Doing Organizational Identity: Earnings Surprises and the Performative Atypicality Premium *

Paul Gouvard, Università della Svizzera italiana

Amir Goldberg, Stanford University

Sameer B. Srivastava, University of California, Berkeley

April 14, 2023

Abstract

How do organizations reconcile the cross-pressures of conformity and differentiation? Existing research predominantly conceptualizes identity as something an organization has by virtue of the products or services it offers. Drawing on constructivist theories, we argue that identity is also dynamically produced through organizational members' interactions with external audiences. We term the extent to which such interactions diverge from audience expectations *performative atypicality*. Applying a novel deep-learning method to conversational text in over 90,000 earnings calls, we find that performative atypicality leads to an evaluation premium by securities analysts, paradoxically resulting in a negative earnings surprise. Moreover, performances that correspond to those of celebrated innovators are received with higher enthusiasm. Our findings suggest that firms that conform to categorical expectations while being performatively atypical can navigate the conflicting demands of similarity and uniqueness, especially if they hew to popular notions of being different.

*We thank participants of the Duke University, Fuqua School of Business Finance Seminar, the Berkeley Haas Culture Conference and the Nagymaros Conference for helpful comments and feedback. The usual disclaimer applies. Direct all correspondence to Paul Gouvard: gouvap@usi.ch.

Target and Trader Joe's are among the most successful retailers in the U.S. Established in 1962, Target is a big box department store chain that sells a wide array of products, ranging from clothes to electronics. Defying the conventional distinction between department and grocery stores, Target began selling fresh produce in 2009, aggressively competing with traditional supermarkets.[1] Trader Joe's, in contrast, focuses almost exclusively on selling groceries. Known for its groovy surfer-inspired store decor, it is also famous for having a cult-like following. Challenging industry norms, moreover, Trader Joe's staff wear Hawaiian shirts and are encouraged to engage in playful conversation with customers.

Existing organizational theory would consider both Target and Trader Joe's somewhat atypical organizations. Different theoretical perspectives would make different predictions about whether this atypicality should appeal to these organizations' intended audiences. Whereas research on the role of categories in markets would likely emphasize the penalties associated with these retailers' categorical noncompliance, optimal distinctiveness theory might predict that these firms' moderate levels of atypicality should result in successful differentiation (Zuckerman, 2016; Zhao, Fisher, Lounsbury, and Miller, 2017; Haans, 2019).

But Target and Trader Joe's are atypical in different ways. Target, especially when it first launched this strategy, differed from its competitors in the kind of products it offers. Selling merchandise that one would find either in a typical department store or supermarket, it provides an unconventional mix of offerings. We refer to this form of noncompliance as *categorical atypicality*. Trader Joe's, on the other hand, differs from its competitors not in what it sells but in how it interacts with outside stakeholders. We term this type of divergence from expectations *performative atypicality*.

The analytical distinction between categorical and performative atypicality is consequential if outside observers react to them differently. To investigate this possibility, we examine securities analysts' reactions to executives' performative atypicality in quarterly earnings calls. Using word embedding models (Mikolov, Sutskever, Chen, Corrado, and Dean, 2013), we develop a method for measuring performative atypicality in these calls' transcripts. Consistent with prior work, we find that performative atypicality breeds disagreement. Yet, contra the predictions of existing theories, we find that performatively atypical organizations receive a valuation boost from differentiation: analysts overestimate these firms' future earnings. We refer to this advantage as the *performative atypicality premium*. Ironically, this premium leads to an adverse outcome—a negative earnings surprise.

Drawing on abductive reasoning (Brandt and Timmermans, 2021), we exploit the possibilities afforded by automated textual analysis to investigate this seemingly surprising finding. Taking a forensic computational approach (McFarland, Lewis, and Goldberg, 2016; Goldberg, 2015), we find that not all forms of performative atypicality generate equal levels of optimism. Rather, firms whose atypicality emulates perceived innovators' performances are evaluated more positively than those that are idiosyncratically atypical. Difference, in other words, is especially rewarded when it conforms to popular expectations about what constitutes novelty.

The remainder of this article is structured as follows. First, we provide a theoretical motivation for the distinction between categorical and performative atypicality. Next, we explain in detail how we measure performative atypicality and what data we use for that purpose. We then present our empirical results. We first demonstrate that performative atypicality is analytically distinct from categorical atypicality (using the conventional analyst overlap method for measuring the latter). Second, we show that, all other things being equal, analysts are overly optimistic about the future earnings of performatively atypical firms. This effect holds whether using between- or within-firm models. Third, taking inspiration from Nelson's (2020) computational grounded theory approach, we inductively chart the semantic dimensions that structure atypical performances.

In the final section of this article, we discuss the theoretical implications of our findings. We argue that they imply a bilateral process, whereby the categorical and performative aspects of an organization's identity catalyze different processes of audience valuation. While audiences penalize categorically atypical organizations, they interpret performative atypicality as a source of competitive advantage. This, we contend, relates to two primary dimensions along which an organization is evaluated: how unusual it is in *what* it produces and in *how* it produces those items. We discuss the scope conditions of this model and its implications for our understanding of the conditions under which being different is penalized or rewarded in markets.

THEORY

Unidimensional Conceptions of Atypicality

How can firms balance the pressures of conformity and differentiation? Existing literature provides two dominant explanations. The first, drawing on Brewer's (1991) "optimal distinctiveness" imagery, argues that organizations gain positive attention when they are moderately different from their competitors. Organizations need to conform to fundamental audience expectations in order to gain legitimacy. Those that manage to do so while remaining distinct from their competitors are judged favorably by outside audiences (Navis and Glynn, 2011). Research relying on the optimal distinctiveness framework normally assumes that external evaluators' judgments about legitimacy and distinctiveness occur simultaneously. It therefore predicts an inverted U-shaped relationship between atypicality and success, such that organizations poised between full conformity and radical deviance gain positive evaluations from outside audiences (Askin and Mauskapf, 2017; Uzzi, Mukherjee, Stringer, and Jones, 2013).

A second approach, focusing on the role of categories in structuring market activity, highlights the disciplining roles of categorical expectations. This work commonly assumes a sequential two-stage process of valuation (Zuckerman, 2016). In the first stage, audience members associate an organization with a recognized category. This association determines which criteria will be used to evaluate the organization and, importantly, what reference group it will be compared to. Only in the second stage, once an organization's categorical identity has been established, do audience members evaluate the extent to which it is distinct from its competitors.

This two-stage process has been used to explain why categorically atypical organizations, especially those that straddle multiple categories, suffer negative consequences. While audiences generally seek and favor distinct organizations, they evaluate such distinction positively only if they can make sense of these organizations. When external evaluators are confused about the categorical identity of an organization, they find it difficult to interpret its performance and to compare it to others. Consequently, categorically atypical organizations, despite their potential appeal, are systematically penalized (Zuckerman, 2017). Although the two-stage model explains how an organization can—in theory—be simultaneously compliant and differentiated, it gives theoretical precedence to categorical conformity in the first stage.

Whereas optimal distinctiveness theory assumes simultaneous evaluation, the two-stage model assumes that evaluators first determine an organization's categorical identity. Consequently, these two theories make different predictions. While optimal distinctiveness predicts a curvilinear relationship between atypicality and audience appeal, the two-stage model predicts that this relationship is monotonically decreasing.

Empirical investigations offer a frustratingly diverse set of mixed, often contradictory, results. Consistent with the two-stage model's linear prediction, a large body of work demonstrates that products and organizations that do not adhere to typical categorical expectations have, on average, lower appeal and exhibit weaker performance across a variety of contexts (Zuckerman, 1999; Leung and Sharkey, 2013). Other studies find support for the optimal distinctiveness model's prediction, demonstrating that products (e.g., Askin and Mauskapf, 2017; Zhao et al., 2017) and organizations (e.g., Deephouse, 1999) that are moderately differentiated are rewarded for their optimal distinctiveness. Neither theory can explain why, as various studies show, atypicality is often rewarded (e.g., Taeuscher, Bouncken, and Pesch, 2021).

In attempting to address this gap, scholars have proposed several mechanisms explaining when identity expectations are stringently enforced and when they are relaxed. The first argues that different audiences have differing levels of tolerance for atypicality because they subscribe to different theories of value (Paolella and Durand, 2015; Cattani, Ferriani, and Allison, 2014). Venture capital firms, for example, see greater value in atypicality than institutional investors do (Pontikes, 2012). Second, expectations of typicality vary by market and domain (Chatterji, Luo, and Seamans, 2021; Carnabuci, Operti, and Kovács, 2015; Zhao, Ishihara, Jennings, and Lounsbury, 2018; Keuschnigg and Wimmer, 2017). The penalties for atypicality are especially muted in emergent (Ruef and Patterson, 2009) and homogeneous markets (Haans, 2019), or for early-stage firms (Taeuscher et al., 2021). Finally, different firms operating within the same market and being evaluated by the same audience might still be rewarded differently for atypicality, depending on their reputation and status. High-status firms enjoy greater latitude to defy categorical conventions (Rao, Monin, and Durand, 2005; Smith, 2011; Sgourev and Althuizen, 2014; Durand and Kremp, 2015).

These mechanisms delineate how atypicality is rewarded in some contexts and penalized in others. But they do not explain why some organizations are more successful than their competitors in managing the conflicting demands of similarity and differentiation when they operate in the same market, cater to the same audience, and have access to similar reputational resources. This limitation, we contend, relates to the unidimensional way by which scholars, regardless of theoretical orientation, have tended to conceptualize organizational atypicality.

A unidimensional conceptualization of atypicality effectively assumes that audiences perceive organizations using a singular taxonomical system. Such an assumption necessarily implies that, at any given moment in time, an organization occupies a fixed location on the atypicality continuum. This means that organizations cannot simultaneously enjoy the benefits of intelligibility that come with typicality and the benefits of differentiation that come with atypicality. Optimal distinctiveness and the two-stage model reach different conclusions about how this tension is resolved.

But human cognition is messier and more complex. People do not perceive objects through unitary taxonomical lenses. Rather, they cognitively represent the world along a multitude of intersecting semantic dimensions (Hannan, Mens, Hsu, Kovács, Negro, Pólos, Pontikes, and Sharkey, 2019; Murphy, 2004). Animals, for example, are not only understood as belonging to different species. As Grand, Blank, Pereira, and Fedorenko (2022) show, people also conceptually sort them along other dimensions such as size, wetness, or how dangerous they are.

Organizations are also interpreted along multiple semantic dimensions. Indeed, researchers studying atypicality in markets often concede that successful organizations only differentiate themselves along a small subset of dimensions (Zhao et al., 2017; Zuckerman, 2016). Nevertheless, such accounts seldom specify what makes some dimensions more conducive to differentiation than others. In practice, moreover, empirical investigations almost exclusively operationalize atypicality as a unidimensional construct. These studies often define atypicality as the overall difference between a firm's product or service features, relative to its competitors (e.g., Askin and Mauskapf, 2017), or the extent to which audiences classify it as spanning multiple categories (e.g., Carnabuci et al., 2015; Goldberg, Hannan, and Kovács, 2016a). When they distinguish favorable and unfavorable dimensions of differentiation, such distinctions are mostly non-generalizable beyond the specific context being studied (e.g. Phillips, Turco, and Zuckerman, 2013; Wry, Lounsbury, and Jennings, 2014). These studies therefore fall short of proposing general principles that delineate which organizational features are important for conformity and which are amenable to differentiation.

Categorical and Performative Atypicality

We argue, in contrast, that audiences determine an organization's identity, and concomitantly infer its atypicality, in two different ways.[2] These relate to two different and mostly tangential sociological approaches to the study of atypicality and its consequences.

The first, which has been widely influential in research on organizations, conceptualizes organizational identity through a *categorical* lens (Zuckerman, 1999; Hannan et al., 2019). This approach understands sense-making as a classificatory process, wherein external observers, drawing on a shared set of taxonomical criteria, divide organizations into distinct groups of similar entities. Prototypical membership in these groups is mutually exclusive: a typical restaurant, for example, is distinctively different from a typical hospital. Organizations that exhibit feature combinations that crosscut categorical boundaries are difficult to classify. We refer to this kind of multi-category membership as *categorical atypicality*.

A categorical approach to atypicality has two implications. First, it orients researchers toward an organization's primary attributes, most commonly those that relate to the products it makes or the services it provides. A restaurant, for example, is defined first and foremost by the fact that it serves food, whereas a hospital's definition is rooted in the services it provides to people in medical need. Second, because it anchors on these primary attributes, a categorical approach tends to see organizational identity as static. Although firms can change their products and business scope, this evolution is mostly incremental and slow. Categorical identity is consequently stable or slow-changing throughout an organization's lifetime.

An alternative approach hails from *constructivist* social identity theories (Berger and Luckmann, 1967), specifically those that emphasize the performative nature of social interaction. Originally applied to gender (West and Zimmerman, 1987) and later extended to social identity more broadly (West and Fenstermaker, 1995), this approach maintains that identity is not a fixed designation but rather an attribution that is established repeatedly through interaction. Unlike the categorical approach, which focuses on fixed attributes, this perspective emphasizes the dynamic and emergent aspects of identity. To be understood by others as having a specific identity—for example, a woman, an economist, or an Evangelical—one's interactional performances need to conform to audiences' expectations about how such an identity is behaviorally enacted. An identity, in other words, is not something one innately has but something one continuously does. Performances that diverge from expectations—for example, a woman exhibiting stereotypically masculine behaviors or an economist behaving like a sociologist—are identity inconsistent. We refer to this type of incongruence as *performative atypicality*.

We argue that, like individual actors, organizations are subject to evaluations of performative atypicality. Indeed, research on organizational identity often analogizes it to how persons construct their self-identity, wherein members of the organization formulate an answer to the question "who are we?" (Albert and Whetten, 1985; Whetten, 2006). Early work in this vein emphasized the enduring aspects of organizational identity. A more recent stream has questioned the assumption of stability, examining instead how organizational identities shift and evolve over time. This work builds on the premise, grounded in symbolic interactionism (Mead, 1934; Goffman, 1959), that individual identity is constructed through interpersonal interaction (e.g., Ibarra and Barbulescu, 2010), and extends it to organizational identity formation (Gioia, Patvardhan, Hamilton, and Corley, 2013; Schultz and Hernes, 2012).

We shift focus from organizational members' to outsiders' perceptions, contending that a similar dynamism extends to how external evaluators form impressions of an organization. Such impressions are formed not only from the attributes of the products or services these organizations offer. Rather, external evaluations also arise through routine interactions between external audiences and organizational members.[3] Whether introducing a new product at a trade show, responding to questions from customers, or participating in a quarterly earnings call with financial analysts, organizational members are engaged in a meaning-laden social performance with external audiences.

Such performances mostly communicate literal information about current or anticipated future performance such as sales forecasts, new products in development, leadership transitions, and impending mergers or divestitures. Performers' subtle and often unconscious word or behavioral choices also convey a wide range of connotative meanings that are not explicitly communicated. These connotative meanings shape audiences' high-level interpretations of speakers' discursive performances. This is where implicit and culturally shared schemas are being invoked (Zilber, 2006).

For example, when Tesla's iconoclast CEO Elon Musk repudiated "moats" in a controversial earnings call in May 2018, audiences interpreted his comments as a rejection of a strategy that is focused on sustaining competitive advantage. Musk was communicating to investors that his company is, instead, pursuing a strategy of dynamic innovation.[4] Recent research demonstrates that connotative meanings communicated in language implicitly affect audience evaluations. The use of generic language in academic abstracts, for example, increases readers' perceptions of the research's importance, holding its substantive content constant (DeJesus, Callanan, Solis, and Gelman, 2019). Similarly, reaffirmations of monetary assumptions in The Federal Reserve Chair's speeches counterintuitively lead investors to question these assumptions, resulting in increased market uncertainty (Harmon, 2019).

Although we empirically focus below on quarterly earnings calls, it is important to note that an organization's performative atypicality is not only communicated by its top executives. Rather, it is manifest in a variety of media, from everyday interactions between employees and outside stakeholders, to the organization's aesthetic and architectural choices (Wasserman and Frenkel, 2011). The personal and unscripted conversations, for example, that call center representatives at Zappos are trained to conduct with customers, are a far cry from the structured and formal experiences characteristic of conventional customer service exchanges. They connote the online shoe retailer's nontraditional customer-focused approach to retail.

The "What" and "How" of Organizational Identity

Categorical and performative atypicality, we contend, correspond to different aspects of sense-making. The former relates predominantly to inferences that outside observers make about *what kind* of an organization a firm is and, consequently, who its competitors are. Categorical atypicality, in other words, relates to the constitutive elements of a firm's identity.

Performative atypicality, in contrast, relates to inferences about *how* an organization goes about doing what it does. These might be fundamental to how it operates, but not to what it, in essence, is. Tesla, for example, would still be seen as an electric vehicle company even if its CEO were to step down and his performatively atypical antics were replaced with more conventional behaviors. But if the company were to shift from manufacturing cars to manufacturing office furniture, its categorical identity would have shifted, irrespective of these antics.

Similarly, organizations can be performatively atypical if they interact with stakeholders in ways that are inconsistent with how their competitors interact, even if they are categorically typical—that is, their products are similar to their competitors'. The British airline Virgin Atlantic, for example, founded in 1984, competes in a clearly defined market with relatively limited heterogeneity. There is no confusion about the type of services that the airline provides or who its competitors are. Nevertheless, Virgin Atlantic's interactions with outside audiences, especially in its early years, have been quite atypical. This atypicality is personified in the public performance of its CEO, Richard Branson, whose adventurous personality stands in stark contrast to the formality of traditional airlines. This informality is prominently reflected in Virgin Atlantic's casual customer service philosophy and playful aircraft design, connoting the airline's unique strategic position (Navis and Glynn, 2011).

Of course, the distinction between the "what" and "how" of organizational identity is much crisper as an analytical abstraction than it is experienced in people's messy cognition.[5] What sets these two inferential processes apart are the different *mediums* on which they depend. An organization's categorical identity is directly inferred from the products and services it offers. Its performative atypicality, in turn, is evaluated on the basis of its communicative interactions with outside stakeholders. These stakeholders do not directly observe how the organization operates.

Rather, they infer the "how" from the meanings connoted performatively.

The distinct ways by which categorical and performative atypicality arise lead to two important differences between them. First, whereas the "what" is directly gleaned from the organization's products and services, the "how" is indirectly inferred from interactional performances. An organization's performative atypicality therefore corresponds to its perceived, but not necessarily objective, operational uniqueness. Second, unlike categorical atypicality, performative atypicality is dynamically produced and is therefore more likely to fluctuate over an organization's lifespan. This does not mean that an organization's performative atypicality is necessarily unstable. Nevertheless, this dynamism suggests that an organization's performative atypicality can change significantly, and dramatically, over time.

Responses to Performative Atypicality

When evaluating an organization, outside audiences are concerned with its quality. Customers seek to ascertain the quality of its products or services, whereas investors aspire to evaluate its potential financial performance. Ultimately, quality judgments depend on an organization's value proposition. Outside stakeholders draw on the various pieces of information available to them in making inferences about that value proposition. How does performative atypicality factor into this process?

A straightforward application of constructivist identity theory to organizational identity would predict that performatively atypical organizations are significantly devalued by external audiences. Indeed, individual atypical performances, such as gender noncompliant behaviors, are normally strongly frowned upon. There are, of course, fundamental differences between how people understand gender and how they interpret organizational identity. Nevertheless, two assumptions motivating the "doing identity" framework appear to be largely applicable to an organizational context.

First, as West and Zimmerman (1987) argue, because identity is fluid, it needs to be continuously displayed. Identity-incongruent performances therefore undermine the audience's perception about the actor's claimed identity. When the actor is an organization, this can lead to skepticism about its ability to perform economically. An eccentric airline like Virgin Atlantic, for example, whose executives at times speak as if they are running an entertainment company, might be perceived as lacking the capabilities necessary for managing a complex aviation fleet (Hsu, 2006; Hsu, Hannan, and Koçak, 2009). Second, performances that defy behavioral expectations also undercut perceived boundaries between different types of organizations, and the markets they operate in. These boundaries are essential cognitive tools that people use for imposing order on an otherwise unstructured terrain of producers. Audiences will therefore react with dismay when atypical performances appear to erode these categorical distinctions (Hollander, 2013).

There are, however, reasons to doubt these negative expectations. Investors, like the ones we investigate below, are primarily motivated by value maximization. They often perceive uniqueness and nonconformity as indications of such value (Durand and Calori, 2006; Haans, 2019). Being different is a source of advantage in markets because it makes an organization distinct in the eyes of audiences (Barney, 1991; Deephouse, 1999). Investors should reward performatively atypical organizations to the extent that they perceive this atypicality as an indication of unique and difficult to imitate capabilities.

Consider Trader Joe's again. In a rare public comment, the supermarket chain's CEO recently reacted to a podcast titled "Should America be Run by Trader Joe's?" "We are pretty sure such work would likely require a coat and tie," he responded, "we like Hawaiian shirts...so we will pass."[6] The company's unusual style—from store decor to executives' public performances—symbolizes its unorthodox customer-focused strategy which refrains from discounts, advertising, or data-driven targeting. Its leadership's willingness to challenge industry conventions appears to signal this unique strategy. If outside observers indeed interpret the CEO's unconventional behavior as an indication of such a strategy, it should lead them to a evaluate Trader Joe's favorably.

Performative Atypicality and Analyst Predictions

To evaluate whether audiences interpret performative atypicality as an indication of organizational incompetence or as a signal of its unparalleled capabilities, we focus our attention on securities analysts. As demonstrated by a range of scholars (Zuckerman, 1999, 2000; Bowers, 2014; Smith, 2011), investors and analysts strongly rely on categorical distinctions when evaluating firms. They are therefore highly sensitized to instances of atypicality. Investigating analysts' reactions has two advantages for our purposes.

First, financial analysts occupy a cross-pressured position in financial markets: They are simultaneously motivated to enforce normative behaviors and reward nonconformity. Analysts rely on established industry categories to cluster firms and thus are often presented as enforcers of the market order (Zuckerman, 1999, 2004). Yet they gain recognition and status based on their ability to introduce novelty in their reports and, in particular, new or emerging categories (Giorgi and Weber, 2015; Pontikes and Kim, 2017). Analysts can therefore benefit from adopting behaviors akin to that of "market makers" (Pontikes, 2012) as they risk losing ground to their peers if they fail to identify "the next big thing." Navigating these contradictory pressures, analysts' predictions provide fertile ground for exploring the implications of performative atypicality.

Second, analysts cover a broad diversity of industries and market contexts, effectively analyzing the full range of activity in the U.S. for-profit economy. Their estimates are not limited to specific market contexts. In our analyses below, we account for this variation. This enables us to evaluate the relationship between performative atypicality and audience reactions while holding constant market dynamics and the audience's theory of value. Moreover, we can hold constant resources (such as reputation) that are uniquely available to a given firm by observing it over time.

Firms can perform their identities in various forms and media, ranging from formal documents submitted to regulatory agencies to stylistic signals made through subtle office design choices. To derive performative atypicality, we focus on quarterly earnings calls: periodic calls that the management teams of most publicly traded firms in the U.S. hold with the financial analysts who cover their stocks. During these calls, managers discuss their recent financial performance, as well as their strategy and prospects for the future. Calls typically unfold in two stages: managers first read prepared statements and then engage in a more informal question and answer (Q&A) session. By all accounts, quarterly earnings calls are highly scripted, tightly controlled, and ritualized (Lee, 2016). Yet managers often reveal new or unexpected information—either deliberately or inadvertently—as they interact with each other and with analysts. Overall, speakers convey both conscious and unselfconscious meanings about the organization.

We evaluate analysts' reactions in two different ways. First, if atypicality affects audiences' perceptions, it should beget uncertainty and ambiguity. Unlike categorical ambiguity, however, performative ambiguity arises not because audiences are unable to identify what kind of organization the one under consideration is or who its competitors are. Rather, ambiguity emerges precisely because the organization communicates meanings that are inconsistent with those typically communicated by similar organizations. These unusual meanings make it more likely that different analysts will reach different conclusions about the firm's future performance. Insofar as analysts pay attention not only to tangible data and facts but also to the subtleties of word choice, performative atypicality should result in greater disagreement in analysts' earnings forecasts.

Second, we focus on earnings surprises, the extent to which a firm's reported quarterly profits diverge from median analyst expectations. As work in accounting and finance demonstrates, deviations from analyst forecasts affect future valuations and are commonly interpreted as a reflection of information-flow inefficiency in the market (e.g., Kasznik and McNichols, 2002). When making their predictions, analysts presumably take into account the variety of information—especially hard data relating to performance—available about a firm. An earnings surprise corresponds to a bias in analysts' estimations above and beyond this information. A positive (negative) earnings surprise occurs when analysts, on average, underestimate (overestimate) a firm's future performance. Systematic prediction error driven by performative atypicality, we argue, indicates that analysts rely on this atypicality to make inferences about a firm's underlying quality.

Previous work on atypicality and firm valuation has tended to focus on investment flows (e.g., Smith, 2011). Because these studies seek to estimate the categorical atypicality discount above and beyond firm fundamentals, they typically employ complex methods of taking these fundamentals into account (e.g., excess value calculations in Zuckerman [1999]). Earnings surprises obviate this need. Analyst performance predictions presumably take into account these analysts' perceptions of how firm fundamentals should affect future performance. The earnings surprise represents the extent to which this consensus estimation is biased.

DATA AND METHODS

Data

Our data, which come from Seeking Alpha (https://seekingalpha.com/), include 99,307 transcripts of quarterly earnings calls for 5,986 firms from 2008 to 2016. We trained a word embedding model (described in greater detail below) on the text of these calls to develop quarterly measures of performative atypicality for each firm. We then merged our measures of performative atypicality with analyst estimates from the Institutional Brokers' Estimate System (I/B/E/S, using unadjusted data) to derive our dependent variables and with firm performance data from Compustat. To model analyst reactions, we use firm-quarter observations for which we could measure performative atypicality, our dependent variables (earnings surprise and analyst disagreement), as well as a host of additional control variables described below. To ensure that our estimates are not driven by outliers or especially small firms and consistent with standard practice, we winsorize analyst disagreement at the top 99% level and earnings surprise as the 99% level on both ends and remove observations for firms' whose stock price is less than \$1 or whose book value is less that \$5M. This results in a total of 61,670 firm-quarter observations.

Measuring Performative Atypicality

Word Embedding Models

We derived our measure of performative atypicality using word embedding models, a neural network-based unsupervised machine learning method for representing words in a high-dimensional vector space. These models are especially well-suited to analyzing connotative information in conversational text and are inspired by the *distributional hypothesis*, which states that the meaning of a word depends on the contexts in which it appears (Harris, 1954; Lenci, 2018). The approach we use in this study relies on the continuous bag-of-words (CBOW) method, wherein a two-layer neural network is trained to predict a word based on its surrounding words (Mikolov et al., 2013). Each word is then projected to a location in a shared vector space with several hundred dimensions. Although these dimensions are often uninterpretable to human observers, the resulting vectors are generally found to capture meaningful semantic relations between words, such that the distance between two words in this high-dimensional space inversely corresponds to their semantic similarity (Mikolov et al., 2013).

Word embedding models are especially useful for our purposes as they are effective at capturing connotative meanings above and beyond the literal meanings of words (Lix, Goldberg, Srivastava, and Valentine, 2022; Kozlowski, Taddy, and Evans, 2019). Previous work demonstrates that implicit gendered associations in the meanings of various occupations track with these occupations' historical gender compositions (Garg, Schiebinger, Jurafsky, and Zou, 2018) or that different lifestyle activities invoke class, race, and gender identities (Kozlowski et al., 2019). These studies identify specific dimensions of meaning—gender, class or race—by measuring the distance between a focal word and exemplars in the relevant meaning dimension (e.g., "woman"). Because we are not focused on specific words or specific dimensions of meaning, we employ a different approach, wherein we measure the similarity between two earnings calls as the distance between their centroids (averaged across all words in each call) in embedding space. This captures the overall similarity in meanings being conveyed in the two calls.

To illustrate the advantage of our approach, consider a situation in which we have three real estate firms—A, B, and C—and three words in the vocabulary—"office," "space," and "personality." Assume further that Firm A uses only the word "office" in its transcript, that Firm B uses both the words "office" and "space" in equal proportions, and that Firm C uses both the words "office" and "personality" in equal proportions. A simple frequentist approach that does not take into account the semantic relationships between words would find that the calls of Firm B and C have the same level of similarity to the call of Firm A. Yet Firm A ought to be considered closer to Firm B than to Firm C given the semantic dissimilarity between "personality" and "office" or "space" relative to the latter two's similarity. Firm C's vocabulary carries meanings that are not common in real estate parlance.

We pre-processed each transcript following usual guidelines in natural language processing (i.e., removing digits, punctuation, and stopwords and then tokenizing the text). After preprocessing, we trained word embedding models on a quarterly basis to account for potential shifts in word meanings that may have occurred over our observation period (Hamilton, Leskovec, and Jurafsky, 2016). Specifically, for each quarter, we trained a model on transcripts representing calls that took place in the focal quarter or in the three preceding ones. For example, the model for Q4 2016 was trained on transcripts of earnings calls that occurred between Q1 2016 and Q4 2016. We use quarter-specific vocabularies containing 10,000 words each. We then represented firms within this semantic space and derived a measure of performative atypicality by considering each firm's distance in this space from its competitors. We provide validations of the word embedding models in Appendix A.

Measure Construction

To measure performative atypicality, we first represented each transcript as the sum over the words it contains of each word's embedding vector by the word's frequency in the transcript. Let $f \in F$ index firms, $q \in Q$ index quarters, and $C_{f,q}$ denote a quarterly earnings call for firm f at quarter q. We represent each call's embedding centroid as follows:

$$V_{f,q} = \sum_{w \in C_{f,q}} W_{f,q}(w) \cdot V_{w,q} \tag{1}$$

where $V_{w,q}$ is the embedding vector for word w at time q and $W_{f,q}(w)$ is the proportion of word w in document $C_{f,q}$.

The centroid $V_{f,q}$ represents the firm's location in embedding space at the time of the earnings call. To evaluate the firm's typicality relative to categorically similar competitors, we measure the distance between this centroid and the centroid of all peer firms in the preceding three quarters as follows:

$$PV_{f,q} = \frac{1}{|P_{f,q}|} \sum_{p \in P_{f,q}} \frac{1}{3} \sum_{t \in (q-3,q-1)} V_{p,t} \tag{2}$$

where $P_{f,q}$ is the set of f's peers.

To determine a firm's set of peers we draw on the Text-based Network Industry Classification (TNIC) developed by Hoberg and Phillips (2016). Drawing on firms' product descriptions in their annual 10-K statements, this classification identifies a set of competitors for each firm in a given year. This classification is particularly suited for our purposes for two reasons. First, because it depends on product descriptions, this classification comes closer to identifying competitors than traditional industry classifications such as SIC or NAICS.[7] Second, because the set of competitors varies by firm, firms are not lumped into mutually exclusive categories. This is especially applicable to multi-category organizations and is more consistent with how audiences classify firms.

We define performative atypicality as the cosine distance between a firm's centroid and its peer centroid. To account for the right-tailed skewness of this measure, we log transform it as follows:

$$PA_{f,q} = \log(1 - \cos(V_{f,q}, PV_{f,q})) \tag{3}$$

Performative atypicality, $PA_{f,q}$, is high (low) for firms that have calls in which the semantic meanings expressed are quite unusual (commonplace) relative to the meanings expressed in calls of peers.[8]

Performative atypicality is sensitive to the length of the earnings call. Longer calls provide an opportunity for a wider range of meanings to be discussed, mechanically reducing performative atypicality. We therefore remove calls that include fewer than 200 words, and include call length as a control variable in multivariate models. Where we report uni- or bivariate distributions, we use the performative atypicality measure adjusted for call length. This measure is calculated as the residual in a linear model wherein performative atypicality is predicted from the logged number of words in a call.

Categorical Atypicality

Given that the relationship between categorical atypicality and analyst valuations has been extensively demonstrated in prior work (though recent work challenges the categorical atypicality discount, see Goldfarb and Yan (2021)), we do not explore it further. Nevertheless, we include it as an independent variable in all our models for two reasons. First, we aim to explore whether categorical and performative atypicality exhibit different patterns and relate differently to analyst valuations. Second, we seek to demonstrate that categorical and performative atypicality are independent of one another; performative atypicality is not merely a proxy for categorical atypicality.

Following Bowers (2014) and Zuckerman (2004), we implement categorical atypicality as an organization's "coherence," inferred from the degree of stock coverage overlap between the analysts covering its stock. This operationalization assumes that an organization's categorical atypicality is best reflected in the perceptions of evaluators. These perceptions, in turn, are inferred from the extent to which a firm draws a varied or homogeneous set of evaluators. Organizations covered by analysts who tend to cover different stocks are, by this construction, categorically atypical.

To construct this measure, we first calculate for each pair of analysts i and j their level of coverage overlap as $p_{ij} = \min(\frac{m_{ij}}{n_i}, \frac{m_{ij}}{n_j})$, where m_{ij} is the number of stocks covered by both analysts and n_i is the number of stocks covered by analyst i. A stock is covered by an analyst when the analyst issued at least one forecast for the focal stock in the year up to and including the current quarter. We then define categorical atypicality for firm f as:

$$CA_f = 1 - \frac{\sum_{i}^{I^{fq}-1} \sum_{j>i}^{J^f} p_{ij} \cdot c_{fi} \cdot c_{fj}}{I^f(I^f - 1)/2} \tag{4}$$

where I^f is the number of analysts covering firm f and $c_{fi} = 1$ if analyst i covers firm f or $c_{fi} = 0$ otherwise. Note that for notation simplicity, we disregard time in equation 4, but construct the variable separately for each firm-quarter pairing.

This measure is sensitive to the number of analysts covering the firm, I^f. As the number of analysts grows, the likelihood of stock coverage overlap between any two analysts increases and thus categorical atypicality decreases mechanically. We therefore include number of estimates as a control variable in multivariate models. Where we report uni- or bivariate distributions, we use the categorical atypicality measure adjusted for number of estimates. This measure is calculated as a the residual in a linear model wherein categorical atypicality is predicted from the number of estimates.

Dependent Variables

Analyst Disagreement. To evaluate the relationship between performative atypicality and analyst disagreement, we use the standard deviation in analysts' estimates for a given quarter. We compute this variable directly based on analysts' estimates, using each analyst's most recent estimate for a given quarter. To mitigate the influence of extreme values, we winsorize this variable at the top 1 percent.

Earnings Surprise. To evaluate the relationship between performative atypicality and analyst bias, we compute earnings surprise for a given quarter. Following standard practice in research on earnings surprises, we use the difference between a firm's reported earnings per share and analysts' consensus estimate (i.e., the median estimate across analysts for a given quarter) divided by the firm's stock price at the end of the preceding quarter (Guo, Sengul, and Yu, 2019; Westphal, Park, McDonald, and Hayward, 2012; Barron, Byard, and Yu, 2008; Livnat and Mendenhall, 2006). We then multiply it by 100 so that an earnings surprise of 1 means that the earnings surprise is 1 percent. For example, for a firm with a reported earnings of 1, a consensus estimate of 0.99 and a stock price of 1, the earnings surprise is then 100x(1-0.99)/1 = 1 percent. To mitigate the influence of extreme values, we winsorize this variable at the top and bottom 1 percent (as for example in Skinner and Sloan, 2002; Bochkay, Hales, and Chava, 2019).[9] The mean earnings surprise is slightly negative in our sample, which is in line with other studies using similar measurement of surprise (such as Akbas, 2016; Lee, 2016; Hartzmark and Shue, 2018; Livnat and Mendenhall, 2006).

Control Variables

We include a variety of control variables to account for additional factors that can affect the dependent variables. The controls fall into three main categories: firm, call, and analyst attributes. Moreover, to control for mean differences between industries, we include industry fixed effects in all models that do not include firm fixed effects. The industry classification is based on the Text-based Fixed Industry Classifications (Hoberg and Phillips, 2016), which is the equivalent of two-digit SIC codes.

Firm Attributes

Assets. We control for firm size using log of assets.

Leverage. We control for leverage, measured as total liabilities over total assets and winsorized at the top and bottom 1 percent. Leveraged firms have limited access to credit and greater cash flow constraints, which makes them more likely to experience a negative earnings surprise. Moreover, as previous research suggests, investors' reactions to the information communicated in earnings calls is contingent on firms' risk profiles (Pan, McNamara, Lee, Haleblian, and Devers, 2018).

Preceding positive surprise. Recent surprises convey signals on future performance that may influence the perception of market participants (e.g. Pfarrer, Pollock, and Rindova, 2010; Shanthikumar, 2012). We thus control for past earnings surprises using a dummy that takes a value of 1 if there was a positive earnings surprise in the preceding quarter and 0 otherwise.[10]

Call Attributes

Order in quarter. Interviews we conducted with communication professionals who advise management teams on how to prepare for quarterly earnings calls suggested that firms sometimes make strategic choices about when to schedule their call relative to other firms. In some situations, firms prefer to go early in the call order so they can shape the industry narrative. In other cases, they prefer to go later so they can hear from their peers before deciding on their own messaging. We therefore control for the order of a firm's call in a given quarter relative to other firms in the same industry.

Positivity. Managers strategically influence the tone of conference calls (D'Augusta and DeAngelis, 2020). As these strategic efforts may correlate both with atypicality and future earnings, we control for the positivity of the earnings call. To do so, we use Loughran and McDonald's (2011) sentiment dictionary for financial disclosures. We compute positivity as the difference between the number of positive and negative words divided by their sum.

Time horizon. The time orientation of an earnings call may convey signals about the firm's subsequent ability to achieve robust performance in the future. We therefore control for the call's time horizon using DesJardine and Bansal's (2019) dictionary of short-term and long-term oriented words. Specifically, we operationalize time horizon as the difference between the number of long-term words and the number of short-term words divided by their sum.

Litigiousness. A high litigation risk may impact subsequent surprise (Matsumoto, 2002). Additionally, firms may purposefully use atypical language to remain ambiguous regarding ongoing litigations. We thus control for the "litigiousness" of calls using the proportion of litigious words in the call. We again used Loughran and McDonald's (2011) sentiment dictionary for financial disclosures to identify litigious words.

Length. As mentioned above, an earnings call's length mechanically correlates with performative atypicality. Call length may also be related to future earning surprises, for example, if it is indicative of firm risk, above and beyond its mechanical relationship with performative atypicality. We therefore include the log of the total number of words in the call after tokenization as a control.

Analyst Attributes

Analysts churn. Analysts have some latitude is deciding which firms to cover. The composition of analysts is likely related to the probability of an earnings surprise and may be spuriously related to performative atypicality. In particular, because analysts specialize by industry, they may be discouraged by performative atypicality, resulting in their decision not to cover such firms. Moreover, atypical firms may attract inexperienced analysts. Both of these mechanisms would lead to larger surprises. To ensure that this is not driving our result, we control for analyst churn—i.e., the proportion of analysts producing an estimate for the current quarter that did not produce an estimate for the preceding one.

Number of estimates. As noted above, the number of analysts covering a firm mechanically correlates with its categorical atypicality. Additionally, firms that draw a smaller number of analysts may be more likely to experience earnings surprise. We thus control for analysts' coverage using the total number of analysts publishing an estimate for the firm's earnings in the current quarter.

Disagreement. In models where earnings surprise is our dependent variable, we control for the standard deviation in analysts' estimates given that surprises are more likely to occur when analysts have divergent expectations of future performance.

RESULTS

Our empirical analysis has three components. First, we explore the distributional properties of performative atypicality. This distributional analysis is aimed at (1) validating our measure by demonstrating that it is high for firms that are known for being performatively atypical, (2) evaluating the extent to which performative and categorical atypicality capture different empirical phenomena, and (3) demonstrating that performative atypicality exhibits within-firm variation as we conjecture. In the second part of the analysis we explore the relationship between performative atypicality and future earnings surprises. Finally, in the third part of our analysis we use the tools of computational linguistics to inductively unpack the relationship between performative atypicality and analysts' evaluations. Our objective is to understand why analysts interpret this form of atypicality the way they do.

Performative Atypicality's Properties

We begin with exploring the distributional properties of performative atypicality. Figure 1 plots the kernel density for performative atypicality (we report descriptive statistics of the main variables of interest in Table 1). As the figure demonstrates, performative atypicality roughly follows a normal distribution.

[TABLE 1 ABOUT HERE]

[FIGURE 1 ABOUT HERE]

Figure 2 plots standardized performative atypicality (adjusted for call length) as a function of standardized categorical atypicality (adjusted for number of estimates). Each dot corresponds to one firm, such that its location on the plot corresponds to the firm's levels of atypicality, averaged across all time periods. Dot sizes are proportional to firm size (in assets, logged). We highlight various firms for illustrative purposes.

[FIGURE 2 ABOUT HERE]

The patterns in Figure 2 support our assumptions about performative atypicality. First, validating the measure, it illustrates that performative atypicality is higher among firms that have a reputation for doing things differently. Consistent with intuitive expectations, iconoclastic technology firms such as Twitter and Facebook are among the highest in performative atypicality overall. Differences within industries also conform to these expectations. Tesla, for example, is significantly more performatively atypical than Ford. Similarly, Nvidia and Google are much higher in performative atypicality than Microsoft or Dell. And whereas major banks such as JPMorgan Chase are below average in performative atypicality, Green Dot—a mobile banking platform—is among the highest. Importantly, differences in performative atypicality are not merely reflections of differences in technological innovation. Sprint, for example, stands out relative to other mobile operators, while General Motors is much more performatively atypical than Ford, despite both having almost identical categorical atypicality levels.

Moreover, the mean levels of performative atypicality substantially vary between industries. Although there is significant variation within the food industry between firms such as Kellogg, Hershey and Kraft Heinz, for example, their mean performative atypicality is low relative to software companies. This comports with naive expectations that technology sectors exhibit greater overall atypicality than traditional industries and underscores the need to account for mean differences between industries when estimating between-firm effects, as we do below.

Second, it is evident that the the two forms of atypicality—performative and categorical—capture different phenomena. Although the two adjusted measures are significantly correlated at the mean firm level ($\rho = 0.092, p < 0.001$), this correlation is weak. Overall, across all quarterly observations, the correlation between the adjusted measures is even weaker ($\rho = 0.035, p < 0.001$).[11] Firms like Akamai (a provider of distributed computing platforms, cybersecurity and cloud computing) and Intuit (a financial services and software company), which are among the highest in categorical atypicality, exhibit below mean levels of performative atypicality. While their product portfolios comprise quite unusual combinations, their performances in quarterly earning calls are fairly standard.

Finally, consistent with our argument that performative atypicality is dynamically produced, it exhibits greater within-firm variance than categorical typicality does. While there is significant variation in performative atypicality between firms, a substantial proportion of the variance is explained by fluctuations within-firm. As the inset in Figure 2 illustrates, even Tesla and Ford, two car manufacturers with, respectively, consistently high and low performative atypicality, exhibit significant within-firm variation. In fact, as Panel A of Figure 3 shows, roughly half of the variance in performative atypicality is explained by differences between firms; the rest is attributable to within-firm fluctuations. In contrast, between-firm differences explain roughly 85% of the variance in categorical atypicality. This is also reflected in Panel B of the Figure, plotting the kernel densities for the standard deviation, by firm, for both types of (adjusted and standardized) atypicality measures. As this plot demonstrates, there is far greater variation within firm for performative atypicality than there is for categorical atypicality.

This is also evident in Panel C of Figure 3, which plots mean (standardized) performative and categorical atypicality over time. Once again, we see that performative atypicality is less stable than categorical atypicality. Changes in mean levels of performative atypicality closely track movement in the S&P 500 index, whereas changes in categorical atypicality do not, suggesting that firms have more latitude to diverge from performative conventions during times of growth. During the first three years of our observation window, when the market was reeling from the 2008 financial crash and the great recession that followed, mean levels of performative atypicality were suppressed. Consistent with research on threat rigidity (Staw, Sandelands, and Dutton, 1981), firms often resort to more conservative actions during times of uncertainty and instability. Whether merely self-presentational or a true reflection of firm behavior, we interpret the relationship between market uncertainty and performative atypicality as an indication that the latter is a signal of a firm's deviation from conventional practices.

[FIGURE 3 ABOUT HERE]

The Performative Atypicality Premium

How do analysts interpret performative atypicality? To answer this question, we examine the relationship between performative atypicality and analyst forecasts using between- and within-firm model specifications. We use ordinary least squares and cluster standard errors by firm in all models to account for within-firm interdependencies. All variables are measured at the quarter level. Given that, as Figures 2 and 3 show, performative atypicality varies by industry and time, we include industry and period fixed effects. Because we cannot identify random sources of variation in performative atypicality, our modeling strategy ultimately does not yield causal estimates. Nevertheless, in addition to including fixed effects, we lag the dependent variables (as well as contemporaneous performance controls) such that the effects of atypicality are estimated for analyst disagreement and earnings surprises in the subsequent quarter. For ease of interpretation, both atypicality measures are standardized.

Tables 2 and 3 report results for between-firm (Models 1-3) and within-firm (Models 4-6) OLS models, where the dependent variable is modeled as a function of performative atypicality. We include categorical atypicality as an independent variable to compare its effects to those of performative atypicality and to explore whether the two forms of atypicality relate differently to analysts' interpretations. Between-firm models include industry-year-quarter fixed effects, to account for variation that is attributable to changes within industries over time. These models should therefore be interpreted as reflecting the effects of differences in atypicality between firms that are competing in the same industry and at the same time.[12] The within-firm models include firm and year fixed effects.[13] They should be interpreted as reflecting the effects of changes in atypicality that occur within a firm over its life course, net of its fixed attributes.

Analyst Disagreement

Table 2 reports models estimating analyst disagreement. The effect is positive and significant for performative atypicality across all specifications. We plot the marginal effects estimated by Models 3 and 6 in Figure 4. As a firm becomes more preformatively atypical—whether compared to other firms or relative to itself—analysts increasingly disagree about how to predict its future performance. As Model 1 demonstrates, replicating established findings, categorical atypicality is equally disorienting, leading to a similarly sized increase in analyst disagreement. Model 3 shows that the effects of performative and categorical atypicality on disagreement are independent of one another, further demonstrating that these two dimensions of atypicality relate to different interpretative pathways. Yet the effect of categorical atypicality becomes insignificant in within-firm specifications (Models 4 and 6). Not only is there far less within-firm variation in categorical atypicality than there is in performative atypicality (Figure 3, Panel A), when firms experience shifts in categorical atypicality, analysts appear to be less responsive to such change. We conjecture that this is because they tend to see firms' categorical identities as fixed.

[TABLE 2 ABOUT HERE]

[FIGURE 4 ABOUT HERE]

Earnings Surprises

Results in Table 3 provide robust evidence that analysts reward performative atypicality. Whether estimating between- or within-firm effects, all specifications demonstrate a significant negative relationship between performative atypicality and surprise, indicating that analysts are optimistic about performatively atypical firms. We plot this relationship, as estimated by Models 3 and 6, in Figure 5. As executives veer from conventional meanings in quarterly earnings calls, analysts' tend to overestimate these executives' firms' future performance. Rather than signaling incompetence, performative atypicality appears to be interpreted as an indication of a firm's unique capabilities. We refer to this advantage as the *performative atypicality premium*.

[TABLE 3 ABOUT HERE]

[FIGURE 5 ABOUT HERE]

Are the rewards to performative atypicality linear? In the models reported in Table 4 we divide performative atypicality into quintiles. As the results illustrate, analysts are particularly optimistic about the future earnings of firms at the upper quintile of performative atypicality.[14] Given the high within-firm variation of performative atypicality (cf. Figure 3), we interpret this finding as indication that firms' performative atypicality has a particularly strong influence on analysts' evaluations when executives express especially unconventional meanings in their interactions with analysts in a given quarter. Optimal distinctiveness theory would have predicted that analysts overestimate the performance of firms with moderate levels of performative atypicality. Yet our results indicate that the more preformatively atypical a firm, the greater the unjustified enthusiasm of analysts covering it.

[TABLE 4 ABOUT HERE]

The systematic relationship between performative atypicality and analysts' biased forecasts may be driven by two different pathways. Performative atypicality might be a credible signal of low future earnings if executives behave atypically without conscious awareness when they are trying to conceal negative information. This would produce a negative earnings surprise if analysts fail to pick up on that signal. Alternatively, if performative atypicality is unrelated to future earnings, negative earnings surprises will ensue if analysts misinterpret it as a positive signal. Consistent with the latter interpretation, we report results in Appendix B which show that performative atypicality is not associated with future low (or high) earnings. We interpret this as evidence that analysts are overly optimistic about the implications of performative atypicality, not that performative atypicality is a signal of low earnings that analysts fail to identify.

Finally, the models reported in Tables 3 and 4 do not find a significant effect for categorical atypicality in any specification. This does not necessarily mean that categorically atypical firms do not suffer an illegitimacy discount; such firms might be discounted by the market, leading analysts to correctly factor that discount into their estimates.[15] Nevertheless, our results clearly indicate that while analysts are overly optimistic about the implications of performative atypicality on future earnings, their estimates are not similarly systematically biased by categorical atypicality. We discuss the implications of this finding below.

What Explains the Performative Atypicality Premium?

Why are analysts swayed by performative atypicality? We posit that there are two general explanations for analysts' overall tendency to be bullish about the prospects of performatively atypical firms. The first maintains that performative atypicality signals executives' private information. Their indifference to audience expectations is interpreted as indication of their confidence about their firm's unique capabilities, regardless of what these capabilities are. The content of their performances is therefore less important than the confidence these atypical performances signal.

Alternatively, analysts may have a theory of value that privileges certain types of atypicality over others. If that is the case, we should find that the relationship between performative atypicality and earnings surprises is patterned along specific dimensions of meaning.

To evaluate the two possibilities, we inductively explore the semantics of atypical performances. We leverage the scale afforded by word embeddings to identify whether certain dimensions of semantic divergence are especially rewarded, or penalized, by analysts. To do so, we use embedding vector subtraction. As Mikolov et al. (2013) showed, the vector subtraction "King"-"Male" captures a semantic difference that is analogous to the meaning "Royal". Building on this rationale, we subtract a firm's peer embedding centroid from its call centroid to capture the call's semantic difference from its peers. We refer to this difference as an earning call's *atypicality centroid*. The atypicality centroid corresponds to the meanings uniquely communicated in a specific call, relative to its peers (see Appendix C.2 for a detailed definition). If atypicality centroids are non-randomly associated with analysts' reactions, we can conclude that analysts are attentive to the content of performative atypicality.

Drawing on Nelson's (2020) computational grounded theory approach, we perform this inductive analysis in three steps. In the first, we use principal component analysis (PCA) to evaluate the extent to which firms' performative atypicality is structured. We find that an overwhelming majority of the variance in atypicality centroids is explained by a handful of PCA dimensions, suggesting that atypical performances are not arbitrarily divergent. Rather, they are structured by a few dominant axes of meaning. Further analyses show that this variance is explained by differences between firms but not between industries. This assuages concerns that performative atypicality merely captures topical variation driven by differences between industries. We report this analysis in full in Appendix C.2.

In the second step, we turn to human-based hermeneutics to systematically interpret these dimensions and explore how they relate to earnings surprises. These results paint a fairly complex picture, as we discuss in detail in Appendix C.3. Rather than falling into the trap of fine-grained nuance (Healy, 2017), however, we point to the two overarching conclusions they afford. First, being performatively divergent does not automatically yield an evaluation premium. Not all types of atypical performances result in systematic analyst overestimation. Second, one underlying axis of meaning appears to dominate the structure of analysts' reactions. One end of this axis is populated by performances that tend to atypically focus on the procedural aspects of financial earnings. This, we conjecture, connotes the firm's staid and circumspect formality. At the other end are performances that deviate from conventions by discussing the firm's innovative competitiveness, its collaborative orientation, and its obligations and opportunities. Together, these suggest that analysts develop an optimistic impression of a firm's earnings prospects when its performative atypicality invokes conventional connotations of innovation and creativity.

If this interpretation is correct, we should find that the performative atypicality premium is more likely to occur as a firm's performance becomes more like the performances of firms that are perceived as innovative and less like its own competitors' performances. To test this proposition, in the third and final step we decompose an earnings call's performative atypicality into two components: *innovation-biased atypicality* and its complement, *non-innovation-biased atypicality*. Innovation-biased atypicality is the portion of atypicality that emulates performances of firms that are perceived as innovative. Non innovation-biased atypicality is the remainder. We operationalize innovative firms either as high-technology firms (using Kile and Phillips' [2009] classification approach) or as firms listed by Fast Company as one of the world's Most Innovative Companies during our window of observation. Additional details hare provided in Appendix C.4.

We include innovation-biased atypicality and non-innovation-biased atypicality as variables in between- and within-firm models predicting earnings surprise (following the same specification reported in Table 3) and using both operationalizations of innovative firms. Results are reported in Table 5. Innovation-biased atypicality significantly predicts a negative earnings surprise in all specifications, irrespective of operationalization or modeling approach. As a firm becomes atypical in a way that connotes innovation, whether relative to other firms or to itself, analysts tend to overestimate its future earnings. Non-innovation-biased atypicality, in contrast, is weaker in its effect on earnings surprise and fails to reach significance in all specifications. Atypicality that does not connote innovation, in other words, has an attenuated effect on analysts' overestimation of firm earnings.

[TABLE 5 ABOUT HERE]

DISCUSSION

Firms that meet, or exceed, earnings expectations are rewarded by the market (Kasznik and McNichols, 2002). A negative earnings surprise is therefore an adverse outcome that executives seek to prevent. Ironically, however, our results suggest that a firm's performative atypicality might lead to a subsequent negative earnings surprise because, counterintuitively, analysts interpret such performances as positive signals about a firm's strategic positioning and future financial performance. While categorical atypicality is, ultimately, a liability leading to an illegitimacy discount (Zuckerman, 1999), performative atypicality appears to generate a uniqueness premium.

Our inductive analysis led us to conclude that not all forms of performative atypicality are created equal. Atypical performances that invoke meanings of innovation and creativity appear to result in especially buoyant analyst forecasts; other types of atypicality are received with less enthusiasm. Although this buoyancy translates into a disadvantageous position in the setting we study empirically, it may very well be the case that it results in significant rewards in other settings if investors, or outside audiences more generally, are commonly swayed by innovation-biased performative atypicality.

The spectacular rise and fall of WeWork, the shared work space management company, provides an instructive example of the innovation-biased performative atypicality premium. Founded in 2010, WeWork's product was by no means categorically unusual. Shared work spaces were not a novel idea at the time, and competitors such as Regus were already managing such spaces across the globe for two decades prior to WeWork's entry into the market. Nevertheless, WeWork was perceived as inherently different. Owing to the eccentric style of its founder, Adam Neumann—who was occasionally spotted walking barefoot on the streets of Manhattan and frequently professing unconventional aspirations, such as living forever, in interviews and public appearances—the company was seen as innovative and pioneering relative to its gray, conventional, and seemingly unambitious competitors. In the eyes of many, WeWork was not a typical real estate company but a "capitalist kibbutz" ushering a new model of work and collaboration.[16]

Indeed, WeWork was named one of the world's most innovative companies by Fast Company in 2015. Leading and experienced investors were tempted by this performative atypicality. As Neumann himself confessed, these investments were based more on "our energy and spirituality than ... on a multiple of revenue."[17] Upon filing its initial public offering prospectus in 2019, however, it became apparent that WeWork's revenue model, profitability strategy, and governance structure were inherently flawed. The IPO was subsequently withdrawn, and the company's valuation, peaking at a staggering $47B, was cut by almost 80%.

A story of excess, delusion and debauchery, WeWork's implosion has been hailed by some as an "astounding moment in business" (Brown and Farrell, 2021, p. xi), its performative atypicality so extreme that many fell prey to the belief that what was, ultimately, no more than an office space provider was a truly trailblazing tech company. Our findings suggest that, while undoubtedly unusual in magnitude, WeWork's tale is an extreme manifestation of a more broadly, if modestly, prevalent principle. Firms performatively behaving like celebrated innovators appear to create an exaggerated impression of ability. This, we contend, has several implications for our understanding of how atypicality relates to audience perceptions.

A Bilateral Model of Valuation

Our results are consistent with the two-stage model of valuation. In the first stage, observers determine the "what" on the basis of the products and services an organization offers. In the second stage, they infer the "how" on the basis of the organization's performative interactions. As we show, the two forms of atypicality have independent effects on earnings surprises (Table 3).[18] This, we posit, suggests that outside observers compartmentalize their inferences about an organization's identity. Each stage catalyzes a different cognitive process.

Importantly, our bilateral conceptualization of atypicality extends the two-stage model. In doing so, it addresses one of that model's major shortcomings. Researchers often acknowledge that organizations differentiate along a small subset of features. Yet, they mostly remain vague as to which features are conducive to differentiation and which are important for gaining legitimacy. The few studies that lay out this distinction provide ad hoc and context-specific explanations. Phillips et al. (2013), for example, demonstrate that high-status law firms face disapproval when diversifying into personal injury law but not into family law. It remains unclear why organizations can, in general, successfully differentiate along certain features but not others.

Our analytical approach offers a different way of thinking about the axes along which organizations are expected to conform or stand out. Rather than focusing on organizational features, it points to the different *mediums* through which organizations communicate their identities: the products and services they offer, or their ongoing interactions with outside stakeholders. Our results suggest that these different mediums correspond to two different interpretative dimensions along which analysts evaluate organizations. One relates to the industry an organization competes in. The other to its degree of innovation. This explains how organizations can be simultaneously typical and distinctive. We refer to this theoretical extension as the *bilateral model of valuation* and illustrate it in Figure 6.

The bilateral model rests on two assumptions. First, like the original two-stage model, it assumes that the first stage of valuation precedes the second. Observers evaluate the "how" only after determining the "what." In fact, our operationalization of performative atypicality assumes that observes first reach a conclusion about a firm's competitors before they can determine the extent to which it diverges from them performatively. This does not mean, however, that the two types of atypicality are contingent on one another. Our second assumption, therefore, is that the effects of categorical and performative atypicality on audience valuations are independent. Our results are consistent with this assumption.

These results imply that successful organizations can enjoy the differentiating benefits of performative atypicality without necessarily paying the price of categorical atypicality. Consider the two organizations labeled *A* and *B* in Figure 6. Both offer conventional products, making them easily classifiable. Existing theory would therefore expect them to be more favorably valued relative to organization *C*, which is categorically atypical. But these two firms appear differently in the second stage. While firm *A* stands out as unique, firm *B*'s interactions with stakeholders are similar to its peers'. Our findings suggest that analysts will interpret the former as more innovative, leading them to overestimate its future earnings.

Two qualifications are in order. First, because we do not directly measure the cognitive mechanisms connecting atypical performances and negative earnings surprises, we cannot determine with certitude that this relationship is driven by analysts' compartmentalized inferences. Yet the evidence is consistent with such an explanation. In particular, performative atypicality is associated with an increase in disagreement between analysts' forecasts (Table 2), suggesting that it induces interpretative uncertainty. At the same time, it is not associated with a decline in earnings (Table B1), suggesting that it is unrelated to firm financial outcomes. Together, these findings suggest that performative atypicality is related to analysts' perceptions.

Moreover, our finding that analysts are especially optimistic about an organization's earnings if it performatively emulates perceived innovators is consistent with the contention that performative atypicality catalyzes inferences about the "how" of organizational identity. Our inductive analyses also indicate that the performative atypicality premium is not driven by discussions of firms' atypical product offerings (as reported in Appendix D), alleviating concerns that it is simply a different manifestation of categorical atypicality. Whether consciously or not, analysts appear to interpret preformative atypicality as an indication about unique firm capabilities.

Second, the distinction between the categorical and performative mediums of valuation is pronounced in some contexts more than others. Securities analysts arrive in quarterly earnings calls with intimate knowledge of the companies they cover. Their impressions of executives' performative atypicality are temporally differentiated from their determinations of the firm's categorical atypicality. In other contexts, however, these inferences occur contemporaneously. A clothing store window, for example, communicates both categorical (what types of clothes it sells) and performative (how unusual is the window dressing relative to other clothing stores) information. The bilateral model of valuation will apply as long as a patron evaluating the window can differentiate between categorical and performative features (e.g., between the products on sale and their arrangement). This will result in a performative atypicality premium if the patron has a preference for uniqueness.

More broadly, this implies scope conditions for our theoretical conclusions. The bilateral model of valuation, and its resultant performative atypicality premium, should generally apply in settings where two conditions hold. First, the categorical and performative channels of information are distinct, and, second, audiences see value in candidates' unique capabilities. These conditions are more likely to apply, for example, when venture capital firms evaluate relatively late-stage (e.g., series C) funding opportunities, than very early stage opportunities (e.g., seed funding) where there is greater ambiguity about a firm's products.[19]

Distinguishing the Categorical and the Performative in Empirical Work

Thinking about atypicality through our bilateral lens sheds new light on previous empirical findings. In a recent paper, for example, Taeuscher et al. (2021) analyze the success of technology crowdfunding campaigns. Apparently contrary to the expectations of the two-stage valuation model, the authors find that higher levels of distinctiveness are associated with greater, rather than smaller, crowdfunding success. Yet they measure this distinctiveness as the level of topical divergence between a founding team's entrepreneurial story and the prototypical entrepreneurship story in the venture's market category. Although these stories often refer to product features—affecting the venture's perceived categorical atypicality—they are also inherently performative. It is therefore quite likely that, consistent with the bilateral model of valuation, funding success is driven by performative, rather than categorical atypicality.

More broadly, we conjecture that, in practice, work examining the illegitimacy discount has often conflated categorical and performative atypicality. Consider the common focus in the categories literature on the penalty accruing to category-spanning restaurants (Goldberg et al., 2016a; Rao et al., 2005). Menus, for example, are frequently used in this research stream as product descriptions for the purpose of inferring atypicality (e.g., Kovács and Johnson, 2014). Restaurants that include terms that are typical of different cuisines—such as ciabatta (Italian) and chapati (Indian)—are considered atypical by this construction. But menus are also performative. Some minimally list ingredients, whereas others include more evocative descriptions about how these ingredients are "tossed in our homemade secret BBQ sauce." These stylistic choices convey information about the restaurant's identity above and beyond its cuisine classification. The mere insinuation of customer choice, for example, connotes the restaurant's lack of culinary sophistication (Jurafsky, 2015).

Future work, our findings suggest, should pay greater attention to differentiating between the categorical and performative dimensions of atypicality. In settings where these dimensions are experientially distinct, like the one we study here, this task is fairly straightforward. In other settings, however, the same object—a menu, a storefront or a sales pitch—can simultaneously convey categorical and performative information. Existing literature mostly treats this information as uniform organizational "features," aggregating them to measure an organization's overall level of atypicality. Our findings highlight the importance of distinguishing features that relate to inferences about "what" from those that relate to inferences about "how." This distinction, we conjecture, will vary in nature as a function of context. In menus, for example, words describing dishes or ingredients will relate to categorical atypicality, while the use of non-culinary terminology will serve a performative role. During a startup pitch, in contrast, categorical information is conveyed in discussions of product features, while performative information is conveyed through speakers' aesthetic and linguistic choices. Where this distinction exists is where the performative atypicality premium should be more pronounced.

Future work might also explore the assumptions about the cognitive mechanisms that connect performative atypicality to a valuation premium. This work should seek to confirm whether the categorical and performative channels affect, respectively, impressions of "what" and "how" as we conjecture, and whether these effects are independent. Rather than using valuations as their dependent variables, such investigations will need to employ tools from cognitive science to measure audience members' interpretations.

Finally, while a voluminous prior literature demonstrates that categorical atypicality is associated with an illegitimacy discount, we do not find empirical support for that dynamic in our data. Counter the intuitions of the two-stage valuation model, analysts do not underestimate the future earnings of categorically atypical firms (Table 3). Indeed recent work has had limited success in replicating Zuckerman's (1999]) original findings (Goldfarb and Yan, 2021) about analysts' illegtimicay discount. Nevertheless, we are hesitant to conclude that categorical nonconformity is inconsequential for analysts evaluations.[20] Because our purpose has been to explore the implications of preformative atypicality, not the scope conditions for the categorical atypicality discount, we leave a more thorough investigation of this result for future work.

Innovation-Biased Performative Atypicality as Conventional Coolness

Existing literature predominantly conceptualizes atypicality in distributive terms, as the magnitude of deviation from normative expectations. Our results, however, suggest that whether or not performative atypicality is interpreted as a positive signal about firm capabilities depends not only on the degree of atypicality but also on its content. Merely being performatively different translates to a small, mostly insignificant, premium (Table 5). Yet, atypical performances that conform to conventional understandings of what being innovative looks like are met with consistent optimism.

In other words, to resolve the conflicting demands of conformity and differentiation, it is not enough for firms to limit atypicality to performative channels. Rather, this performative atypicality needs, ironically, to resonate with conventional images of uniqueness. Idiosyncratic departures from normative expectations are not interpreted as signals of competence in their own right. Only when they evoke recognizable scripts of successful heterodoxy they are perceived as indications of quality. Our results suggest that what makes these scripts symbols of quality is not so much their substance as the identities of those that originate them.

This conventional performative atypicality, we conjecture, is not limited to for-profit organizations vying for the attention of investors. We speculate that it emerges in other contexts in which actors need to balance the conflicting demands of conformity and uniqueness. In fact, the innovation-biased performative atypicality that is interpreted by analysts as indication of strategic competence shares strong affinities with the elusive concept of "coolness" that is pervasive in Western culture (Quartz and Asp, 2016). As Zuckerman (2016) points out, nonconformist performances tread a thin line between being interpreted as cool or as incompetent. Our findings suggest that coolness emerges when actors deviate from expectations in ways that connote familiar images of success.

CONCLUSION

In one of the most memorable scenes in "The Life of Brian," British comedy troupe Monty Python's celebrated religious satire, the protagonist, who is mistaken for the Messiah, tells his thousands of followers that they are all individuals. "We are all individuals!" they respond in unison, with the exception of one screechy voice shouting "I'm not!" This brilliantly comic exchange epitomizes a perennial social conundrum: the conflicting need for individuality and desire to fit in (Goldberg, Srivastava, Manian, Monroe, and Potts, 2016b; Chan, Berger, and Van Boven, 2012; Brewer, 1991). Organizations' ability to balance these dual pressures is a matter of survival.

Those that are too unique are dismissed as unintelligible or unappealing, whereas those too conformist struggle to get audiences' attention.

Dominant theories either argue that successful organizations strike a fine balance between conformity and differentiation or that they are ultimately forced to comply with categorical expectations. These conclusions, we argue, relate to these theories' undimensional and static conceptualizations of atypicality. Building on constructivist theories of identity, we propose, in contrast, that organizational members interactionally "do" organizational identity. We analytically distinguish between categorical and performative atypicality and demonstrate that the latter results in a premium in the eyes of external stakeholders. Organizations can meet the need for differentiation performatively, while maintaining categorical conformity. Yet, harnessing the forensic affordances of computational linguistics, we also find that to result in positive reactions performative atypicality needs to heed to conventional scripts of being different. Like the lone anonymous dissenter in "The Life of Brian," being idiosyncratically different is often greeted with indifference.

Notes

[1]Target was not the first or only retailer to do so. Other large American retailers began combining groceries and non-food items in the late 1980s. Yet these mixed retail spaces were limited to superstores.

[2]The term identity has been used by organizational scholars in a variety ways. Many use the term as reference to the ways by which members of an organization understand its core and enduring attributes (e.g. Whetten, 2006; Gioia, 1998). Our focus, in contrast, is on perceptions of external audiences (Hsu and Hannan, 2005). We define identity as the various meanings that outside observers typically associate with an organization. Our theoretical focus also stands in contrast to the concept of organizational image, which is commonly conceived as the ways by which organizational members imagine that outside stakeholders view their organization (Gioia, Schultz, and Corley, 2000).

[3]A related literature on organizational impression management has also emphasized external stakeholders' perceptions. That work often focuses on the purposeful actions that organizational leaders take in order to influence their status and approval in the eyes of outside audiences (Highhouse, Brooks, and Gregarus, 2009). Our approach, instead, focuses on the interactional ways by which these impressions are formed, emphasizing the role of typicality, or lack thereof, in shaping these impressions.

[4]https://hbr.org/2018/05/a-40-year-debate-over-corporate-strategy-gets-revived-by-elon-musk-and-warren-buffett

[5]In fact, in some cases the "what" and "how" are inherently intertwined. The handmade processes used by a craft chocolatier, for example, are integral to the product's value proposition.

[6]https://gen.medium.com/should-america-be-run-by-trader-joes-22e3e3f6190

[7]This is especially the case in industries in which different firms offer differentiated products. As Hoberg and Phillips (2016) show, for example, the Business Services industry is, in effect, differentiated into multiple submarkets.

[8]The results reported below are robust to an alternative construction of this variable whereby peer firms are determined on the basis of their 2-digit SIC classification.

[9]Note that our use of unadjusted I/B/E/S data addresses the "rounding problem" identified by Payne and Thomas (2003). We use CRSP adjustment factors to account for cases where stock splits occur in between a forecast and earnings announcement.

[10]In unreported models, we also include a dummy variable for past negative earnings surprise. Our results are robust to this specification, which we do not report for the sake of brevity.

[11]Note, moreover, that the raw correlation between the two variables, as reported in Table 1, is misleadingly high. This correlation is mechanically driven by the number of analysts covering a firm, which affects both the call's length and the number of estimates produced by analysts. These two latter variables are negatively correlated with performative and categorical atypicality, respectively. Consequently, the correlation between performative and categorical atypicality drops from 0.156 to 0.035, once they are adjusted for call length and number of estimates.

[12]Specifically, we test between-firm models of the form:

$$Y_{f,q+1} = \beta_1 PA_{f,q} + \beta_2 CA_{f,q} + \beta_3 X_{f,q+1} + \beta_4 Z_{f,q} + \alpha_{i_f,q} + u_{f,q} \tag{5}$$

where $Y_{f,q+1}$ is our dependent variable for firm f in quarter-year q+1, $PA_{f,q}$ is performative atypicality for firm f in quarter q, $CA_{f,q}$ is categorical atypicality for firm f in quarter q, $X_{f,q+1}$ is a vector of controls for firm f in quarter q+1, $Z_{f,q}$ is a vector of controls for firm f in quarter q, $\alpha_{i_f,q}$ is the fixed-effect for i, the industry of firm f, in quarter q and $u_{f,q}$ is the error term.

[13]Specifically, we test within-firm models of the form:

$$Y_{f,q+1} = \beta_1 PA_{f,q} + \beta_2 CA_{f,q} + \beta_3 X_{f,q+1} + \beta_4 Z_{f,q} + \alpha_f + \theta_{y_q} + u_{f,q} \tag{6}$$

where $Y_{f,q+1}$ is our dependent variable for firm f in quarter-year q+1, $PA_{f,q}$ is performative atypicality for firm f in quarter q, $CA_{f,q}$ is categorical atypicality for firm f in quarter q, $X_{f,q+1}$ is a vector of controls for firm f in quarter q+1, $Z_{f,q}$ is a vector of controls for firm f in quarter q, α_f is the fixed-effect for firm f, θ_{y_q} is the fixed-effect for year y, the year of quarter q, and $u_{f,q}$ is the error term. We do not include quarter fixed effects as that would absorb too much variation.

[14]These results are robust to dividing performative atypicality into deciles.

[15]We do not, however, find evidence for such a market discount. In additional analyses we do not find that categorical atypicality is significantly associated with lower median estimates, lower earnings, or lower returns on assets.

[16]https://www.nytimes.com/2019/11/02/business/adam-neumann-wework-exit-package.html

[17]https://www.forbes.com/sites/stevenbertoni/2017/10/02/the-way-we-work/?sh=30044b521b18

[18]In additional analyses we do not find evidence for an interaction effect between categorical and performative atypicality in producing the performative atypicality premium.

[19]This would also suggest that audience members with different perspectives about the same organization might reach divergent conclusions about its appeal.

[20]We do so for two reasons. First, while previous work predominantly explored the relationship between categorical atypicality and market returns, our analysis focuses on earnings forecasts. Second, consistent with prior work, we operationalize categorical atypicality as the inverse of the mean overlap in analyst coverage. This measure is one step removed from categorical atypicality in that it relates to analysts' perceptions of atypicality rather than objective atypicality per-se. We also constructed an alternative measure of categorical atypicality using the TNIC (Hoberg and Phillips, 2016). We operationalize a firm's categorical atypicality as its weighted clustering coefficient in the product similarity network. The analyst overlap and TNIC-based measures of categorical atypicality are weakly (<0.09 in

all specifications), but significantly correlated. This alternative operationalization is also insignificantly associated with earnings surprises. Moreover, the relationship between performative atypicality and negative earnings surprise is robust to this implementation.

REFERENCES

Akbas, F.

2016 “The Calm before the Storm.” The Journal of Finance, 71: 225–266.

Albert, S. and D. A. Whetten

1985 “Organizational Identity.” Research in Organizational Behavior, 7: 263–295.

Askin, N. and M. Mauskapf

2017 “What Makes Popular Culture Popular? Product Features and Optimal Differentiation in Music.” American Sociological Review, 82: 910–944.

Barney, J.

1991 “Firm Resources and Sustained Competitive Advantage.” Journal of Management, 17: 99–120.

Barron, O. E., D. Byard, and Y. Yu

2008 “Earnings Surprises That Motivate Analysts to Reduce Average Forecast Error.” The Accounting Review, 83: 303–325.

Berger, P. L. and T. Luckmann

1967 The Social Construction of Reality: A Treatise in the Sociology of Knowledge. New York: Anchor.

Bochkay, K., J. Hales, and S. Chava

2019 “Hyperbole or Reality? Investor Response to Extreme Language in Earnings Conference Calls.” The Accounting Review, 95: 31–60.

Bowers, A.

2014 “Relative Comparison and Category Membership: The Case of Equity Analysts.” Organization Science, 26: 571–583.

Brandt, P. and S. Timmermans

2021 "Abductive Logic of Inquiry for Quantitative Research in the Digital Age." Sociological Science, 8: 191–210.

Brewer, M. B.

1991 "The Social Self: On Being the Same and Different at the Same Time." Personality and Social Psychology Bulletin, 17: 475–482.

Brown, E. and M. Farrell

2021 The Cult of We: WeWork, Adam Neumann, and the Great Startup Delusion. New York: Crown.

Carnabuci, G., E. Operti, and B. Kovács

2015 "The Categorical Imperative and Structural Reproduction: Dynamics of Technological Entry in the Semiconductor Industry." Organization Science, 26: 1734–1751.

Cattani, G., S. Ferriani, and P. D. Allison

2014 "Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A Core-Periphery Perspective." American Sociological Review, 79: 258–281.

Chan, C., J. Berger, and L. Van Boven

2012 "Identifiable but Not Identical: Combining Social Identity and Uniqueness Motives in Choice." Journal of Consumer Research, 39: 561–573.

Chatterji, A. K., J. Luo, and R. C. Seamans

2021 "Categorical Competition in the Wake of Crisis: Banks vs. Credit Unions." Organization Science, .

D'Augusta, C. and M. D. DeAngelis

2020 "Does Accounting Conservatism Discipline Qualitative Disclosure? Evidence From Tone Management in the MD&A." Contemporary Accounting Research, 37: 2287–2318.

Deephouse, D. L.

1999 “To Be Different, or to Be the Same? It’s a Question (and Theory) of Strategic Balance.” Strategic Management Journal, 20: 147–166.

DeJesus, J. M., M. A. Callanan, G. Solis, and S. A. Gelman

2019 “Generic Language in Scientific Communication.” Proceedings of the National Academy of Sciences, 116: 18370–18377.

DesJardine, M. and P. Bansal

2019 “One Step Forward, Two Steps Back: How Negative External Evaluations Can Shorten Organizational Time Horizons.” Organization Science, 30: 761–780.

Durand, R. and R. Calori

2006 “Sameness, Otherness? Enriching Organizational Change Theories With Philosophical Considerations On The Same And The Other.” Academy of Management Review, 31: 93–114.

Durand, R. and P.-A. Kremp

2015 “Classical Deviation: Organizational and Individual Status as Antecedents of Conformity.” Academy of Management Journal, 59: 65–89.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou

2018 “Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes.” Proceedings of the National Academy of Sciences, 115: E3635–E3644.

Gioia, D. A.

1998 “From Individual to Organizational Identity.” In Identity in Organizations: Building Theory through Conversations Foundations for Organizational Science 17–31. Thousand Oaks, CA, US: Sage Publications, Inc.

Gioia, D. A., S. D. Patvardhan, A. L. Hamilton, and K. G. Corley

2013 “Organizational Identity Formation and Change.” Academy of Management Annals, 7: 123–193.

Gioia, D. A., M. Schultz, and K. G. Corley

2000 “Organizational Identity, Image, and Adaptive Instability.” Academy of Management Review, 25: 63–81.

Giorgi, S. and K. Weber

2015 “Marks of Distinction: Framing and Audience Appreciation in the Context of Investment Advice.” Administrative Science Quarterly, 60: 333–367.

Goffman, E.

1959 The Presentation of Self in Everyday Life. New York, NY: Anchor 1 edition edition.

Goldberg, A.

2015 “In Defense of Forensic Social Science.” Big Data & Society, 2: 2053951715601145.

Goldberg, A., M. T. Hannan, and B. Kovács

2016a “What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption.” American Sociological Review, 81: 215–241.

Goldberg, A., S. B. Srivastava, V. G. Manian, W. Monroe, and C. Potts

2016b “Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness.” American Sociological Review, 81: 1190–1222.

Goldfarb, B. and L. Yan

2021 “Revisiting Zuckerman’s (1999) Categorical Imperative: An Application of Epistemic Maps for Replication.” Strategic Management Journal, 42: 1963–1992.

Grand, G., I. A. Blank, F. Pereira, and E. Fedorenko

2022 “Semantic Projection Recovers Rich Human Knowledge of Multiple Object Features from Word Embeddings.” Nature Human Behaviour, 6: 975–987.

Guo, W., M. Sengul, and T. Yu

2019 “Rivals’ Negative Earnings Surprises, Language Signals, and Firms’ Competitive Actions.” Academy of Management Journal, 63: 637–659.

Haans, R. F. J.

2019 “What’s the Value of Being Different When Everyone Is? The Effects of Distinctiveness on Performance in Homogeneous versus Heterogeneous Categories.” Strategic Management Journal, 40: 3–27.

Hamilton, W. L., J. Leskovec, and D. Jurafsky

2016 “Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change.” Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing, 2016: 2116–2121.

Hamilton, W. L., J. Leskovec, and D. Jurafsky

2018 “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” arXiv:1605.09096 [cs], .

Hannan, M. T., G. L. Mens, G. Hsu, B. Kovács, G. Negro, L. Pólos, E. Pontikes, and A. J. Sharkey

2019 Concepts and Categories: Foundations for Sociological and Cultural Analysis. Columbia University Press.

Harmon, D. J.

2019 “When the Fed Speaks: Arguments, Emotions, and the Microfoundations of Institutions.” Administrative Science Quarterly, 64: 542–575.

Harris, Z. S.

1954 “Distributional Structure.” *WORD*, 10: 146–162.

Hartzmark, S. M. and K. Shue

2018 “A Tough Act to Follow: Contrast Effects in Financial Markets.” The Journal of Finance, 73: 1567–1613.

Healy, K.

2017 “Fuck Nuance.” Sociological Theory, 35: 118–127.

Highhouse, S., M. E. Brooks, and G. Gregarus

2009 "An Organizational Impression Management Perspective on the Formation of Corporate Reputations." Journal of Management, 35: 1481–1493.

Hoberg, G. and G. Phillips

2016 "Text-Based Network Industries and Endogenous Product Differentiation." Journal of Political Economy, 124: 1423–1465.

Hollander, J. A.

2013 ""I Demand More of People": Accountability, Interaction, and Gender Change." Gender & Society, 27: 5–29.

Hsu, G.

2006 "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." Administrative Science Quarterly, 51: 420–450.

Hsu, G. and M. T. Hannan

2005 "Identities, Genres, and Organizational Forms." Organization Science, 16: 474–490.

Hsu, G., M. T. Hannan, and Ö. Koçak

2009 "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." American Sociological Review, 74: 150–169.

Ibarra, H. and R. Barbulescu

2010 "Identity as Narrative: Prevalence, Effectiveness, and Consequences of Narrative Identity Work in Macro Work Role Transitions." The Academy of Management Review, 35: 135–154.

Jurafsky, D.

2015 The Language of Food: A Linguist Reads the Menu. New York: W. W. Norton & Company reprint edition edition.

Kasznik, R. and M. F. McNichols

2002 “Does Meeting Earnings Expectations Matter? Evidence from Analyst Forecast Revisions and Share Prices.” Journal of Accounting Research, 40: 727–759.

Keuschnigg, M. and T. Wimmer

2017 “Is Category Spanning Truly Disadvantageous? New Evidence from Primary and Secondary Movie Markets.” Social Forces, 96: 449–479.

Kile, C. O. and M. E. Phillips

2009 “Using Industry Classification Codes to Sample High-Technology Firms: Analysis and Recommendations.” Journal of Accounting, Auditing & Finance, 24: 35–58.

Kovács, B. and R. Johnson

2014 “Contrasting Alternative Explanations for the Consequences of Category Spanning: A Study of Restaurant Reviews and Menus in San Francisco.” Strategic Organization, 12: 7–37.

Kozlowski, A. C., M. Taddy, and J. A. Evans

2019 “The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings.” American Sociological Review, 84: 905–949.

Lee, J.

2016 “Can Investors Detect Managers’ Lack of Spontaneity? Adherence to Predetermined Scripts during Earnings Conference Calls.” Accounting Review, 91: 229–250.

Lenci, A.

2018 “Distributional Models of Word Meaning.” Annual Review of Linguistics, 4: 151–171.

Leung, M. D. and A. J. Sharkey

2013 “Out of Sight, Out of Mind? Evidence of Perceptual Factors in the Multiple-Category Discount.” Organization Science, 25: 171–184.

Livnat, J. and R. R. Mendenhall

2006 “Comparing the Post–Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts.” Journal of Accounting Research, 44: 177–205.

Lix, K., A. Goldberg, S. B. Srivastava, and M. A. Valentine

2022 "Aligning Differences: Discursive Diversity and Team Performance." Management Science, .

Loughran, T. and B. Mcdonald

2011 "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks." The Journal of Finance, 66: 35–65.

Matsumoto, D. A.

2002 "Management's Incentives to Avoid Negative Earnings Surprises." The Accounting Review, 77: 483–514.

McFarland, D. A., K. Lewis, and A. Goldberg

2016 "Sociology in the Era of Big Data: The Ascent of Forensic Social Science." The American Sociologist, 47: 12–35.

Mead, G. H.

1934 Mind, Self and Society. Chicago: University of Chicago Press.

Mikolov, T., I. Sutskever, K. Chen, G. S. Corrado, and J. Dean

2013 "Distributed Representations of Words and Phrases and Their Compositionality." In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger (eds.), Advances in Neural Information Processing Systems 26 3111–3119. Curran Associates, Inc.

Murphy, G.

2004 The Big Book of Concepts. Cambridge: A Bradford Book revised edition edition.

Navis, C. and M. A. Glynn

2011 "Legitimate Distinctiveness and The Entrepreneurial Identity: Influence on Investor Judgments of New Venture Plausibility." Academy of Management Review, 36: 479–499.

Nelson, L. K.

2020 “Computational Grounded Theory: A Methodological Framework.” Sociological Methods & Research, 49: 3–42.

Pan, L., G. McNamara, J. J. Lee, J. J. Haleblian, and C. E. Devers
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: 2204–2225.

Paolella, L. and R. Durand
2015 “Category Spanning, Evaluation, and Performance: Revised Theory and Test on the Corporate Law Market.” Academy of Management Journal, 59: 330–351.

Payne, J. L. and W. B. Thomas
2003 “The Implications of Using Stock-Split Adjusted I/B/E/S Data in Empirical Research.” The Accounting Review, 78: 1049–1067.

Pfarrer, M. D., T. G. Pollock, and V. P. Rindova
2010 “A Tale of Two Assets: The Effects of Firm Reputation and Celebrity on Earnings Surprises and Investors’ Reactions.” Academy of Management Journal, 53: 1131–1152.

Phillips, D. J., C. J. Turco, and E. W. Zuckerman
2013 “Betrayal as Market Barrier: Identity-Based Limits to Diversification among High-Status Corporate Law Firms.” American Journal of Sociology, 118: 1023–1054.

Pontikes, E. G.
2012 “Two Sides of the Same Coin: How Ambiguous Classification Affects Multiple Audiences’ Evaluations.” Administrative Science Quarterly, 57: 81–118.

Pontikes, E. G. and R. Kim
2017 “Strategic Categorization.” In From Categories to Categorization: Studies in Sociology, Organizations and Strategy at the Crossroads volume 51 of Research in the Sociology of Organizations 71–111. Emerald Publishing Limited.

Quartz, S. and A. Asp

2016 Cool: How the Brain's Hidden Quest for Cool Drives Our Economy and Shapes Our World. Farrar, Straus and Giroux reprint edition edition.

Rao, H., P. Monin, and R. Durand

2005 "Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy." American Sociological Review, 70: 968–991.

Ruef, M. and K. Patterson

2009 "Credit and Classification: The Impact of Industry Boundaries in Nineteenth-Century America." Administrative Science Quarterly, 54: 486–520.

Schultz, M. and T. Hernes

2012 "A Temporal Perspective on Organizational Identity." Organization Science, 24: 1–21.

Sgourev, S. V. and N. Althuizen

2014 ""Notable" or "Not Able": When Are Acts of Inconsistency Rewarded?" American Sociological Review, 79: 282–302.

Shanthikumar, D. M.

2012 "Consecutive Earnings Surprises: Small and Large Trader Reactions." The Accounting Review, 87: 1709–1736.

Skinner, D. J. and R. G. Sloan

2002 "Earnings Surprises, Growth Expectations, and Stock Returns or Don't Let an Earnings Torpedo Sink Your Portfolio." Review of Accounting Studies, 7: 289–312.

Smith, E. B.

2011 "Identities as Lenses: How Organizational Identity Affects Audiences' Evaluation of Organizational Performance." Administrative Science Quarterly, 56: 61–94.

Staw, B. M., L. E. Sandelands, and J. E. Dutton

1981 “Threat Rigidity Effects in Organizational Behavior: A Multilevel Analysis.” Administrative Science Quarterly, 26: 501–524.

Taeuscher, K., R. Bouncken, and R. Pesch
2021 “Gaining Legitimacy by Being Different: Optimal Distinctiveness in Crowdfunding Platforms.” Academy of Management Journal, 64: 149–179.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones
2013 “Atypical Combinations and Scientific Impact.” Science, 342: 468–472.

Wasserman, V. and M. Frenkel
2011 “Organizational Aesthetics: Caught Between Identity Regulation and Culture Jamming.” Organization Science, 22: 503–521.

West, C. and S. Fenstermaker
1995 “Doing Difference.” Gender & Society, 9: 8–37.

West, C. and D. H. Zimmerman
1987 “Doing Gender.” Gender & Society, 1: 125–151.

Westphal, J. D., S. H. Park, M. L. McDonald, and M. L. A. Hayward
2012 “Helping Other CEOs Avoid Bad Press: Social Exchange and Impression Management Support among CEOs in Communications with Journalists.” Administrative Science Quarterly, 57: 217–268.

Whetten, D. A.
2006 “Albert and Whetten Revisited: Strengthening the Concept of Organizational Identity.” Journal of Management Inquiry, 15: 219–234.

Wry, T., M. Lounsbury, and P. D. Jennings
2014 “Hybrid Vigor: Securing Venture Capital by Spanning Categories in Nanotechnology.” Academy of Management Journal, 57: 1309–1333.

Zhao, E. Y., G. Fisher, M. Lounsbury, and D. Miller

2017 "Optimal Distinctiveness: Broadening the Interface between Institutional Theory and Strategic Management." Strategic Management Journal, 38: 93–113.

Zhao, E. Y., M. Ishihara, P. D. Jennings, and M. Lounsbury

2018 "Optimal Distinctiveness in the Console Video Game Industry: An Exemplar-Based Model of Proto-Category Evolution." Organization Science, 29: 588–611.

Zilber, T. B.

2006 "The Work of the Symbolic in Institutional Processes: Translations of Rational Myths in Israeli High Tech." Academy of Management Journal, 49: 281–303.

Zuckerman, E. W.

1999 "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." American Journal of Sociology, 104: 1398–1438.

Zuckerman, E. W.

2000 "Focusing the Corporate Product: Securities Analysts and De-diversification." Administrative Science Quarterly, 45: 591–619.

Zuckerman, E. W.

2004 "Structural Incoherence and Stock Market Activity." American Sociological Review, 69: 405–432.

Zuckerman, E. W.

2016 Optimal Distinctiveness Revisited 183–199.

Zuckerman, E. W.

2017 The Categorical Imperative Revisited: Implications of Categorization as a Theoretical Tool.

Figures

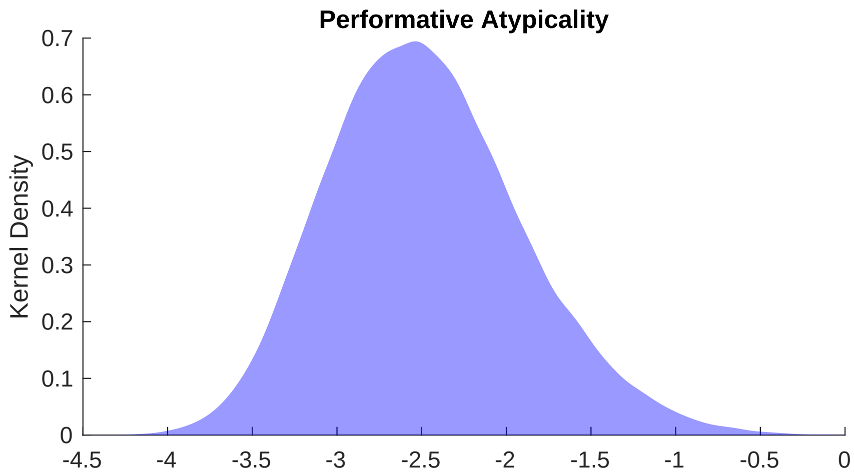

Figure 1: Kernel density for Performative Atypicality

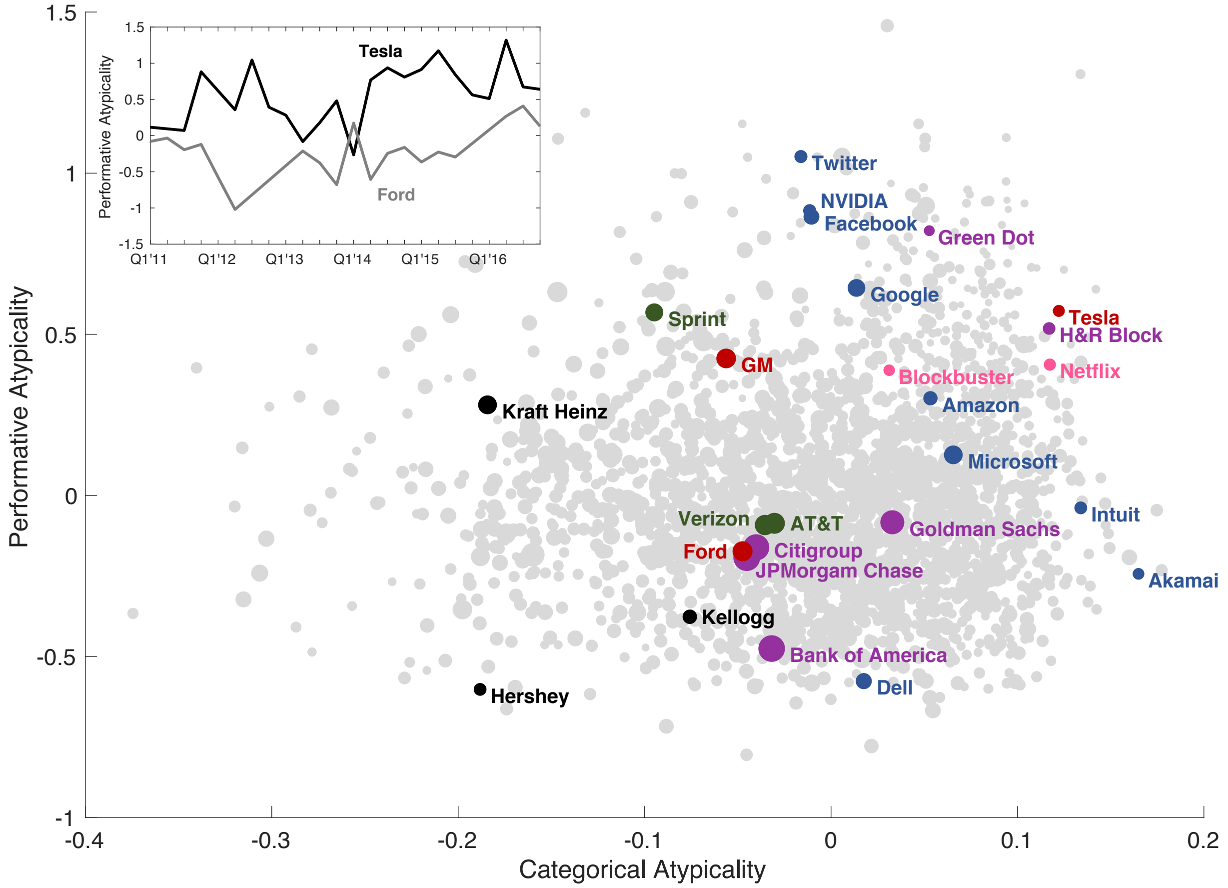

Figure 2: Atypicality by firm. Each dot represents one firm's mean (standardized) performative and categorical atypicality (for firms with a minimum of 10 quarterly observations). Highlighted firms are color coded by FIC200 industry. The inset plots Tesla's and Ford's performative atypicality over time.

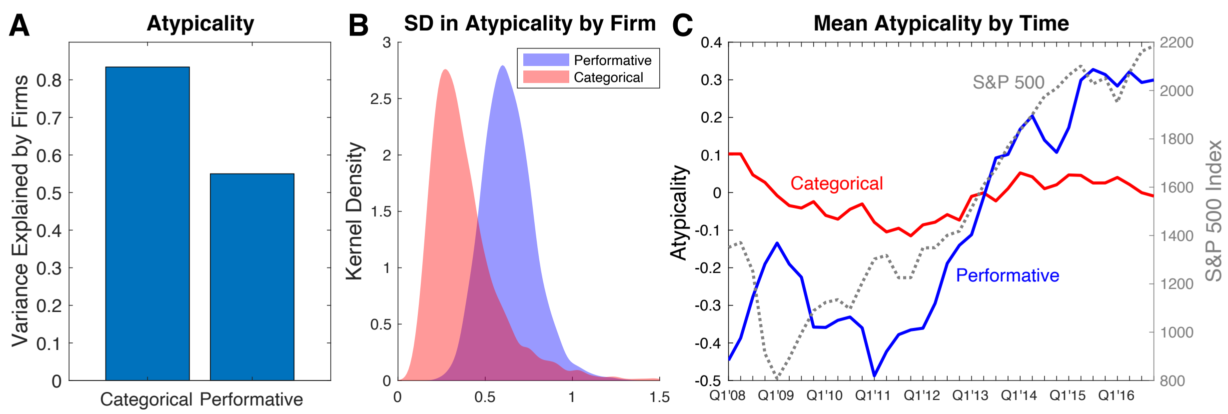

Figure 3: (A) The proportion of variance in performative and categorical atypicality explained by fixed firm differences. (B) Kernel densities for the standard deviation, by firm, in performative and categorical atypicality. (C) Mean performative and categorical atypicality by quarter. The dotted line corresponds to the S&P 500 index.

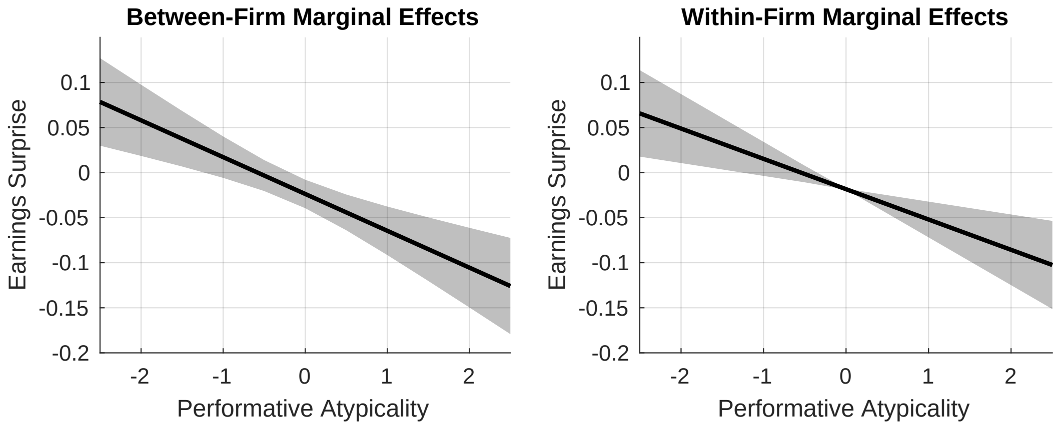

Figure 4: Marginal effects of between-firm (left) and within-firm (right) performative atypicality on analyst disagreement (Models 3 and 6, Table 2).

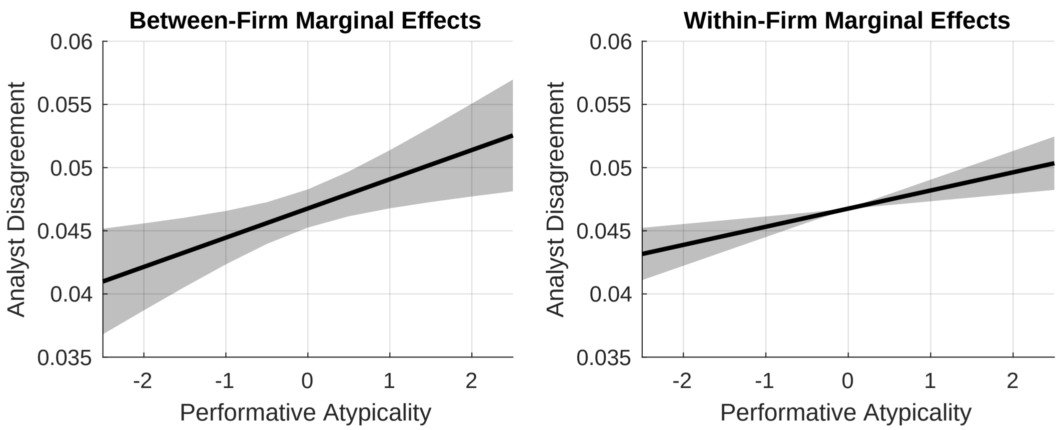

Figure 5: Marginal effects of between-firm (left) and within-firm (right) performative atypicality on earnings surprise (Models 3 and 6, Table 3).

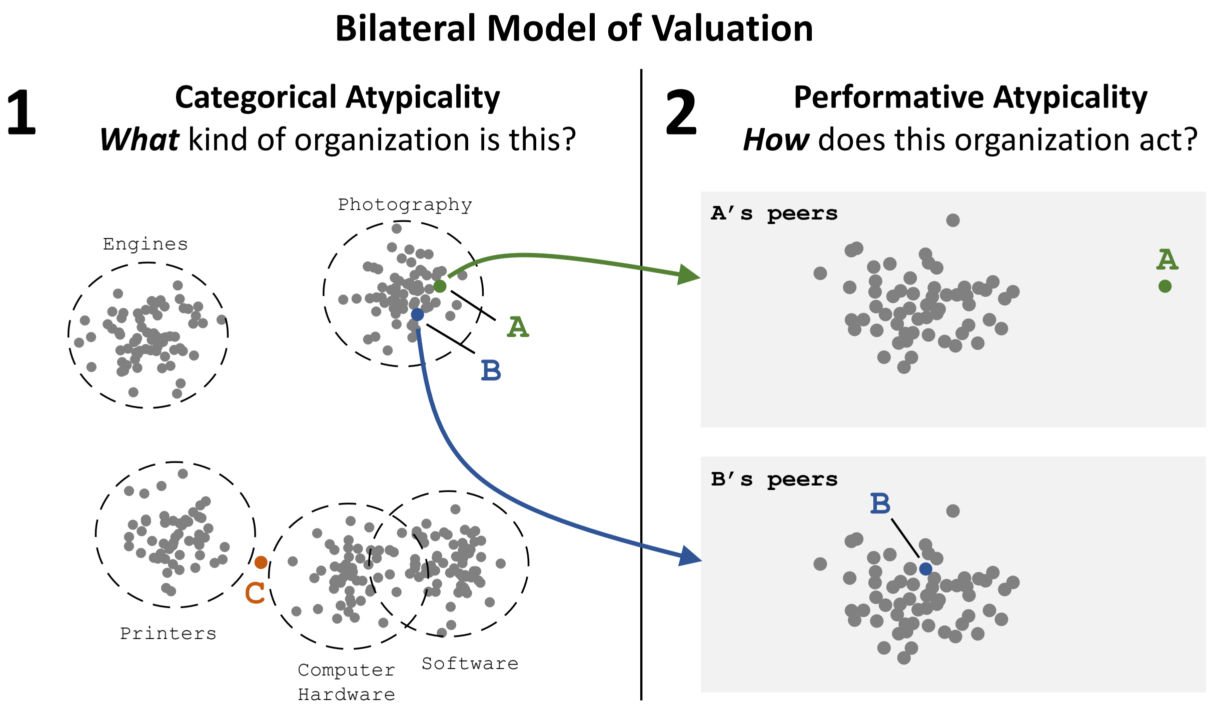

Figure 6: An illustration of the bilateral model of valuation, depicting a hypothetical audience member evaluating three firms labeled A, B and C. Grayed dots represent other firms. In the first stage of valuation, each firm is located in a stylized categorical space, where labels correspond to categories and dashed lines represent perceived categorical boundaries. In the second stage, each firm is located in a stylized performative space populated by its perceived group of peers.

Tables

Table 1: Descriptive Statistics and Correlations

	VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	Min	Max	Mean	Median	Std. Dev.
(1)	Earnings surprise	1.000													-11.39	5.525	-0.023	0.0533	1.459
(2)	Disagreement	-0.093	1.000												0	0.444	0.0467	0.0247	0.0677
(3)	Performative Atyp.	-0.028	0.028	1.000											-4.440	0.0995	-2.490	-2.525	0.571
(4)	Categorical Atyp.	-0.020	-0.081	0.156	1.000										0	0.993	0.774	0.785	0.104
(5)	Log of assets	0.019	0.195	-0.248	-0.481	1.00									1.920	14.76	7.637	7.629	1.890
(6)	Leverage	-0.057	0.123	-0.102	-0.253	0.498	1.000								0.0765	1.015	0.541	0.548	0.223
(7)	Prec. pos. surp.	0.114	-0.053	-0.062	-0.056	0.078	-0.025	1.000							0	1	0.611	1	0.488
(8)	Order in quarter	-0.020	-0.033	0.097	0.122	-0.238	-0.105	-0.053	1.000						0	68	24.65	23	10.98
(9)	Positivity	0.066	-0.150	-0.103	0.040	-0.031	-0.038	0.132	0.039	1.000					-1	0.857	0.0943	0.107	0.219
(10)	Horizon	0.017	0.085	0.014	-0.117	0.223	0.077	0.029	-0.019	0.075	1.000				-1	0.714	-0.648	-0.673	0.192
(11)	Litigious	-0.022	0.062	0.099	-0.047	0.055	0.075	-0.029	-0.010	-0.175	0.077	1.000			0	1	0.248	0	0.432
(12)	Log of length	0.033	0.051	-0.519	-0.251	0.378	0.084	0.070	-0.115	0.126	0.159	-0.049	1.000		5.298	9.243	7.470	7.568	0.514
(13)	Analysts churn	0.012	-0.080	0.059	0.103	-0.123	-0.057	-0.017	0.018	0.029	-0.018	0.005	-0.087	1.000	0	1	0.187	0.143	0.201
(14)	No of estimates	0.050	0.100	-0.238	-0.392	0.551	0.086	0.101	-0.141	0.048	0.133	-0.032	0.489	-0.118	1	48	9.938	8	7.400

Table 2: OLS of Analyst Disagreement (lagged)

	Between-Firm†			Within-Firm†		
	(1)	(2)	(3)	(4)	(5)	(6)
Performative Atypicality		0.002**	0.002**		0.001***	0.001***
		(2.86)	(2.83)		(3.38)	(3.37)
Categorical Atypicality	0.002*		0.002*	-0.000		-0.000
	(2.16)		(2.14)	(-0.48)		(-0.44)
Firm Attributes						
Leverage†	0.009*	0.009*	0.009*	0.045***	0.045***	0.045***
	(2.25)	(2.19)	(2.31)	(10.13)	(10.12)	(10.11)
Log of assets†	0.005***	0.005***	0.005***	0.021***	0.021***	0.021***
	(6.68)	(6.70)	(6.73)	(11.55)	(11.56)	(11.55)
Preceding pos. surprise	-0.004***	-0.004***	-0.004***	-0.003***	-0.003***	-0.003***
	(-5.04)	(-5.09)	(-4.99)	(-4.93)	(-4.90)	(-4.90)
Call Attributes						
Order in quarter	-0.000	-0.000	-0.000	0.000	0.000	0.000
	(-0.33)	(-0.31)	(-0.42)	(0.62)	(0.70)	(0.70)
Positivity	-0.025***	-0.024***	-0.025***	-0.015***	-0.015***	-0.015***
	(-9.55)	(-9.25)	(-9.31)	(-10.33)	(-10.20)	(-10.19)
Horizon	0.011***	0.010***	0.010***	0.004**	0.003*	0.003*
	(4.48)	(4.21)	(4.21)	(2.64)	(2.42)	(2.42)
Litigiousness	0.002	0.002	0.002	0.001	0.001	0.001
	(1.47)	(1.37)	(1.35)	(0.96)	(0.90)	(0.90)
Log of length	0.000	0.002	0.002	-0.000	0.001	0.001
	(0.22)	(1.54)	(1.60)	(-0.38)	(1.19)	(1.19)
Analyst Attributes						
Analysts churn†	-0.017***	-0.017***	-0.017***	-0.001	-0.001	-0.001
	(-8.51)	(-8.50)	(-8.55)	(-1.12)	(-1.12)	(-1.11)
No. of estimates†	0.000*	0.000*	0.000*	0.001***	0.001***	0.001***
	(2.52)	(2.18)	(2.44)	(6.17)	(6.17)	(6.15)
Constant	0.011	-0.002	-0.006	-0.137***	-0.146***	-0.146***
	(0.99)	(-0.16)	(-0.48)	(-9.44)	(-9.60)	(-9.60)
Industry/Year/Quarter FE	Yes	Yes	Yes	No	No	No
Firm FE	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes
Observations	60688	60688	60688	61440	61440	61440
R^2	0.217	0.216	0.217	0.516	0.516	0.516

t statistics in parentheses, Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: OLS of Earnings Surprise (lagged)

	Between-Firm†			Within-Firm†		
	(1)	(2)	(3)	(4)	(5)	(6)
Performative Atypicality		-0.041***	-0.041***		-0.034***	-0.034***
		(-4.16)	(-4.16)		(-3.43)	(-3.40)
Categorical Atypicality	-0.011		-0.011	0.015		0.014
	(-0.95)		(-0.92)	(0.66)		(0.61)
Firm Attributes						
Leverage†	-0.362***	-0.364***	-0.366***	-0.534***	-0.533***	-0.532***
	(-6.81)	(-6.84)	(-6.87)	(-4.77)	(-4.77)	(-4.76)
Log of assets†	0.035***	0.035***	0.034***	-0.023	-0.023	-0.022
	(4.51)	(4.59)	(4.38)	(-0.69)	(-0.68)	(-0.67)
Preceding pos. surprise	0.243***	0.243***	0.242***	0.075***	0.074***	0.075***
	(15.12)	(15.13)	(15.08)	(5.58)	(5.56)	(5.56)
Call Attributes						
Order in quarter	-0.002***	-0.002***	-0.002***	-0.000	-0.000	-0.000
	(-3.46)	(-3.39)	(-3.36)	(-0.33)	(-0.39)	(-0.40)
Positivity	0.158***	0.145***	0.148***	0.278***	0.271***	0.271***
	(4.16)	(3.85)	(3.90)	(7.16)	(7.01)	(6.99)
Horizon	0.033	0.045	0.045	-0.022	-0.015	-0.015
	(0.91)	(1.24)	(1.24)	(-0.58)	(-0.38)	(-0.38)
Litigiousness	-0.050**	-0.047*	-0.047*	-0.056**	-0.055**	-0.055**
	(-2.70)	(-2.56)	(-2.55)	(-3.15)	(-3.10)	(-3.10)
Log of length	-0.001	-0.038	-0.039	-0.015	-0.045	-0.045
	(-0.07)	(-1.63)	(-1.64)	(-0.61)	(-1.79)	(-1.79)
Analyst Attributes						
Analysts churn†	0.093*	0.093*	0.094**	0.098**	0.098**	0.098**
	(2.56)	(2.57)	(2.59)	(3.01)	(3.01)	(3.00)
No. of estimates†	0.006***	0.006***	0.006***	0.002	0.002	0.002
	(3.53)	(3.93)	(3.70)	(0.66)	(0.67)	(0.69)
Analyst Disagreement†	-1.343***	-1.333***	-1.330***	-1.939***	-1.932***	-1.931***
	(-5.40)	(-5.38)	(-5.36)	(-6.55)	(-6.54)	(-6.53)
Constant	-0.164	0.108	0.125	0.547	0.779**	0.774**
	(-1.00)	(0.61)	(0.70)	(1.93)	(2.67)	(2.66)
Industry/Year/Quarter FE	Yes	Yes	Yes	No	No	No
Firm FE	No	No	No	Yes	Yes	Yes
Year FE	No	No	No	Yes	Yes	Yes
Observations	60688	60688	60688	61440	61440	61440
R^2	0.117	0.117	0.117	0.214	0.215	0.215

t statistics in parentheses, Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: OLS of Earnings Surprise (lagged)

	(1)	(2)
Performative Atypicality		
2^{nd} quintile	-0.017	-0.024
	(-0.94)	(-1.49)
3^{rd} quintile	-0.027	-0.033
	(-1.44)	(-1.86)
4^{th} quintile	-0.042*	-0.029
	(-1.97)	(-1.37)
5^{th} quintile	-0.111***	-0.095***
	(-4.16)	(-3.62)
Categorical Atypicality	-0.011	0.014
	(-0.95)	(0.62)
Analyst Disagreement†	-1.333***	-1.934***
	(-5.37)	(-6.54)
Constant	0.128	0.772**
	(0.73)	(2.66)
Firm Controls	Yes	Yes
Call Controls	Yes	Yes
Analyst Controls	Yes	Yes
Industry/Year/Quarter FE	Yes	No
Firm FE	No	Yes
Year FE	No	Yes
Observations	60688	61440
R^2	0.117	0.215

t statistics in parentheses

Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: OLS of Earnings Surprise (lagged)

	High-Technology		Innovative Companies	
	(1)	(2)	(3)	(4)
Performative Atypicality				
Innovation-Biased	-0.574***	-0.697***	-0.500***	-0.508**
	(-3.38)	(-4.44)	(-3.31)	(-3.23)
Non Innovation-Biased	-0.185*	-0.129	-0.204*	-0.178*
	(-2.01)	(-1.49)	(-2.22)	(-2.03)
Categorical atypicality	-0.011	0.015	-0.011	0.014
	(-0.92)	(0.67)	(-0.93)	(0.62)
Constant	-0.170	0.519	-0.166	0.548
	(-1.03)	(1.83)	(-0.99)	(1.94)
Firm Controls	Yes	Yes	Yes	Yes
Call Controls	Yes	Yes	Yes	Yes
Analyst Controls	Yes	Yes	Yes	Yes
Industry/Year/Quarter	Yes	No	Yes	No
Firm FE	No	Yes	No	Yes
Year FE	No	Yes	No	Yes
Observations	60688	61440	60654	61354
R^2	0.117	0.215	0.117	0.214

t statistics in parentheses

Standard errors clustered by firm

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A Validating the Word Embedding Model

There are two main techniques for validating word embedding models: most-similar queries and word analogy tasks. The idea with most-similar queries is to find words that are closest in semantic space to a focal word and assess whether it makes sense for these words to be in close proximity to one another. For example, in a corporate setting, the word "board" might refer to a board of directors, whereas in construction a "board" might reference a physical object. In our data, the words closest to "boards" in Q4 2016 were: "committee," "directors," and "CEOs." In the same time period, the words closes to "drilling" were "completions," "exploration," and "fracking."

Because we fit different word embedding models for different time periods, we can also recover subtle changes in word meanings that occurred during our observation period. As an illustration, in the model for Q4 2006, the word closest to "phones" was "cell," given that cell phones were still common and smartphones had not yet come on the scene. In Q4 2016, the word closest to "phones" was "smartphones," which by then had become ubiquitous. Our queries also revealed that the models capture context-specific semantic relationships. For example, the word "color" is often used by analysts when they ask managers to "give some more color" on a given topic. Consistent with this meaning of the word in the context of analyst calls, we found that the words closest to "color" throughout the observation period were: "granularity," "detail," and "insight."

To further establish model validity, we examined whether mathematical operations in the vector space produced by our embeddings model could solve analogical reasoning problems. For example, Mikolov et al. (2013) showed that "King" - "Man" + "Woman" = "Queen." That is, their model captured the notion that man is to king as woman is to queen. Applying this approach to our embedding models, we found, for example, that "Boeing" - "USA" + "Europe" = "Airbus." Examples of other analogy tasks we tested on our models are shown in Table A1. Overall, these analyses indicated that our embedding models captured semantically meaningful relationships between words used in quarterly earnings calls.

Table A1: Sample Analogy Tasks Applied to the Word Embedding Model for Q4 2016

Analogy Task	Answer
Toyota - Japan + Germany = ?	BMW
Boeing - USA + Europe = ?	Airbus
Huawai - China + Korea = ?	Samsung
Amazon - America + China = ?	Alibaba
Youtube + Series = ?	Netflix
Amazon - Stores = ?	AWS (Amazon Web Services)
Google + Finance = ?	Yahoo
Microsoft - Windows = ?	Dell
Employees - Managers + Parents = ?	Children
Stakeholders - Stakes + Stocks = ?	Stockholders
CEO - Organization + Finance = ?	CFO
Shareholders - Shares + Property = ?	Landlord
Managers - Management + Consulting = ?	Consultants

Appendix B Does performative atypicality predict future earnings?

Results reported in Table B1, estimating the effect of performative atypicality on earnings per share, show that performative atypicality is not a predictor of future earnings. Whether using a between- or within-firm specification, the estimators are insignificant. We produce these estimations using specifications that only include firm size as a control (Models 1 and 3), as well as specifications that control for other firm, call, and analyst attributes (Models 2 and 4). If performative atypicality were correlated with other indicators of lower future earnings (such as a call's order in the quarter), then the inclusion of such indicators would have undermined the effect of performative atypicality. Yet this effect is insignificant even in Models 1 and 3 in which no potentially collinear indicators are included.

Table B1: OLS of Earnings Per Share (lagged)

	Between-Firm		Within-Firm	
	(1)	(2)	(3)	(4)
Performative Atypicality	-0.002	0.006	-0.003	0.007
	(-0.21)	(0.43)	(-0.85)	(1.59)
Categorical Atypicality	0.007	0.010	-0.003	-0.004
	(0.60)	(0.81)	(-0.27)	(-0.40)
Firm Attributes				
Leverage†		-0.404***		-0.346***
		(-5.74)		(-6.14)
Log of assets†	0.177***	0.179***	0.192***	0.199***
	(13.61)	(12.86)	(5.64)	(5.85)
Preceding pos. surprise		0.146***		0.072***
		(15.01)		(10.58)
Call Attributes				
Order in quarter		-0.004***		-0.003***
		(-4.72)		(-5.65)
Positivity		0.042		0.154***
		(1.43)		(9.35)
Long-term time horizon		0.067		-0.014
		(1.95)		(-0.74)
High proportion of litigious words		-0.008		-0.002
		(-0.55)		(-0.26)
Log of length		0.008		0.031**
		(0.50)		(2.95)
Analyst Attributes				
Analysts churn†		-0.112***		0.037***
		(-5.60)		(3.47)
Constant	-0.924***	-0.717***	-1.041***	-1.143***
	(-9.95)	(-5.12)	(-3.99)	(-4.31)
Industry/Year/Quarter FE	Yes	Yes	No	No
Firm FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes
Observations	60688	60688	61440	61440
R^2	0.253	0.269	0.604	0.609

t statistics in parentheses, Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix C Inductive Analysis of Performative Atypicality

This appendix provide a detailed description of the inductive analysis of performative atypicality that is summarized in the main text. Here, we provide full technical details of analyses we conducted. We follow the three step structure reported in the main text.

C.1 Aligning Word Embedding Models

To construct performative atypicality at the firm-year-quarter level, we produce separate word embedding models for each year-quarter period independently. Because of the stochastic nature of word embedding models, these separate models are not naturally aligned (namely, an embedding location is not equivalent across two time periods). In analyzing what explains performative atypicality we seek to identify time-invariant patterns that underlie atypicality centroids across all observations. To do so, we rotate each time period using orthogonal Procrustes applied to words that appear in all periods, as explained by Hamilton, Leskovec, and Jurafsky (2018). This procedure produces a rotational alignment wherein positions across time periods are comparable.

Let $\mathbf{W}^{(t)}$ be the word embedding model for time t. The aligned model is obtained using:

$$\mathbf{R}^{(t)} = argmin_{\mathbf{Q^TQ=I}} \|\mathbf{QW^{(t)}} - \mathbf{W^{(t+1)}}\|_F \tag{C1}$$

C.2 Step 1: Semantic Pattern Detection

In the first step we detect patterns in performative atypicality. We define an earnings call's *atypicality centroid*, $\alpha_{f,q}$, as its corresponding industry embedding centroid, subtracted from the call's embedding centroid:

$$\alpha_{f,q} = V_{f,q} - PV_{f,q} \tag{C2}$$

where f and q index firms and year-quarters, respectively.

We use principle component analysis (PCA) to identify the dominant dimensions of variability in atypicality centroids. PCA is a dimensionality reduction technique that transforms a dataset into a new coordinate system where dimensions are descendingly ordered by the variation they explain in the data. An observation's loading is its position on a dimension in this new coordinate system. Note that these dimensions do not correspond to the meanings discussed in the earnings calls as such; rather, they represent the atypical meanings, relative to industry conventions, that executives express.

We apply the PCA analysis to atypicality centroids derived from the aligned model $\mathbf{R}^{(t)}$, as described in eq. C1. We produce the atypicality centroid matrix $\mathbf{A} \in \mathbb{R}^{N\times|V|}$, as defined in eq. C2, where N is the number of all firm-year-quarter observations. We apply PCA to $\mathbf{A}$ to produce $\Lambda \in \mathbb{R}^{|V|\times|V|}$, which corresponds to the loadings of each PCA dimension.

Panel A of Figure C1 plots the variance in atypicality centroids explained by the first twenty PCA dimensions. As it illustrates, the first four PCA dimensions cumulatively explain roughly 46% of the variance in quarterly earnings calls' atypicality. This suggests that atypical performances are structured by a few dominant axes of meaning.

Panel B summarizes the amount of variance in these four leading PCA dimensions explained by firms and industries. Each bar represents the proportion of variance in PCA loadings explained by either firms or industries for each dimension. As the graph illustrates, while differences between firms explain, on average, 40.5% of the variance in earning calls' loadings on the four

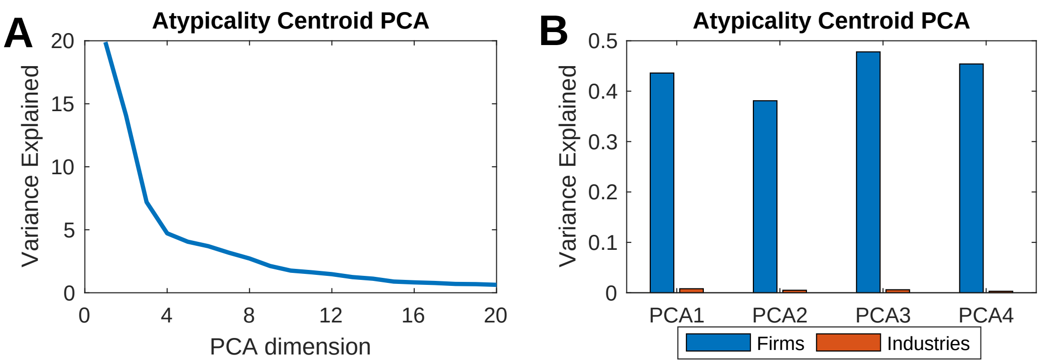

Figure C1: PCA of atypicality centroids. (A) Percent variance explained by leading PCA dimensions. (B) Proportion of variance in four leading PCA dimensions, explained by firms (blue) or industries (red).

leading PCA dimensions, differences between industries explain effectively none of that variance (0.58% on average). This strongly suggests that performantive atypicality does not merely capture topical variation driven by differences between industries. While Tesla's atypical performances have a characteristic flavor, for example, it is not unlikely that other firms in other industries express similar meanings when they diverge from their peers. These analyses of variance also rule out the concern that atypicality centroids merely measure call content. If that were the case, industry would have explained a significantly higher proportion of the variance.

C.3 Step 2: Interpreting the Semantic Axes of Atypicality

Supervised Variance Decomposition

In the second step of the analysis we interpret the meanings these PCA dimension represent. To do so, we use cosine similarity to identify the n closest words to each dimension. To rank words, we compare the cosine similarity between each dimension in Λ and all the words contained in the dictionary, based on their positions in the last period in $\mathbf{R}^{(t)}$. The twenty closest words and their corresponding similarities are listed in Table C1.

Examining these words provides insight into the different meanings that the four leading dimensions of variability in performative atypicality are structured on. Two of these dimensions are stylistic: one relating to expressions of politeness (PCA1), the other to expressions of emotion (PCA4). The remaining two dimensions are substantive in nature. The first (PCA2) appears to relate to financial reporting terminology; these are calls that are atypical in their excessive focus on reporting financial results. The second substantive dimension (PCA3) is not as readily interpretable, however, relating to legal terminology and high-level linguistic modality.

In Figure C2 we visualize the semantic concentration of these 20 closest words to each PCA dimension. To do so, we apply another PCA analysis to the last period in $\mathbf{R}^{(t)}$ to produce Λ'. The twenty closest words for the leading four dimensions in Λ are plotted as a function of the two leading PCA dimensions in Λ'. Each dot represents the location of one closest word. Overall, Figure C2 illustrates that the meanings encompassed by each dimension are fairly broad (especially PCA1) and that there appears to be significant semantic overlap between dimensions (especially between PCA3 and PCA4). This suggests that the high-level summary of semantic variance produced by the PCA analysis masks more subtle but potentially important variations

Table C1: Closest Words to Leading PCA Dimensions of Aligned Atypicality Centroids

PCA1		PCA2		PCA3		PCA4	
Token	$cos(\theta)$	Token	$cos(\theta)$	Token	$cos(\theta)$	Token	$cos(\theta)$
thank	0.791	quarter	0.655	recourse	0.499	funny	0.520
evening	0.657	fullyear	0.612	entitled	0.491	know	0.509
thanks	0.651	sequentially	0.603	theoretically	0.478	crazy	0.500
afternoon	0.635	yearoveryear	0.599	amended	0.469	boy	0.498
morning	0.621	fourth	0.598	legally	0.469	strange	0.495
hello	0.598	expecting	0.582	holdco	0.466	weird	0.475
hi	0.597	yeartodate	0.576	contingencies	0.465	anyways	0.467
welcome	0.588	yearonyear	0.572	permitted	0.459	honest	0.467
goodbye	0.584	sequential	0.558	penalties	0.456	gut	0.461
questions	0.581	fiscal	0.545	refinanced	0.455	hearing	0.461
beth	0.579	flattish	0.544	repay	0.452	nervous	0.457
conference	0.577	annualize	0.541	triggered	0.448	choppy	0.455
listening	0.573	implies	0.541	theoretical	0.447	headlines	0.454
julie	0.571	guiding	0.534	escrow	0.445	caught	0.453
webcast	0.569	guided	0.534	calculation	0.443	figured	0.452
harold	0.563	quarteroverquarter	0.525	repaid	0.443	mean	0.444
kelly	0.557	quarteronquarter	0.522	warrants	0.443	gosh	0.441
gentlemen	0.557	modestly	0.520	maximum	0.438	anecdotally	0.440
appreciate	0.555	guidance	0.519	amendment	0.437	mess	0.438
anne	0.549	normalize	0.518	calculated	0.433	surprising	0.438

Each columns lists the twenty words with the highest cosine similarity to the respective dimension's loadings in atypical meanings expressed in quarterly earnings calls. This heterogeneity is likely driven, to some extent at least, by time trends (for example, if certain technological or economic shifts influence the content of performative atypicality). We therefore proceed by analyzing each period independently.

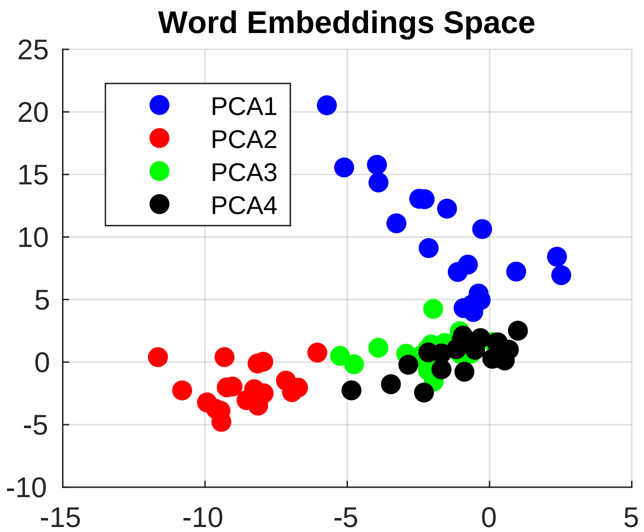

Figure C2: Top 20 closest words to four leading atypicality PCA centroids, positioned by leading two PCA dimensions derived from the rotated word embedding space.

To do so, we apply a separate PCA analysis to the (unaligned) atypicality centroid matrix $\mathbf{A^{(t)}}$ for each period t, 36 in total. We follow the same procedure described above to identify the n closest words to each of the four leading PCA dimensions produced for $\mathbf{A^{(t)}}$. This process produces 144 PCA dimensions. We then manually examine the 30 leading words in each of these 144 dimensions. Our process has two stages. In the first, we describe each word set with a label. In the second stage, we collapse these word sets into higher order sets based on semantic similarity. This leads us to identify eleven dimensions of meaning that structure variation in atypical performances. Each dimension is described by a set of words. These are listed in Figure C3.

Three of these dimensions correspond to the politeness, emotion and financial reporting dimensions discovered by the time-invariant PCA above (Table C1). Our qualitative interpretation identified eight additional dimensions. Three of these dimensions relate to strategy: one discussing differentiation and innovation, the second partnerships, and the third growth. Two refer to industry specifics: one is about product features (ranging from fragrance to footware) and the other to seasonal characteristics (presumably related to demand cycles). Two additional dimensions relate to high-level categories: linguistic modality (e.g., “permitted”) and extension (e.g., “include”). A final dimension appears to be specifically about financial valuation.

Interpretative Analysis

Atypical performances are organized along a variety of meanings. Some (like the “politeness” dimension) are stylistic, whereas others (like the “differentiation” dimension) are substantive. Are analysts equally responsive to these different types of atypical performances, or are some dimensions more strongly associated with earnings surprises than others?

To answer this question, we explore how these dimensions relate to analyst evaluations. Based on our interpretative analysis, we construct a centroid for each dimension d using the geometric mean of the words listed in Figure C3 for that dimension. This centroid is computed separately for each quarter, using its corresponding word embedding model. Using the cosine similarity, we compute the extent to which an earnings call's atypicality centroid is similar to each of these dimensions. We refer to this similarity as *dimensional atypicality*. Formally, a firm's dimensional atypicality is defined as:

$$DA_{f,q,d} = \cos(V_{f,q} - PV_{f,q}, V_{d,q}) \tag{C3}$$

where f and q index firms and year-quarters, and $V_{d,q}$ refers to the word embedding centroid for dimension d at time q. The higher $\beta_{f,q,d}$, the more executives of firm f atypically express meanings close to dimension d during time q.

We estimate the effect of each dimension on earnings surprises using OLS, extending the models 3 and 6 reported in Table 3. Specifically, we estimate between-firm models using the following specification:

$$Y_{f,q+1} = \beta_0 + \beta_1 DA_{f,q,d} + \beta_2 PA_{f,q} + \beta_3 CA_{f,q} + \beta_4 X_{f,q+1} + \beta_5 Z_{f,q} + \alpha_{i_f,q} + u_{f,q} \tag{C4}$$

where f indexes firms, q indexes quarter-year pairs, i indexes industries and d indexes dimensions (as listed in Figure C3). Y is earnings surprise, PA is performative atypicality, CA is categorical atypicality, X is a vector of contemporaneous controls, Z is a vector of lagged controls (see Table 3 for details) and u is the error term. Within-firm models are specified as follows:

$$Y_{f,q+1} = \beta_0 + \beta_1 DA_{f,q,d} + \beta_2 PA_{f,q} + \beta_3 CA_{f,q} + \beta_4 X_{f,q+1} + \beta_5 Z_{f,q} + \alpha_f + \theta_{y_q} + u_{f,q} \tag{C5}$$

Both models include performative atypicality as an additional independent variable. The coefficient estimates therefore correspond to the extent to which specific dimensional atypicality relates to earnings surprise, above and beyond the baseline effect of performative atypicality on surprise. When performative atypicality is not included in the model, similar results are obtained (with only one difference: the effect for dimensional atypicality for the extension dimension in the between-firm model becomes significant at t=-2.02)).

Figure C3 reports coefficients estimated by these models. Each coefficient is estimated in a separate model. As the figure illustrates, several findings emerge. First, we see that three dimensions of atypical meaning are strongly predictive of negative earnings surprises both in between- and within-firm models: differentiation, partnership, and modality. As executives diverge from industry conventions by connoting their firm's differentiated and innovative strategy, or its partnerships, analysts respond by being overly optimistic about its future performance. The same occurs when executives use linguistic modality, presumably as they discuss what their firm has, or does not have, latitude to do, possibly invoking actions that go against the grain. We are cautious not to over-speculate about the reasons for this optimism. One potential explanation is that deontic modality is expressed when executives discuss regulatory or legal approval, or intention to defy such restrictions, and that analysts overreact to such approval or defiance.

Second, the only dimension consistently and strongly predicting analysts' underestimation is the financial reporting dimension. When executives atypically discuss financial results, analysts

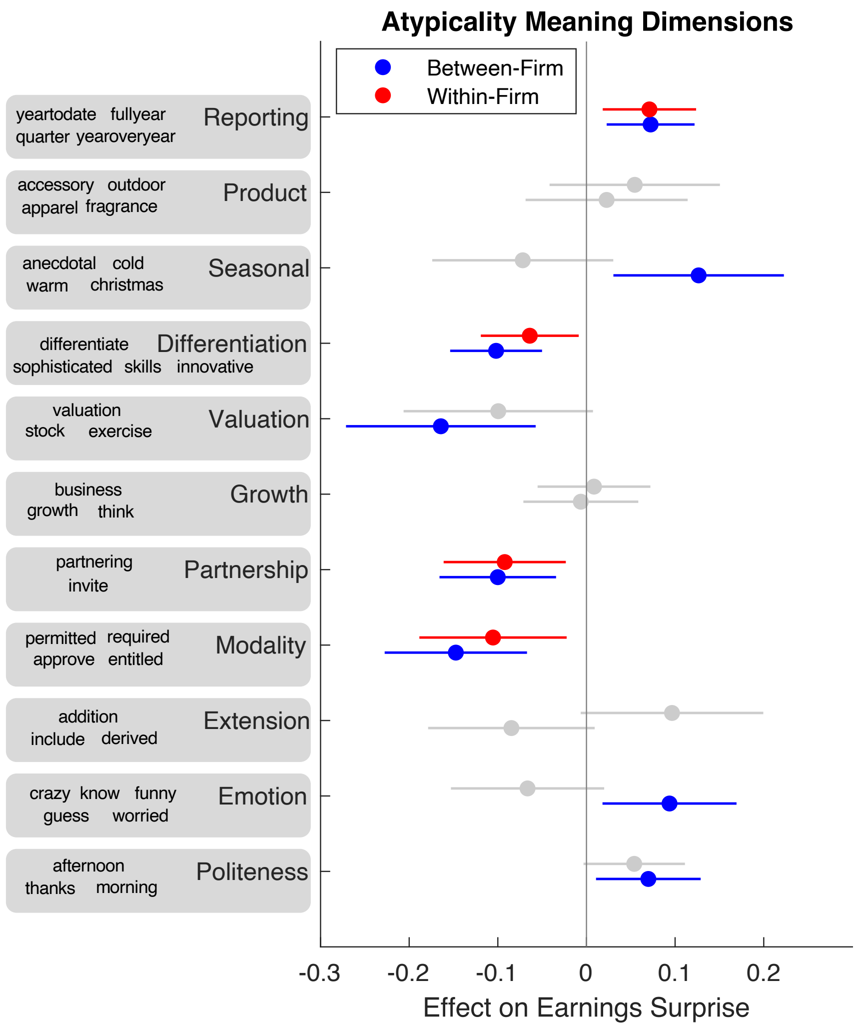

Figure C3: Dimensional Atypicality and its effect on earnings surprise. Coefficients for dimensional atypicality, and their confidence intervals, as estimated by between-firm (blue) and within-firm (red) models of earnings surprise. Each row corresponds to the centroid comprised of the lowercase words in the grayed box on the left. Models additionally include all variables specified in Table 3, models 3 and 6 respectively.

systematically become overly pessimistic about the firm's future performance. A potential explanation for this relationship may be that analysts interpret such a focus as defensiveness, leading them to conclude that managers are concealing private information. Yet, the strong negative correlation between this dimension and the differentiation (-0.79), partnership (-0.89) and modality (-0.63) dimensions suggest a different possibility. They indicate that atypical performances' focus on financials, or on firm strategy and opportunities, are mutually exclusive. Analysts might therefore interpret the latter as an indication of a firm's unique capabilities, while interpreting the former as indication of its formulism.

Finally, dimensions that are insignificantly or weakly related to earnings surprises provide additional insight. That not all coefficients are significant highlights that only some dimensions of atypicality provoke systematic responses: whereas analysts become optimistic when calls atypically discuss differentiation and innovation, atypical discussions of growth, for example, are unrelated to analysts' tendency to over or underestimate earnings. Not all forms of atypicality, in other words, influence perceptions of future performance.

The fact that the product dimension is insignificant is particularly informative (this is the case whether performative atypicality is or is not included in the model). It suggests that the relationship between performative atypicality and analyst reactions is not driven by atypical performances merely alluding to categorical atypicality, as manifest in firm atypical product offerings (see Appendix D for additional analyses). Stylistic atypicality is associated with earnings surprises, but only in between-firm models. Calls that are atypically polite or emotional are greeted with analyst pessimism. Unlike the enthusiasm generated by calls that atypically discuss differentiation or partnerships, non-substantive deviations from conventions appear to be generally interpreted as negative signals about firm performance. Finally, atypical discussions of firm valuation and investment are associated with analyst overestimation, but only in between-firm models. This may relate to firm-specific financials.

C.4 Step 3: Testing the Innovation-Biased Performative Atypicality Premium Hypothesis

Overall, the interpretative analysis above leads to one overarching conclusion: that analysts respond positively to performative atypicality when it conveys meanings associated with innovation and creativity. We test this conclusion in the third and final step of the inductive analysis, by exploring the relationships between innovation-biased and non innovation-biased atypicality on earnings surprises. Building on Garg et al. (2018), we define innovation-biased atypicality as the similarity between the meanings expressed in a call and those expressed in calls by innovative companies, minus the similarity between the meanings expressed in a call and those expressed in its peers calls. Innovation-biased atypicality grows as a call becomes more semantically similar to calls by innovative companies, and less semantically similar to a firm's immediate competitors. Non innovation-biased atypicality, in turn, is defined as the difference between the meanings expressed in a call and those expressed in peer calls, minus the similarity between the meanings expressed in a call and those expressed by companies that are perceived as innovative. This is the portion of atypicality not invoking innovation. Non-innovation biased atypicality increases as a call becomes different from its peers' and from innovative companies.

Formally, innovation-biased atypicality is defined as:

$$IBA_{f,q} = \cos(V_{f,q}, I_q) - \cos(V_{f,q}, PV_{f,q}) \tag{C6}$$

where, as previously, $V_{f,q}$ represents a firm's f embedding centroid at time q and $PV_{f,q}$ represents the firm's peer embedding centroid at time q (eq. 2). I_q represents the embedding centroid for firms considered innovative during time q. The higher $IBA_{f,q}$, the more a firm behaves like an innovative company while being different from its peers.

Non innovation-biased atypicality, in turn, is defined as the difference between the meanings expressed in a call and those expressed in peer calls, minus the similarity between the meanings expressed in a call and those expressed by innovative companies. Formally, non-innovation biased atypicality is defined as:

$$NBA_{f,q} = (1 - \cos(V_{f,q}, PV_{f,q})) - \cos(V_{f,q}, I_q) \tag{C7}$$

We employ two approaches for identifying innovative firms and creating the embedding centroid for firms considered innovative in a particular quarter. The first approach assumes that high-technology firms are generally considered innovative. To determine which firms are high-technology firms, we follow Kile and Phillips' (2009) classification approach (using SIC codes). Roughly 29% of firm-quarter observations are classified as high-technology firms in this way.

Second, we draw on the Fast Company Most Innovative Companies list. Fast Company, a leading American business magazine, identifies the 50 global enterprises "at the forefront of their industries" each year. Though many firms on the list are high-technology firms, the award's philosophy is explicitly oriented toward innovation in business practices broadly defined, not exclusively technological innovation. Indeed, many of these honorees are not technology firms by any definition, ranging from retail (e.g., WalMart) to apparel (e.g., Nike) firms. Overall, we identify 71 publicly-traded firms (included in our dataset) that appear on any of the annual lists published between 2008-2016 (see Figure C4). Many firms (such as Apple or GE) appear multiple times. In constructing the centroid for innovative firms, we weigh each firm by its frequency of appearance. We do not restrict the inclusion of a firm to its year of nomination, as this may result in too few firms comprising the innovation centroid that year.

Figure C4 lists all firms included in our sample and named by Fast Company among the 50 most innovative companies in the world, and the years in which they were named.

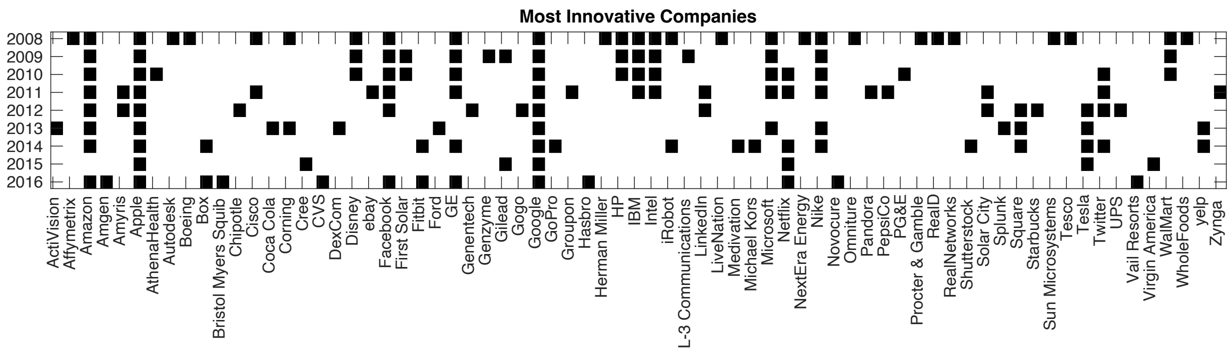

Figure C4: Most Innovative Companies by Fast Company

Appendix D Are Results Driven by Product Features?

A possible concern about our measurement strategy is that atypical earnings calls predominantly focus on firms' product and service offerings. If that were the case, then our operationalization of performative atypicality would be picking up on features related to a firm's categorical rather than performative identity. To explore this possibility, we investigate the extent to which our results are driven by five meaning dimensions: technology, product, customer, market and culture. We construct measures of dimensional atypicality (eq. C3) corresponding to each of these five words, where the dimensional centroid of reference for each dimension is its corresponding word's embedding position. For example, the dimensional atypicality for "product" is the cosine similarity between a call's atypicality centroid and the embedding position of the word "product." The greater the dimensional atypicality, the more the earning call is atypically focused on meanings semantically similar to that word. We investigate the extent to which earnings surprises are driven by these five different forms of dimensional atypicality.

Tables D1 and D2 report, respectively, the results of between- and within-firm models of earnings surprise as a function of dimensional atypicality (with specifications similar to models 3 and 6 of Table 3). Two results are worth noting. First, the dimensions "product" and "market" are not significant in any specification. Moreover, "customer" is weakly significant in the between-firm model, and is insignificant in the within-firm one. We interpret this as reassuring evidence that our results are not driven by quarterly earnings calls that are discussing atypical products, customers or markets.

Second, dimensional atypicality for "technology" significantly predicts a negative earnings surprise between-firm but not within-firm, while dimensional atypicality for "culture" is significant in both specifications. Consistent with our argument, analysts appear to be responding to semantic dimensions related to how firms pursue their strategies. This is especially true for culture, an aspect of firm capabilities that is independent of its product or service offerings. Indeed, culture is often perceived as an especially sticky source of sustaining competitive advantage. Overall, we interpret these findings as supporting of our contention that performative atypicality influences analysts' evaluation primarily through shaping their perceptions of "how", as opposed to "what," firms do.

Table D1: Earnings Surprise (lagged) by Dimensional Atypicality, Between-Firm

	(1) Technology	(2) Product	(3) Customer	(4) Market	(5) Culture
Dimensional Atypicality	-0.127***	-0.074	-0.097*	-0.002	-0.180***
	(-4.07)	(-1.89)	(-2.16)	(-0.04)	(-4.26)
Analyst Disagreement†	-1.340***	-1.347***	-1.348***	-1.342***	-1.342***
	(-5.40)	(-5.42)	(-5.42)	(-5.40)	(-5.40)
Categorical Atypicality	-0.010	-0.011	-0.011	-0.011	-0.010
	(-0.88)	(-0.94)	(-0.91)	(-0.96)	(-0.90)
Constant	-0.179	-0.188	-0.206	-0.167	-0.203
	(-1.08)	(-1.13)	(-1.23)	(-1.00)	(-1.22)
Firm Controls	Yes	Yes	Yes	Yes	Yes
Call Controls	Yes	Yes	Yes	Yes	Yes
Analyst Controls	Yes	Yes	Yes	Yes	Yes
Industry/Year/Market FE	Yes	Yes	Yes	Yes	Yes
Observations	60654	60654	60654	60654	60654
R^2	0.117	0.117	0.117	0.117	0.117

t statistics in parentheses

Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D2: Earnings Surprise (lagged) by Dimensional Atypicality, Within-Firm

	(1) Technology	(2) Product	(3) Customer	(4) Market	(5) Culture
Dimensional Atypicality	-0.064	0.000	-0.040	-0.062	-0.092*
	(-1.91)	(0.00)	(-0.91)	(-1.69)	(-2.02)
Analyst Disagreement†	-1.934***	-1.935***	-1.935***	-1.939***	-1.933***
	(-6.53)	(-6.53)	(-6.54)	(-6.54)	(-6.53)
Categorical Atypicality	0.014	0.014	0.014	0.014	0.014
	(0.63)	(0.63)	(0.64)	(0.64)	(0.63)
Constant	0.535	0.552	0.533	0.529	0.527
	(1.89)	(1.94)	(1.87)	(1.87)	(1.86)
Firm Controls	Yes	Yes	Yes	Yes	Yes
Call Controls	Yes	Yes	Yes	Yes	Yes
Analyst Controls	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Observations	61354	61354	61354	61354	61354
R^2	0.214	0.214	0.214	0.214	0.214

t statistics in parentheses

Standard errors clustered by firm

†Lagged variables, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$