section Sd (average of all reporting periods)	0.36
Im-Pesaran-Shin t-bar (AIC)	-24.76
Fisher-type Inverse Chi-squared (dfuller, lag1)	1402.04

Table 3: Sentence-level Probit Regression

This table presents results from a sentence-level probit model that examines whether performance contributors (compared to detractors) are more likely to be attributed to internal factors. Specifically, I estimate the following regression:

$$Internal_{s,i} = \beta Contributor_{s,i} + \gamma FinBERT\_positive_{s,i} + \delta LM\_sent_{s,i} + \mu_i + \varepsilon_{s,i},$$

where the dependent variable *Internal* is a dummy variable that equals one if a performance-related sentence s in N-CSR filing i is internally attributed and equals zero if externally attributed. *Contributor* equals one if sentence s concerns what contributes to the fund's performance and equals zero if it concerns what detracts from the fund's performance. $FinBERT\_positive$ equals 1 if the tone of sentence s measured by FinBERT (Araci, 2019). LM_sent is a set of textual control variables including *LM_Positive, LM_Negative, LM_Uncertainty*, *LM_Litigious*, *LM_Strong*, *LM_Weak*, and *LM_Constraining*, which measure the *tf_idf* (Term Frequency-Inverse Document Frequency) of relevant relevant LM sentiment keywords (Loughran and McDonald, 2011) in sentence s. I include filing fixed effects μ_i to rule out the effects of potential unobserved filing-level factors such as writing style. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. *t* statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *Internal*	
	(1)	(2)
Contributor	0.4811***	0.4062***
	(24.91)	(17.11)
FinBERT_positive		0.1171***
		(5.04)
LM_Positive		0.8068*
		(1.73)
LM_Negative		-0.2915***
		(-2.68)
LM_Uncertainty		0.7621
		(1.57)
LM_Litigious		0.9023
		(1.19)
LM_Strong		1.4832
		(0.78)
LM_Weak		-4.2990**
		(-2.33)
LM_Constraining		0.0322
		(0.03)
Filing fixed effects	Y	Y
N	121,815	112,242
R-squared	0.550	0.598

Table 4: *Self-attribution Score (SAS)* and Future Performance

This table displays results that regress fund's future performance on *Self-attribution Score (SAS)*. Specifically, I estimate the following panel regression:

$$Perf_{i,[t+1,t+6]} = \beta SAS_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variable *Perf* is the cumulative alphas of fund *i* from one month after a shareholder report was filed (*t*+1) to six months after the filing (*t*+6), and alphas are computed from various factor models using beta coefficients obtained from a rolling regression over the prior twenty-four-month period. *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. *FilingControls* is a set of filing-level control variables including *FinBERT_positive, FinBERT_negative*, and *FinBERT_neutral*, which are the weighted average tone scores of performance-related sentences in the shareholder report determined by FinBERT (Araci, 2019), and *LM_Positive*, *LM_Negative*, *LM_Uncertainty, LM_Litigious, LM_Strong, LM_Weak,* and *LM_Constraining,* which measure the *tf_idf* (Term Frequency-Inverse Document Frequency) of relevant words in an LM sentiment dictionary (Loughran and McDonald, 2011). *FundControls* is a set of fund characteristics including *Age*, *Size*, *Expense Ratio*, *Past Flow*, *Past Performance*, and *Turnover*. I include fund style * reporting period fixed effects $\mu_{style,[t-6,t-1]}$ to partial out potential style-period-level macro confounding factors and only compare funds within the same style and reporting period. I also include fund fixed effects π_i to exclude any impact of potential fund-level confounding characteristics. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *Perf*					
	(1)	(2)	(3)	(4)	(5)	(6)
SAS	-0.0040***	-0.0018*	-0.0064**	-0.0432**	-0.0443**	-0.0206*
	(-3.30)	(-1.88)	(-2.01)	(-2.21)	(-2.05)	(-1.80)
FinBERT_Positive	0.0009***	0.0003***	0.0012**	-0.0055**	0.0023	0.0009
	(6.74)	(3.13)	(2.24)	(-2.12)	(1.15)	(0.53)
FinBERT_Negative	-0.0001	0.0003	0.0003	-0.0107**	0.0046*	0.0008
	(-0.26)	(1.52)	(1.09)	(-2.10)	(1.71)	(0.50)
FinBERT_Neutral	0.0001	0.0001	-0.0001	0.0032**	0.0028**	0.0014
	(1.32)	(1.24)	(-0.30)	(2.22)	(2.12)	(1.37)
LM_Positive	-0.0131	0.0131	0.1100	-0.164	0.8820**	0.8370***
	(-0.36)	(0.48)	(1.34)	(-0.57)	(2.26)	(2.89)
LM_Negative	-0.0328	-0.1030***	0.0543	-0.7750**	-0.7750***	0.3600
	(-1.63)	(-6.27)	(0.82)	(-2.56)	(-3.07)	(1.62)
LM_Uncertainty	-0.0608	0.0690**	0.0709	-2.1990**	0.6810	-0.4390
	(-1.49)	(2.28)	(0.62)	(-2.57)	(0.79)	(-0.86)

(cont'd)

	(1)	(2)	(3)	(4)	(5)	(6)
LM_Litigious	-0.0681	0.1950***	0.1520	-1.0600	-0.0662	0.4770
	(-0.98)	(3.61)	(1.46)	(-1.51)	(-0.08)	(1.35)
LM_Strong	0.1780	-0.0057	0.5980	1.9420	1.0720	0.7980
	(1.28)	(-0.05)	(1.44)	(1.37)	(0.60)	(0.56)
LM_Weak	0.1650	0.0032	0.0096	4.8130**	-1.8220	2.1580
	(1.25)	(0.03)	(0.04)	(2.28)	(-1.29)	(1.37)
LM_Constraining	-0.1620**	-0.2390***	0.1490	0.9630*	0.7590	1.0930**
	(-2.51)	(-4.28)	(1.00)	(1.72)	(0.81)	(2.12)
Size	-0.0172***	-0.0140***	-0.0195***	-0.0998	0.0127	-0.0213
	(-8.65)	(-9.22)	(-3.77)	(-1.65)	(0.39)	(-1.51)
Expense Ratio	-0.8040	-0.3860	-1.5110	-3.2660	8.2850	-2.3420
	(-0.94)	(-0.56)	(-1.05)	(-0.22)	(0.67)	(-0.49)
Age	0.0054	0.0045	0.0254	0.335	-0.3960*	0.1730**
	(1.14)	(1.19)	(0.96)	(1.16)	(-1.69)	(2.17)
Turnover	0.0022	-0.0001	0.0048	-0.0007	0.0747**	-0.0488
	(1.11)	(-0.06)	(1.25)	(-0.03)	(2.05)	(-1.02)
Flow[m+1, m+6]	0.0036***	0.0059***	0.0034	-0.0142	0.0277	0.0053
	(2.78)	(5.64)	(0.95)	(-0.96)	(1.25)	(0.47)
Flow[m-6, m-1]	-0.0021*	-0.0041***	-0.0063	0.0096	-0.0131	-0.0186
	(-1.77)	(-4.26)	(-1.23)	(1.05)	(-0.53)	(-1.12)
Perf[m-6, m-1]	-0.3820***	-0.1580***	-0.1570***	-0.0163	-0.6900***	-0.0814
	(-34.50)	(-18.81)	(-12.48)	(-0.10)	(-4.00)	(-1.49)
Perf Specification	*Raw*	*Excess Mkt*	*CAPM*	*CAPMSP*	*FF3*	*FFCarhart*
Fund style-by-reporting period fixed effects	Y	Y	Y	Y	Y	Y
Fund fixed effects	Y	Y	Y	Y	Y	Y
N	15,434	15,434	15,434	15,434	15,434	15,434
R-squared	0.652	0.420	0.170	0.110	0.235	0.137

Table 5: *Self-attribution Score (SAS)* and Future Portfolio Turnover

This table reports the results on the relation between a fund's future portfolio turnover and its *Self-attribution Score (SAS)*. Specifically, I estimate the following panel regression:

$$Turnover_{i,[t+1,t+6]} = \beta SAS_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variable *Turnover* is the portfolio turnover of fund i from one month after a shareholder report was filed (t+1) to six months after the filing (t+6). I follow Ben-David, Franzoni, and Moussawi (2010) to calculate three types of turnover measures using 13F holding data. In column (1), the dependent variable *Turnover1* is calculated as the minimum of total buys and sales divided by portfolio size; in column (2), *Turnover2* is calculated as the minimum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size; in column (3), *Turnover3* is calculated as the sum of total buys and sales adjusted by net flows and redemptions and divided by portfolio size. *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *Turnover*		
	(1)	(2)	(3)
SAS	0.0078***	0.0040	0.0083**
	(3.17)	(1.55)	(2.33)
Filing controls	Y	Y	Y
Fund controls	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y
Fund fixed effects	Y	Y	Y
N	11,279	11,279	11,279
R-squared	0.834	0.371	0.839

Table 6: *Self-attribution Score (SAS)* and Future Idiosyncratic Risk

This table investigates the relationship between a fund's future portfolio idiosyncratic risk and its *Self-attribution Score (SAS)*. Specifically, I estimate the following panel regression:

$$IdioRisk_{i,[t+1,t+6]} = \beta SAS_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variable *IdioRisk* is the portfolio idiosyncratic risk on fund *i*'s returns from one month after a shareholder report was filed (t+1) to six months after the filing (t+6). To be specific, in columns (1) and (2), *IdioRisk* is calculated as the standard deviation of a fund's raw return and excess return to market; in columns (3) to (6), *IdioRisk* is calculated as the standard deviation of model residuals using beta coefficients estimated using *CAPM*, *CAPMSP*, *FF3*, and *FFCarhart* models from a rolling regression over the prior twenty-four-month period. *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *IdioRisk*					
	(1)	(2)	(3)	(4)	(5)	(6)
SAS	0.0002	0.0000	0.0002***	0.0005***	0.0042**	0.0060**
	(1.37)	(1.21)	(3.22)	(2.81)	(2.17)	(1.99)
Perf Specification	*Raw*	*Excess Mkt*	*CAPM*	*CAPMSP*	*FF3*	*FFCarhart*
Filing controls	Y	Y	Y	Y	Y	Y
Fund controls	Y	Y	Y	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y	Y	Y	Y
Fund fixed effects	Y	Y	Y	Y	Y	Y
N	15,432	15,432	15,431	15,431	15,431	15,431
R-squared	0.766	0.791	0.768	0.738	0.557	0.225

Table 7: *Self-attribution Score (SAS)* and Investor Flows

This table reports the results on investor flow sensitivity on *Self-attribution Score (SAS)*. Specifically, I estimate the follow regression:

$$Flow_{i,t+1} = \beta_1 SAS_{i,t} + \beta_2 LowPerf_{i,[t-6,t-1]} + \beta_3 HighPerf_{i,[t-6,t-1]} + \beta_4 SAS_{i,m} * LowPerf_{i,[t-6,t-1]} + \beta_5 SAS_{i,m} * HighPerf_{i,[t-6,t-1]} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,m}$$

where *Flow* is the monthly flow of fund *i* in month (*t*+1), month *t* being the month in which the shareholder report was filed. *LowPerf* and *HighPerf* are dummies that take a value of one if fund *i*'s performance over the reporting period (*t*-6 to *t*-1) is in the bottom or top quantile (within the same fund style) as defined in Sirri and Tufano (1998). Performance alphas are computed from *CAPM* using beta coefficients obtained from a rolling regression over the past twenty-four-month period. *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *Flow*					
	(1)	(2)	(3)	(4)	(5)	(6)
SAS	0.0022	-0.0032	-0.0029	-0.0015	-0.0000	-0.0023
	(0.44)	(-1.00)	(-0.95)	(-0.47)	(-0.01)	(-0.82)
LowPerf	-0.0219***	-0.0116	-0.0368***	-0.0263***	-0.0093	-0.0158***
	(-2.75)	(-1.55)	(-6.68)	(-4.11)	(-1.34)	(-2.59)
HighPerf	0.0124*	0.0052	0.0167***	0.0117**	0.0093*	0.0133**
	(1.77)	(0.77)	(3.05)	(2.11)	(1.71)	(2.25)
*SAS * LowPerf*	0.0118**	0.0115*	0.0186***	0.0131*	0.0182**	0.0115***
	(2.08)	(1.83)	(2.85)	(1.71)	(2.23)	(3.16)
*SAS * HighPerf*	0.0053	0.0066	0.0016*	0.0018	0.0033	0.0020
	(0.79)	(1.04)	(1.76)	(0.30)	(0.54)	(0.30)
Perf Specification	*Raw*	*Excess Mkt*	*CAPM*	*CAPMSP*	*FF3*	*FFCarhart*
Filing controls	Y	Y	Y	Y	Y	Y
Fund controls	Y	Y	Y	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y	Y	Y	Y
Fund fixed effects	Y	Y	Y	Y	Y	Y
N	15,711	15,711	15,711	15,711	15,711	15,711
R-squared	0.276	0.277	0.277	0.275	0.273	0.274

Table 8: *Self-attribution Score (SAS) and Past Performance*

This table examines the relation between fund's *Self-attribution Score (SAS)* and its past performance. Specifically, I estimate the following panel regression:

$$SAS_{i,t} = \beta_1 Perf_{i,[t-6,t-1]} + \beta_2 AS_{i,[t-6,t-1]} + \beta_3 Perf_{i,[t-6,t-1]} \times AS_{i,[t-6,t-1]} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variable *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. *Perf* is the cumulative alphas of fund *i* from six months before a shareholder report was filed (t-6) to one month before the filing (t-1), where alphas are computed from various factor models using beta coefficients obtained from a rolling regression over the prior twenty-four-month period, and *AS* is the active share for fund *i* over past six months (*t*-6 to *t*-1), which is used to proxy for fund-specific deviation and calculated as in Cremers and Petajisto (2009). I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *SAS*					
	(1)	(2)	(3)	(4)	(5)	(6)
Perf	0.0231	0.4460**	0.0068**	0.0054*	0.0023*	0.0058*
	(1.08)	(2.35)	(2.37)	(1.67)	(1.75)	(1.66)
AS	-0.0412	-0.0350	-0.0401	-0.0421	-0.0407	-0.0408
	(-0.25)	(-0.21)	(-0.25)	(-0.26)	(-0.25)	(-0.25)
*Perf * AS*	0.0959	-0.0612	-0.0810	-0.0580	0.0040	-0.0180
	(0.28)	(-0.55)	(-1.09)	(-0.92)	(0.14)	(-0.44)
Perf Specification	*Raw*	*Excess Mkt*	*CAPM*	*CAPMSP*	*FF3*	*FFCarhart*
Filing controls	Y	Y	Y	Y	Y	Y
Fund controls	Y	Y	Y	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y	Y	Y	Y
Fund fixed effects	Y	Y	Y	Y	Y	Y
N	14,932	14,932	14,932	14,932	14,932	14,932
R-squared	0.226	0.227	0.226	0.226	0.226	0.226

Table 9: *Self-attribution Score (SAS)* and Cognitive Bias Indicators

This table presents results on the relation between fund's *Self-attribution Score (SAS*) and potential cognitive bias indicators. Specifically, I estimate the following panel regression:

$$SAS_{i,t} = \beta CBI_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \varepsilon_{i,t},$$

where where the dependent variable *SAS* is the *Self-attribution Score* measured by the GPT3-based NLP model (described in Section 2.2) on performance-related sentences in fund *i*'s shareholder report filed in month *t*. *CBI* is a set of congnitive bias indicators including *Male*, *Manager Tenure*, *Portfolio Concentration*, and *SMF*. *Male* is a dummy that equals one if the fund is managed by male managers only; *Manager Tenure* is the average tenure of managers who manage the fund; *Portfolio Concentration* is the maximum weight of holdings in portfolio; *SMF* is a dummy that equals one if the fund is single-managed fund. I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *SAS*				
	(1)	(2)	(3)	(4)	(5)
Male	0.0329				0.0325
	(1.15)				(1.14)
Manager Tenure		-0.0035**			-0.0056**
		(-2.54)			(-2.93)
Portfolio Concentration			0.3840*		0.3620*
			(1.65)		(1.71)
SMF				0.0156	0.0211
				(0.91)	(0.70)
Filing controls	Y	Y	Y	Y	Y
Fund controls	Y	Y	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y	Y	Y
N	12,282	14,218	12,372	15,434	9,675
R-squared	0.227	0.215	0.234	0.215	0.245

Internet Appendix: Components of *Self-attribution Score (SAS)* and Future Performance

This table presents results that regresses the fund future performance on decomposed components of *Self-attribution Score (SAS)*. Specifically, I estimate the panel regression:

$$Perf_{i,[t+1,t+6]} = \beta SASComp_{i,t} + \gamma FilingControls_{i,t} + \delta FundControls_{i,t} + \mu_{style,[t-6,t-1]} + \pi_i + \varepsilon_{i,t},$$

where the dependent variable *Perf* is the cumulative alphas of fund *i* from one month after a shareholder report was filed (t+1) to six months after the filing (t+6), and alphas are computed from various factor models using beta coefficients obtained from a rolling regression over the prior twenty-four-month period. $SASComp$ is a set of decomposed components of SAS including *Contribution, Internal Attribution*, *SAS_contributors, and SAS_detractors. Contributors* is the proportion of contributors in performance-related sentences in fund *i*'s shareholder report filled in month *t*; *Internal Attributiors* is the proportion of internally attributed factors in performance-related sentences in fund *i*'s shareholder report filled in month *t*. *SAS_contributors* is the *Self-attribution Score* measured from performance contributors only, and *SAS_detractors* is the *Self-attribution Score* measured from performance detractors only. I include the same set of control variables and fixed effects as specified in Table 4. Standard errors are adjusted for heteroskedasticity and clustered at the fund level. t statistics are in parentheses. * indicates significance at the 10% level; ** at the 5% level; and ***, at the 1% level.

	Dep. variable = *Perf*			
	(1)	(2)	(3)	(4)
Contributors	0.0012		-0.0030	
	(0.38)		(-0.86)	
Internal Attributors		-0.0007	0.0028	
		(-0.30)	(0.78)	
SAS			-0.0070**	
			(-2.16)	
SAS_contributors				-0.0059**
				(-2.05)
SAS_detractors				-0.0071*
				(-1.73)
Filing controls	Y	Y	Y	Y
Fund controls	Y	Y	Y	Y
Fund style-by-reporting period fixed effects	Y	Y	Y	Y
Fund fixed effects	Y	Y	Y	Y
N	15,434	15,434	15,434	15,434
R-squared	0.169	0.169	0.170	0.170