Received: 2 March 2020 | Revised: 26 June 2023 | Accepted: 17 July 2023

DOI: 10.1002/smj.3546

RESEARCH ARTICLE

SMS | Strategic Management Journal

WILEY

The tangled webs we weave: Examining the effects of CEO deception on analyst recommendations

Steven J. Hyde[1] | Eric Bachura[2] | Jonathan Bundy[3] | Richard T. Gretz[4] | Wm. Gerard Sanders[5]

[1]College of Business and Economics, Boise State University, Boise, Idaho, USA

[2]Department of Business, University of Texas San Antonio, San Antonio, Texas, USA

[3]Department of Management, WP Carey School of Business, Arizona State University, Tempe, Arizona, USA

[4]College of Business, University of Texas San Antonio, San Antonio, Texas, USA

[5]College of Business, University of Nevada Las Vegas, Las Vegas, Nevada, USA

Correspondence

Steven J. Hyde, College of Business and Economics, Boise State University, 1910 W University Drive, Boise, ID 83725, USA.

Email: steven.hyde@utsa.edu

Abstract

Research Summary: Organizations are punished by analysts and investors when material deceit by their CEO is uncovered. However, few studies examine analysts' responses to deceptive CEOs before their deceit is publicly known. We use machine learning (ML) models to operationalize the likelihood of CEO deception as well as analysts' suspicion of CEO deception on earnings calls. Controlling for analysts' suspicion of deception, we show that analysts are prone to assigning superior recommendations to deceptive CEOs, particularly those deemed as All-Star analysts. We find that the benefits of CEO deception are lower for habitual deceivers, pointing to diminishing returns of deception. This study contributes to corporate governance research by enhancing our understanding of analysts' reactions to CEO deception prior to public exposure of any fraud or misconduct.

Managerial Summary: Undetected deception by CEOs can impact the stock market by influencing analysts' recommendations. Using an advanced ML model, our study measures the likelihood of deception more accurately than previous methods and identifies a tendency among financial analysts to favor deceptive CEOs, particularly high-status analysts. However, deception is less effective with analysts who are repeatedly exposed to deception. These findings underscore the importance of awareness of potential deception in CEO communications and the need for continuous scrutiny, learning, and adaptability among analysts.

KEYWORDS

agency theory, deception, governance, machine learning, upper echelons

1 | INTRODUCTION

Financial analysts, as information intermediaries, wield significant influence on the stock market (Boivie, Graffin, & Gentry, 2016; Brown et al., 2013; Graham et al., 2005; Womack, 1996). They provide earnings forecasts and recommendations to investors by analyzing the financial health and potential growth of firms. In doing so, they often rely on statements and conversations with CEOs, who may not always be truthful or forthcoming in their communications (Larcker & Zakolyukina, 2012; Soltani, 2014). This dynamic is evident in cases like Enron, WorldCom, Wells Fargo, and Theranos, where CEOs defrauded the market. Interestingly, analysts often accept CEOs' deceptive statements at face value rather than questioning them (for some notable examples, see Roberts, 2002; Wendel, 2014).

This study aims to assess the extent of analysts' susceptibility to CEO deception, thereby illuminating their role in firm governance. We develop a novel machine learning (ML) measure to detect deception, enhancing the reliability and precision of deception measurement and study in real-world contexts. This method addresses the ecological validity criticisms (O'Sullivan et al., 2009) often faced in deception research, and provides a practical tool for testing deception theories, such as the truth–default theory, in real-world scenarios.

Our findings suggest that analysts are indeed vulnerable to CEO deception. We posit that this susceptibility is due to analysts' natural tendency to bias toward truth, or rather, subconsciously presume that others are generally honest in their communications—a bias that is shared by most people (Levine, 2014). Despite analysts' duty to spot fraud, this truth bias can often lead them to overlook signs of managerial deception. Historical cases like the Enron scandal demonstrate such oversight, where analysts missed apparent red flags (Oppel, 2002). Consequently, we propose that analysts may unwittingly accept CEO deception, defaulting to a presumption of truth in their communications.

We also consider moderating conditions to this general susceptibility to deception. Specifically, we argue and show that there are diminishing returns to deception (Bolino et al., 2014). That is, a CEO's repeated use of deception weakens analysts' susceptibility to deception. Over time, analysts can compile enough evidence of deception to question their default assumption of truth, particularly as deception often does not result in the benefits that were promised. In other words, there is a learning effect to the truth–default bias, and analysts' motivation to be accurate will eventually overcome their tendency to assume the truth as they are repeatedly exposed to deception.

Furthermore, we argue that high-reputation analysts—or those hailed as the best by investors and other analysts—are more susceptible to deception from CEOs, and particularly from those who habitually use deception. As analysts' reputation grows, they become increasingly confident in their abilities and opinions. Consequently, their biases become more ingrained, resulting in a stronger truth bias. They may be less critical of CEOs in general, falling prey to hubristic tendencies or a belief in their own hype, which can reduce their level of scrutiny (Lovelace et al., 2018). Thus, analysts with the best reputations are the most likely to reward CEOs who are deceptive.

This study contributes to our understanding of deception and the role of analysts in several ways. First, we contribute to the literature by developing a novel ML approach to detect deception in the context of CEOs in conference calls. We have made versions of this model available on GitHub.[1] Our approach allows for a systematic and objective examination of analysts' susceptibility to deception. Prior studies in this context have obtained an accuracy rate of ∼65% using linguistic-based measures of deception detection (Larcker & Zakolyukina, 2012). The accuracy of our ML approach in predicting deception in our holdout sample is 84.2%. Our reproducible approach offers scholars a more accurate and efficient way to measure and test outcomes associated with deception.

Next, our study highlights the inherent limitations of analysts, whose innate truth bias can compromise their evaluation of CEO statements. Our study suggests that analysts can be manipulated more easily and cheaply than previously thought. There have been numerous studies showing that analysts may be influenced by firms. Westphal and Clement (2008), for example, find that CEOs "co-opt" analysts through personal and professional favors as well as increased access. However, our results suggest that such actions may not necessarily be the main factor at play. Humans have a natural tendency to default to the belief that others are telling the truth. This truth bias can interfere with analysts' ability to provide accurate and reliable information to investors, potentially leading to biased or inaccurate recommendations.

Finally, we contribute to corporate governance literature by introducing a nuanced understanding of deception. Prior work finds executives are penalized for deceit upon discovery (Arthaud-Day et al., 2006; Collins et al., 2009; Hennes et al., 2008; Persons, 2006). However, our study shows that analysts tend to assume truth in their communications, rewarding firms with positive recommendations rather than punishing, questioning, or doubting them. Our study demonstrates the temporal aspect of this bias, as analysts become aware of the CEO's deceit over time and stop rewarding dishonest CEOs. However, we also find that reputation can magnify the truth bias, with All-Star analysts being slower to catch on to deception. This implies that industry classification of All-Star analysts may need reconsideration, and investors should be cautious of relying on All-Stars versus Non-All-Stars when there is suspicious CEO behavior. Overall, our study highlights the need for better monitoring and accountability in corporate governance.

2 | LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1 | Deception

Deception is a broad category that includes many different behaviors, and not all are relevant to this study. For example, deceptive practices that are considered trivial or even legitimate, such as "white lies," or small exaggerations, are outside the bounds of this article (Bansal & Clelland, 2004; Graham et al., 2005). Rather, we focus on deception that is used to intentionally

[1] https://github.com/ebachUTSA/DeceptionClassifier.

mislead stakeholders, or in other words, language that is indicative of fraud and malintent. Our view of deception as purposeful and malevolent is consistent with prior research (Craig et al., 2013; Hobson et al., 2012). As Larcker and Zakolyukina stated, "We assume that these executives either intentionally manipulated the financial reports or they knew that investors were being provided with false accounting information during the conference call" (Larcker & Zakolyukina, 2012, p. 534). Hence, deception includes both sins of omission—intentionally leaving out critical information—and sins of commission—in which active mistruths are being told. However, being vague or dodging a question is not the same as purposefully withholding vital information or telling an outright lie.

We also distinguish deception from other forms of communication like impression or perception management (Busenbark et al., 2017; Westphal & Clement, 2008). These tactics aim to shape stakeholder perceptions favorably, without necessarily intending to deceive or mislead investors. They may border on deception, much like a politician spinning a policy favorably, but are generally seen as legitimate due to their nonegregious nature (Clatworthy & Jones, 2006; Merkl-Davies et al., 2011). Our focus, however, is on deception regarded as dishonest, egregious, and often illegal. It is noteworthy that executive deception is not always as deliberate or high-stakes as in infamous cases like Wells Fargo or Enron. Research shows that executive deception is more widespread than just the severe cases exemplified above and is still damaging. Indeed, prior research supports the prevalence of deception in public CEO communication (Hobson et al., 2012; Larcker & Zakolyukina, 2012), suggesting our theory and findings apply to all illegitimate or serious forms of deception, not limited to sensational fraud cases.

Finally, we assume that people tend to deceive for a reason. According to Levine et al. (2010), people are more likely to lie when their goals are not consistent with reality. When goals are consistent with truth, people will almost always be honest (Levine et al., 2010). Related to the context at hand, this suggests that CEOs use deception to garner positive evaluations and/or avoid negative evaluations when their goals are inconsistent with reality (Grover, 2005; Westphal & Clement, 2008; Zahra et al., 2005). As such, we assume that CEOs use deception to exaggerate or mislead analysts in a positive direction, as they seek to meet or beat analysts' expectations and advance positive recommendations for their firms' securities.

3 | DIFFICULTY OF DECEPTION DETECTION—TRUTH-BY-DEFAULT

With an accuracy just above chance, people generally perform poorly in detecting deception (Bond & DePaulo, 2006). This is particularly evident in distinguishing deceptive statements, where accuracy hovers around 47%, compared to 61% for honest ones (Bond & DePaulo, 2006). According to Truth-by-Default-Theory (TDT), this is because people either (1) fail to consider the possibility of deceit or (2) fail to obtain enough evidence to confirm the suspicion of deception (Levine, 2014). This results in a general assumption or bias that most messages are truthful. Studies have shown this truth–default bias in many different contexts, including politics, news, entertainment, education, and across cultures (Clementson, 2018; Levine et al., 2021; Park & Levine, 2017).

Thus, the question arises: should we expect this same truth–default bias among informed analysts? Their career progression relies heavily on the accuracy of their information and credibility (Gleason & Lee, 2003; Groysberg et al., 2011; Hong & Kubik, 2003). Hence, they are incentivized to uncover deception. Furthermore, their comprehensive knowledge of firms and industries, frequent interactions with CEOs, and access to data and advanced technology suggest that they should be better equipped to identify deception than the average person (Blair et al., 2010; Levine et al., 2014; Mann et al., 2004).

Yet, despite analysts' motivation and capabilities, they likely remain susceptible to truth–default bias. Indeed, studies suggest that analysts may struggle to accurately detect CEO deception (Brown et al., 2015; Hartwig et al., 2006). For example, Brown et al. (2015) found that analysts often spend minimal time verifying the accuracy of a firm's earnings due to the challenging nature of uncovering deception, and thus they tend to accept the data they receive as truthful. Given the low base rate of fraud in public firms (Dyck et al., 2013), analysts might find it more practical to presume honesty, and this bias toward truth is generally validated.

There are also significant repercussions for wrongly accusing a CEO of deception, including loss of access to privileged information (Westphal & Clement, 2008), which may far outweigh the consequences of being deceived. Furthermore, analysts face conformity pressures (Zhu & Westphal, 2011), making them less likely to deviate from consensus opinions, even when deception is suspected. This makes analysts less likely to risk their reputations and relationships without substantial evidence of deception.

We argue that the challenges of detecting deception, the low fraud prevalence, and the high cost of false deception detection (Boivie et al., 2016; Westphal & Clement, 2008; Womack, 1996) reinforce the truth–default bias among analysts. This makes it likely that deception will go undetected and that deceptive content will influence their recommendations. Given the benefits of positive evaluations and the associated risks of negative ones, we argue that analysts are ideal targets for CEO deception and are likely to be manipulated into providing positive recommendations.

Hypothesis 1. *A CEO's use of deception will be positively related to analysts' subsequent recommendations.*

So far, we have focused on the truth–default bias, which we cannot directly measure. However, we can examine the mechanism indirectly through two key moderators: history of deception and reputation. Below we argue that the more someone has used deception in the past, the less likely it will be influential—weakening the bias toward truth. Conversely, we propose that reputation should strengthen one's bias since it increases access to executives and hubristic thinking.

4 | THE DIMINISHING RETURNS OF DECEPTION

It is important to remember that analyst and CEO interactions extend beyond a phone call. Analysts typically follow a small group of firms and executives throughout their careers (Brochet et al., 2013; O'Brien & Bhushan, 1990). For this reason, it is critical to consider how an analyst's susceptibility to deception from a CEO is impacted by the temporal component of their relationship.

Research suggests that repeated deception results in distrust and damage to credibility (Farber, 2005; Grover, 2005; O'Sullivan, 2003), potentially leading analysts to intensify their scrutiny and limit positive endorsements. While initial deception might sway analysts due to striking claims and conformity pressures, increased suspicion may elevate suspicion (Buller & Burgoon, 1996) and analysts may withhold additional positive recommendations as they become more uncertain of the credibility of the CEO's statement. Essentially, the marginal returns of deception diminish over time (Anderson & Smith, 2013; Bolino et al., 2014). For example, if a CEO's deceitful claims persuade analysts to upgrade their recommendations from Sell to Hold, but the firm's performance does not improve, the analysts may be unable to justify additional upgrades.[2] Over time, the gap between the CEO's deceptive promises and the reality of the firm's outcomes might expose the CEO's unreliability and lessen the impact of future deception.

Considering the above arguments, we suggest that deception has diminishing returns. An initially honest CEO could elicit a positive response when resorting to deception. However, a history of dishonesty may exhaust deception's benefits as analysts stop justifying continuous positive evaluations—particularly when reality contradicts the CEO's message (Bolino et al., 2014). Recurring deception could eventually lead to a "trigger event" as defined by Levine (2019)—a moment where indications of deception override the truth–default, prompting analysts to abandon their initial trust. Thus, a CEO's previous deceit might catch up with them as analysts become more aware and potentially expose their deception. We propose the following:

> **Hypothesis 2.** *The positive relationship between a CEO's use of deception and analysts' recommendations will be negatively moderated by their history of deception. Prior use of deception will weaken the positive effects of subsequent deception.*

5 | HIGH REPUTATION INCREASES SUSCEPTIBILITY TO DECEPTION

Institutional Investor's annual "All-Star" analysts list recognizes those analysts with exceptional reputations in their respective industries (Rindova et al., 2005; Stickel, 1992). These All-Stars, known for their market influence, experience greater benefits including working at larger firms, have higher compensation, and have privileged access to firms and executives (Boivie et al., 2016; Gleason & Lee, 2003; Groysberg et al., 2011; Paik et al., 2022; Soltes, 2014). However, their high standing may paradoxically make them more susceptible to deception.

As they gain stature and prestige in their social networks, individuals with a high reputation are prone to increased hubris and a feeling of confidence and entitlement (Hayward et al., 2004; Lovelace et al., 2018). In other words, they start to "believe their own press" (Hayward et al., 2004, p. 645). Such an effect is often accompanied by an increased sense of authority, additional deference from others, a reduced willingness to deliberate and consider alternatives in decision making, and an increased sense of trust in others (Brandt et al., 2015; Lovelace et al., 2018; Paik et al., 2022).

As such, All-Star analysts may be more confident that CEOs will not deceive them and thus more likely to take their words at face value. That is, they are more likely to default to truth. Additionally, scholars have demonstrated that relationship closeness between a message sender and receiver increases truth bias (Levine, 2019). McCornack and Parks (1986), for example, argued that as individuals become closer, they tend to grow in confidence that they can detect each other's deception—further increasing their truth bias and lowering their accuracy.

[2]We recognize that analyst recommendations have an upper and lower boundary (i.e., Strong Sell and Strong Buy). We will be sensitive to this possibility as an alternative explanation in our empirical analyses as follows.

All-Stars' increased access to CEOs will likely further bolster their truth bias making them more vulnerable to the effects of deception. Hence, we posit:

> **Hypothesis 3.** *The positive relationship between a CEO's use of deception and analysts' recommendations will be positively moderated by the reputation of the analyst. That is, All-Star analysts will give better recommendations to deceptive CEOs than Non-All-Star analysts.*

We have argued that the truth-by-default phenomena can be weakened by prior deception and strengthened by analyst reputation. Combining this logic, we suggest that the weakening effect of prior deception on truth bias will be mitigated as analyst reputation increases. High-reputation analysts may be more likely to trust executives and exhibit hubristic thinking, which can lead to an increased truth bias. This can lead to an individual's "blindness" to the actions and incentives of others, including their behaviors and motivations (Conniff, 2005). This decreased awareness limits high-reputation analysts' ability to detect prior deception over time. As a result, All-Star analysts are less likely than Non-All-Star analysts to have a triggering event, and may be slower to learn from historical deception. Given this, All-Star analysts will be more susceptible to deception when the CEO has a greater history of deception, compared to their Non-All-Star counterparts.

> **Hypothesis 4.** *The negative moderation of a CEO's history of deception on the relationship between deception and analysts' recommendations is weakened by analyst reputation. That is, All-Star analysts will give better recommendations to deceptive CEOs who have a greater history of being deceptive, compared to Non-All-Star analysts.*

5.1 | Methodology

Our study focuses on CEO deception during quarterly earnings calls. We assume that these executives are acutely aware of the financial health of the firm and have a financial incentive to obtain the highest recommendations possible (Larcker & Zakolyukina, 2012). We assembled a CEO–analyst dyad sample from S&P 1500 firms' call transcripts from 2008 to 2016, obtained from SeekingAlpha.com. Each participant and their respective role are identified each time they speak, allowing us to discern who attended each call and separate the transcripts by speaker. Following Larcker and Zakolyukina (2012), who found no difference in deceptive cues between managerial discussions and Q&A sections, we included the entire transcript in our analysis for a comprehensive portrayal of the analysts' experience.

For consistency and validity, we used a 150-word minimum CEO speech threshold and excluded calls with dual CEOs (Hope & Wang, 2018; Larcker & Zakolyukina, 2012). We integrated data from COMPUSTAT, EXECUCOMP, IBES, *Institutional Investor*, and SEC for financial, executive, and analyst metrics. Additionally, we used the Miami University Deception Detection database (MU3D) to control for analysts' suspicion of CEO deception (Lloyd et al., 2019).

All transcripts used in the development of our deception measure are not included in our final analysis. This ensures that the results are not an artifact from the ML process or driven by publicized deceit. Our final sample comprised 64,107 dyadic observations of analysts and CEOs.

5.2 | Dependent variable

5.2.1 | Analyst recommendations

We used *analyst recommendations* as our dependent variable, scoring them from 1 (Strong Sell) to 5 (Strong Buy; König et al., 2017). To minimize the impact of extraneous factors, we collected these recommendations 3 days after each conference call, a window where the bulk of revisions typically occur (Yezegel, 2015). We considered a recommendation inactive and hence removed it from our model if it had not been updated within the preceding 24 months, the maximum period an analyst typically deems a recommendation active.[3]

5.3 | Independent and moderating variables

5.3.1 | Deception

To measure deception, we leveraged literature spanning two decades, showcasing consistent links between specific language patterns and intentional deception (e.g., increased negative emotions and decreased references to self; Newman et al., 2003; Pennebaker, 2011). These patterns are subconscious and challenging to manipulate or detect without computer assistance, rendering them reliable markers of deception (Alaskar et al., 2022; Newman et al., 2003; Pennebaker, 2011). Furthermore, the linguistic patterns of deception have been consistently identified across multiple disciplines such as psychology, linguistics, communications, law, criminal justice, computer science, accounting, and finance, underscoring its reliability for measurement (Arciuli et al., 2010; Burgoon et al., 2016; Fornaciari & Poesio, 2013; Ho et al., 2016; Larcker & Zakolyukina, 2012; Ten Brinke & Porter, 2012; Toma & Hancock, 2012). However, traditional text-based deception detection has limitations, with meta-analysis revealing only 62%–65% accuracy (Hauch et al., 2015). While there has been skepticism about the effectiveness of these methods, our independent work has achieved an accuracy of approximately 80%, similar to or better than recent advances in ML as reported by Alaskar et al. (2022). This underscores the robustness of our own ML-based deception measure, which was developed and validated prior to these findings.

To develop a linguistic deception measure, we develop a Deep Neural Net (DNN)[4] classifier using an Adam solving function (Abadi et al., 2016; Maas et al., 2013). We accomplished this through three steps: (1) creating a sample of text from CEOs known to be deceptive, (2) training an algorithm to identify critical patterns of speech in the sample, and (3) applying the algorithm to all other transcripts related to the study. Importantly, our model's development, testing, and application were conducted on independent samples. To clarify, the algorithm was constructed with a training sample, verified with a separate holdout sample, and ultimately used to score CEOs on a third distinct sample to test our hypotheses. The independence of these samples ensures the robustness of our model.

[3]For a robustness check, we calculated the change in recommendations—the analyst's postconference call recommendation minus the precall recommendation. The results align with our original model.

[4]We chose a DNN classification model because it tends to outperform other ML approaches with data of this kind (Moraes et al., 2013).

For our training and validation sample, we first identified cases from 2008 to 2015 where the SEC issued an Accounting and Auditing Enforcement Release (AAER) due to irregularities within a firm's financial statements and were subject to litigation. AAERs are a common proxy for fraud since they are used to punish fraudulent behavior or reckless neglect of accounting principles (Hennes et al., 2008; Koch-Bayram & Wernicke, 2018). Consistent with prior research (Larcker & Zakolyukina, 2012), we used the description of the misconduct within the AAER to narrow our sample to only those described as fraudulent (e.g., material restatement, fraud, and falsification).[5] Hence, we can infer a "sample of firms that intentionally violated the reporting requirements … and directly align the misstating to the respective CEO" (Koch-Bayram & Wernicke, 2018, p. 2948). Given that our sample occurs post-Sarbanes–Oxley, and that CEOs are now required to certify financial results, we believe that this inference of intentionality is realistic. Thus, the conference calls that occurred during the firm's period of fraud were labeled deceptive. For example, Monsanto was fined 80 million dollars for materially misstating their earnings from 2009 to 2011. The conference calls corresponding with these misstated earning reports were labeled deceptive.

Next, we created a matched sample[6] of cases where there was no evidence of detected deception by the SEC. We selected matched firms with similar aged and tenured CEOs, firm size (firm sales), financial performance metrics,[7] and industry (4-digit SIC code). To optimize learning rate and accuracy, we oversampled the minority (deceptive) cases to balance the dataset and preserve matched sample equivalencies (i.e., 50% of cases were deceptive and the other 50% were not; Chawla et al., 2002). The final training dataset consisted of 19,654 transcripts by 3721 CEOs, which included 48,296,007 spoken words by CEOs.

In building our ML algorithm, we adhered to established research methodology, using 80% of our training sample to train the algorithm and the remaining 20% for accuracy evaluation (Harrison et al., 2019).[8,9] For our input layer, we selected 22 linguistic features linked to deception, thereby capturing the cognitive and emotional variations that occur during deceptive discourse.[10] This allowed us to capture differences in the call's content and in the CEO's affect and cognition that are a result of deception. Deception is cognitively demanding and often emotional (Ekman & Friesen, 1969; Newman et al., 2003). We therefore included word groups such as analytic, clout, words larger than six letters, negations, cognitive process, and positive and negative emotions (i.e., anxiety, anger, and sadness). Deceivers typically use fewer

[5]These terms are commonly used to describe violations of the Securities Act of 1933 section 17 or Securities Act of 1934 section 10.

[6]Our sampling approach is designed to capture both within- and between-person variance. We matched fraudulent quarters with nonfraudulent quarters with the same CEO as well as their equivalent peers. This is done to ensure our model is exposed to linguistic differences between as well as within speakers.

[7]All financial information used within this study was based on the information that was originally reported by the firm and not any revised financial metrics.

[8]Minority class oversampling was separated from holdout to preserve evaluations against unseen data.

[9]We used 5K-Fold hyperparameter comparison over the initial 80% holdout to test alternative hyperparameter configurations. Our initial configuration of two hidden layers of equal input space node count was the best performer across the training split with minimal variance in each fold's model instance performance. Each fold iteration split the initial 80% into a separate 80/20 split for the fold training/eval. After hyperparameter verification, a model was trained on the initial 80% training split with the confirmed hyperparameter configuration and then evaluated on the 20% holdout (that was never seen during the K-Fold iterations for hyperparameter performance examination).

[10]Our ML model is developed differently from the standard data-driven approach, where algorithms start with all features and trim them based on accuracy shifts. We are cautious of this approach's potential to threaten validity and generalizability by highlighting idiosyncratic linguistic relationships. Hence, we prune features in alignment with the existing literature prior to training the model (Hyde et al., 2023).

self-references, are inconsistent with verb tenses, and use fewer sensory related words. Hence the inclusion of categories like personal pronouns, I-pronouns, motion, space, time, and tense-focused words (i.e., focus past, present, and future; Hauch et al., 2015; Newman et al., 2003). We also incorporated non-fluences (e.g., "er", "hmm") and filler words (e.g., "I mean", "you know") to reflect speech hesitations (Arciuli et al., 2010; Larcker & Zakolyukina, 2012), and risk-related words to capture the inherent risks associated with engaging in deception during conference calls. Lastly, we included Newman et al. (2003) original deception equation, called authentic within LIWC2015.

Next, we processed these features through two hidden layers with 22 nodes each, and early stopping enabled via a tolerance value of 0.0001. This establishes the model training process by defining a stopping point to be when performance does not improve at the tolerance threshold within 10 iterations. The resulting model is trained on data that allows it to detect the distinct language patterns of CEOs who engage in deception. That is, the language differences between CEOs who committed financial fraud and those that did not.

For our final step, we assessed the model's accuracy using the remaining 20% of the sample to predict fraud likelihood, indicated by the SEC issuing an AAER. The model converged with an Area Under the Curve (AUC) score of 84.20%, which reflects the model's true positive rate against its false positive rate. An AUC score above 80 percent is considered "good accuracy" (Gallop et al., 2003). The overall accuracy of the model was 84.18%, meaning the model correctly identified transcripts of firms that were later charged with fraud with 84.18% accuracy. These scores indicate that the classification model achieved a false positive and true positive rate much better than chance, or rather that the algorithm detects deception with a high degree of accuracy. Moreover, our model is approximately 22 percentage points (or 29.5%) more accurate in detecting deception than traditional text-based approaches (Larcker & Zakolyukina, 2012; Newman et al., 2003). In sum, the algorithm reliably detects deception based on the distinct language patterns of CEOs (see Appendix A for a selection of quotes our model identified as deceptive or honest).[11]

The final output from the ML algorithm is our key independent variable labeled *deception*. This is a percentage score from 0 to 1 that indicates the model's confidence in detecting deceptive CEOs during earnings calls, based on the similarity of the speaker's linguistic patterns to those of known fraudulent individuals. Lower scores imply honesty, while higher ones suggest potential deception. Unlike most deception research, which dichotomously operationalizes deception, our approach uses a continuous approximation due to the lack of certainty inherent in nonexperimental detection. Moreover, using a continuous approximation will enable us to better identify the impact of deception, as opposed to a dichotomous one. That said, for consistency with extant literature, our interpretation of the impact of *deception* reflects the dichotomous standard (i.e., deceptive CEO vs. honest CEO). We applied this algorithm to the CEO-related text on earnings calls between 2008 and 2016. It is important to reiterate that none of the transcripts used to train or evaluate the algorithm were used in our hypotheses testing estimations.

5.3.2 | History of deception

We measured the *history of deception* through a running average of the CEO's *deception* score throughout their tenure. The measure reflects their average *deception* score prior to the

[11]Versions of this model are publicly available at the second author's academic GitHub: https://github.com/ebachUTSA/DeceptionClassifier.

conference call of interest, ranging between 0 and 1. As a robustness check, we operationalized this measure as the running average of the CEO's *deception* score while the analyst has been following the firm. The results remained consistent regardless of which operationalization we used.

5.3.3 | All-Star analyst

Institutional Investor conducts an annual worldwide survey of money managers to assess the best analysts in each region (i.e., America, Europe, Asia, and Latin America). *Institutional Investor* then publishes their list of top analysts by region. We coded the variable *All-Star analyst* as 1 for analysts who were recognized by the magazine as a top analyst in a given year, regardless of region. Otherwise, we coded the variable as 0.

5.4 | Validity checks of the measure of deception

To validate our measure, we assessed the accuracy and AUC of our ML model in various samples and contexts. First, we tested our model's ability to identify fraudulent peer-reviewed articles using Retraction Watch, a database that tracks journal retractions. We restricted our sample to articles available on PubMed, ensuring that the context was distinct enough from our training context (financial fraud vs. academic fraud) to minimize alternative explanations for our model's performance. We begin with a sample of 3602 PubMed articles that were retracted, of which 112 were labeled by Retraction Watch as fraudulent ("Error in Data," "Error in Methods," "Investigation by Journal/Publisher," "Manipulation of Results," "Unreliable Results") and investigated by the journal/publisher, confirming data/result manipulation by authors. To create a sample of honest articles for comparison, we included all PubMED articles with a PubMED ID that is equal to 1 plus the PubMED ID of the retracted articles. In total, we collected 3602 honest articles using this method. We randomly selected 112 honest articles to construct a balanced sample of 224 articles. We repeated this process 1000 times, sampling with replacement from the 3602 honest cases, and tested our algorithm's performance on each sample. Our model achieved an average accuracy of 80.47% and an average AUC of 93.75%, indicating that our algorithm can effectively detect deception in peer-reviewed articles.[12]

Second, we assessed our model with a convenience sample of YouTube videos of known examples of deception. This sample includes videos of famous business executives, athletes, politicians, and criminals lying in television interviews as well as deception within a high-stakes game show (e.g., Lance Armstrong, Elizabeth Holmes, the game show Jubilee). In total, our sample contained 62 honest and 23 deceptive statements for a total of 85 statements. Our model was able to detect deception in this sample with an accuracy of 73% and an AUC of 80%. This AUC is similar to the AUC of 84.20% from our original hold-out data.

Third, we examine *deception*'s association with relevant factors that motivate deception, such as *AAER violations* that occurred after our initial data collection, *data breaches*, and *consumer protection fines*. Additionally, we explored the correlation between *deception* and CEO

[12]The standard deviation of AUC and accuracy over the 1000 samples is 1.6% and 1.6%, respectively. Also, the minimum AUC and accuracy over the 1000 samples is 87.9% and 75.9%, respectively.

gender, given that males are more likely to be deceptive (Dreber & Johannesson, 2008; Gupta et al., 2020). We ran this validity check using annual data on S&P 500 firms from 2008 to 2016.

We obtained an indicator variable for AAER violations for a firm in the focal year (*AAER violation* = 1) that came to light after our initial data collection through the same process outlined above. *Consumer protection fines* were collected through Violation Tracker (goodjobsfirst.org) and are measured as the natural logarithm of the number of fines for a firm in the focal year. We obtained data breach information from Privacy Clearinghouse with a dummy variable that indicates if the firm experiences a data breach in the focal year (*data breaches* = 1). Finally, we measured gender using a dummy variable (*Female CEO* = 1) from ExecuComp. We assessed relationships via OLS estimation of *deception* on *AAER violations*, *consumer protection fines*, *data breaches*, *Female CEO*, standard controls, and random firm effects, with standard errors robust to firm-level clustering. The results provided further evidence of the validity of our measure (*AAER violations*: $\beta = .092$, $p < .001$; *data breaches*: $\beta = .032$, $p = .063$; *consumer protection fines*: $\beta = .002$, $p = .012$; *Female CEO*: $\beta = -0.057$, $p = .033$).

Finally, we note that the within- and between-CEO standard deviations for our measure of deception are notably similar (0.30 and 0.27, respectively). This is important because it suggests that our measure of deception captures deception itself rather than a disposition toward a particular language pattern. If our measure was predominantly capturing a consistent language disposition, we would expect to see a much tighter standard deviation within subjects. This parity in deviations, therefore, reinforces the validity of our deception measure, demonstrating its robustness in identifying genuine instances of deceptive speech.

In sum, our validity checks give us high confidence that the *deception* measure from our ML model is appropriate to test hypotheses concerning the influence of CEO deception on analysts' recommendations; results are unlikely to be unduly influenced by spurious relationships in our context. Additionally, our results suggest that our algorithm may identify firms that deceive but were never formally detected. While our model excels in predicting officially recognized instances of deception such as AAER violations, data breaches, and consumer-related fines, it may also be accurately capturing deception in undetected instances. Although we cannot explicitly test this, the model's high degree of accuracy suggests that it is capable of spotting linguistic patterns indicative of such covert cases of deception.

5.5 | Control variables

We controlled for executive characteristics that could influence risk-seeking behavior and analyst perceptions, including *CEO age* and *CEO tenure*, *CEO compensation* (logarithmically adjusted), and *CEO awards* measured by total awards won from the following outlets *Barrons*, *Business Week*, *Chief Executive*, *Electronic Business*, *Financial World*, *Forbes*, *Fortune*, *Industry Week*, *Institutional Investor*, *Morning Star*, *Time* and *CNN* (Hambrick & Mason, 1984; Harris & Bromiley, 2007; Li et al., 2022).

We added controls for several elements of the CEO's language during earnings calls that could affect analysts' recommendations. We controlled for *CEO Positive Tone* through the ratio of positive affective words to total affective words used by the CEO (Mayew & Venkatachalam, 2012). We also controlled for *CEO concrete language*, *CEO tentative language*, and *CEO charisma*. Following extant literature (Pan et al., 2018), we measured concrete language using a composite score of standardized LIWC2015 categories (concrete language = verbs + numbers + past tense − adjectives − nonspecific quantifiers − future tense words). *CEO tentative language* is the proportion of words categorized as tentative in LIWC2015, which indicates the level of uncertainty or hesitation. *CEO charisma* was captured using the charismatic vision statement dictionary developed by Fanelli et al. (2008). Analysts may be similarly influenced by the confidence of the CEO (Zyung & Shi, 2022). For this reason, we include *CEO clout* measured by the portion of words that relate to clout as indicated by LIWC2015.

We developed a measure for *CEO obfuscation* to account for instances where a CEO may mislead analysts by avoiding their questions rather than using deception. Our measure is adapted from Lee (2016) measure of scripting, which calculates the linguistic similarity between the CEO's prepared remarks and their answers to questions using cosine similarity. To determine the degree of obfuscation, we calculate the linguistic similarity between the analyst's question and the CEO's response. This gives us the CEO's cosine similarity score in the focal call. Then, we compare this score to the CEO's average cosine similarity score across their tenure. The difference between the two scores reveals whether the CEO is obfuscating more or less than usual. Alternatively, a CEO may obfuscate by limiting their contribution during the call by speaking as little as possible. For this reason, we controlled for the natural log of the CEO's word count on the call (*CEO word count*).

We also use the MU3D dataset[13] (Lloyd et al., 2019), encompassing 320 videos of honest and deceitful statements. The dataset incorporated responses from 405 mTurk participants who had been incentivized to correctly identify the deceptive videos. These responses helped construct an ML model for suspiciousness. The model used similar input features to our deception measure, with an added factor for tentative language. After a training/holdout split of 80/20, the model attained an accuracy of 82.8% and an AUC of 81.6%. To ensure that the MU3D data could generalize to our transcripts, we performed a modified adversarial validation which showed that there were no notable linguistic differences between the two datasets. We then applied the suspiciousness model to our dataset to create *CEO suspiciousness*, ranging between 0 and 1, which captures analysts' suspicion of CEO deception, independent of actual deception by the CEO. To be as thorough as possible we also obtain *CEO history of suspiciousness* with a running average of *CEO suspiciousness* throughout their tenure. The measure reflects their average suspiciousness score prior to the focal conference call.

5.5.1 | Analyst characteristics

We controlled for several analyst characteristics that reflect their experience and knowledge. We include the count of the *number of prior All-Star recognitions* by *Institutional Investor* before the current conference call. This controls for the impact that past experience with recognition as a top analyst might have on recommendations. It also ensures the *All-Star analysts* variable reflects the impact of top analyst recognition in the current year holding the effect of any prior recognition constant. Additionally, we used several measures to control for their experience through their overall *analyst tenure* and the number of *regulatory exams* they have passed. Since

[13]The effectiveness of our deception detection algorithm is constrained by corpus size, needing at least 150 words. Most MU3D cases did not meet this criteria. Still, when tested against nine sufficiently large cases, the algorithm achieved 79% AUC and 78% accuracy, surprisingly showing generalizability even to low-stakes deception scenarios found in the MU3D data. However, we remain skeptical of this finding given the small sample.

analysts vary in how frequently they change their recommendation, we controlled for the number of *days since* (*the*) *last recommendation* provided by the analyst. This variable was *z*-scored for ease of interpretation. Given that the analyst's prior recommendation impacts their current recommendation, we controlled for the analyst's most recent recommendation for the firm prior to the conference call, *prior recommendation*, with dummy coding (1–5) similar to the dependent variable.

Since research indicates that analysts impact each other's recommendations, we also included analyst controls at the group level (Clement & Tse, 2005). First, we controlled for the *change in median recommendation* by subtracting the group's median recommendation before the call from the median recommendation following the call. Next, we controlled for the dispersion of recommendations through the *standard deviation of recommendations* in the group prior to the conference call. We controlled for the *number of analysts* following the firm through the number of recommendations provided following the conference call.

We controlled for the *forecast accuracy* (FA) of the analysts by taking the absolute value of the difference between analysts' consensus forecast of earnings per share (F) by the firm's actual earnings per share (A), divided by the actual earnings per share (Ertimur, Sunder, & Sunder, 2007), as seen in the equation below.

$$FA = \left| \frac{F - A}{A} \right|$$

5.5.2 | Firm characteristics

We controlled for *firm size* through the natural log of the total number of employees. We controlled for performance through return on equity, or *ROE*, and *earnings surprise*. We measure *earnings surprise* with a dummy variable, where firms that meet or exceed analysts' consensus expectations were coded 1, 0 otherwise. We also include *year dummies* and *month dummies*, which are month- and year-specific indicators to address seasonal and yearly shocks common to all CEO–analyst dyads. We note that we winsorize *compensation*, *number of analysts*, *forecast accuracy*, and *ROE* at the 99th and 1st percentile to address outliers in these variables.

5.6 | Endogeneity issues

We account for time-invariant factors, including CEO, analyst, and CEO–analyst relationship characteristics, with CEO–analyst dyad fixed effects. The Hausman test ($\chi^2 = 469.31$ (47), $p < .001$) suggests fixed effects are preferred to a random-effects approach. Despite fixed effects controlling for certain time-invariant attributes (e.g., analyst gender), *deception* remains potentially endogenous. For example, if *deception* is a result of an unobserved time-varying attribute, such as a negative shock that may influence *analyst recommendations*, our estimates will be biased. In response to this, we use a control function approach (Petrin & Train, 2010), which necessitates valid (i.e., uncorrelated with the error term in the second stage) and relevant instruments (i.e., related to *deception*). In our context, the validity requirement means that the impact of the instruments on *analyst recommendations* is completely mediated by the instruments' effect on *deception*. We regress the endogenous variable on instruments and the exogenous independent variables then include the residuals from this regression (*control function*) in the second-stage estimation. This approach ensures consistent second-stage estimates by directly controlling for the endogenous component of the problematic variable (Wooldridge, 2015).

We use two instruments in the first stage that are related to *deception*, but otherwise not related to *analyst recommendations*. To use deception in such a high-stakes setting, the individual would need to feel somewhat confident in their ability to avoid detection. Otherwise, they would not attempt this risky strategy. Prior research demonstrates that leaders who feel confident are more likely to use *reward words* (Akstinaite et al., 2020). Additionally, deceitful individuals will also seek to distance themselves from their dishonest statements (Newman et al., 2003). CEOs are responsible for everything the firm does. As a result, they will seek to distance themselves from responsibility or power over others when deceiving, thus likely avoiding words included in the LIWC power dictionary. This suggests that *deception* should be negatively associated with *power words* and positively related to *reward words*.

Negative shocks should not directly relate to changes in the use of these word categories. First, *power words* are not typically associated with negative shocks (Scheuerlein & Chládková, 2019). Second, the use of *reward words* varies based on a CEO's confidence in overcoming a crisis (Akstinaite et al., 2020), hence the relationship will be positive for some and negative for others. Consequently, *reward words* and *power words* are apt instruments as they lack systematic links to negative shocks. We provide additional evidence of validity below.

5.7 | Method of analysis

Our dependent variable is an ordinal categorical variable. Scholars have predominantly used a linear fixed-effect panel estimation when dealing with this kind of data since, under most circumstances, results are asymptotically unbiased (Riedl & Geishecker, 2014). However, a linear regression applied to an ordered dependent variable can produce distorted effect sizes especially when observed outcomes are skewed toward a few categories, which is the case for our dataset as shown in Table B1 in Appendix B (Wooldridge, 2010). For this reason, we test our hypotheses using an ordered probit regression that takes into account the ordinal nature of the dependent variable. Still, we show linear fixed-effects panel estimations in Appendix B as a robustness check, which are consistent with the ordered probit results.

We can obtain predicted probabilities of analyst recommendation categories (Strong Sell to Strong Buy) for any combination of covariate values with ordered probit estimates. This helps address difficulties in interpreting coefficients in nonlinear estimators and produces realistic effect sizes (Hoetker, 2007; Shaver, 2007). However, integrating fixed effects in an ordered probit can lead to the incidental-parameters problem, making maximum likelihood estimation biased and inconsistent (Greene, 2017). We address this by using the technique described in Wooldridge (2010) where we include panel averages of all independent variables as additional regressors. We note this approach reduces efficiency due to lost degrees of freedom. However, it controls for time-invariant attributes of the CEO–analyst dyad by allowing for correlation between the panel averages and the time-variant regressors, similar to linear fixed effects, with some reasonable restrictions on the distribution of panel-specific effects (see equation (15.67) in Wooldridge (2010)).

We bootstrap both stages to account for additional sampling error from incorporating the *control function* in the second-stage estimation (Wooldridge, 2010). We generate cluster robust standard errors with the bootstrap by sampling with replacement at the CEO–analyst dyad to address heteroscedasticity between panels and error correlation within panels.[14]

6 | DESCRIPTIVE STATISTICS AND RESULTS

Table 1 provides descriptive statistics and correlations; we note the variance inflation factors (VIFs) suggest that multicollinearity is not an issue.[15] Average *deception* is 31.0 percent, which suggests that CEOs are honest most of the time, although not always honest, as is commonly found in prior studies of deception (Levine, 2014).

6.1 | First-stage results

Table 2 shows the first-stage linear fixed-effects panel estimation and the second-stage ordered probit results for all models. Model 1 is the first-stage estimation for the full second-stage estimation in Model 6, which includes an interaction term between *history of deception* and *All-Star analyst*. *Reward words* positively (Table 2, Model 1, $\beta = .020$, $p < .001$) and *power words* negatively (Table 2, Model 1, $\beta = -.007$, $p = .009$) relate to deception, consistent with expectations. The first-stage F statistic of 21.88 exceeds the benchmark of 10, indicating compelling instrument strength (Wooldridge, 2010).

We test overidentification restrictions in Model 2 because we have more instruments than endogenous variables. The appropriate test statistic from Lung-Fei (1992) for an ordered probit estimate like ours is $\chi^2 = .089$ ($p = .765$), which provides additional evidence of instrument validity.

Also, the Durbin–Wu–Hausman test for endogeneity is given by the *control function* in the second-stage (Wooldridge, 2010). Failure to address the endogenous nature of *deception* (e.g., Table 2, Model 6, *control function*: $\beta = -2.178$, $p = .025$) results in biased and inconsistent estimates. Table B2, Model A1 in Appendix B shows the outcome of our main model without correcting for endogeneity. The findings align with Table 2, but the positive impact of *deception* on *analyst recommendations* is substantially smaller. This indicates a negative bias, where time-variant shocks that encourage *deception* (e.g., unexpected poor firm performance) are correlated with lower *analyst recommendations*. The impact of CEO deception is much larger after addressing the endogeneity issue.

6.2 | Hypothesis testing and magnitude of effects

Models 2–6 show results from the ordered probit estimations. Model 2 reports main effects without any interactions. Models 3–5 display interactions of *deception* with *history of*

[14]All estimations are produced with STATA/MP 15.1 for Windows (64-bit x86-64). We use STATA's "xtreg" command in the first stage, then the "oprobit" command in the second stage. Note that we use the "xtreg" command in the second stage for the linear fixed-effects panel estimation robustness checks shown in Appendix B. We run the entire procedure on 1000 bootstrap samples obtained by sampling with replacement over CEO-analyst dyads using STATA's "bootstrap" command. The code to run the procedure is available from the authors by request.

[15]The condition number is 3.033, which falls under the threshold of 20 suggesting our estimations are stable (Greene, 2017). Also, the average VIF score is 1.33, and the highest VIF score of 2.24, falling under the generally accepted threshold of 10 (Ryan, 1997).

TABLE 1 Descriptive statistics and correlations for all variables in the model.

Variables	Mean	SD	Correlations (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) Analyst recommendations	3.625	0.907	1.000												
(2) Deception	0.310	0.314	−0.010	1.000											
(3) History of deception	0.313	0.173	−0.015	0.542	1.000										
(4) All-Star analyst	0.098	0.298	−0.070	−0.002	0.011	1.000									
(5) ROE	0.014	0.021	0.063	−0.008	−0.013	−0.023	1.000								
(6) Firm size	2.218	1.708	−0.008	0.031	0.046	0.078	0.118	1.000							
(7) Earnings surprise	0.652	0.476	0.038	0.013	0.016	−0.001	0.167	0.070	1.000						
(8) CEO tenure	7.384	6.499	0.014	0.006	0.019	−0.036	0.034	−0.108	0.024	1.000					
(9) CEO age	56.250	6.271	−0.010	−0.044	−0.058	0.013	−0.013	0.094	−0.037	0.366	1.000				
(10) Compensation	8.843	0.802	0.020	0.021	0.037	0.094	0.036	0.518	0.094	−0.041	0.101	1.000			
(11) CEO awards	0.456	1.395	0.032	0.012	0.035	0.018	0.072	0.284	0.042	0.177	0.090	0.283	1.000		
(12) CEO positive tone	0.873	0.068	0.040	−0.076	−0.103	−0.015	0.100	0.040	0.140	−0.106	−0.096	0.056	−0.026	1.000	
(13) CEO suspiciousness	0.000	0.326	−0.007	0.080	0.144	0.006	−0.014	−0.050	−0.002	−0.075	−0.060	−0.031	−0.092	0.011	1.000
(14) CEO history of suspiciousness	0.000	0.235	−0.003	0.096	0.181	0.007	−0.013	−0.101	−0.002	−0.105	−0.091	−0.058	−0.121	0.014	0.726
(15) CEO concrete language	0.005	2.425	−0.019	0.015	−0.017	−0.001	−0.057	−0.006	−0.039	−0.006	0.059	0.002	−0.020	−0.084	−0.154
(16) CEO charisma	0.125	0.019	−0.004	−0.148	−0.155	−0.022	0.039	0.153	0.055	−0.084	−0.029	−0.024	−0.071	0.283	0.020
(17) CEO tentative language	2.515	0.758	−0.009	0.071	0.076	0.000	−0.033	−0.112	−0.025	0.113	0.008	−0.080	0.007	−0.221	−0.291
(18) CEO Clout	85.599	6.527	0.010	−0.136	−0.162	−0.017	−0.011	0.080	0.037	−0.121	−0.050	0.022	−0.066	0.320	−0.120
(19) CEO obfuscation	0.005	0.072	−0.002	0.063	0.022	−0.005	0.013	−0.015	−0.021	0.002	0.005	−0.010	−0.006	−0.076	−0.058
(20) CEO word count	7.858	0.508	0.000	0.092	0.080	0.001	0.016	0.048	−0.005	−0.065	−0.153	0.082	0.041	−0.106	−0.244
(21) Number of prior All-Star recognitions	0.291	0.967	−0.063	−0.012	−0.019	0.593	−0.030	0.062	−0.010	−0.039	0.024	0.097	0.023	0.010	0.003
(22) Analyst tenure	12.930	5.465	0.031	−0.017	−0.030	0.076	−0.005	0.045	−0.005	−0.007	0.035	0.021	0.021	0.059	0.006

TABLE 1 (Continued)

Variables	Mean	SD	Correlations (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(23) Regulatory exams	3.983	0.832	0.031	−0.015	−0.026	−0.041	−0.034	−0.074	−0.027	0.001	−0.006	−0.036	−0.016	−0.029	−0.009
(24) Analyst firms	3.394	2.081	0.027	−0.005	−0.008	−0.072	0.005	−0.026	0.016	0.037	−0.015	0.033	0.046	0.045	0.000
(25) Days since last recommendation	0.054	1.006	0.066	−0.001	−0.008	0.018	0.017	0.016	0.022	0.007	0.014	0.022	0.017	0.047	0.005
(26) Prior recommendation	0.004	0.911	0.949	−0.012	−0.016	−0.069	0.059	−0.009	0.029	0.012	−0.011	0.015	0.031	0.033	−0.006
(27) Number of analysts	18.501	9.415	−0.021	0.018	0.049	0.046	0.034	0.362	0.044	−0.041	−0.037	0.445	0.309	0.024	−0.089
(28) Standard deviation of recommendations	0.829	0.201	−0.013	−0.013	−0.015	0.011	0.025	0.057	0.008	−0.010	−0.031	0.001	0.000	−0.011	−0.015
(29) Change in median recommendation	−0.003	0.300	0.012	−0.007	0.003	−0.002	0.008	−0.007	0.041	0.006	−0.004	0.017	−0.002	0.028	0.014
(30) Forecast accuracy	0.277	0.607	−0.036	−0.001	−0.015	0.008	−0.206	−0.152	−0.313	−0.021	−0.012	−0.110	−0.052	−0.093	0.031
(31) Reward words	1.881	0.496	0.009	−0.035	−0.050	0.002	0.014	0.155	0.046	−0.006	0.026	0.076	0.040	0.280	−0.147
(32) Power words	2.533	0.621	−0.020	0.052	0.131	0.000	−0.028	−0.003	0.018	0.063	0.068	−0.030	0.000	−0.101	0.199

Variables	Correlations (14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
(15) CEO concrete language	−0.141	1.000													
(16) CEO charisma	0.019	−0.114	1.000												
(17) CEO tentative language	−0.274	−0.047	−0.414	1.000											
(18) CEO clout	−0.086	−0.020	0.680	−0.418	1.000										
(19) CEO obfuscation	−0.020	0.014	−0.196	0.155	−0.148	1.000									
(20) CEO word count	−0.191	0.044	−0.303	0.329	−0.184	0.291	1.000								
(21) Number of prior All-Star recognitions	0.001	−0.003	−0.015	−0.015	0.000	−0.005	0.022	1.000							
(22) Analyst tenure	0.006	0.008	0.042	−0.037	0.047	0.001	0.001	0.159	1.000						

TABLE 1 (Continued)

Variables	Correlations (14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)
(23) Regulatory exams	−0.017	−0.015	−0.035	0.026	−0.010	0.013	−0.010	0.002	0.198	1.000					
(24) Analyst firms	−0.003	−0.034	−0.014	−0.013	0.010	0.003	0.006	−0.060	0.448	0.180	1.000				
(25) Days since last recommendation	−0.002	0.002	0.013	−0.022	0.018	−0.001	0.002	0.026	0.081	−0.007	−0.002	1.000			
(26) Prior recommendation	−0.003	−0.016	−0.004	−0.008	0.008	−0.001	0.001	−0.063	0.029	0.031	0.028	0.072	1.000		
(27) Number of analysts	−0.113	−0.071	−0.085	0.015	−0.034	−0.004	0.084	0.068	−0.025	0.020	0.088	0.004	−0.021	1.000	
(28) Standard deviation of recommendations	−0.011	−0.002	0.007	−0.011	0.004	0.012	0.021	−0.013	−0.036	−0.016	−0.005	−0.024	−0.012	0.128	1.000
(29) Change in median recommendation	0.012	−0.011	−0.007	−0.006	−0.005	−0.008	0.011	−0.002	−0.003	−0.008	0.001	−0.005	−0.022	−0.038	0.009
(30) Forecast accuracy	0.034	0.018	−0.029	0.020	−0.008	0.005	−0.038	0.003	−0.032	0.023	−0.014	−0.035	−0.036	−0.066	0.017
(31) Reward words	−0.158	−0.003	0.322	−0.106	0.283	−0.058	−0.035	0.013	0.022	−0.017	−0.023	0.025	0.005	0.000	0.008
(32) Power words	0.186	−0.014	0.105	−0.162	−0.010	−0.091	−0.187	−0.018	0.005	−0.012	−0.014	0.006	−0.018	−0.076	−0.041

Variables	Correlations (29)	(30)	(31)
(30) Forecast accuracy	0.016	1.000	
(31) Reward words	0.002	−0.038	1.000
(32) Power words	0.011	−0.014	−0.008

Note: $N = 64{,}107$. All correlations greater than or equal to $|0.008|$ are significant at a .05 level.

TABLE 2 Deceptive CEOs obtain superior recommendations compared to honest CEOs.

Dependent variable	Model 1 First stage Deception	Model 2 Second stage Analyst recommendations	Model 3 Second stage Analyst recommendations	Model 4 Second stage Analyst recommendations	Model 5 Second stage Analyst recommendations	Model 6 Second stage Analyst recommendations
Estimation approach	Linear fixed effects	Ordered probit	Ordered probit	Ordered probit	Ordered probit	Ordered probit
Deception		2.198 (.027)	2.215 (.026)	2.195 (.027)	2.187 (.028)	2.197 (.027)
Deception × History of deception			−0.328 (.030)		−0.331 (.029)	−0.420 (.011)
Deception × All-star analyst				0.124 (.036)	0.125 (.034)	0.060 (.363)
Deception × History of deception × All-star analyst						0.852 (.009)
History of deception	1.574 (.000)	−3.166 (.043)	−3.150 (.044)	−3.181 (.042)	−3.127 (.046)	−3.153 (.045)
All-Star analyst	−0.006 (.488)	0.018 (.731)	0.018 (.733)	0.017 (.748)	0.017 (.752)	−0.008 (.875)
History of deception × All-Star analyst	−0.051 (.167)					0.172 (.425)
ROE	−0.099 (.226)	2.250 (.000)	2.257 (.000)	2.275 (.000)	2.253 (.000)	2.255 (.000)
Firm size	−0.003 (.749)	0.082 (.152)	0.080 (.162)	0.082 (.148)	0.080 (.161)	0.081 (.157)
Earnings surprise	0.008 (.002)	0.049 (.014)	0.048 (.015)	0.049 (.013)	0.049 (.014)	0.049 (.014)
CEO tenure	−0.109 (.248)	−0.402 (.519)	−0.394 (.528)	−0.411 (.509)	−0.394 (.528)	−0.387 (.535)
CEO age	0.106 (.032)	−0.198 (.467)	−0.197 (.469)	−0.196 (.471)	−0.193 (.476)	−0.195 (.472)
Compensation	−0.016 (.000)	0.098 (.001)	0.098 (.001)	0.098 (.001)	0.098 (.001)	0.098 (.001)
CEO awards	0.005 (.070)	−0.015 (.363)	−0.015 (.361)	−0.015 (.363)	−0.015 (.367)	−0.015 (.372)
CEO positive tone	−0.076 (.002)	0.631 (.000)	0.628 (.000)	0.637 (.000)	0.627 (.000)	0.624 (.000)
CEO suspiciousness	0.005 (.371)	−0.089 (.021)	−0.090 (.020)	−0.090 (.019)	−0.091 (.019)	−0.091 (.019)
CEO history of suspiciousness	−0.017 (.203)	0.090 (.338)	0.092 (.329)	0.092 (.329)	0.093 (.325)	0.094 (.319)
CEO concrete language	0.006 (.000)	−0.017 (.015)	−0.017 (.014)	−0.017 (.014)	−0.017 (.015)	−0.017 (.015)

TABLE 2 (Continued)

Dependent variable	Model 1 First stage Deception	Model 2 Second stage Analyst recommendations	Model 3 Second stage Analyst recommendations	Model 4 Second stage Analyst recommendations	Model 5 Second stage Analyst recommendations	Model 6 Second stage Analyst recommendations
CEO charisma	−1.301 (.000)	2.091 (.156)	2.109 (.153)	2.095 (.155)	2.093 (.156)	2.106 (.154)
CEO tentative language	0.017 (.000)	−0.040 (.120)	−0.040 (.118)	−0.040 (.116)	−0.040 (.120)	−0.040 (.121)
CEO clout	−0.002 (.000)	0.010 (.003)	0.010 (.003)	0.010 (.003)	0.010 (.003)	0.010 (.003)
CEO obfuscation	0.061 (.001)	−0.092 (.509)	−0.092 (.510)	−0.092 (.512)	−0.090 (.518)	−0.090 (.521)
CEO word count	0.019 (.000)	−0.045 (.210)	−0.045 (.215)	−0.046 (.203)	−0.045 (.214)	−0.045 (.212)
Number of prior All-Star recognitions	0.001 (.817)	0.029 (.225)	0.029 (.224)	0.029 (.223)	0.029 (.226)	0.031 (.185)
Analyst tenure	−0.036 (.041)	−0.013 (.894)	−0.013 (.897)	−0.012 (.899)	−0.012 (.902)	−0.011 (.914)
Regulatory exams	−0.004 (.605)	−0.130 (.013)	−0.132 (.012)	−0.132 (.012)	−0.131 (.012)	−0.131 (.013)
Analyst firms	−0.008 (.084)	0.017 (.566)	0.016 (.582)	0.017 (.566)	0.016 (.588)	0.016 (.590)
Days since last recommendation	0.001 (.383)	−0.024 (.009)	−0.024 (.009)	−0.024 (.007)	−0.024 (.008)	−0.024 (.008)
Prior recommendation	−0.001 (.624)	2.656 (.000)	2.657 (.000)	2.657 (.000)	2.657 (.000)	2.657 (.000)
Number of analysts	−0.001 (.092)	0.005 (.087)	0.005 (.084)	0.005 (.086)	0.005 (.086)	0.005 (.085)
Standard deviation of recommendations	−0.014 (.082)	−0.174 (.010)	−0.175 (.009)	−0.177 (.009)	−0.174 (.009)	−0.174 (.010)
Change in median recommendation	−0.007 (.094)	0.523 (.000)	0.524 (.000)	0.528 (.000)	0.523 (.000)	0.523 (.000)
Forecast accuracy	0.007 (.003)	−0.028 (.065)	−0.029 (.062)	−0.029 (.061)	−0.029 (.061)	−0.029 (.061)
Reward words	0.020 (.000)					
Power words	−0.007 (.009)					
Control function		−2.189 (.028)	−2.189 (.028)	−2.199 (.027)	−2.174 (.029)	−2.178 (.029)

TABLE 2 (Continued)

Dependent variable	Model 1 First stage Deception	Model 2 Second stage Analyst recommendations	Model 3 Second stage Analyst recommendations	Model 4 Second stage Analyst recommendations	Model 5 Second stage Analyst recommendations	Model 6 Second stage Analyst recommendations
Panel averages of all independent variables		X	X	X	X	X
Month dummies	X	X	X	X	X	X
Year dummies	X	X	X	X	X	X
Cut 1[a]		−5.775 (.000)	−5.788 (.000)	−5.783 (.000)	−5.777 (.000)	−5.784 (.000)
Cut 2		−3.267 (.000)	−3.279 (.000)	−3.269 (.000)	−3.268 (.000)	−3.275 (.000)
Cut 3		0.663 (.037)	0.654 (.040)	0.667 (.035)	0.663 (.037)	0.658 (.038)
Cut 4		3.277 (.000)	3.269 (.000)	3.286 (.000)	3.278 (.000)	3.273 (.000)
N	64,107	64,107	64,107	64,107	64,107	64,107
R^2 within	0.175					
Log-pseudolikelihood		−21,833.174	−21,831.104	−21,832.340	−21,830.005	−21,827.041
AIC	−9827.804	43,870.350	43,872.210	43,874.680	43,872.010	43,876.080

Note: *p* values for coefficient significance in parenthesis. All *p* values calculated using standard errors robust to clustering at the CEO–analyst dyad level. Fixed effects at the CEO–analyst dyad level for linear fixed-effects estimation approach.

[a]Estimated cut points in the ordered probit to differentiate recommendation type.

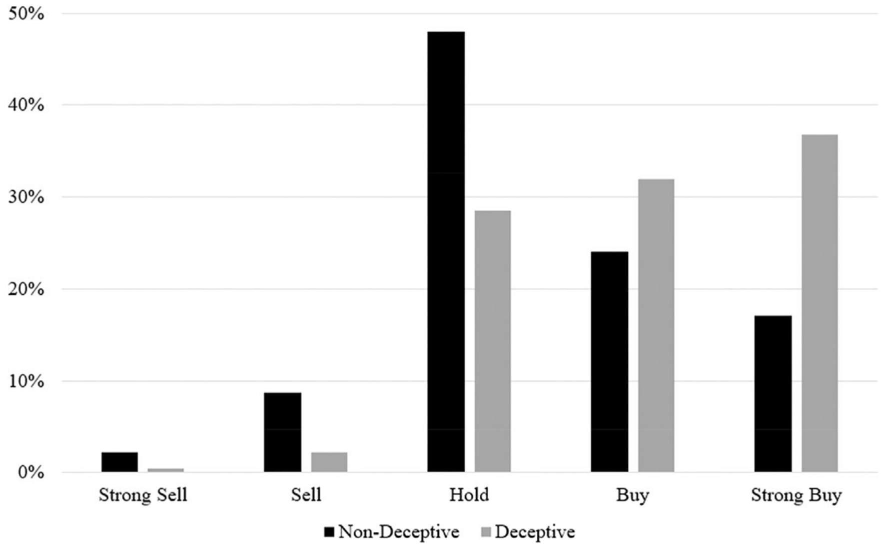

FIGURE 1 Predicted probability of recommendation type for deceptive versus nondeceptive CEOs. *Averaged using observed values of all other variables but setting deception = 0 versus 1.

deception and *All-Star analyst*. Model 6 includes all two-way interactions with *deception* and the three-way interaction of *deception*, *history of deception*, and *All-Star analyst*. We center all continuous variables before creating related interaction terms to ease interpretation of main effects. We use Model 6 to test hypotheses and to show the magnitude of effects, though we note our results are consistent across all models.

6.2.1 | Hypothesis 1

The results provide support for H1: A positive relationship exists between *deception* and *analyst recommendations* (Table 2, Model 6, $\beta = 2.197$, $p = .027$). Assessing the magnitude from the coefficient proves challenging due to the nonlinear nature of our ordered probit estimation (Shaver, 2007). Hoetker (2007) recommends analyzing the average predicted probability of each dependent variable category.[16] Comparing the average predicted probabilities for two different values of a focal variable demonstrates the influence on the average probability of a recommendation. However, this may understate how the change in a focal variable impacts recommendation decisions at the individual analyst level as averaging does not normalize the reference recommendation categories from which to compare a change. To address this issue we also report the average probability of an upgrade in each observation as a result of changing the variable of interest, which holds constant the base of comparison. Refer to Appendix C for more details on calculating the average probability of an upgrade and comparisons to average predicted probabilities.

We show the average predicted probability of each recommendation for deceptive CEOs (*deception* = 1) and honest CEOs (*deception* = 0) in Figure 1, which suggests that CEOs who

[16]This is obtained by averaging predicted probabilities over all observations holding the focal variable at a fixed value but using all other observed variable values (Greene, 2017).

engage in deception can have a meaningful impact on analyst recommendations. Compared to honest CEOs, deceptive CEOs have a 27.7 percentage point drop in the average predicted probability of receiving a Hold, Sell, or Strong Sell (59.0% for honest CEOs compared to 31.3% for deceptive CEOs) and an equivalent increase in the average predicted probability of a Buy or Strong Buy (41.0% for honest CEOs compared to 68.7% deceptive CEOs).

The impact of deception is more pronounced when we look at changes in recommendation probabilities relative to reference recommendation probabilities for each observation. A CEO who engages in deception increases their probability of receiving an upgrade relative to honesty by an average of 47.7 ($p = .030$) probability points.

6.2.2 | Hypothesis 2

We find support for H2: *History of deception* negatively moderates the positive impact of *deception* on *analyst recommendations* (Table 2, Model 6, $\beta = -.420$, $p = .011$). To gauge the magnitude of this moderation, we show the average predicted probabilities for deceptive CEOs with a greater *history of deception* (1.5 standard deviations above the mean) versus an average *history of deception* in Figure 2. CEOs with an average *history of deception* benefit much more from deception on the focal call. They increase the average probability of receiving a Buy or better by 12.4 percentage points (68.5% with an average *history of deception* vs. 56.1% with a greater *history of deception*) while reducing the average probability of a Hold or worse by the same amount (31.5% with an average *history of deception* vs. 43.9% with a greater *history of deception*).

The effect is more prominent when considering changes in recommendation probabilities in each observation. CEOs who use deception with an average *history of deception* are more likely

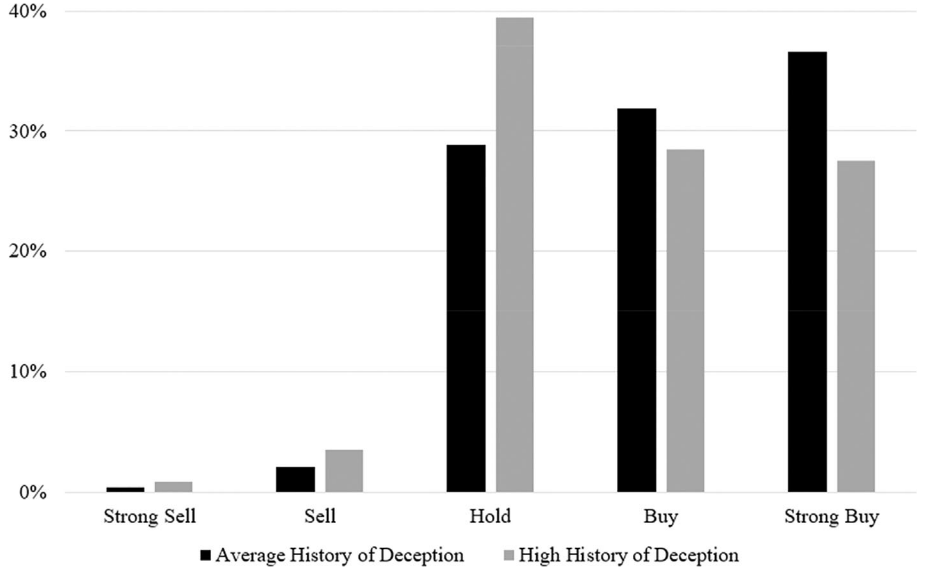

FIGURE 2 Predicted probability of recommendation type for deceptive CEOs with average history of deception versus high history of deception. *Averaged using observed values of all other variables but setting deception = 1; history of deception = mean versus mean + 1.5 SD.

to receive an upgrade by an average of 21.6 ($p = .024$) percentage points compared to engaging in deception with a greater *history of deception*. Taken together, our results suggest there are diminishing returns to continued deceptiveness; CEOs who have a history of honesty receive greater benefits from being deceptive on the focal call.

6.2.3 | Hypothesis 3

We find weak support for H3 at average variable values as shown by the positive moderating effect of *All-Star analyst* on *deception* in Model 6 (Table 2, Model 6, $\beta = .060$, $p = .363$). Recall that we center all continuous variables before generating interaction terms, so this two-way interaction is the moderating effect of *All-Star analyst* on *deception* holding *history of deception* at its mean. This means that the influence of CEO deception on an analyst's recommendation is similar for All-Star and Non-All-Star analysts when the CEO has an average history of using deception. However, note that the moderation effect of *All-Star analyst* on *deception* is larger in both Model 4 (Table 2, Model 4, $\beta = .124$, $p = .036$) and Model 5 (Table 2, Model 5, $\beta = .125$, $p = .034$), when we do not incorporate the influence of *history of deception* on the two-way interaction. This suggests that the positive moderating effect of *All-Star analyst* on *deception* becomes more apparent at certain values for *history of deception*. We explore this further when discussing H4 next.

6.2.4 | Hypothesis 4

We find support for H4. We test this with the three-way interaction of *deception*, *All-Star analyst*, and *history of deception*, which is positive (Table 2, Model 6, $\beta = .852$, $p = .009$). Deceptive CEOs who also have a greater *history of deception* will receive better recommendations from All-Star analysts compared to Non-All-Stars. Note that the two-way interaction of *All-Star analyst* and *deception* discussed for H3 also indicates that there is little difference in the moderation of *deception* and *history of deception* for All-Stars and Non-All-Stars when *history of deception* is at its mean. However, floodlight analysis (Spiller et al., 2013) reveals that differences in the moderation effect of *deception* on *history of deception* become more meaningful for All-Star and Non-All-Star analysts when *history of deception* is greater than 0.3925 ($p < .050$). We highlight this relationship in Figures 3 and 4.

Figure 3 shows the average predicted probability of recommendation type for deceptive CEOs by All-Star and Non-All analysts when CEOs have an average *history of deception*. Consistent with our findings for H3, All-Stars are more likely to give better recommendations than Non-All-Stars, but the difference is minimal. On average, All-Star analysts are 0.5 percentage points more likely to give a Buy or better compared to Non-All-Stars (68.9% for All-Star analysts vs. 68.4% for Non-All-Star analysts) and an equivalent amount less likely to give a Hold or worse (31.1% for All-Star analysts vs. 31.6% for Non-All-Star analysts).

The difference in All-Star and Non-All-Star analysts' responses to deceptive CEOs becomes more pronounced when CEOs have a greater *history of deception*. Figure 4 shows the average predicted probability of recommendation type for deceptive CEOs by All-Star and Non-All-Star analysts with *history of deception* at 1.5 standard deviations above its mean. This value for *history of deception* is in the range where the difference in the moderation effect of *deception* on

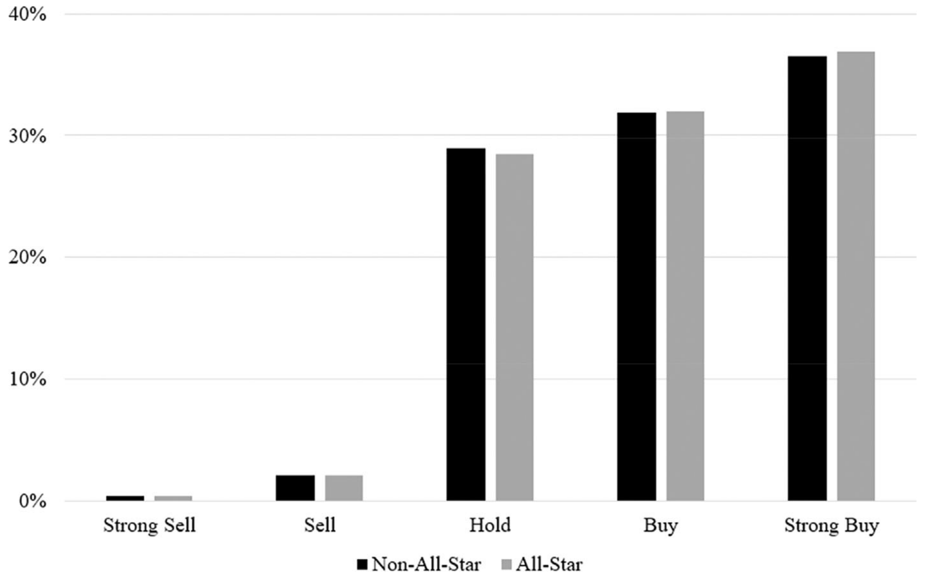

FIGURE 3 Predicted probability of recommendation type for deceptive CEOs with average history of deception by non-All-Star versus All-Star analysts. *Averaged using observed values of all other variables but setting deception = 1: history of deception = mean; All-Star analyst = 0 versus 1.

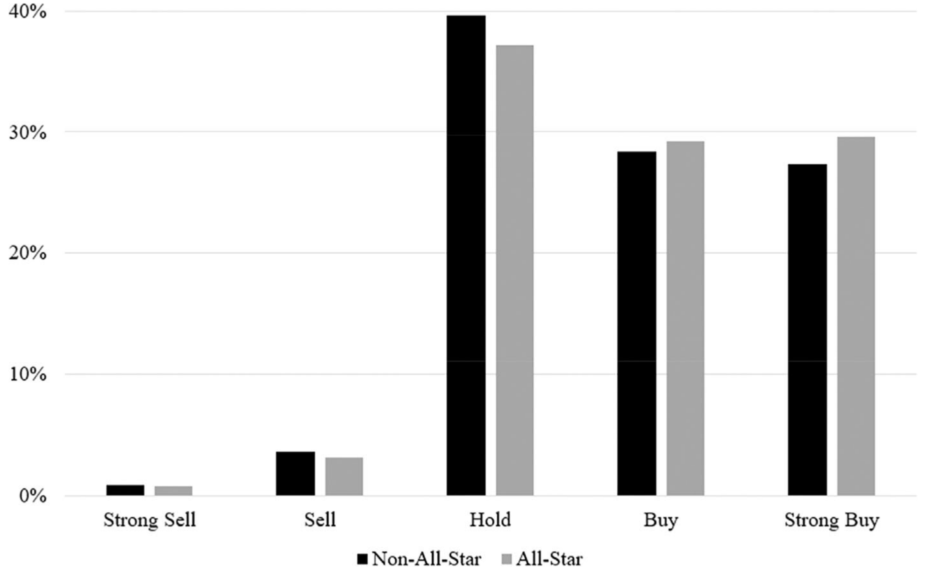

FIGURE 4 Predicted probability of recommendation type for deceptive CEOs with a high history of deception by non-AU-Star versus All-Star analysts. *Averaged using observed values of all other variables but setting deception = 1: history of deception = mean +1.5 SD: All-Star analyst = 0 versus 1.

history of deception for All-Stars and Non-All-Stars is more meaningful as indicated by our floodlight analysis. Here, All-Stars are 3.0 percentage points less likely to give a Hold or worse on average (41.2% for All-Star analysts vs. 44.2% for Non-All-Star analysts) and an equal amount more likely to give a Buy or better compared to Non-All-Stars (58.8% for All-Star analysts vs. 55.8% for Non-All-Star analysts).

This relationship is more apparent when we look into relative changes in recommendation probabilities in each observation. On average, CEOs who engage in deception with a greater *history of deception* are 5.3 percentage points ($p = .007$) more likely to receive an upgrade from an All-Star analyst relative to a Non-All-Star. These findings indicate that Non-All-Star analysts are more likely than All-Stars to give a deceptive CEO a lower recommendation when the CEO has a greater history of continued deceptiveness.

Lastly, though not hypothesized, we find that analysts' suspicion of CEO deception led to worse analyst recommendations (e.g., Table 2, Model 6: $\beta = -.091$, $p = .019$). Regardless of whether engaging in deception or not, CEOs who appear to be deceptive are more likely to receive lower recommendations.

6.3 | Robustness checks

We run several additional analyses to assess the robustness of our findings (see Appendix D). First, we compare analysts' recommendations during fraudulent and nonfraudulent quarters using the matched sample we created for our training dataset (Table D1, Model A9). We anticipate similar outcomes as analysts' reactions to *deception* likely mirror reactions to fraudulent quarters. We include the moderating effect of *All-Star analyst*; however, we could not include a main or moderating effect consistent with our *history of deception* measure since no firm had multiple AAER violations in our training sample. We find that firms obtain better recommendations during fraudulent compared to nonfraudulent quarters, while controlling for (fraudulent) reported firm performance metrics ($\beta = .181$, $p = .016$). Additionally, All-Star analysts give higher recommendations to firms during fraudulent quarters compared to Non-All-Stars ($\beta = .241$, $p = .046$). These results are consistent with estimations in Table 2 (see Table 2 Models 2, 4, and 5).

Second, the literature suggests analysts may stop covering a firm instead of downgrading it during fraudulent periods (Young & Peng, 2013). Analysts may be aware of deception but are unable to prove it and thus choose to drop coverage rather than face the associated downgrading penalties (Westphal & Clement, 2008). As such, the more a CEO uses deception over time, the more likely analysts are to drop coverage. Using a Cox proportional hazard model (Table D2, Model A10), we find that *history of deception* is positively related to an analyst's propensity to drop coverage (Hazard Ratio = 1.284, $p = .005$). In contrast, *deception* exhibits a negative effect (Hazard Ratio = 0.910, $p = .064$), reinforcing our argument about deception's diminishing benefits.

Finally, a censoring restriction in the supplemental linear fixed-effects models in Appendix B could impact results given analysts' limitations to recommend beyond Strong Buy or Sell (note the ordered probit analysis does not have this issue). We re-run the linear fixed-effects models removing cases where an analyst's prior recommendation is Strong Buy or Strong Sell and they did not change their current recommendation. Results are consistent with our analysis (see Appendix C, Table D3, Model A11). To further ensure that our analysis is not the result of a spurious mathematical phenomenon, we also examine the distribution of the *history of deception* (broken into quartiles) over analysts' recommendations. The distribution is similar across quartiles (see Appendix B, Table B1).

7 | DISCUSSION

The primary objective of this study was to examine analysts' reactions to executive deception and identify moderating factors. Our findings aid scholars and practitioners in understanding how and why deception tends to go undetected. Despite numerous corporate scandals and extensive literature on corporate fraud, our understanding of analyst responses to CEO deception, especially before public disclosure, is limited. Our findings suggest that while CEOs may gain short-term benefits from deception, these advantages potentially wane due to trust deterioration between analysts and CEOs. Moreover, the results indicate that analysts with the highest reputation are more susceptible to deception when CEOs have a greater history of deception. Overall, the results provide strong support for our theoretical perspective and hypotheses.

Our first hypothesis examines the impact of CEO deception during the focal earnings call on analysts' recommendations. Our findings reveal that deceptive CEOs are more likely to receive favorable recommendations. We argue that deceptive CEOs receive better recommendations because analysts have a default-to-truth bias. Due to the low base rate of deception and difficulty of detection, analysts tend to assume that the CEO is being truthful (Levine, 2014). As a result, analysts are influenced by CEO deception and provide more positive recommendations.

Our findings highlight the diminishing returns of deception, demonstrating that sustained dishonesty can adversely affect recommendations. CEOs who have been honest in previous earnings calls initially gain significant advantages from deception. However, these benefits diminish with each subsequent act of deception. Thus, a history of deceit leads to a threshold where analysts' positive recommendations plateau, potentially turning negative as they become more aware of the deceit. This aligns with research suggesting those who feel deceived grow to distrust the deceiver (Grover, 2005; O'Sullivan, 2003), a pattern mirrored in analysts' responses to persistently deceptive CEOs.

Finally, we find that All-Star analysts are more susceptible to historically deceptive CEOs than their peers. Their increased hubristic thinking and executive access leads them to accept CEO statements as truth, despite the CEO's history of deception. Our results suggest that reputation can have a negative effect on information intermediaries, in that it may bolster their confidence and hamper an individual's ability to recognize deceptive behavior.

8 | CONTRIBUTIONS

This study advances our understanding of governance, reputation, and deception in several ways. Previous studies primarily scrutinized reactions to fraud postexposure (Arthaud-Day et al., 2006; Hennes et al., 2008; Persons, 2006), restricting their implications due to uncertainty if responses were to the fraud itself or the repercussions of detection (Bednar, 2012). This study illuminates the distinction, suggesting analysts penalize (or at least do not reward) repeat offenders despite initially falling prey to deception. Notably, All-Star analysts demonstrate a slower response to deception, signifying a higher degree of bias. This suggests that they may unintentionally contribute to the escalation of fraud-related damage. That is, there are times when not having an information intermediary would result in less harm than having one.

Finally, the study develops and applies ML methods to improve the accuracy of detecting deception within a business context by approximately 29.5% compared to previous approaches (Larcker & Zakolyukina, 2012; Newman et al., 2003). ML's iterative nature enables us to account for more complex and hidden relationships, resulting in a more accurate measure than those developed through a frequency approach. Using this methodology, scholars can more reliably measure and study the impacts of deception in a real-world setting, and consequently aid deception scholars in overcoming criticisms of ecological validity (O'Sullivan et al., 2009). With this more accurate measure of deception, scholars can examine the effectiveness of other governing mechanisms in discouraging unethical behavior.

8.1 | Limitations and future research

A notable challenge in this study pertains to the measurement of deception. Direct assessment is not possible due to deception's internal psychological nature, leading us to infer it via linguistic cues. This introduces validity concerns inherent to any measure of this kind. To address these concerns we rigorously adhere to methodological best practices (Alaskar et al., 2022) and provide evidence of the measure's validity through its ability to accurately predict AAER violations, data breaches, and consumer-related fines. Additional validation comes from a nonbusiness deception sample (medical publications and publicly available instances of deception) and a negative correlation with female CEOs, consistent with theory (Ho et al., 2015). We have also accounted for potential confounding factors, like charisma, obfuscation, and CEO characteristics, in our analysis. Nevertheless, validity concerns remain due to the black-box nature of ML algorithms.

The black box issue arises due to the opaque decision-making process of ML models. These atheoretical models prioritize prediction accuracy by randomly combining input features for maximum efficiency. As a result, it is hard to discern why our model categorizes specific texts as deceptive or honest (see Appendix A).

Given the limitations of our understanding of the model's decision-making process, it is challenging to predict potential biases or inaccuracies in specific contexts. For example, the model's accuracy might diminish for certain demographics like minority groups, non-native English speakers, or neurodivergent individuals. Further, our training data includes those who were caught lying. It is possible that individuals who were not caught have different linguistic patterns, resulting in lower accuracy for that population. Therefore, to enhance measurement precision, we call upon future researchers to apply ML measures in diverse contexts and populations. Moreover, exploring alternative methods of operationalizing deception that are not based on linguistic patterns, such as facial expressions, vocal tone, and body language, can help mitigate any uncontrolled or overlooked linguistic biases potentially driving our results. Thus, it is vital for future users of this measure to make concerted efforts in ensuring its validity for their respective contexts and populations.

Building on these considerations, we also encourage scholars to delve into the causes of deceptive behavior. Unraveling the triggers of deception may provide insights into its prevention and containment. Additionally, understanding mechanisms that effectively deter deceptive practices can augment our collective knowledge and inspire more robust governance strategies. As we navigate these challenges, we look forward to the critical insights that future research will undoubtedly unveil.

8.1.1 | Practical concerns

Given the attention deception has received within the information systems literature (Alaskar et al., 2022; Şen et al., 2020), it seems inevitable that measures like ours will be employed in practical applications. As such, we provide a word of caution for the use of this technology.

Misclassifying an individual as deceptive or honest could have a significant impact on their life, such as in the case of predictive policing or credit scoring (Hassani, 2021; Shapiro, 2017). While our measures may increase the accuracy of predicting deception, algorithms like ours will inevitably have Type I and Type II errors. Therefore, we cannot assume guilt based on a deception detection algorithm alone. This is why our measure produces a probability score and not a categorical indicator of deception. Additionally, with the recent advancements in generative AI (such as GPT3; Floridi & Chiriatti, 2020), individuals will likely be able to adjust their written content to beat deception detection algorithms like ours. To summarize, we advise considerable caution when using deception detection algorithms as they alone cannot determine guilt or innocence.

9 | CONCLUSION

Security analysts remain an essential form of external governance, and the metrics they use to influence the market (Womack, 1996). Unfortunately, they have a default-to-truth bias and are likely to believe deception as a result. This article examines how this truth bias impacts the recommendations they provide. Our findings indicate that deceptive CEOs obtain better recommendations, at least in the short term. Additionally, we found that the benefits of using deception will degrade from overuse. Lastly, we demonstrate that analysts become more susceptible to repeated deception as their reputation increases.

DATA AVAILABILITY STATEMENT

Data available on request from the authors

ORCID

Jonathan Bundy https://orcid.org/0000-0001-7821-121X

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Google Brain. (2016). Tensorflow: A system for large-scale machine learning. *Operating Systems Design and Implementation*, *16*, 265–283.

Akstinaite, V., Robinson, G., & Sadler-Smith, E. (2020). Linguistic markers of CEO hubris. *Journal of Business Ethics*, *167*, 687–705.

Alaskar, H., Sbaï, Z., Khan, W., Hussain, A., & Alrawais, A. (2022). Intelligent techniques for deception detection: A survey and critical study. *Soft Computing*, *27*, 3581–3600. https://doi.org/10.1007/s00500-022-07603-w

Anderson, A., & Smith, L. (2013). Dynamic deception. *American Economic Review*, *103*(7), 2811–2847.

Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied PsychoLinguistics*, *31*(3), 397–411.

Arthaud-Day, M. L., Certo, S. T., Dalton, C. M., & Dalton, D. R. (2006). A changing of the guard: Executive and director turnover following corporate financial restatements. *Academy of Management Journal*, *49*(6), 1119–1136.

Bansal, P., & Clelland, I. (2004). Talking trash: Legitimacy, impression management, and unsystematic risk in the context of the natural environment. *The Academy of Management Journal*, *47*(1), 93–103.

Bednar, M. K. (2012). Watchdog or lapdog? A behavioral view of the media as a corporate governance mechanism. *The Academy of Management Journal*, *55*(1), 131–150.

Blair, J. P., Levine, T. R., & Shaw, A. S. (2010). Content in context improves deception detection accuracy. *Human Communication Research*, *36*(3), 423–442.

Boivie, S., Bednar, M. K., Aguilera, R. V., & Andrus, J. L. (2016). Are boards designed to fail? The implausibility of effective board monitoring. *Academy of Management Annals*, *10*(1), 319–407.

Boivie, S., Graffin, S. D., & Gentry, R. J. (2016). Understanding the direction, magnitude, and joint effects of reputation when multiple actors' reputations collide. *Academy of Management Journal*, *59*(1), 188–206.

Bolino, M. C., Klotz, A. C., & Daniels, D. (2014). The impact of impression management over time. *Journal of Managerial Psychology*, *29*(3), 266–284.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

Brandt, M. J., Wetherell, G., & Henry, P. J. (2015). Changes in income predict change in social trust: A longitudinal analysis. *Political Psychology*, *36*(6), 761–768.

Brochet, F., Miller, G. S., & Srinivasan, S. (2013). Do analysts follow managers who switch companies? An analysis of relationships in the capital markets. *Accounting Review*, *89*(2), 451–482.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research*, *53*(1), 1–47.

Brown, N. C., Wei, K. D., & Wermers, R. (2013). Analyst recommendations, mutual fund herding, and overreaction in stock prices. *Management Science*, *60*(1), 1–20.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242.

Burgoon, J., Mayew, W. J., Giboney, J. S., Elkins, A. C., Moffitt, K., Dorn, B., Byrd, M., & Spitzley, L. (2016). Which spoken language markers identify deception in high-stakes settings? Evidence from earnings conference calls. *Journal of Language and Social Psychology*, *35*(2), 123–157.

Busenbark, J. R., Lange, D., & Certo, S. T. (2017). Foreshadowing as impression management: Illuminating the path for security analysts. *Strategic Management Journal*, *38*(12), 2486–2507.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Clatworthy, M. A., & Jones, M. J. (2006). Differential patterns of textual characteristics and company performance in the chairman's statement. *Accounting, Auditing & Accountability Journal*, *19*(4), 493–511.

Clement, M. B., & Tse, S. Y. (2005). Financial analyst characteristics and herding behavior in forecasting. *Journal of Finance*, *60*(1), 307–341.

Clementson, D. E. (2018). Truth bias and partisan bias in political deception detection. *Journal of Language and Social Psychology*, *37*(4), 407–430.

Collins, D., Masli, A. D. I., Reitenga, A. L., & Sanchez, J. M. (2009). Earnings restatements, the Sarbanes-Oxley Act, and the disciplining of chief financial officers. *Journal of Accounting, Auditing and Finance*, *24*(1), 1–34.

Conniff, R. (2005). Crass commanders. *Psychology Today*, *38*(5), 60–62.

Craig, R., Mortensen, T., & Iyer, S. (2013). Exploring top management language for signals of possible deception: The words of Satyam's chair Ramalinga Raju. *Journal of Business Ethics*, *113*(2), 333–347.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197–199.

Dyck, I. J., Morse, A., & Zingales, L. (2013). How pervasive is corporate fraud? Rotman School of Management Working Paper No. 2222608.

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, *32*(1), 88–106.

Ertimur, Y., Sunder, J., & Sunder, S. V. (2007). Measure for measure: The relation between forecast accuracy and recommendation profitability of analysts. *Journal of Accounting Research*, *45*(3), 567–606.

Fanelli, A., Misangyi, V. F., & Tosi, H. L. (2008). In charisma we trust: The effects of CEO charismatic visions on securities analysts. *Organization Science*, *20*(6), 1011–1033.

Farber, D. B. (2005). Restoring trust after fraud: Does corporate governance matter? *The Accounting Review*, *80*(2), 539–561.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*(4), 681–694.

Fornaciari, T., & Poesio, M. (2013). Automatic deception detection in Italian court cases. *Artificial Intelligence and Law*, *21*(3), 303–340.

Gallop, R. J., Crits-Christoph, P., Muenz, L. R., & Tu, X. M. (2003). Determination and interpretation of the optimal operating point for ROC curves derived through generalized linear models. *Understanding Statistics*, *2*(4), 219–242.

Gleason, C. A., & Lee, C. M. (2003). Analyst forecast revisions and market price discovery. *The Accounting Review*, *78*(1), 193–225.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics*, *40*(1–3), 3–73.

Greene, W. (2017). *Econometric analysis* (8th ed.). New York.

Greene, W. H., & Hensher, D. A. (2010). *Modeling ordered choices: A primer*. Cambridge University Press.

Grover, S. L. (2005). The truth, the whole truth, and nothing but the truth: The causes and management of workplace lying. *Academy of Management Executive*, *19*(2), 148–157.

Groysberg, B., Healy, P. M., & Maber, D. A. (2011). What drives sell-side analyst compensation at high-status investment banks? *Journal of Accounting Research*, *49*(4), 969–1000.

Gupta, V. K., Mortal, S., Chakrabarty, B., Guo, X., & Turban, D. B. (2020). CFO gender and financial statement irregularities. *Academy of Management Journal*, *63*(3), 802–831.

Hambrick, D. C., & Mason, P. A. (1984). Upper Echelons: The organization as a reflection of its top managers. *Academy of Management Review*, *9*(2), 193–206.

Harris, J., & Bromiley, P. (2007). Incentives to cheat: The influence of executive compensation and firm performance on financial misrepresentation. *Organization Science*, *18*(3), 350–367.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2019). Measuring CEO personality: Developing, validating, and testing a linguistic tool. *Strategic Management Journal*, *40*(8), 1316–1330.

Hartwig, M., Granhag, P. A., Strömwall, L. A., & Kronkvist, O. (2006). Strategic use of evidence during police interviews: When training to detect deception works. *Law and Human Behavior*, *30*(5), 603–619.

Hassani, B. K. (2021). Societal bias reinforcement through machine learning: A credit scoring perspective. *AI and Ethics*, *1*(3), 239–247.

Hauch, V., Blandón-Gitlin, I., Masip, J., & Sporer, S. L. (2015). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review*, *19*(4), 307–342.

Hayward, M. L., Rindova, V. P., & Pollock, T. G. (2004). Believing one's own press: The causes and consequences of CEO celebrity. *Strategic Management Journal*, *25*(7), 637–653.

Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The importance of distinguishing errors from irregularities in restatement research: The case of restatements and CEO/CFO turnover. *The Accounting Review*, *83*(6), 1487–1519.

Ho, S. M., Hancock, J. T., Booth, C., & Liu, X. (2016). Computer-mediated deception: Strategies revealed by language-action cues in spontaneous communication. *Journal of Management Information Systems*, *33*(2), 393–420.

Ho, S. S., Li, A. Y., Tam, K., & Zhang, F. (2015). CEO gender, ethical leadership, and accounting conservatism. *Journal of Business Ethics*, *127*(2), 351–370.

Hobson, J. L., Mayew, W. J., & Venkatachalam, M. (2012). Analyzing speech to detect financial misreporting. *Journal of Accounting Research*, *50*(2), 349–392.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343.

Hong, H., & Kubik, J. D. (2003). Analyzing the analysts: Career concerns and biased earnings forecasts. *The Journal of Finance*, *58*(1), 313–351.

Honts, C. R., & Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *Journal of Applied Psychology*, *79*(2), 252.

Hope, O.-K., & Wang, J. (2018). Management deception, big-bath accounting, and information asymmetry: Evidence from linguistic analysis. *Accounting, Organizations and Society*, *70*, 33–51.

Hyde, S. J., Bachura, E., & Harrison, J. S. (2023). Garbage in, garbage out: A theory-driven approach to improve data handling in supervised machine learning. In D. A. Hill, A. F. McKenny, P. O'Kane, & S. Paroutis (Eds.), *Methods to improve our field, research methodology in strategy and management* (Vol. 14, pp. 101–132). Emerald Publishing Limited. https://doi.org/10.1108/S1479-838720220000014006

Koch-Bayram, I. F., & Wernicke, G. (2018). Drilled to obey? Ex-military CEOs and financial misconduct. *Strategic Management Journal*, *39*(11), 2943–2964.

König, A., Mammen, J., Luger, J., Fehn, A., & Enders, A. (2017). Silver bullet or ricochet? CEOs' use of metaphorical communication and infomediaries' evaluations. *Academy of Management Journal*, *61*(4), 1196–1230.

Larcker, D. F., & Zakolyukina, A. A. (2012). Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, *50*(2), 495–540.

Lee, J. (2016). Can investors detect managers' lack of spontaneity? Adherence to predetermined scripts during earnings conference calls. *The Accounting Review*, *91*(1), 229–250.

Levine, T. R. (2014). Truth-default theory (TDT): A theory of human deception and deception detection. *Journal of Language and Social Psychology*, *33*(4), 378–392.

Levine, T. R. (2019). *Duped: Truth-default theory and the social science of lying and deception*. University Alabama Press.

Levine, T. R., Blair, J. P., & Clare, D. D. (2014). Diagnostic utility: Experimental demonstrations and replications of powerful question effects in high-stakes deception detection. *Human Communication Research*, *40*(2), 262–289.

Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People lie for a reason: Three experiments documenting the principle of veracity. *Communication Research Reports*, *27*(4), 271–285.

Levine, T. R., Punyanunt-Carter, N. M., & Moore, A. (2021). The truth-default and video clips: Testing the limits of credulity. *Communication Studies*, *72*(2), 133–145.

Li, J., Shi, W., Connelly, B. L., Yi, X., & Qin, X. (2022). CEO awards and financial misconduct. *Journal of Management*, *48*(2), 380–409.

Lloyd, E. P., Deska, J. C., Hugenberg, K., McConnell, A. R., Humphrey, B. T., & Kunstman, J. W. (2019). Miami University deception detection database. *Behavior Research Methods*, *51*(1), 429–439.

Lovelace, J. B., Bundy, J., Hambrick, D. C., & Pollock, T. G. (2018). The shackles of CEO celebrity: Sociocognitive and behavioral role constraints on "star" leaders. *Academy of Management Review*, *43*(3), 419–444.

Lung-Fei, L. (1992). Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables. *In Econometric Reviews*. Taylor & Francis, *11*(3), 319–328.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA. JMLR: W&CP, Vol. 30, pp. 1–6.

Mann, S., Vrij, A., & Bull, R. (2004). Detecting true lies: Police officers' ability to detect suspects' lies. *Journal of Applied Psychology*, *89*(1), 137–149.

Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), 1–43.

McCornack, S. A., & Parks, M. R. (1986). Deception detection and relationship development: The other side of trust. *Annals of the International Communication Association*, *9*(1), 377–389.

Merkl-Davies, D. M., Brennan, N. M., & McLeay, S. J. (2011). Impression management and retrospective sense-making in corporate narratives: A social psychology perspective. *Accounting, Auditing & Accountability Journal*, *24*(3), 315–344.

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621–633.

Newman, M. L., Pennebaker, J. W., Berry, D. S., & Richards, J. M. (2003). Lying words: Predicting deception from linguistic styles. *Personality and Social Psychology Bulletin*, *29*(5), 665–675.

O'Brien, P. C., & Bhushan, R. (1990). Analyst following and institutional ownership. *Journal of Accounting Research*, *28*, 55–76.

Oppel, R. A. (2002, February 28). Wall Street analysts faulted on Enron. *The New York Times*.

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The boy-who-cried-wolf effect. *Personality and Social Psychology Bulletin*, *29*(10), 1316–1327.

O'Sullivan, M., Frank, M. G., Hurley, C. M., & Tiwana, J. (2009). Police lie detection accuracy: The effect of lie scenario. *Law and Human Behavior*, *33*(6), 530–538.

Paik, E. T., Pollock, T. G., Boivie, S., Lange, D., & Lee, P. M. (2022). A star is born: The relationship between performance and achieving status through certification contests in the context of equity analysts. *Organization Science*, *34*(1), 75–99.

Pan, L., McNamara, G., Lee, J. J., Haleblian, J., & Devers, C. E. (2018). Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. *Strategic Management Journal*, *39*(8), 2204–2225.

Park, H. S., & Levine, T. R. (2017). The effects of truth–lie base-rates on deception detection accuracy in Korea. *Asian Journal of Communication*, *27*(5), 554–562.

Pennebaker, J. W. (2011). *The secret life of pronouns: What our words say about us*. Bloomsbury Press.

Persons, O. S. (2006). The effects of fraud and lawsuit revelation on U.S. executive turnover and compensation. *Journal of Business Ethics*, *64*(4), 405–419.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research*, *47*(1), 3–13.

Riedl, M., & Geishecker, I. (2014). Keep it simple: Estimation strategies for ordered response models with fixed effects. *Journal of Applied Statistics*, *41*(11), 2358–2374.

Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal*, *48*(6), 1033–1049.

Roberts, J. (2002, February 28). Analysts: No signs of trouble at Enron. https://www.cbsnews.com/news/analysts-no-signs-of-trouble-at-enron/

Ryan, T. P. (1997). *Modern regression analysis*. Wiley.

Scheuerlein, J., & Chládková, H. (2019). Linguistic differences between successful and non-successful CEOs during the financial crisis. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *67*(2), 583–595.

Şen, M. U., Perez-Rosas, V., Yanikoglu, B., Abouelenien, M., Burzo, M., & Mihalcea, R. (2020). Multimodal deception detection using real-life trial data. *IEEE Transactions on Affective Computing*, *13*(1), 306–319.

Shapiro, A. (2017). Reform predictive policing. *Nature*, *541*(7638), 458–460.

Shaver, M. J. (2007). Interpreting empirical results in strategy and management research. In A. Hill, A. McKenny, J. Lê, A. Smith, P. O'Kane, & S. Paroutis (Eds.), *Research methodology in strategy and management* (pp. 273–293). Emerald Group Publishing Limited.

Soltani, B. (2014). The anatomy of corporate fraud: A comparative analysis of high profile American and European corporate scandals. *Journal of Business Ethics*, *120*(2), 251–274.

Soltes, E. (2014). Private interaction between firm management and sell-side analysts. *Journal of Accounting Research*, *52*(1), 245–272.

Spiller, S. A., Fitzsimons, G. J., Lynch, J. G., Jr., & McClelland, G. H. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*, *50*(2), 277–288.

Stickel, S. E. (1992). Reputation and performance among security analysts. *The Journal of Finance*, *47*(5), 1811–1836.

Ten Brinke, L., & Porter, S. (2012). Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception. *Law and Human Behavior*, *36*(6), 469–477.

Toma, C. L., & Hancock, J. T. (2012). What lies beneath: The linguistic traces of deception in online dating profiles. *Journal of Communication*, *62*(1), 78–97.

Wendel, C. (2014, November 21). Cross selling as secret sauce for success. *Banking Strategies*. https://www.bai.org/banking-strategies/article-detail/cross-selling-as-secret-sauce-for-success

Westphal, J. D., & Clement, M. B. (2008). Sociopolitical dynamics in relations between top managers and security analysts: Favor rendering, reciprocity, and analyst stock recommendations. *Academy of Management Journal*, *51*(5), 873–897.

Wetschoreck, F., Krabel, T., & Krishnamurthy, S. (2020). 8080labs/ppscore: zenodo release, October 2020. https://doi.org/10.5281/zenodo.4091345

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *Journal of Finance*, *51*(1), 137–167.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445.

Yezegel, A. (2015). Why do analysts revise their stock recommendations after earnings announcements? *Journal of Accounting and Economics*, *59*(2–3), 163–181.

Young, S. M., & Peng, E. Y. (2013). An analysis of accounting frauds and the timing of analyst coverage decisions and recommendation revisions: Evidence from the US. *Journal of Business Finance & Accounting*, *40*(3–4), 399–437.

Zahra, S. A., Priem, R. L., & Rasheed, A. A. (2005). The antecedents and consequences of top management fraud. *Journal of Management*, *31*(6), 803–828.

Zhu, D. H., & Westphal, J. D. (2011). Misperceiving the beliefs of others: How pluralistic ignorance contributes to the persistence of positive security analyst reactions to the adoption of stock repurchase plans. *Organization Science*, *22*(4), 869–886.

Zyung, J. D., & Shi, W. (2022). In retrospect: The influence of chief executive officers' historical relative pay on overconfidence. *Strategic Organization*, *20*(3), 627–651.

How to cite this article: Hyde, S. J., Bachura, E., Bundy, J., Gretz, R. T., & Sanders, W. G. (2024). The tangled webs we weave: Examining the effects of CEO deception on analyst recommendations. *Strategic Management Journal*, *45*(1), 66–112. https://doi.org/10.1002/smj.3546

APPENDIX A: DECEPTIVE AND HONEST CEO QUOTES

LIWC input features used for the machine learning (ML) model are underlined.

CEO 1—Deceptive
Charged with bribery related to their international sales and securities fraud.
"Yes I would—this is [*name redacted*]. A number of pieces to that. The reality is the European markets continue to be challenging particularly the more developed Western markets. But having said that we continue to benefit significantly by emerging developing markets China Brazil and so on. Also generally speaking the plasma proteins across the board including developed Europe were very strong in the second quarter. So, it's tough to generalize regarding all markets outside the U.S. We are pleased by the overall strength and growth of our sales outside the U.-S. but it is really—the dynamics are very different between developed Europe which continues and we anticipate it's going to continue to be challenging versus our expanding investment and presence in emerging and developing markets which consistently has been a growth priority for us."

CEO 2—Deceptive
Charged with kickbacks and false claims.
"You know I don't think we really see—on the regulatory outcome I mean I would rather predict the weather than predict a regulatory outcome. Right now, as I look at it, I don't particularly see a higher risk on any of them and actually I think everything so far we've seen everything is progressing well. I mean they are all in different categories. [*name redacted*] obviously doesn't have the same size population as a [*name redacted*]. At the same time, I would suspect that that's probably a little bigger than most people think. This is a drug that really has a very clear meaningful benefit in those patients that need the drug and there isn't anything else there today. I actually think I think [*name redacted*] is probably underestimated. [*name redacted*] is extraordinarily important for our diabetes franchise and getting that—and then I think particularly because—I think it's fair to say that we see big synergy with [*name redacted*] especially because you can either add GLP1 to insulin or add insulin to this which [*name redacted*] can't. So, I think that one is a big one."

CEO 3—Honest

"The proceeds are not included in consumer. On a GAAP basis the $2.2 million gain and we in our release adjusted in our commentary and release adjusted our results to include that $2.2 million gain. [*name redacted*] comments also excluded that gain so that we could really discuss the underlying results of our consumer businesses which are positive but not anything to write home about. On a relative basis they're actually quite good versus a lot of what's going on in the consumer markets.

CEO 4—Honest

"Yeah we have all of the factory machinery in place we have the vast majority of the tooling in place. There are a few stamping dies (*inaudible*) due to late changes and those are coming in next month. But really the machinery is in place and it's just a question of ironing out any bugs with the overall manufacturing system. The factory is sort of like big machine with many subcomponents effectively. So, it's getting that machine so work effectively at the subsystem level and then in the transition for one subsystem to next and then being able to kind of spool it up and go at a greater and greater speed. And we really want to be supersensitive to the quality of the product and our aspiration is to deliver costs that have zero defects. So, it will take sometime to be able to have the whole system move with equal cadence to achieve that goal."

Discussion of the ML algorithm

To the naked eye, there may not appear to be obvious differences between the deceptive messages and honest messages provided above, even with the deceptive cues highlighted. This is expected since our measure is based on the combination of linguistic cues rather than the simple presence of a cue. In this way, our measure differs from other common detection approaches.

For example, traditional polygraph tests use multiple physiological measures (such as blood pressure and skin conductivity) to capture one construct related to deception: arousal. As mentioned above, deception tends to incite strong negative emotions (e.g., anxiety), resulting in increased arousal. Hence, these physiological measures are used to capture the affective nature of deception. This makes the polygraph simple to understand, but limited in its accuracy (Honts & Kircher, 1994). Arousal does not always suggest deception and its absence does not always indicate honesty. Below is a graphical representation of the polygraph test (Figure A1).

In contrast, our model's prediction is based on multiple linguistic measures to capture not only the affective state of the CEO, but also their cognition and other contextual cues that are indicative of deception. Yet, our measure is not simply based on the presence or absence of these constructs within a call, but rather the unique combination of these constructs on the call. This is important since it reduces the likelihood of a Type I or Type II error, since the model has the flexibility to detect multiple types of deception and honesty. For example, there may be instances where the deceiver does not feel increased anxiety. The polygraph test will fail to detect this type of deception since its prediction is purely based on the speaker's affect. In contrast, our approach is far more likely to accurately detect this type of deception since it incorporates affect and nonaffect-based constructs. Furthermore, it incorporates the multiple combinations of the constructs to inform its prediction. Hence, our model has the flexibility to

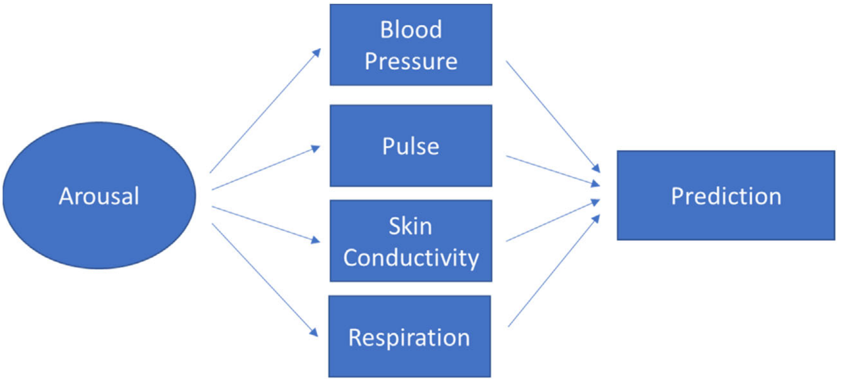

FIGURE A1 Polygraph model.

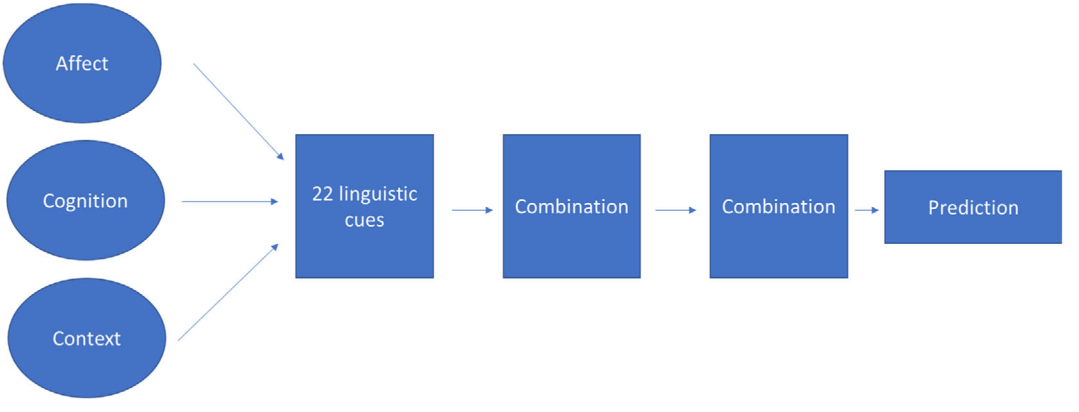

FIGURE A2 Linguistic deception model.

detect a wide variety of types or patterns of deception. Below is a graphical representation of our approach (Figure A2).

However, the downside of using an ML approach, like ours, is that it is difficult to determine the relative importance of any given linguistic feature for our prediction. As a demonstration, we ran a Power Prediction Score analysis, which runs a series of random forest regressions to determine the relationship strengths between our features and deception (Wetschoreck et al., 2020). The resulting values range from 0 to 1, representing the degree of predictive power one variable has over another (with 1 being perfectly predictive). We found that the feature with the highest score, hence highest individual contributor irrespective of other relationships, was nonfluences (i.e., err, em, uh) at 0.013, which is a small effect. These results indicate that none of the features are strong signals of deception alone. One consequence of this is that our model is more difficult to understand, since its prediction is not simply based on the presence of linguistic cues in a call, but rather their unique combination.

Because our model has 22 input nodes and two hidden layers with 22 nodes each, there are 484 connections between the input layer and layer 1. Likewise, layer 1 and layer 2 have 484 connections between them. In sum, there are 4,194,304 possible arrangements, as represented in Figure A3. Visualizing and understanding the complexity of this structure, and the total

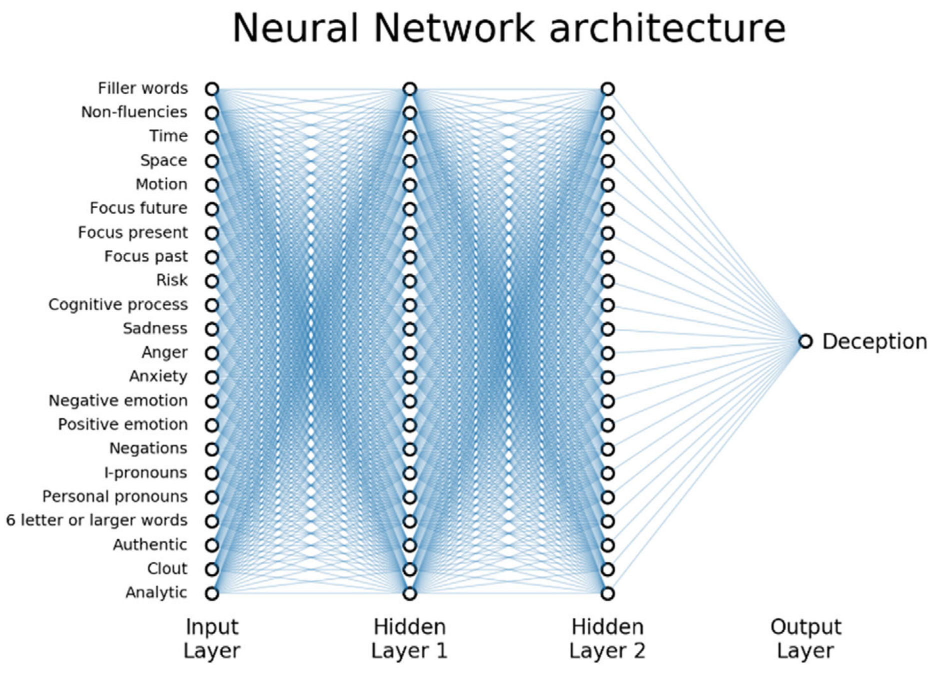

FIGURE A3 Neural network architecture of linguistic deception model.

number of different parameters that are learned, highlights both the capability of a neural network to learn complex tasks as well as the difficulty in interpreting its structure and parameters. Furthermore, this also highlights the complexity required to capture deception since the individual cues themselves provide little predictive power compared to the combination of those cues. Hence, deception is best detected through examining the relationship between linguistic cues rather than the individual linguistic indicators in isolation.

APPENDIX B

TABLE B1 History of deception by recommendations.

	History of deception				
	First quartile	Second quartile	Third quartile	Fourth quartile	Total
Strong sell	197	207	177	185	766
Sell	805	911	840	813	3369
Hold	6889	6928	7136	7254	28,207
Buy	4592	4680	4643	4635	18,550
Strong buy	3539	3307	3226	3143	13,215

Estimations without endogeneity correction and linear fixed-effects panel estimation robustness checks

TABLE B2 Deceptive CEOs obtain superior recommendations compared to honest CEOs—Linear models.

Estimation approach	Model A1 Ordered probit	Model A2 Linear fixed effects	Model A3 Linear fixed effects	Model A4 Linear fixed effects	Model A5 Linear fixed effects	Model A6 Linear fixed effects	Model A7 Linear fixed effects
Deception	0.021 (.506)	0.004 (.471)	0.420 (.014)	0.422 (.014)	0.415 (.015)	0.418 (.014)	0.420 (.014)
Deception × History of deception	−0.421 (.012)	−0.077 (.009)		−0.060 (.027)		−0.061 (.026)	−0.077 (.009)
Deception × All-Star analyst	0.062 (.358)	0.011 (.344)			0.022 (.038)	0.023 (.036)	0.011 (.357)
Deception × History of deception × All-Star analyst	0.847 (.013)	0.157 (.009)					0.158 (.009)
History of deception	0.272 (.023)	0.046 (.032)	−0.611 (.023)	−0.607 (.024)	−0.608 (.023)	−0.604 (.024)	−0.609 (.024)
All-Star analyst	−0.021 (.664)	−0.002 (.774)	0.005 (.594)	0.005 (.594)	0.005 (.602)	0.005 (.603)	0.000 (.996)
History of deception × All-Star analyst	0.061 (.748)	0.011 (.734)					0.033 (.395)
ROE	2.045 (.000)	0.367 (.000)	0.406 (.000)	0.406 (.000)	0.407 (.000)	0.407 (.000)	0.407 (.000)
Firm size	0.072 (.193)	0.011 (.293)	0.012 (.252)	0.012 (.254)	0.012 (.254)	0.012 (.256)	0.012 (.252)
Earnings surprise	0.068 (.000)	0.013 (.000)	0.009 (.009)	0.009 (.009)	0.009 (.009)	0.009 (.009)	0.009 (.009)
CEO tenure	−0.615 (.287)	−0.114 (.257)	−0.074 (.497)	−0.071 (.509)	−0.074 (.493)	−0.072 (.505)	−0.071 (.510)
CEO age	0.039 (.862)	0.002 (.964)	−0.043 (.330)	−0.043 (.332)	−0.043 (.334)	−0.042 (.337)	−0.043 (.333)
Compensation	0.064 (.008)	0.013 (.002)	0.020 (.000)	0.020 (.000)	0.019 (.000)	0.020 (.000)	0.020 (.000)
CEO awards	−0.004 (.762)	0.000 (.976)	−0.002 (.474)	−0.002 (.477)	−0.002 (.481)	−0.002 (.484)	−0.002 (.495)
CEO positive tone	0.529 (.000)	0.090 (.001)	0.109 (.000)	0.108 (.000)	0.109 (.000)	0.108 (.000)	0.108 (.000)
CEO suspiciousness	−0.084 (.021)	−0.015 (.020)	−0.016 (.026)	−0.016 (.023)	−0.016 (.025)	−0.016 (.023)	−0.016 (.023)
CEO history of suspiciousness	0.057 (.502)	0.011 (.465)	0.018 (.287)	0.018 (.277)	0.018 (.284)	0.018 (.274)	0.018 (.274)
CEO concrete language	−0.005 (.252)	−0.001 (.262)	−0.003 (.011)	−0.003 (.011)	−0.003 (.012)	−0.003 (.011)	−0.003 (.011)
CEO charisma	−0.596 (.455)	−0.096 (.496)	0.415 (.105)	0.420 (.101)	0.412 (.107)	0.417 (.103)	0.420 (.101)
CEO tentative language	−0.001 (.939)	0.000 (.937)	−0.008 (.087)	−0.008 (.088)	−0.008 (.087)	−0.008 (.088)	−0.008 (.087)

TABLE B2 (Continued)

Estimation approach	Model A1 Ordered probit	Model A2 Linear fixed effects	Model A3 Linear fixed effects	Model A4 Linear fixed effects	Model A5 Linear fixed effects	Model A6 Linear fixed effects	Model A7 Linear fixed effects
CEO clout	0.005 (.030)	0.001 (.030)	0.002 (.002)	0.002 (.002)	0.002 (.002)	0.002 (.002)	0.002 (.002)
CEO obfuscation	0.040 (.732)	0.009 (.661)	−0.016 (.524)	−0.016 (.524)	−0.016 (.528)	−0.016 (.527)	−0.016 (.529)
CEO word count	−0.001 (.981)	−0.001 (.881)	−0.009 (.154)	−0.009 (.160)	−0.009 (.153)	−0.009 (.159)	−0.009 (.158)
Number of prior All-Star recognitions	0.033 (.130)	0.008 (.049)	0.007 (.102)	0.007 (.101)	0.007 (.102)	0.007 (.101)	0.007 (.079)
Analyst tenure	−0.087 (.259)	−0.015 (.268)	−0.001 (.971)	−0.001 (.971)	−0.001 (.974)	−0.001 (.974)	0.000 (.986)
Regulatory exams	−0.139 (.006)	−0.023 (.009)	−0.022 (.020)	−0.022 (.019)	−0.022 (.020)	−0.022 (.020)	−0.022 (.020)
Analyst firms	−0.001 (.960)	−0.001 (.835)	0.002 (.647)	0.002 (.667)	0.002 (.651)	0.002 (.671)	0.002 (.675)
Days since last recommendation	−0.021 (.013)	−0.004 (.009)	−0.004 (.007)	−0.004 (.007)	−0.004 (.007)	−0.004 (.006)	−0.004 (.006)
Prior recommendation	2.655 (.000)	0.895 (.000)	0.896 (.000)	0.896 (.000)	0.896 (.000)	0.896 (.000)	0.896 (.000)
Number of analysts	0.003 (.177)	0.001 (.123)	0.001 (.055)	0.001 (.054)	0.001 (.055)	0.001 (.054)	0.001 (.054)
Standard deviation of recommendations	−0.204 (.001)	−0.026 (.018)	−0.020 (.091)	−0.020 (.088)	−0.020 (.091)	−0.020 (.088)	−0.020 (.089)
Change in median recommendation	0.508 (.000)	0.096 (.000)	0.099 (.000)	0.099 (.000)	0.099 (.000)	0.099 (.000)	0.099 (.000)
Forecast accuracy	−0.014 (.292)	−0.002 (.293)	−0.005 (.057)	−0.005 (.055)	−0.005 (.057)	−0.005 (.055)	−0.005 (.054)
Control function			−0.418 (.015)	−0.417 (.015)	−0.416 (.015)	−0.416 (.015)	−0.416 (.015)
Panel averages of all independent variables	X						
Month dummies	X	X	X	X	X	X	X
Year dummies	X	X	X	X	X	X	X
Cut 1[a]	−5.708 (.000)						
Cut 2	−3.200 (.000)						
Cut 3	0.733 (.012)						
Cut 4	3.348 (.000)						

TABLE B2 (Continued)

Estimation approach	Model A1 Ordered probit	Model A2 Linear fixed effects	Model A3 Linear fixed effects	Model A4 Linear fixed effects	Model A5 Linear fixed effects	Model A6 Linear fixed effects	Model A7 Linear fixed effects
N	64,107	64,107	64,107	64,107	64,107	64,107	64,107
R^2 within		0.795	0.795	0.795	0.795	0.795	0.795
Log-pseudolikelihood	−21,829.669						
AIC	43,879.340	−365.236	−365.430	−369.910	−366.725	−371.303	−371.915

Note: p values for coefficient significance in parenthesis. All p values calculated using standard errors robust to clustering at the CEO–analyst dyad level. Fixed effects at the CEO–analyst dyad level for linear fixed-effects estimation approach. DV = analyst recommendations.

[a]Estimated cut points in the ordered probit to differentiate recommendation type.

APPENDIX C: THE AVERAGE PROBABILITY OF AN UPGRADE AND COMPARISONS TO AVERAGE PREDICTED PROBABILITIES

Comparing changes in the average predicted probability of recommendation categories due to a change in a focal variable may understate the impact of the focal variable at the individual observation level. This is because averaging predicted probability of the recommendation category over all observations combines instances where the recommendation category is highly likely and highly unlikely.

Consider the following example to illustrate this issue. We average predicted recommendations involving a different analyst, firm, and period for two observations. For simplicity assume that at original variable values observation 1 yields 100% predicted probability of a Sell recommendation (0% everywhere else) while observation 2 yields 100% predicted probability of a Hold recommendation. We obtain an average predicted probability of 50% Sell and 50% Hold. Now assume we change a variable of interest and this shifts the predicted recommendation category to one greater by 100 percentage points in each observation: We obtain a 100% predicted probability of a Hold (up from Sell) for observation 1 and a 100% predicted probability of a Buy (up from Hold) for observation 2. The average predicted probability is now 50% Hold and 50% Buy. However, note that the predicted probability of an upgrade is 100 percentage points in each observation. Comparing average predicted probabilities, the change in the variable of interest results in a 50-percentage point increase in the average probability of a Buy recommendation. We do not capture the 100 percentage point increase in predicted probability of an upgrade for each observation by looking at average predicted probability because each observation started at a different reference recommendation from which to compare the impact of changing the variable of interest.

We obtain the probability of an upgrade by comparing how the distribution of the predicted probabilities of recommendation categories changes when we change a variable of interest for each observation, which is also the cumulative positive partial effect (Greene, 2017; Greene & Hensher, 2010; Wooldridge, 2010). This captures how a change in the variable of interest changes predicted probabilities of recommendation categories relative to the predicted probabilities calculated at original variable values in each observation (i.e., reference predicted probabilities). This allows us to assess the relative impact of changing a variable of interest in observations with different reference predicted probabilities, as in the above example. Averaging over all observations gives the average probability of an upgrade due to changing the focal variable.

To compare how the distribution of predicted probabilities of recommendation categories change for a change in a variable of interest, we only consider positive partial effects because the sum of the changes in recommendation probabilities for a single observation will be 0 since predicted probabilities of each recommendation category must sum to 1 (Greene, 2017; Greene & Hensher, 2010; Wooldridge, 2010). The direction of the change given by the cumulative positive partial effect is consistent with the sign of the coefficient on the variable of interest in the ordinal probit estimation by construction (Greene, 2017; Greene & Hensher, 2010; Wooldridge, 2010). The average cumulative positive partial effect captures the average percentage point change in predicted analysts' recommendations relative to their prior predictions. In our example above, the cumulative positive partial effect is 100 percentage points in each observation. The average cumulative positive partial effect shows that the average probability of an upgrade in each observation is 100 probability points, which is our main quantity of interest.

APPENDIX D: ADDITIONAL ROBUSTNESS CHECKS

TABLE D1 Matched sample robustness check—Firms obtain superior recommendations during fraudulent quarters compared to nonfraudulent quarters.

	Model A8 Ordered probit			Model A9 Ordered probit		
Estimation approach	***β***	**SE**	***p***	***β***	**SE**	***p***
AAER violation[a]	.484	0.097	.000	.471	0.092	.000
AAER violation × All-Star analyst				.390	0.158	.013
All-Star analyst	−.005	0.081	.948	−.049	0.088	.579
ROE	.001	0.000	.004	.001	0.000	.005
Firm size	−.004	0.013	.760	−.004	0.013	.775
Earnings surprise	−.113	0.027	.000	−.111	0.027	.000
CEO tenure	.003	0.003	.284	.003	0.003	.301
CEO age	−.003	0.003	.354	−.003	0.003	.373
Compensation	.074	0.022	.001	.074	0.022	.001
CEO awards	.055	0.023	.018	.054	0.023	.019
CEO positive tone	.047	0.025	.062	.048	0.025	.058
CEO suspiciousness	.024	0.079	.762	.022	0.079	.782
CEO history of suspiciousness	.006	0.107	.959	.006	0.107	.957
CEO concrete language	.001	0.007	.912	.001	0.007	.911
CEO charisma	−2.709	1.174	.021	−2.711	1.175	.021
CEO tentative language	−.001	0.021	.970	−.001	0.021	.958
CEO clout	.002	0.003	.555	.002	0.003	.599
CEO obfuscation	.007	0.153	.966	.007	0.152	.964
CEO word count	−.055	0.042	.188	−.057	0.042	.171
Number of prior All-Star recognitions	−.006	0.003	.039	−.005	0.003	.124
Analyst tenure	.007	0.003	.016	.007	0.003	.015
Regulatory exams	−.010	0.017	.551	−.009	0.017	.600
Analyst firms	0.003	0.006	.583	.003	0.006	.548
Days since last recommendation	0.033	0.020	.109	.032	0.020	.114
Prior recommendation	0.127	0.021	.000	.126	0.021	.000
Number of analysts	0.000	0.004	.979	.000	0.004	.992
Standard deviation of recommendations	−0.173	0.067	.010	−.173	0.067	.010
Change in median recommendation	−0.106	0.037	.004	−.107	0.037	.004
Forecast accuracy	0.015	0.011	.173	.015	0.011	.180
Month dummies	X	X	X	X	X	X
Year dummies	X	X	X	X	X	X
Cut 1[b]	−1.593	0.659		−1.580	0.647	
Cut 2	−0.784	0.656		−.772	0.644	

TABLE D1 (Continued)

	Model A8 Ordered probit			Model A9 Ordered probit		
Estimation approach	***β***	**SE**	***p***	***β***	**SE**	***p***
Cut 3	0.687	0.659		.700	0.648	
Cut 4	1.404	0.661		1.418	0.649	
N	26,084			26,084		
Log-pseudolikelihood	−33,443.8			−33,433		
AIC	66,985.66			66,965.8		

Note: Standard errors and associated *p* values are robust to clustering at the firm level. DV = analyst recommendations.
[a]AAER Violation: dummy variable = 1 for the CEO in the period an AAER violation occurred, 0 otherwise.
[b]Estimated cut points in the ordered probit to differentiate recommendation type.

TABLE D2 Cox proportional hazard robustness check—History of deception increases the hazard that an analyst drops coverage.

	Model A10		
	Cox proportional hazard		
Estimation approach	**Hazard ratio**	**SE**	***p***
Deception	0.910	0.047	.064
History of deception	1.282	0.114	.005
All-Star analyst	0.695	0.043	.000
ROE	0.520	0.357	.341
Firm size	1.007	0.010	.451
Earnings surprise	0.998	0.030	.949
CEO tenure	1.011	0.002	.000
CEO age	1.002	0.002	.332
Compensation	0.948	0.020	.009
CEO awards	1.016	0.010	.088
CEO positive tone	0.989	0.213	.961
CEO suspiciousness	1.131	0.068	.042
CEO history of suspiciousness	0.726	0.059	.000
CEO concrete language	1.002	0.006	.788
CEO charisma	1.046	1.105	.966
CEO tentative language	1.007	0.023	.755
CEO clout	1.003	0.003	.326
CEO obfuscation	0.892	0.173	.558
CEO word count	0.990	0.030	.730
Number of prior All-Star recognitions	1.013	0.018	.463
Analyst tenure	0.991	0.003	.002
Regulatory exams	1.020	0.016	.223
Analyst firms	1.017	0.007	.017
Days since last recommendation	1.682	0.021	.000
Prior recommendation	0.945	0.014	.000
Number of analysts	1.002	0.002	.226
Standard deviation of recommendations	1.042	0.068	.532
Change in median recommendation	0.957	0.047	.370
Forecast accuracy	1.014	0.024	.566
Month dummies	X	X	X
Year dummies	X	X	X
N	64,107		
Log-pseudolikelihood	−44,592.87		0
AIC	89,218.6		

Note: Standard errors and associated *p* values are robust to clustering at the CEO–analyst dyad level. DV = analyst drops coverage of the firm.

TABLE D3 OLS regression removing all instances where an analyst previously recommended either a "strong buy" or "strong sell" position and did not change their current recommendation—Deceptive CEO obtained superior recommendations compared to honest CEOs.

	Model A11		
	OLS		
	β	SE	*p*
Deception	.004	0.006	.475
Deception × History of deception	−.081	0.030	.007
Deception × All-Star analyst	.011	0.012	.355
Deception × History of deception × All-Star analyst	.172	0.062	.006
History of deception	.046	0.022	.035
All-Star analyst	−.003	0.009	.755
History of deception × All-Star analyst	.004	0.035	.906
ROE	.364	0.095	.000
Firm size	.013	0.011	.216
Earnings surprise	.013	0.003	.000
CEO tenure	−.111	0.103	.282
CEO age	.013	0.045	.777
Compensation	.013	0.004	.003
CEO awards	.000	0.003	.937
CEO positive tone	.091	0.027	.001
CEO suspiciousness	−0.015	0.007	.024
CEO history of suspiciousness	.010	0.016	.538
CEO concrete language	−.001	0.001	.321
CEO charisma	−.078	0.144	.587
CEO tentative language	.000	0.003	.914
CEO clout	.001	0.000	.027
CEO obfuscation	.013	0.022	.560
CEO word count	−.001	0.005	.898
Number of prior All-Star recognitions	.007	0.004	.110
Analyst tenure	−.014	0.014	.322
Regulatory exams	−.022	0.009	.017
Analyst firms	−.002	0.005	.689
Days since last recommendation	−.004	0.002	.007
Prior recommendation	.894	0.004	.000
Number of analysts	.001	0.000	.073
Standard deviation of recommendations	−.027	0.011	.017
Change in median recommendation	.099	0.007	.000
Forecast accuracy	−.002	0.002	.330
Panel averages of all independent variables	X	X	X

TABLE D3 (Continued)

	Model A11		
	OLS		
	β	SE	p
Month dummies	X	X	X
Year dummies	X	X	X
N	62,342		
R^2(within)	.790		

Note: Standard errors and associated p values are robust to clustering at the CEO–analyst dyad level. DV = analyst recommendations.