

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

August 9, 2021

Abstract

Research on the role of categories in markets encapsulates a puzzle: prior work has theorized and generally found that organizations suffer negative evaluations when they deviate from categorical expectations; however, organizations are, in many cases, rewarded for being atypical. We propose that this puzzle's roots lie in the fact that prior research has conceptualized atypicality solely through the lens of categorical membership. We instead propose that atypicality can also arise performatively—based on organizational members' interactions with external audiences. Integrating the performative approach to identity with the logic of categorization in markets, we introduce the construct of *performative atypicality*—the extent to which organizational members' identity performances diverge from audience expectations. Drawing on the two-stage valuation model of the candidate-audience interface, we theorize that, while categorical atypicality leads audiences to discount an organization, performative atypicality is instead rewarded. Applying a deep-learning method to conversational text in over 90,000 earnings calls that firms hold with financial analysts, we develop a novel, language-based measure of performative atypicality. We find that performative atypicality leads to: (a) greater variance in analyst forecasts; and (b) more inflated forecasts. Paradoxically, this premium of performative atypicality results in the adverse outcome of a negative earnings surprise.

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

Research on the role of categories in markets presents a seeming paradox. On the one hand, it consistently demonstrates that organizations that defy categorical expectations suffer negative consequences because audiences pay less attention to them, struggle to make sense of them, and, ultimately, find them inherently less appealing (Hannan, Mens, Hsu, Kovács, Negro, Pólos, Pontikes, and Sharkey, 2019; Zuckerman, 1999; Leung and Sharkey, 2013; Goldberg, Hannan, and Kovács, 2016; Pontikes, 2012; Smith, 2011). At the same time, however, organizations that explore novel opportunities outside established categorical boundaries are often rewarded with outsized success (Deephouse, 1999; Durand and Calori, 2006; Zuckerman, 2017). Why do organizations sometimes suffer penalties for subverting the categorical order, while in other cases they reap benefits from doing so?

In answering this question, existing literature has tended to conceptualize atypicality through a narrow and singular lens of category membership. This literature generally sees organizations as fixed entities that are evaluated against shared classification criteria and infers categorical atypicality from primary attributes such as an organization's product features. Firms with attributes that deviate from these criteria exhibit what we refer to as *categorical atypicality*. But organizational identity is not static. Drawing on constructivist theories of social identity (Berger and Luckmann, 1967; West and Zimmerman, 1987), we argue that organizational identity is not something that an organization essentially *has* but rather a routine accomplishment that it needs to continuously *do*. Firms enact their identities through their constant and ongoing interactions with external audiences. Organizations that, in their interactions with outside stakeholders, diverge from the meanings commonly expressed in such interactions by their peers exhibit what we term *performative atypicality*.

Building on the two-stage model of valuation (Zuckerman, 2016), we argue that categorical and performative atypicality relate to different stages of the valuation process. Whereas categorical atypicality is especially salient in the first stage, when audiences associate an organization with a recognized category, performative atypicality is salient in the second stage, when audiences assess the extent to which the organization differs from its categorical peers. Consequently, while audiences find it difficult to infer the identity of categorically atypical organizations, leading them

to discount such organizations, performatively atypical organizations enjoy the benefits of differentiation. We refer to this advantage as the *performative atypicality premium*.

Using word embedding models, we develop a novel measure of performative atypicality from the spoken texts of firm leaders during quarterly earnings calls. Consistent with our theory, we demonstrate that performative atypicality varies within-firm and that stock analysts respond to this variation in formulating their predictions about firm performance. Moreover, we show that performative atypicality is associated with ambiguity, and that, all things equal, performatively atypical firms enjoy a premium, leading analysts to systematically over-estimate these firms' future earnings. Thus, the premium of performative atypicality paradoxically leads to the adverse outcome of a negative earnings surprise.

THEORY

Categorical and Performative Atypicality

What does it mean to be an atypical organization? We draw on two different and mostly tangential approaches to the study of atypicality and its consequences. The first, which has been influential in research on organizations, conceptualizes organizational identity through a *categorical* lens (Zuckerman, 1999; Hannan et al., 2019).¹ This approach understands sense-making as a classificatory process, wherein external observers (often referred to abstractly as the “audience”), drawing on a shared set of 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.² 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. Shifts in classification are therefore relatively rare. When dramatic changes do occur, such as IBM's transition from computer manufacturing to business consulting (Harreld, O'Reilly, and Tushman, 2007), they normally unfold over long periods of time. 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—e.g., 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—e.g., a woman exhibiting stereotypically masculine behaviors or an economist behaving like an historian—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—predominantly its leaders—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 interaction-

ism (Mead, 1934; Goffman, 1959), that individual identity is constructed through interpersonal interaction (e.g., Ibarra and Barbulescu, 2010). Extending this view to organizational identity formation, work taking a dynamic view has predominantly focused on the interactional processes by which organizational members develop their own perceptions of their organization's identity (Corley and Gioia, 2004; Gioia, Patvardhan, Hamilton, and Corley, 2013; Schultz and Hernes, 2012; Hatch and Schultz, 2002). These studies often emphasize the deliberative ways by which organizational leaders respond to outside influences in fashioning their firm's image (e.g., Gioia and Thomas, 1996).

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 the business press, or participating in a quarterly earnings call with financial analysts, business leaders are engaged in a meaning-laden social performance with external audiences.

Such performances communicate both denotative and connotative information. *Denotative* information relates to the literal meanings being enunciated, namely, functional 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 categorical and performative atypicality are related, they are distinct forms of atypicality. Each, we contend, corresponds 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 more 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'. Until the early 2000s, for example, Apple Computer, as it was still known at the time, had a fairly conventional product portfolio primarily focused on personal computer hardware and software. From a categorical perspective it was not especially difficult to classify. Nevertheless, its interactions with outside audiences were quite atypical. This atypicality was personified in the public performances of its founder and CEO, Steve Jobs, whose iconic jeans and black turtleneck stood in stark contrast to the formality of companies like Microsoft and IBM, connoting Apple's nimbleness and innovation relative to its competitors' staid bureaucracy.

We recognize that the distinction between the categorical and performative elements of or-

ganizational identity can be blurry, and that these two dimensions of atypicality are potentially intertwined. The National Rifle Association, for example, a non-profit gun rights advocacy group, was mostly focused on marksmanship and hunting until the late 1970s, when it shifted focus to partisan political influence. This shift was both categorical, relating to the organization's core mission and activities, and performative, as reflected in its leaders' rhetoric. Nevertheless, as we show below, the two forms of atypicality are largely uncorrelated, at least in the for-profit data we examine. In the majority of cases, shifts in performative identity do not necessarily relate to shifts in categorical identity.

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; in fact, as we show below, a large proportion of the variance in performative atypicality is attributable to stable differences between organizations. Nevertheless, this dynamism suggests that an organization's performative atypicality can change significantly, and dramatically, over time. To the same extent that organizational members' perceptions of their own organization's identity can evolve over time, particularly through the interactions they have with external stakeholders (Gioia, Schultz, and Corley, 2000; Gioia et al., 2013), we expect this dynamism to apply also to how these stakeholders make sense of the organization.

The Performative Atypicality Premium

A common assumption in the literature on categorical membership and its consequences is that outside observers draw on a two-stage process when evaluating an organization (Zuckerman, 2016). In the first stage, they assess what kind of organization it is by associating it with a category. This association determines which criteria will be used to evaluate the organization and, importantly, what reference group it will be compared to. In the second stage, once an organization's categorical identity has been established, 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 despite the fact that differentiation is a fundamental source of competitive advantage and that markets reward novelty (Barney, 1991; Deephouse, 1999). 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).

With little exception, this work perceives atypicality to be a function of an organization's categorical membership. Because categorical identity is very slow to change, this implies that an organization's atypicality is inferred on the basis of the same features during both stages of the valuation process. An organization, therefore, cannot simultaneously enjoy the benefits of intelligibility that come with being perceived as typical in the first stage and the benefits of differentiation that come with being perceived as atypical in the second. Organizations consequently need to make a choice where to locate on the categorical atypicality continuum.

Yet, despite the strong incentive for categorical conformity, atypicality is not only prevalent, it is also often rewarded with success. Apple Inc., for example, has increasingly diversified its product portfolio since the early 2000s, launching convention-defying products such as the iPad. Contra baseline expectations of category theory, this atypical strategy was met with persistent success, making Apple the world's largest company in terms of market capitalization. Existing research normally explains this seeming empirical abnormality as the result of relaxed classification standards during the first valuation stage. This easing can stem from the fact that certain audience members are more tolerant of atypicality than others (Pontikes, 2012; Bowers, 2014; Goldberg et al., 2016), that audiences are less stringent about classification in certain domains (Chatterji, Luo, and Seamans, 2021; Carnabuci, Operti, and Kovács, 2015; Haans, 2019), or that certain organizations—such as Apple—enjoy greater latitude to defy categorical conventions due to their history or status (Rao, Monin, and Durand, 2005; Smith, 2011; Sgourev and Althuizen,

2014; Durand and Kremp, 2015).

We propose, instead, that while categorical atypicality influences audience impressions mostly during the first stage of the valuation process, performative atypicality is especially salient in the second stage. Organizations can therefore enjoy the benefits of differentiation stemming from performative atypicality without necessarily paying the costs of being difficult to classify. This does not mean that performative atypicality is necessarily inconsequential during the first stage of valuation. We assume, however, that when outside observers have access to concrete information about the organization's products or services—as is the case with investors evaluating for-profit firms—they primarily draw on this information in making inferences about the organization's type.⁵ Its performative atypicality can then facilitate its perceived distinctiveness.

Figure 1 provides an illustration of the two-stage valuation model and its relationship with the two forms of atypicality. It depicts a hypothetical audience member—for example, a securities analyst—evaluating three firms, labeled *A*, *B* and *C*. In the first stage, the analyst assesses each firm's categorical atypicality by determining its identity on the basis of its products and services. Whereas firms *A* and *B* are easily classifiable, firm *C* exhibits attributes that make it categorically ambiguous. Consistent with existing theory, we expect firm *C* to be devalued relative to the other two. In the second stage, each firm's performative atypicality is assessed relative to the typical performances of its peers. Because firms *A* and *B* are perceived as near categorically identical, they are compared to the same set of peers. But while firm *A* stands out as unique, firm *B*'s performances are similar to its peers. Consequently, we contend, firm *A* will be seen more favorably.

That performative atypicality can be interpreted as a positive signal may seem counterintuitive at first. After all, individual atypical performances, such as gender noncompliant behaviors, are normally strongly frowned upon. This is the case because the typical behaviors associated with gender, or other social identities, are central to the maintenance of social order. Norm defying behaviors are seen as threats to this perceived natural order and are therefore opposed, especially by those who benefit most from maintaining the status quo (Meadow, 2018).

In most market contexts, however, audiences are concerned with maximizing value rather

than maintaining normative order.⁶ They commonly perceive uniqueness and nonconformity as indications of such value (Durand and Calori, 2006; Haans, 2019). This is particularly the case for investors in the setting we empirically investigate below. As strategy scholars have persistently demonstrated, being different is a source of advantage in markets because it makes an organization distinct in the eyes of audiences and provides it with resources that are difficult to imitate (Barney, 1991; Deephouse, 1999). While categorical atypicality is, ultimately, a liability leading to an “illegitimacy discount” (Zuckerman, 1999), performative atypicality, we contend, is interpreted by audiences as a predominantly positive indicator about strategic positioning and future performance. We refer to this advantage as the *performative atypicality premium*.

The spectacular rise and fall of WeWork, the shared work space management company, provides an instructive example of the performative atypicality premium. Founded in 2010, WeWork was by no means a categorically unusual organization. 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 its founder’s, Adam Neumann, eccentric style—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.⁷ Leading and experienced investors, most prominently SoftBank, 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.”⁸ 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%.

Performative Atypicality and Analyst Predictions

We contend that the performative atypicality premium is prevalent in many market contexts but focus our attention on securities analysts. As Zuckerman (1999, 2000) and a variety of studies since (Bowers, 2014; Smith, 2011) demonstrate, investors and analysts strongly rely on categorical distinctions when evaluating firms. They are therefore highly sensitized to instances of atypicality. At the same time, for reasons we discuss below, analysts often reward atypicality. We therefore expect the performative atypicality premium to be pronounced in their valuations.

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 propose two hypotheses. First, we argue that, like categorical atypicality, performative atypicality produces ambiguity for audiences. 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. We therefore anticipate that performative atypicality will lead to greater disagreement in analysts' earnings forecasts. Thus, we expect:

Hypothesis 1: *Performative atypicality will be associated with greater variability in analysts' earnings forecasts in the following period.*

Second, as we argued above, we propose that performatively atypical firms will enjoy a valuation premium. 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 negative earnings surprise occurs when analysts, on average, overestimate a firm's future performance. We consider such a surprise as indicative of a premium in response to meanings communicated during a quarterly earnings call.

We expect the performative atypicality premium to be prevalent in analysts' predictions for two main reasons. First, analysts occupy a cross-pressured position in financial markets. They 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, we expect, will be especially responsive to performative atypicality as it signals novelty without necessarily challenging existing categorical distinctions.

Additionally, securities analysts have good reasons to interpret performative atypicality as a positive signal of future firm performance. As research by strategy scholars shows, top executives often use language strategically to hinder their rivals' actions. For instance, incumbents' use of vague language reduces threat of entry as interested entrants have more difficulty figuring

out incumbents' strategy (Guo, Sengul, and Yu, 2019). Similarly, firms framing their strategic moves using vague or distant timelines successfully delay the response of their rivals (Nadkarni, Pan, and Chen, 2018). To the extent that performative atypicality begets ambiguity, analysts may interpret it as an indication of strategic obfuscation, formulating higher expectations regarding future performance as a result. Thus, analysts have incentives to promote atypical firms and to be more optimistic about them. Overall, we expect that:

***Hypothesis 2:** Performative atypicality will be associated with negative earnings surprises in the following period.*

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 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

test our hypotheses, 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. This results in a total of 68,178 firm-quarter observations. To ensure that our estimates are not driven by outliers or especially small firms and consistent with standard practice, we winsorize the dependent variables at the 99% level on both ends and remove observations for firms' whose stock price is less than \$1 or whose book value is less than \$5M.

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, Sutskever, Chen, Corrado, and Dean, 2013). Each word is then projected to a location in a shared vector space with several hundred dimensions. Although these dimensions are often uninterruptible 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. In a powerful demonstration of this connotative ability, Garg, Schiebinger, Jurafsky, and Zou (2018) show that implicit gendered associations in the meanings of various occupations track with these occupations' historical gender compositions. Kozlowski, Taddy, and Evans (2019) similarly illustrate

how different lifestyle activities invoke class, race, and gender identities. 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 “space” 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 B ought to be considered closer to Firm A than to Firm C given the semantic dissimilarity between “personality” and “office” or “space” relative to the latter two’s similarity. Like the WeWork example we highlighted above, 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 pre-processing, 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). The word “onboarding,” for example, originally used to denote employee socialization, has begun shifting in meaning over the last decade to describe the process by which users are educated and integrated into using a digital platform (Vicinanza, Goldberg, and Srivastava, 2020). 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.

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:

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

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:

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

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 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.⁹ 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:

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

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.¹⁰

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.

Validating the Word Embedding Model

We first sought to validate our word embedding models. There are two main techniques for doing so: 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 closest 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 1. Overall, these analyses indicated that our embedding models captured semantically meaningful relationships between words used in quarterly earnings calls.

Categorical Atypicality

We do not directly hypothesize about categorical atypicality given that its relationship with analyst valuations has been extensively demonstrated in prior work. Nevertheless, we include it as an independent variable in all our models for two reasons. First, we aim to demonstrate that, consistent with our theorizing, categorical and performative atypicality exhibit different patterns and relate differently to analyst valuations. In particular, while we expect performative atypicality to be correlated with negative earning surprises, we do not expect that to be the case for categorical atypicality. Second, we seek to demonstrate that categorical and performative atypicality are related to valuations independently 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 reflected in the extent to which it 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:

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

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 test Hypothesis 1, 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 and bottom 1 percent.

Earnings Surprise. To test Hypotheses 2, 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 et al., 2019; Westphal, Park, McDonald, and Hayward, 2012; Barron, Byard, and Yu, 2008; Livnat and Mendenhall, 2006). We then multiply it by 100 so that a 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 $100 \times (1 - 0.99) / 1 = 1$ 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).¹¹ 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, Haleblan, 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.¹²

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 in 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

Performative Atypicality’s Properties

Before directly testing our hypotheses, we begin with exploring the distributional properties of performative atypicality. Figure 2 plots the kernel density for performative atypicality. As the figure demonstrates, performative atypicality roughly follows a normal distribution.

Figure 3 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). We highlight various firms for illustrative purposes.

The patterns in Figure 3 support our baseline assumptions about performative atypicality. First, consistent with intuitive expectations, innovative 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.

Second, 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, their mean performative atypicality is low relative to software companies. This underscores the need to account for mean differences between industries when estimating between-firm effects, as we do below.

Third, it is evident that 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 ($\rho = 0.035, p < 0.001$). 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 prod-

uct portfolios comprise quite unusual combinations, their performances in quarterly earning calls are fairly standard.

Finally, 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 3 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 4 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. Consistent with our theorizing, categorical atypicality is a significantly more static firm attribute than performative atypicality. In other words, what a firm constitutes changes less frequently than what it does in its public performances.

This is also evident in Panel C of Figure 4, 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, consistent with our theory, that the latter is a signal of a firm's deviation from conventional practices.

Main Results

We test the two hypotheses by examining the relationship between performative atypicality and the dependent variables 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 3 and 4 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 performance controls) such that the effects of atypicality are estimated for analyst disagreement and earnings surprises in the subsequent quarter. While not fully addressing endogeneity concerns, this brings us closer to a causal estimate. For ease of interpretation, both atypicality measures are standardized.

Table 3 reports results for between-firm OLS models, where the two dependent variables—analyst disagreement and earnings surprise—are modeled as a function of performative and categorical atypicality. We include industry-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.

Consistent with Hypothesis 1, Models 1 to 3 demonstrate that performative atypicality is related to an increase in analyst disagreement. This effect is comparable in size to the effect of categorical atypicality on disagreement. As Model 3 shows, these two effects are independent. Holding industry-time effects constant, ambiguity grows with both forms of atypicality. Models 4-6 follow the same specifications in predicting earnings surprises. Here we see that only performative atypicality is significantly (and strongly) related to the outcome. Consistent with Hypothesis 2, analysts tend to overestimate the future performance of firms that exhibit high performative atypicality. We contend that this is because they associate this form of atypicality with innovation and creativity. The marginal effects of performative atypicality on analyst disagreement (Model 3) and earnings surprise (Model 6) are plotted in Figure 5.

Table 4 reports analogous models with within-firm specifications. These models include firm fixed effects, as well as separate industry and year fixed effects.¹³ They also control for the level of analyst disagreement to ensure that the effects we model are attributable to analysts' tendencies to overestimate future performance above and beyond their level of disagreement. Results for performative atypicality mirror those produced by between-firm specifications: in support of Hypothesis 1, Models 1 to 3 demonstrate that analyst disagreement increases with performative atypicality, and, in support of Hypothesis 2, Models 4 to 6 demonstrate that performative atypicality is associated with a negative earnings surprise. The marginal effects on analyst disagreement (Model 3) and earnings surprise (Model 6) are plotted in Figure 6.

Effects for categorical atypicality, however, are no longer significant in all models. This is because, as we argued above, categorical atypicality is a more stable attribute than performative atypicality. Not only is there far less within-firm variation in categorical atypicality than there is in performative atypicality (Figure 4, 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.

Overall, the results reported in Tables 3 and 4 provide support for our two hypotheses. Whether focused on time- and industry-invariant differences between firms or on changes within a firm throughout its lifespan, they demonstrate that performative atypicality is associated with greater ambiguity and with a valuation premium that results in the adverse outcome of a negative earnings surprise.

Extension: Moderation Analysis

To understand potential sources of heterogeneity in the performative atypicality premium, we conducted additional analyses that focused on firms' R&D expenditures. Previous research indicates that, like performative atypicality, R&D spending is associated with greater perceived uncertainty about a firm's future performance and more volatile earnings. Moreover, as is the case with performative atypicality, investors appear to overestimate the extent to which R&D investments signal a firm's growth potential, as reflected in the excess value of R&D intense firms

(Chan, Lakonishok, and Sougiannis, 2001). Firms that invest in research and development are perceived as growth-focused.

If R&D intensity and performative atypicality have similar effects on investors' interpretation, we expect them to offset one another. More specifically, we expect the performative atypicality premium to be attenuated for firms that are R&D active, as these firms already enjoy a favorable outlook in the eyes of investors.

There are a variety of ways to estimate R&D expenditures (e.g., as a proportion of sales, book value, or profits), and each yields very different estimates. We opted to construct a simple binary variable, which is set to 1 if a firm reports any R&D expenditures and to 0 otherwise. This variable represents whether a firm engages or does not engage in research and development.¹⁴

Table 5 reports the results of between- (Model 1) and within-firm (Model 2) models of earnings surprise as a function of an interaction between R&D activity and performative atypicality. These models use the same specifications reported in the main results. In both models, performative atypicality is associated with a negative earnings surprise, but this effect is attenuated for R&D active firms. The marginal effects from both models are plotted in Figure 7. As these plots show, performative atypicality significantly predicts analysts' optimism only for firms that are not R&D active. Although the interaction between R&D activity and performative atypicality is significant for both models, the differences between R&D active and inactive firms are less pronounced in within-firm models, as Figure 7 shows. Our results are robust to different operationalizations of R&D intensity as reported in the Appendix.

DISCUSSION

Organizational research has overwhelmingly conceptualized atypicality as a unidimensional construct: an organization can only occupy a single location on the typicality-atypicality continuum. We argued, instead, that atypicality arises in two distinct and potentially orthogonal ways. Audiences make inferences about an organization's categorical atypicality by examining the products and services it provides and about its performative atypicality by continuously observing

how its members' interact with external stakeholders. While extant literature has consistently demonstrated the negative consequences of categorical atypicality, we showed that performative atypicality is instead associated with a valuation premium. Thus, counter to the perspective that prevails in the literature, atypicality can sometimes be an asset rather than a liability. Yet, in the context of financial analysts' making earnings forecasts, this premium can paradoxically result in the adverse outcome of a negative earnings surprise.

We speculate that the performative atypicality premium operates in other contexts in which it does not necessarily result in a disadvantageous outcome. From venture capitalists' decisions about investing in early stage companies to consumers' product choices, we expect that performatively atypical organizations enjoy an advantage relative to their typical competitors. More broadly, our theoretical developments and findings make several contributions to the literatures on categories in markets and on organizational identity.

Contributions

Although existing work has exclusively theorized atypicality through a categorical lens, we conjecture that, in practice, it 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., 2016; 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.” The mere insinuation of customer choice, for example, connotes the restaurant's lack of culinary sophistication (Jurafsky, 2015). Seen in this light, our findings make several contributions to the literature on categories and markets.

First, our results shed new light on the dual pressures of conformity and differentiation that organizations face. As others have argued (Schneiberg and Berk, 2010), organizational literature

has mostly paid attention to the disciplinary role that categories play in markets. To explain why it is the case that organizations are nevertheless sometimes rewarded for atypicality, this literature has either argued that some organizations have more leeway than others to defy categorical expectations or that some audiences and contexts enforce categorical codes less stringently than others. In contrast, our findings suggest that the same organization can enjoy the benefits of atypicality without necessarily suffering the penalties of categorical ambiguity, even in the eyes of the same audience. It can do so by being performatively atypical but categorically compliant. This suggests that organizations have more agency than conventionally assumed in resolving the tension between conformity and differentiation.

Second, performative atypicality points attention to the dynamic nature of organizational identity construction. Whereas prior work has conceptualized categorical atypicality as a relatively stable organizational attribute, we demonstrate that performative atypicality is fluid (Figure 4). Indeed, our results show that analysts respond to within-firm changes in performative atypicality but are insensitive to similar fluctuations in categorical atypicality (Tables 3 and 4).

The implications of this dynamism go beyond the temporal shifts that organizations exhibit in their performative atypicality. The categorical and performative aspects of organizational identity relate to different types of identity formation processes. While categorical identity is an attribute that an organization has by virtue of the kinds of markets it competes in, performative identity is something that an organization continuously does. Doing organizational identity does not necessarily imply temporal changes in identity; organizational performances can be repeatedly consistent. Rather, unlike the unilateral nature of categorical decisions made by organizations, performative dynamism emerges through the multilateral interactions between an organization and its audiences (Berger and Luckmann, 1967).

The process by which organizational identity is constituted performatively likely has implications for organizational success. For example, are stable performative identities, irrespective of their atypicality or perhaps contingent on it, more conducive to success than fluid ones? To what extent is the process by which performative change unfolds consequential for firm performance? For example, Leung (2014) finds that individuals that exhibit erratic career trajectories

suffer in the labor market relative to those who experience the same scope of change but in a more gradual manner. A similar dynamic might be applicable to an organization's performative atypicality. Addressing such questions will require extending our new measure to estimate changes in an organization's performative identity relative to itself.

While we leave these efforts for future work, we note that our findings and methodological innovations echo recent developments in the work on organizational identity, as perceived from the point of view of organizational members. Whereas early work in this vein emphasized the enduring nature of identity (Albert and Whetten, 1985), a more recent literature has highlighted that organizational identity is malleable and can evolve over time. Some of these accounts emphasize the actions of organizational leaders. Schultz and Hernes (2012), for example, document the process by which the LEGO Group reconstructed its identity between 2000 and 2008 by invoking memories of the past and transforming them into claims about the future. Studies such as these have generally subscribed to a teleological view of organizational change, which assumes that change is purposefully adaptive and slow. Other work has theorized how organizational identity emerges and adapts through members' interactions with outsiders (e.g., Gioia et al., 2013).

Consistent with the latter, we bring the performative aspect of identity, which has been studied extensively at the individual level of analysis, to the organizational level. In our conceptualization, organizational members adapt their collective identity through everyday interactions with outside stakeholders. In some cases, these identity performances can be purposeful—for example, when members seek to proactively shape the organization's image in the minds of external audiences. Yet, in many cases, including the context of highly scripted quarterly earnings calls, identity performances can be enacted without conscious awareness or intent—for example, when members inadvertently “leak” meanings or, more importantly, when they adjust their own interpretations in reaction to these interactive experiences. This constitutive aspect has often been emphasized in research on identity “performativity” (e.g., Butler, 2006). Like the work on categorical atypicality, the work on identity performativity has tended to overemphasize the disciplining aspects of social performances, for example, how girls internalize societal expectations through their interactive performances of gender identity. Market, audiences, however, are less motivated by

normative incentives. This suggests that performative channels can be an avenue through which new identities are negotiated and solidified, rather than merely old ones being reinforced.

Finally, we make a methodological contribution to organizational identity research. Prior work has employed qualitative research methods (e.g., Glynn, 2000) or surveys (e.g., Brickson, 2005) to measure different facets of organizational identity. In similar fashion to recent methodological advances in the study of organizational culture (Srivastava, Goldberg, Manian, and Potts, 2018; Corritore, Goldberg, and Srivastava, 2020; Li, Mai, Shen, and Yan, 2020), we harness the tools of computational linguistics and deep learning to develop fine-grained and time-varying measures of organizational identity. We believe that this approach can readily be extended to other domains in which organizational members enact collective identity—for example, when organizational leaders engage with clients on social media (Lee, Hosanagar, and Nair, 2018; Liu, Shin, and Burns, 2021) or respond to activist challenges from social movement organizations (King, 2008; Weber, Rao, and Thomas, 2009; McDonnell, King, and Soule, 2015).

Limitations and Future Directions

We recognize that this study has certain limitations, which point to useful avenues for future research. First, one of the downsides of using word embedding models is that the vectors representing firms' calls are uninterpretable. Thus, it is not possible to characterize the atypicality we observe in earnings calls. We conjecture, based on our interviews with professionals who advise management teams on how to prepare for these calls, that atypicality may reflect various attempts by management to frame potentially negative news in a more positive light (cf. Suslava, 2018). Future work can profitably supplement word embedding models with other natural language processing tools that can characterize the content of calls.

Like most prior work on this topic, we examined the relationship between organizational atypicality and audience behaviors but did not directly measure the cognitive mechanisms channeling these outcomes. Some of the analyses we conducted shed light on these mechanisms. For example, we find that performative atypicality is associated with greater disagreement, suggesting that it produces ambiguity. Similarly, the attenuated effects for R&D active firms (Table 5), which are

generally more confusing to investors but are also more likely to be interpreted as being innovative and creative, are congruent with our assumptions that performative atypicality is interpreted as a positive signal about a firm's potential growth. Moreover, by controlling for firm, call, and analyst attributes, we rule out certain alternative explanations (for example, that the relationship between a firm's performative atypicality and its future earnings surprise is fully attributable to its past performance). Overall, these findings are consistent with our cognitive arguments about the two-stage valuation process and its relationship with performative atypicality.

We also acknowledge that our study focused on just one domain in which performative atypicality is expressed—i.e., quarterly earnings calls with financial analysts—and its impact on one form of valuation—i.e., earnings forecasts in the subsequent quarter. Yet we recognize that organizational members engage in identity performances in a wide range of contexts—for example, when senior executives are interviewed by the business press, when new products are introduced at trade shows, and when an entrepreneurial team makes a fundraising pitch to a group of venture capitalists—and these different performances might have varying implications for a broader set of outcomes such as a firm's share price or equity valuation. We leave to future research the task of measuring performative atypicality across multiple domains of audience interaction and of understanding how it relates to different forms of valuation.

Next, despite our use of models that include firm fixed effects and thus account for unobserved, time-invariant heterogeneity between firms, as well as our use of lagged variables, we are not able to make strong causal claims with our empirical setup. For example, it is possible that an anticipated negative earnings surprise leads managers to communicate in performatively atypical ways and that this atypicality is interpreted by analysts differently than we posit. One way to pin down the causal link from performative atypicality to negative earnings surprise would be to conduct a lab study. Participants could, for example, be randomly assigned to conditions in which they are exposed to the same substantive information about a firm that is presented in ways that are more or less performatively atypical relative to other firms. Participants could then be asked to provide specific quantitative forecasts of the firm's earnings in the next quarter. Such a study would serve as a useful complement to this one—though questions about external validity would

arise unless actual financial analysts could be recruited as participants.

Another limitation of our approach is that it does not account for the process by which analysts select into covering particular firms or attending specific calls. It is possible that the composition of analysts covering a stock could shift in ways that are related to performative atypicality and earnings surprise. Our use of analyst churn as a control variable partially addresses this concern but does not rule the potentially confounding role of selection. We leave for future research the task of disentangling the effects of selection from that of performative atypicality.

Finally, our results paint previous results related to the effects of categorical atypicality in a different color. While our results are consistent with the claim that categorical atypicality leads to confusion, we do not find evidence for a categorical atypicality discount: categorical atypicality is insignificant in all of the earnings surprise models. We speculate that this relates to the fact that we explore earnings surprises, whereas previous work examined stock performance. It could be the case that we do not find an effect because categorical atypicality has offsetting effects on analysts' predictions.¹⁵ Our inability to reproduce the categorical atypicality discount could also be related to our operationalization of categorical atypicality, consistent with prior work, 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.¹⁶

CONCLUSION

Extant literature has tended to focus on the negative implications of organizational atypicality. Building on constructivist theories of identity, we propose that organizational members “do” organizational identity and introduce a novel form of atypicality—performative atypicality. Contra prevailing wisdom, we demonstrate that performative atypicality can result in a premium in the eyes of external stakeholders, wherein securities analysts overestimate a firm's future success. Organizational members' performances can be detected in the natural language they use in interactions with audiences. As we show, these subtle semantic deviations can have profound

economic consequences.

Notes

¹The term identity has been used by organizational scholars in a variety of ways, at times at odds with one another. 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 et al., 2000).

²In fact, this research is often founded on assumptions about the prototypical nature of categorical cognition (Hannan et al., 2019; Paoletta and Durand, 2015).

³A 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>

⁵Consider the following scenario as illustration. A passerby inadvertently walking into an Apple Store in London shortly after its opening in 2004 might have been confused about its purpose. With products on display as if they were art, employees dressed as docents and no checkout, it looks more like a museum or an art gallery than an electronics store. This confusion would likely cause discomfort. In contrast, a shopper knowing this is an Apple retail store would have experienced no such confusion and instead would have perceived the store's performative atypicality as yet another indication of Apple's uniqueness. The setting we explore below is closer to the latter than the former scenario.

⁶There are obvious exceptions to this rule. In market domains pervaded by authenticity, compliance with normative expectations is part of the value proposition. For example, consumers undervalue Chinese restaurants that adhere to health regulations because contemporary hygiene standards can be inconsistent with traditional Chinese food preparation practices (Lehman, Kovács, and Carroll, 2014).

⁷<https://www.nytimes.com/2019/11/02/business/adam-neumann-wework-exit-package.html>

⁸<https://www.forbes.com/sites/stevenbertoni/2017/10/02/the-way-we-work/?sh=30044b521b18>

⁹This 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.

¹⁰The 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.

¹¹Note 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.

¹²In 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.

¹³We do not include quarter fixed effects as that would absorb too much variation.

¹⁴We use this binary operationalization for two reasons. First, partly due to accounting reasons, R&D expenditure reporting tends to fluctuate significantly during the year and is often a poor representation of true R&D intensity, which tends to be relatively stable. Second, while there is obviously significant variance in R&D intensity between firms, ultimately the most pertinent distinction for our purposes is that between R&D active and inactive firms. The latter's investment in research makes their future performance more difficult to predict.

¹⁵We conducted a variety of additional between-firm analyses in which we relaxed the number of controls and the industry-quarter fixed effects. We found no statistically significant relationship between categorical atypicality and earnings surprise in any of these models.

¹⁶An alternative operationalization might focus on the actual products a firm makes. Such an approach might be better suited for a model, like ours, that includes both categorical and performative atypicality, as it obviates confounding perceptions in the construction of the measure. We conducted a set of preliminary analyses that operationalized categorical atypicality as the overlap between the competitors of a focal firm's competitors, as derived from the Text-based Network Industry Classification. The relationship between performative atypicality and negative earnings surprise is robust to this implementation. Given that this operationalization introduces a set of choices that require further investigation, we do not report these results in this paper and intend to explore them in subsequent work.

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.

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.

Brickson, S. L.

2005 “Organizational Identity Orientation: Forging a Link between Organizational Identity and Organizations’ Relations with Stakeholders.” *Administrative Science Quarterly*, 50: 576–609.

Butler, J.

2006 *Gender Trouble: Feminism and the Subversion of Identity*. New York: Routledge 1st edition edition.

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.

Chan, L. K. C., J. Lakonishok, and T. Sougiannis

2001 “The Stock Market Valuation of Research and Development Expenditures.” *The Journal of Finance*, 56: 2431–2456.

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

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

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

2004 “Identity Ambiguity and Change in the Wake of a Corporate Spin-Off.” *Administrative Science Quarterly*, 49: 173–208.

Corritore, M., A. Goldberg, and S. B. Srivastava

2020 “Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance.” *Administrative Science Quarterly*, 65: 359–394.

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.

Gioia, D. A. and J. B. Thomas

1996 “Identity, Image, and Issue Interpretation: Sensemaking During Strategic Change in Academia.” *Administrative Science Quarterly*, 41: 370–403.

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.

Glynn, M. A.

2000 “When Cymbals Become Symbols: Conflict over Organizational Identity within a Symphony Orchestra.” *Organization Science*, 11: 285–298.

Goffman, E.

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

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

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

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.

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.

Harreld, J. B., C. A. O’Reilly, and M. L. Tushman

2007 “Dynamic Capabilities at IBM: Driving Strategy into Action.” *California Management Review*, 49: 21–43.

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.

Hatch, M. J. and M. Schultz

2002 “The Dynamics of Organizational Identity.” *Human Relations*, 55: 989–1018.

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.

Hsu, G. and M. T. Hannan

2005 “Identities, Genres, and Organizational Forms.” *Organization Science*, 16: 474–490.

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.

King, B. G.

2008 “A Political Mediation Model of Corporate Response to Social Movement Activism.” *Administrative Science Quarterly*, 53: 395–421.

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, D., K. Hosanagar, and H. S. Nair

2018 “Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook.” *Management Science*, 64: 5105–5131.

Lee, J.

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

Lehman, D. W., B. Kovács, and G. R. Carroll

2014 “Conflicting Social Codes and Organizations: Hygiene and Authenticity in Consumer Evaluations of Restaurants.” *Management Science*, 60: 2602–2617.

Lenci, A.

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

Leung, M. D.

2014 “Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market.” *American Sociological Review*, 79: 136–158.

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.

Li, K., F. Mai, R. Shen, and X. Yan

2020 “Measuring Corporate Culture Using Machine Learning.” *The Review of Financial Studies*, .

Liu, X., H. Shin, and A. C. Burns

2021 “Examining the Impact of Luxury Brand’s Social Media Marketing on Customer Engagement: Using Big Data Analytics and Natural Language Processing.” *Journal of Business Research*, 125: 815–826.

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.

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.

McDonnell, M.-H., B. G. King, and S. A. Soule

2015 “A Dynamic Process Model of Private Politics: Activist Targeting and Corporate Receptivity to Social Challenges.” *American Sociological Review*, 80: 654–678.

Mead, G. H.

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

Meadow, T.

2018 *Trans Kids: Being Gendered in the Twenty-First Century*. Univ of California 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.

Nadkarni, S., L. Pan, and T. Chen

2018 “Only Timeline Will Tell: Temporal Framing of Competitive Announcements and Rivals’ Responses.” *Academy of Management Journal*, 62: 117–143.

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.

Paoella, 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.

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.

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.

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.

Schneiberg, M. and G. Berk

2010 “From Categorical Imperative to Learning by Categories: Cost Accounting and New Categorical Practices in American Manufacturing, 1900–1930.” In G. Hsu, G. Negro, and Ö. Koçak (eds.), *Categories in Markets: Origins and Evolution* volume 31 of *Research in the Sociology of Organizations* 255–292. Emerald Group Publishing Limited.

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.

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

2018 “Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations.” *Management Science*, 64: 1348–1364.

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.

Suslava, K.

2018 Three Essays on Equity Valuation and the Predictive Ability of Quantitative and Qualitative Corporate Disclosures. Ph.D. thesis Rutgers University - Graduate School - Newark.

Vicinanza, P., A. Goldberg, and S. Srivastava

2020 “Who Sees the Future? A Deep Learning Language Model Demonstrates the Vision Advantage of Being Small.”, .

Weber, K., H. Rao, and L. G. Thomas

2009 “From Streets to Suites: How the Anti- Biotech Movement Affected German Pharmaceutical Firms.” *American Sociological Review*, 74: 106–127.

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.

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

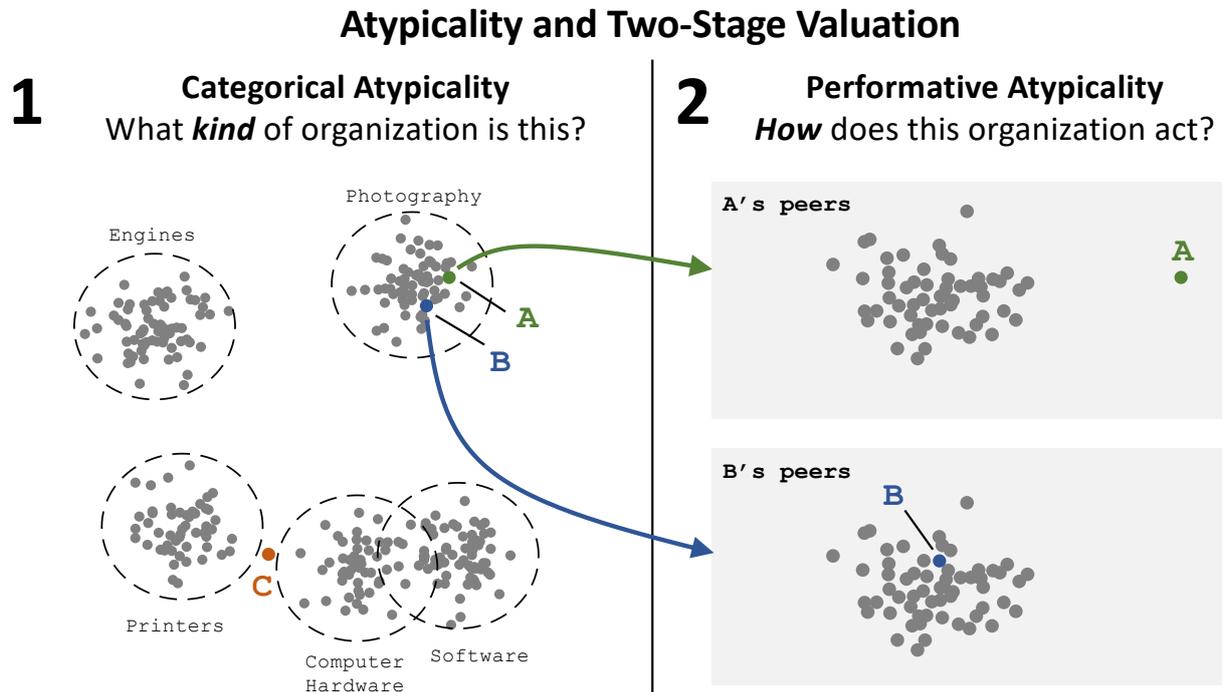


Figure 1: An illustration of our theoretical model, depicting a hypothetical audience member evaluating three firms labeled A, B and C. Grayed dots represent other firms. In the first stage each firm is located in a stylized categorical space, where labels correspond to categories and dashed lines represent perceived categorical boundaries. In the second stage each firm is located in a stylized performative space populated by its perceived group of peers.

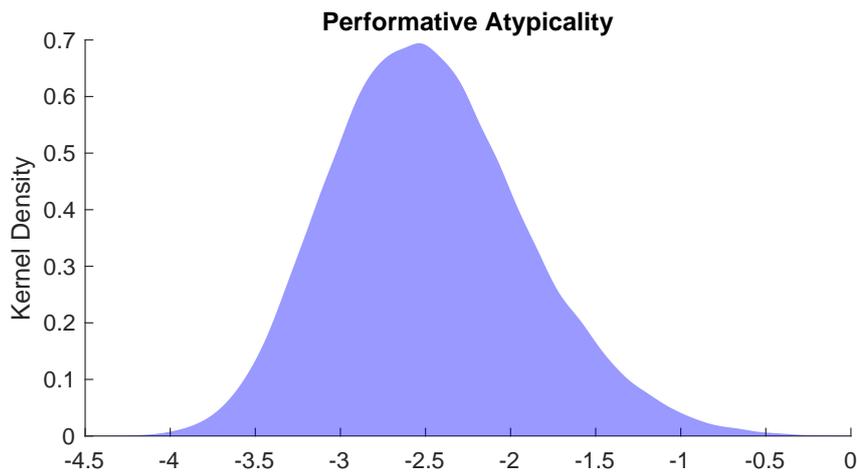


Figure 2: Kernel density for Performative Atypicality

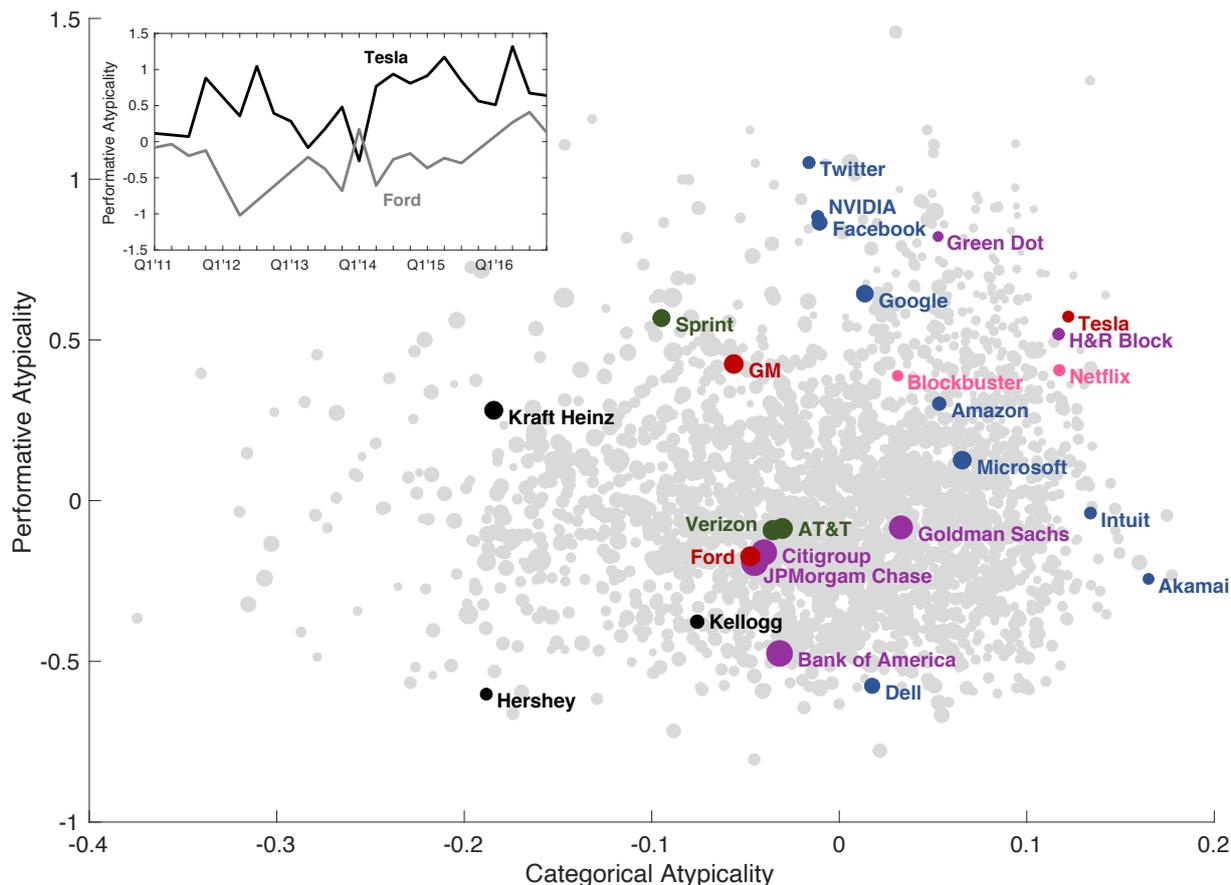


Figure 3: Atypicality by firm. Each dot represents one firm’s mean (standardized) performative and categorical atypicality (for firms with a minimum of 10 quarterly observations). Highlighted firms are color coded by FIC200 industry. The inset plots Tesla’s and Ford’s performative atypicality over time.

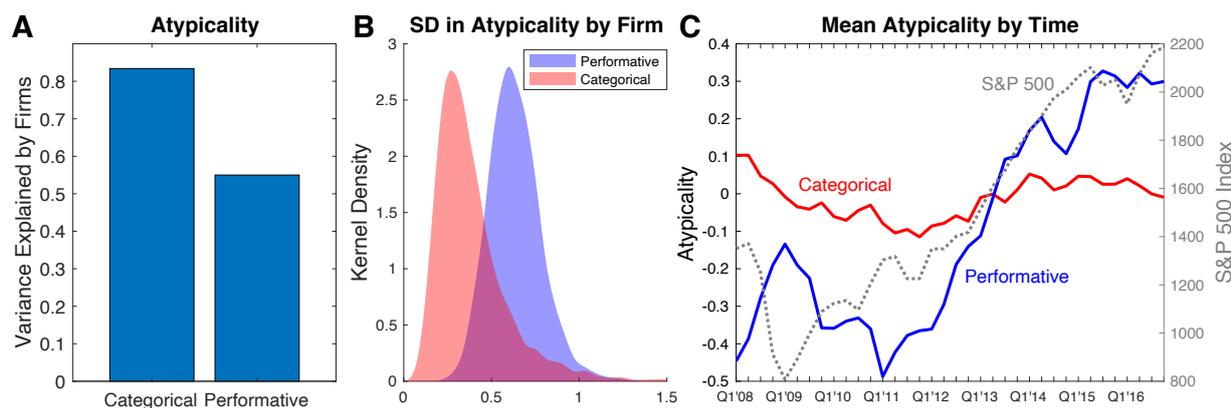


Figure 4: (A) The proportion of variance in performative and categorical atypicality explained by fixed firm differences. **(B)** Kernel densities for the standard deviation, by firm, in performative and categorical atypicality. **(C)** Mean performative and categorical atypicality by quarter. The dotted line corresponds to the S&P 500 index.

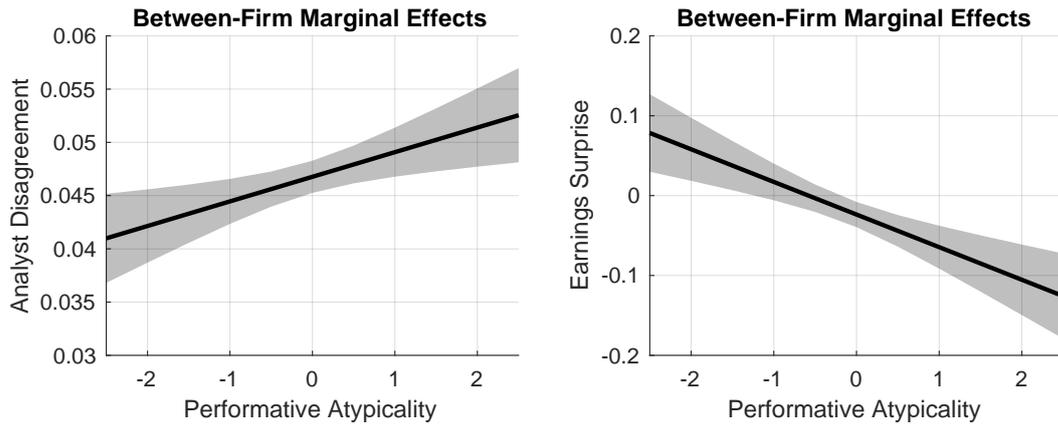


Figure 5: Marginal effects of between-firm performative atypicality on analyst disagreement (left) and earnings surprise (right) (Models 3 and 6, Table 3).

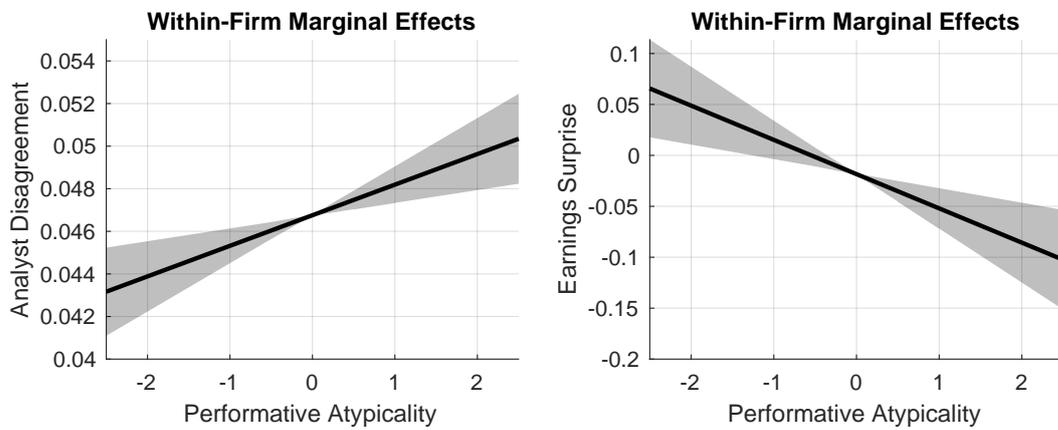


Figure 6: Marginal effects of within-firm performative atypicality on analyst disagreement (left) and earnings surprise (right) (Models 3 and 6, Table 4).

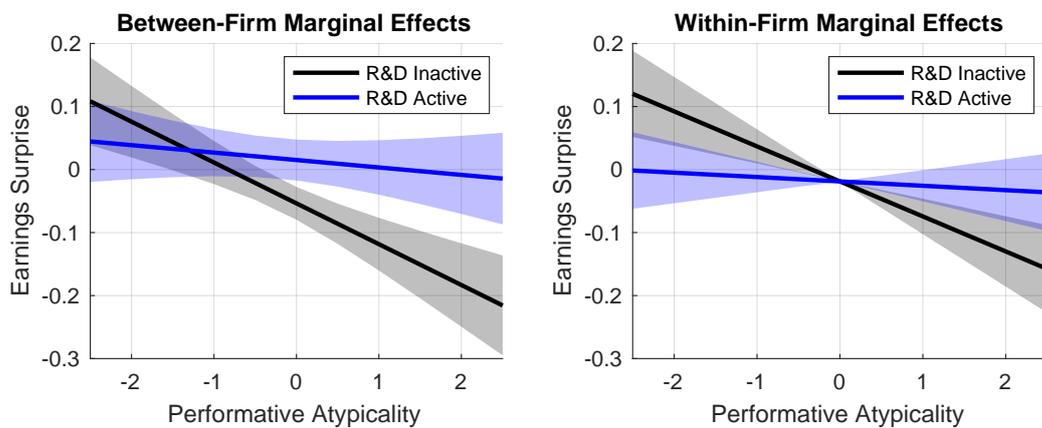


Figure 7: Marginal effects of between-firm (left) and within-firm (right) performative atypicality on earnings surprise by firm R&D activity (Table 5).

Tables

Table 1: Sample Analogy Tasks Applied to the Word Embedding Model for Q4 2016

Analogy Task	Answer
Toyota - Japan + Germany = ?	BMW
Boeing - USA + Europe = ?	Airbus
Huawei - 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

Table 2: 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 3: Between-Firm Models

	Analyst Disagreement [†]			Earnings Surprise [†]		
	(1)	(2)	(3)	(4)	(5)	(6)
Categorical Atypicality	0.002*		0.002*	-0.011		-0.010
	(2.14)		(2.13)	(-0.95)		(-0.91)
Performative Atypicality		0.002**	0.002**		-0.041***	-0.041***
		(2.84)	(2.81)		(-4.13)	(-4.13)
<i>Firm Attributes</i>						
Leverage [†]	0.009*	0.009*	0.009*	-0.362***	-0.364***	-0.366***
	(2.23)	(2.17)	(2.29)	(-6.76)	(-6.78)	(-6.82)
Log of assets [†]	0.005***	0.005***	0.005***	0.035***	0.035***	0.034***
	(6.63)	(6.65)	(6.67)	(4.47)	(4.56)	(4.35)
Preceding pos. surprise	-0.004***	-0.004***	-0.004***	0.243***	0.243***	0.242***
	(-5.00)	(-5.05)	(-4.95)	(15.00)	(15.00)	(14.96)
<i>Call Attributes</i>						
Order in quarter	-0.000	-0.000	-0.000	-0.002***	-0.002***	-0.002***
	(-0.33)	(-0.31)	(-0.42)	(-3.43)	(-3.37)	(-3.33)
Positivity	-0.025***	-0.024***	-0.025***	0.158***	0.145***	0.148***
	(-9.48)	(-9.17)	(-9.24)	(4.13)	(3.82)	(3.86)
Horizon	0.011***	0.010***	0.010***	0.033	0.045	0.045
	(4.45)	(4.17)	(4.17)	(0.90)	(1.23)	(1.23)
Litigious	0.002	0.002	0.002	-0.050**	-0.047*	-0.047*
	(1.46)	(1.36)	(1.34)	(-2.68)	(-2.54)	(-2.53)
Log of length	0.000	0.002	0.002	-0.001	-0.038	-0.039
	(0.22)	(1.53)	(1.58)	(-0.06)	(-1.62)	(-1.63)
<i>Analyst Attributes</i>						
Analysts churn	-0.017***	-0.017***	-0.017***	0.093*	0.093*	0.094*
	(-8.44)	(-8.43)	(-8.48)	(2.54)	(2.55)	(2.57)
No. of estimates [†]	0.000*	0.000*	0.000*	0.006***	0.006***	0.006***
	(2.50)	(2.17)	(2.43)	(3.50)	(3.90)	(3.67)
Disagreement [†]				-1.343***	-1.333***	-1.330***
				(-5.36)	(-5.33)	(-5.32)
Constant	0.011	-0.002	-0.006	-0.165	0.107	0.124
	(0.97)	(-0.17)	(-0.48)	(-0.99)	(0.60)	(0.69)
Industry-Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	61670	61670	61670	61670	61670	61670
R ²	0.225	0.225	0.225	0.127	0.128	0.128

t statistics in parentheses, Standard errors clustered by firm

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

Table 4: Within-Firm Models

	Analyst Disagreement [†]			Earnings Surprise [†]		
	(1)	(2)	(3)	(4)	(5)	(6)
Categorical Atypicality	-0.000 (-0.48)		-0.000 (-0.44)	0.014 (0.66)		0.013 (0.61)
Performative Atypicality		0.001*** (3.38)	0.001*** (3.37)		-0.034*** (-3.43)	-0.034*** (-3.40)
<i>Firm Attributes</i>						
Leverage [†]	0.045*** (10.13)	0.045*** (10.12)	0.045*** (10.11)	-0.534*** (-4.77)	-0.533*** (-4.77)	-0.532*** (-4.76)
Log of assets [†]	0.021*** (11.55)	0.021*** (11.56)	0.021*** (11.55)	-0.023 (-0.69)	-0.023 (-0.68)	-0.022 (-0.67)
Preceding pos. surprise	-0.003*** (-4.93)	-0.003*** (-4.90)	-0.003*** (-4.90)	0.075*** (5.58)	0.074*** (5.56)	0.075*** (5.56)
<i>Call Attributes</i>						
Order in quarter	0.000 (0.62)	0.000 (0.70)	0.000 (0.70)	-0.000 (-0.33)	-0.000 (-0.39)	-0.000 (-0.40)
Positivity	-0.015*** (-10.33)	-0.015*** (-10.20)	-0.015*** (-10.19)	0.278*** (7.16)	0.271*** (7.01)	0.271*** (6.99)
Horizon	0.004** (2.64)	0.003* (2.42)	0.003* (2.42)	-0.022 (-0.58)	-0.015 (-0.38)	-0.015 (-0.38)
Litigious	0.001 (0.96)	0.001 (0.90)	0.001 (0.90)	-0.056** (-3.15)	-0.055** (-3.10)	-0.055** (-3.10)
Log of length	-0.000 (-0.38)	0.001 (1.19)	0.001 (1.19)	-0.015 (-0.61)	-0.045 (-1.79)	-0.045 (-1.79)
<i>Analyst Attributes</i>						
Analysts churn	-0.001 (-1.12)	-0.001 (-1.12)	-0.001 (-1.11)	0.098** (3.01)	0.098** (3.01)	0.098** (3.00)
No. of estimates [†]	0.001*** (6.17)	0.001*** (6.17)	0.001*** (6.15)	0.002 (0.66)	0.002 (0.67)	0.002 (0.69)
Disagreement				-1.939*** (-6.55)	-1.932*** (-6.54)	-1.931*** (-6.53)
Constant	-0.137*** (-9.44)	-0.146*** (-9.60)	-0.146*** (-9.60)	0.548 (1.93)	0.779** (2.67)	0.775** (2.66)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	61440	61440	61440	61440	61440	61440
R ²	0.516	0.516	0.516	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 5: Earnings Surprise by R&D Activity

	(1) Between-Firm	(2) Within-Firm
Performative Atypicality	-0.065*** (-4.54)	-0.056*** (-3.94)
Categorical Atypicality	-0.013 (-1.11)	0.011 (0.54)
R&D Active	0.069** (2.74)	
R&D Active \times Performative Atypicality	0.053** (2.92)	0.049** (2.69)
Analyst Disagreement	-1.315*** (-5.27)	-1.960*** (-6.61)
Constant	0.126 (0.71)	0.756** (2.60)
Firm Attributes	Yes	Yes
Call Attributes	Yes	Yes
Analyst Attributes	Yes	Yes
Industry-Quarter FE	Yes	No
Firm FE	No	Yes
Year FE	No	Yes
Observations	61392	61199
R^2	0.129	0.213

t statistics in parentheses, Standard errors clustered by firm

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

APPENDIX: Robustness Checks

In table 5 we report the relationship between performative atypicality, interacted with R&D activity, and earnings surprise. Here we report additional analyses with different operationalizations of the R&D activity variable. In the main paper we operationalize R&D activity as a binary variable, distinguishing between firms that engage in R&D and those that do not (where firms that report any R&D expenditures during our observation window are defined as R&D active). Here, we produce two additional operationalizations of R&D activity which further differentiate between R&D active firms. As is conventional in the finance and accounting literatures, we define R&D intensity as the R&D expenditure divided by sales, averaged by year to smooth seasonal fluctuations. The first measure differentiates between low and high R&D activity firms at the median of R&D intensity, and the second between low, medium and high R&D intensity firms by tercile. A firm's level of R&D activity is determined as its average level.

Table A1 reproduces the models reported in Table 5 with these variables, with Models 1 and 3 reporting the median-based variable, and Model 2 and 4 the tercile-based variable. The omitted category in all models is firms with no R&D activity. Models 1 and 2 report between-firm specifications, and models 3 and 4 within-firm ones. Our finding that performative atypicality premium is attenuated for R&D active firms is robust to these specifications. The effect is concentrated in highly active firms when examining between-firm effects (Model 1), but this is less the case when examining effects within firm. Overall, these results demonstrate that R&D activity attenuates the preformative atypicality premium.

Table A1: Robustness Checks - Earnings Surprise by R&D Intensity

	Between-Firm		Within-Firm	
	Median (1)	Tercile (2)	Median (3)	Tercile (4)
Performative Atypicality	-0.065*** (-4.52)	-0.065*** (-4.54)	-0.055*** (-3.88)	-0.055*** (-3.89)
Categorical Atypicality	-0.012 (-1.10)	-0.012 (-1.08)	0.011 (0.53)	0.011 (0.54)
<i>R&D Intensity by Median</i>				
Low R&D	0.062* (2.34)			
High R&D	0.063 (1.93)			
Low R&D × Performative Atypicality	0.032 (1.65)		0.043* (2.09)	
High R&D × Performative Atypicality	0.073*** (3.32)		0.051* (2.25)	
<i>R&D Intensity by Tercile</i>				
Low R&D		0.072* (2.47)		
Med R&D		0.048 (1.44)		
High R&D		0.066 (1.68)		
Low R&D × Performative Atypicality		0.046* (2.06)		0.061** (2.65)
Med R&D × Performative Atypicality		0.060** (2.67)		0.050* (2.18)
High R&D × Performative Atypicality		0.055* (2.20)		0.027 (0.97)
Analyst Disagreement	-1.318*** (-5.28)	-1.314*** (-5.26)	-1.965*** (-6.63)	-1.965*** (-6.63)
Constant	0.113 (0.63)	0.128 (0.71)	0.749** (2.58)	0.756** (2.60)
Firm Controls	Yes	Yes	Yes	Yes
Call Controls	Yes	Yes	Yes	Yes
Analyst Controls	Yes	Yes	Yes	Yes
Industry-Quarter FE	Yes	Yes	No	No
Firm FE	No	No	Yes	Yes
Year FE	No	No	Yes	Yes
Observations	61395	61395	61201	61201
R^2	0.129	0.129	0.213	0.213

t statistics in parentheses, standard errors clustered by firm

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