Throwing Curveballs: Unpacking Surprising Questions in Evaluative Settings and Probing their Origins*

Nandil Bhatia
Columbia University

Wei Cai
Columbia University

Sameer B. Srivastava
University of California, Berkeley

July 1, 2024

Abstract

From the interview room to the press room, much of organizational life unfolds in evaluative contexts wherein evaluatees present information that positions themselves in a favorable light, while evaluators ask penetrating questions to evaluate these claims. Although some questions are readily addressed, others are surprising in ways that can unsettle even a carefully crafted presentation. We propose that questions can be surprising in two analytically distinct ways: when they are off-topic and when they are unexpected. We argue that questions that are *on-topic* but *unexpected* are most likely to be disruptive. We refer to such questions as *curveballs* and examine the situations under which they arise. Whereas prior work on interpersonal evaluation focuses on actor- and interaction-level explanations, we consider the role of a structural property: the information environment. We theorize that evaluators are more likely to pose curveball questions when there is a dearth, rather than abundance, of public information about the evaluatee. To evaluate these ideas, we develop a novel measure of curveball questions using natural language processing techniques. Using a corpus of quarterly earnings calls and data on newspaper closures, which induce exogenous variation in a locally headquartered firm's information environment, we find support for our theory.

*We thank Sarayu Vyakaranam for excellent research assistance support and research seminar participants at Columbia Business School, the Computational Culture Lab, and the Berkeley Haas Macro Research Lunch for helpful feedback on this paper. Paul Gouvard, Daniel Keum, and Shiva Rajgopal provided valuable feedback on prior drafts. The usual disclaimer applies. Direct all correspondence to Nandil Bhatia: nb3029@gsb.columbia.edu

INTRODUCTION

Much of organizational life unfolds in an interactional context that Goffman (1970) referred to as an "expression game:" One party communicates information that aims to position itself in a favorable light, while another evaluates the quality of the expressed information. Examples of such "games" include a job applicant meeting with an interviewer (e.g., Rivera 2012), an entrepreneur making a pitch to an angel investor (e.g., Huang & Pearce 2015), an academic presenting new research in a seminar (e.g., Blair-Loy, Rogers, Glaser, Wong, Abraham, & Cosman 2017; Dupas, Modestino, Niederle, Wolfers et al. 2021), and a management team conducting a quarterly earnings call with financial analysts (e.g., Gouvard, Goldberg, & Srivastava 2023).

Yet evaluators are typically not just passive recipients of information provided by evaluatees. Rather, given that evaluatees often have an interest in withholding or obfuscating negative information about themselves (e.g., Bourdage, Roulin, & Tarraf 2018; Bushee, Gow, & Taylor 2018; Garud, Snihur, Thomas, & Phillips 2023) or in dodging difficult or uncomfortable topics (e.g., Hollander, Pronk, & Roelofsen 2010; Rogers & Norton 2011; Barth, Mansouri, & Woebbeking 2023), evaluators tend to pose questions that are designed to penetrate the communication veneer. Although evaluatees may try to anticipate and prepare for such queries, evaluators' questions sometimes genuinely surprise evaluatees and can unsettle even a carefully crafted presentation. These question-and-response dynamics can influence such consequential outcomes as who gets hired, funded, favorably evaluated, or rewarded in the financial markets. What makes a question surprising in ways that it can be disruptive to evaluatees and under what conditions are such questions most likely to be posed?

To shed light on these matters, we first develop a conceptual framework that clarifies the ways in which a question can be surprising in the context of an expression game. We propose that questions posed by evaluators can be surprising to evaluatees in two analytically distinct ways: when they are off-topic and when they are unexpected. We further argue that a particular form of surprising question—one that is *on-topic* but *unexpected*—is most likely to be disruptive to evaluatees. Borrowing from the language of baseball and the types of pitches that get thrown to batters, we refer to such questions as *curveballs*.

Next, we consider the conditions under which curveball questions are most likely to arise. Whereas prior work on expression games focuses on characteristics of actors or the nature of their interactions, we highlight the role of a structural property in which such an interaction is embedded: the quality of the information environment in which evaluators form judgments about an evaluatee. We theorize that, when information about evaluatees is readily accessible—for example, when star ratings exist of job seekers on an online platform or when a regional news outlet exists that reports on a locally headquartered firm—evaluators will tend to have common knowledge about, and will therefore be likely to converge in their interpretations of, a given evaluatee (Chwe 2013). As a result, the questions they pose during an expression game to pierce an evaluatee's communication facade will tend to be on a more similar and more predictable set of topics. Hence, curveball questions are unlikely to arise in such an information environment. Conversely, when information about evaluatees is scant or costly to obtain, there will be less common knowledge about evaluatees and evaluators' interpretations will tend to diverge. The questions evaluators pose during an expression game will thus tend to surface a wider array of issues that are more difficult for evaluatees to predict. It is under these conditions that they are most likely to throw curveballs at evaluatees.

We evaluate these ideas in the context of quarterly earnings calls (QECs), wherein the management team of a publicly traded firm (the evaluatee in our context) presents strategic and financial information and then engages in a question-and-answer session with the financial analysts who cover their stock (the evaluators in our setting). Our sample includes 126,910 earnings call transcripts between 2011 and 2021 for 5,609 listed firms from the United States. Building on the burgeoning literature that employs the tools of computational linguistics in organizational research (Corritore, Goldberg, & Srivastava 2020; Vicinanza, Goldberg, & Srivastava 2023; Kovács, Hsu, & Sharkey 2023; Aceves & Evans 2023), we use the transformer family of models—specifically, a sentence transformer model (SBERT) that is trained on 20 million sentence pairs and a transformer model that is fine-tuned to the context of financial discourse (FinBERT)—to derive measures of the degree to which a question posed by an analyst is off-topic (based on SBERT) and unexpected (based on FinBERT). We combine these indicators to derive our measure of curveball questions. Given that this a novel measure, we report a battery of analyses that help to validate it, showing that that our main measure relates to such consequential behaviors as CEO non-responses and that a call-level version of the measure is positively related to same-day volatility.

We then employ an empirical strategy that has been used in the finance and accounting literature (Allee, Cating, & Rawson 2023; Gao, Lee, & Murphy 2020; Kim, Stice, Stice, & White 2021) to identify exogenous shifts in the quality of the information environment about a focal firm: the closure of a regional newspaper that is located in close proximity to the firm's headquarters. Specifically, we combine data from multiple sources to create a composite dataset of 103 newspaper closures between 2012 and 2017 across a broad set of U.S. counties. Using a staggered difference-in-differences design, we find support for our main hypothesis. We discuss implications of these findings for research on expression games, social evaluations of economic actors, and use of computational methods to measure different facets of interactional dynamics in groups.

SOCIAL EVALUATIONS AND EXPRESSION GAMES

Across such diverse domains as feature films (Hsu 2006; Cattani, Ferriani, & Allison 2014), commercial music (Askin & Mauskapf 2017; Shi 2023), financial investing (Stuart, Hoang,

& Hybels 1999; Zott & Huy 2007; Giorgi & Weber 2015; Hallen & Pahnke 2016), books (Kovács & Sharkey 2014), restaurants (Kovács, Carroll, & Lehman 2014; Goldberg, Hannan, & Kovács 2016; Botelho & Gertsberg 2022), and beverages (Negro & Leung 2013; Verhaal, Khessina, & Dobrev 2015), an extensive literature has examined the social processes through which an audience evaluates the worth of an entity or actor based on its self-presentation (for a review, see Lamont 2012; Zuckerman 2012; Kovács, Negro, Pólos, Pontikes, Sharkey, Le Mens, & Hsu 2019). In many cases, such self-presentational displays and audience evaluations happen passively and indirectly—for example, when a firm chooses a set of labels to describe itself or its product offerings on a website or in a press release and, separately, a venture capitalist reviews such labels in making an assessment of the firm's worth (Pontikes 2012).

We focus instead on evaluative contexts in which an actor actively seeks to shape an audience's perceptions of worth through real-time communication—for example, in a job interview, a fund-raising pitch, a research seminar, or—as in our setting—a quarterly earnings call between a firm's management team and the financial analysts who cover its stock. It is this class of interactions and what they unearth about "the individual's capacity to acquire, reveal, and conceal information" that Goffman (1970, p. 4) examines in his essay, "Expression Games." Using the motivating example of an intelligence agent conducting an interview with a suspected spy, Goffman conceptualizes such interactions as a type of strategic game. Evaluatees can make a number of moves to position themselves in a favorable light—for example, "covering" moves that seek to mask or conceal unfavorable information or "feigning" moves that aim to misrepresent the subject's beliefs, attitudes, and preferences. Similarly, evaluators can engage in "uncovering" moves that are designed to "crack, pierce, penetrate, and otherwise get behind the apparent facts to uncover the real ones" (Goffman 1970, pp. 17-18). Evaluatees can, of course, also anticipate evaluators' uncovering moves and pursue "counter-uncovering" strategies.

Building on this conceptual apparatus, we consider a particular form of uncovering move made by evaluators: the posing of probing questions that can reveal information an evaluatee might otherwise seek to cover. In many cases, such queries are easily handled because evaluatees have not only anticipated but also formulated a response to them even before the evaluative interaction takes place. Yet on occasion such questions not only surprise evaluatees but can also throw off self-presentational display, even it has been carefully crafted and thoroughly rehearsed. To clarify the conditions under which questions might have such an effect, we develop a framework of surprising questions using the motivating example of an academic "job talk."

FORMS OF SURPRISE IN QUESTIONS

In our hypothetical example, a PhD candidate is giving a presentation that aims to position her research and her skills in the most favorable light to her audience—the faculty at the institution she hopes will hire her. Her job market paper is solid but not without some flaws and limitations. To prepare for the talk, she crafts and rehearses her presentation, thinks through the questions and objections she is most likely to receive, and preformulates responses to those queries. The faculty attending her talk seek to pressure test the validity of the claims she is making, the depth of her thinking, and her creativity. Over the course of her presentation, they interject periodically with questions designed to assess these qualities.

We propose that the questions audience members pose to the job candidate could surprise her in two analytically distinct ways. First, they could be *off-topic* rather than *on-topic*. For example, she might be discussing the relevant literature that informs a core hypothesis when someone interjects with a question about her causal identification strategy. Second, they could be *unexpected* rather than *expected*, independent of whether they are on-topic or off-topic. For example, someone might raise an endogeneity threat that she had previously failed to consider.

Juxtaposing these two dimensions, we develop a framework, shown in Figure 1, that clarifies the forms of surprise in evaluators' questions that are most likely to disrupt an evaluatee's self-presentational display. We use a metaphor from the domain of baseball—the types of pitches that might get thrown to a batter and the ease with which they can be handled—to characterize these archetypes. We begin with a type of question that is unlikely to produce much, if any, surprise: one that is both on-topic and expected. We refer to these as "meatball pitch" questions. In baseball, a meatball pitch is one thrown directly over home plate at relatively low velocity and with little movement. Such a pitch is relatively easy for a batter to put into play. Similarly, in the context of our illustrative academic job talk, when a job candidate confronts a question that is both on-topic and expected, she is likely to already have back-up slides prepared in response and may even have hyperlinks to these slides embedded in the places of her talk where each question is most likely to arise.

At the other extreme are "wild pitch" questions, which are both off-topic and unexpected. In baseball, such an errant pitch follows a trajectory that never comes close to the strike zone over home plate. It is therefore relatively easy for batters to simply avoid swinging at it. In a similar vein, a question in an academic job talk that is both off-topic and unexpected can sometimes be readily dismissed with a polite response such as: "That's an interesting idea that might be useful to explore in future research but is outside the scope of my present study." Of course, not every such question can be so parried. Yet, we contend, it is relatively easy to turn away or at least defer an oddball question when it is off-topic.

More complicated are the two off-diagonal positions in our framework. We refer to questions that are expected but off-topic as "paint-the-corners pitch" questions. In baseball, such a pitch is likely to be thrown when the count of balls and strikes strongly favors the pitcher over the batter (e.g., zero balls and two strikes). In such cases, the pitcher will tend to throw pitches not directly over home plate but at the edges of the plate, thereby increasing the chances that the batter swings and misses or fails to swing at a pitch that is just barely a strike. Yet batters can anticipate that such pitches will be thrown when they face an unfavorable count and therefore tend to make defensive swings that redirect the ball out of play, thus spoiling the pitch and forcing the pitcher to try again. A parallel in an academic job talk would be deflecting a question that is expected but either not directly relevant or posed at an inopportune time. The job candidate might, for example, tell a person posing a question about an alternative explanation for her findings to hold off until later in the talk when she will discuss a robustness check she has already run to help rule it out. Alternatively, the job candidate could simply take a short detour and recite the prepared response to the question that she was prepared to address later in the talk.

Finally, we turn to questions that are on-topic but unexpected, which we label "curveball pitch" questions. In baseball, a curveball is a pitch thrown with a spin that causes it to dive as it approaches the plate. The specific trajectory is often difficult to predict and can therefore lead the batter to take a swing but miss the ball. Colloquially, throwing someone a curveball refers to the act of posing a surprising query that is difficult to respond to. In a similar vein, we propose that, in the context of an interrogation game, questions that are on-topic but unexpected are the ones that are most likely to function as curveballs. Because they are on topic, they cannot be readily dismissed or deflected. Yet, because they are unexpected, the evaluatee is forced to improvise a response, which may inadvertently reveal gaps and shortcomings in her crafted presentation. In our hypothetical example, the job candidate may be presenting her robustness checks when an audience member raises a question about a different alternative explanation she had not previously thought of. She may then be forced to either concede the point or come up with a real-time response, which may in turn trigger a cascade of additional curveball questions from the questioner or others in the audience. Thus we propose that, relative to the other three question types, curveball questions are most likely to disrupt and potentially undermine a presenter's flow.

Insert **Figure 1** here

WHEN CURVEBALLS ARE MOST LIKELY TO BE THROWN

Although we are not aware of any literature that has sought to examine the origins of surprising questions per se, prior work on interrogation games has highlighted the factors that shape the quality of moves and countermoves made by evaluatees and evaluators. For example, Goffman (1970) traced the quality and relative efficacy of the two parties' actions in part to attributes of their physical setting, their skills and knowledge, their psychological dispositions, and prevailing social norms. More recent work has called attention to how the framing of question—for example, whether it is directly or indirectly stated (Bitterly & Schweitzer 2020), open- or closed-ended (Sternberg, Lamb, Hershkowitz, Esplin, Redlich, & Sunshine 1996), or stated with the assumption that a problem does or does not exist (Minson, VanEpps, Yip, & Schweitzer 2018)—can influence an evaluatee's likelihood of telling the truth or revealing compromising information. Yet these studies are silent on the question of whether such framing choices matter for the degree of surprise experienced by evaluatees or the conditions under which evaluators tend to resort to such tactics.

Whereas interdisciplinary research on the broad class of "expression games" has focused on on attributes of actors and their interpersonal dynamics, we highlight the role of the external environment in which such an interaction is embedded. We build on a long line of research in organizational theory that has examined how properties of the external environment—for example, market competition or resource abundance—shape different facets of an organization's inner workings (Hannan & Freeman 1984; Davis, Eisenhardt, & Bingham 2009; Alexy, Poetz, Puranam, & Reitzig 2021).

To identify the conditions under which curveball questions are most likely to be thrown, we focus on a different contextual attribute—the quality of the information environment in which evaluators form judgments about an evaluatee—and consider its impact not on the focal entity but on the external audiences with which it communicates. For example, in the context of freelance work, the information environment about job candidates shifted considerably with the arrival of online platforms that provided employers with ratings and reviews of each candidate's past work (Leung 2018). A similar change occurred in the restaurant industry with the development of online reviews (Kovács et al. 2014).

While the previous two examples represent positive shocks to the information environment, it is also possible for the information environment to become more impoverished. An example that is salient in our empirical context is the closure of a regional newspaper, which covers (among other topics) the firms that have headquarters in its vicinity. Recent work has documented the steady drumbeat of newspaper closures throughout the United States (Gao et al. 2020). Given that local papers serve as the primary source of "accountability journalism" in a community, these closures can have consequences for organizational behavior—for example, the propensity of a firm to engage in wrongdoing absent the watchdog function of a local paper (Choi & Valente 2023). (In the Empirical Setting and Data section below, we provide further evidence of how a newspaper closure represents a negative shock to the information environment of a proximally headquartered firm.)

Building on this intuition, we argue that shocks to an information environment, ones that change the ease or cost of access to salient information about an evaluatee, can influence the extent to which evaluators and evaluatees have *common knowledge* about the evaluatee. A fact is common knowledge when all parties in a group know it, when everyone knows that everyone knows it, when everyone knows that everyone knows that everyone knows it, and so on (Chwe 2013). The presence or absence of common knowledge, in turn, shapes how evaluators understand and interpret information about an evaluatee and thereby affect the types of questions they pose. In our context, the analysts who cover a given stock are united in simultaneously seeking to make sense of a given firm's past performance and future prospects.

When information about evaluatees is widely available—for example, when a local paper reports on a local firm's activities, investments, and personnel decisions—evaluators and evaluatees will tend to have common knowledge about the evaluatee. Thus, evaluators will be apt to pose questions on more similar topics—ones that are informed by their shared understanding of the firm—and evaluatees, knowing that evaluators have these shared interpretations, will more readily anticipate these questions. Thus, it is unlikely that a question posed by evaluators to an evaluatee will be a curveball.

Conversely, when the information environment becomes impoverished—for example, because a regional newspaper that historically reported on a local firm shutters—there will be a concomitant loss of common knowledge about the firm. Evaluators will then tend to ask questions about a wider variety of topics. Morevover, evaluatees will be left guessing about the information sources evaluators might have consulted beforehand, thus making it harder for them to anticipate and prepare for the broader set of topics that might be covered. It is under these conditions, we contend, that curveball questions are most likely to arise. Thus, we propose:

Main Hypothesis: *In an expression game, evaluators are more likely to pose a curveball question (i.e., one that is on-topic but unexpected) to an evaluatee when there is a dearth, rather than abundance, of common knowledge about the evaluatee.*

DATA AND METHODS

Quarterly Earnings Calls as a Form of an Expression Game

We test our main hypothesis in the context of quarterly earnings calls (QECs) conducted by publicly traded firms that are based in the United States. In these calls, a firm's management team first presents information about its strategy, financial performance, and outlook and then engages in a question-and-answer session with the sell-side analysts who cover its stock (Chen, Nagar, & Schoenfeld 2018; Mayew, Sethuraman, & Venkatachalam 2020). The former typically follows a well-rehearsed script, while the latter is more open-ended and, despite the management team's best intentions, may veer into topics and territory that the management team would prefer not to discuss.

QECs are a classic form of what Goffman (1970) referred to as an expression game. The firm's management team, the *evaluatee* in the game, seeks to present the firm's performance and its outlook in as favorable a light as possible to its shareholders, sell-side analysts, and the broader market. The analysts, the *evaluators* in the game, ask probing questions of executives to clarify, challenge, or induce them to reveal additional information. The interests of the two parties are not completely aligned: The management team aims to leave a positive impression of the firm, whereas the analysts strive to gain an accurate assessment. Both parties are aware of each other's incentives and anticipate their likely moves and countermoves.

QECs are an ideal setting in which to examine the emergence and potential consequences of curveball questions. Prior work has used QECs to study executive behavior, highlighting how they can be high-pressure situations that require executives to respond in real-time to difficult lines of questioning (Hobson, Mayew, & Venkatachalam 2012; Larcker & Zakolyukina 2012; Li, Minnis, Nagar, & Rajan 2014; Cai, Rouen, & Zou 2022; Gouvard et al. 2023). Although the management team typically possesses more information about the firm than do the analysts who cover the firm, they cannot plausibly anticipate every contingency that might arise and, forced to respond in real-time, may be caught off-guard by a given analyst's query. Such a dynamic is not, of course, limited to QECs and is also present in such other evaluative contexts as interviews or depositions or police interrogations.

Empirical Setting and Data

Our sample consists of 126,910 QEC transcripts for 5,609 listed firms in the United States and spans the period from 2011 to 2021. We downloaded this corpus from Capital IQ, which partially structures the data by identifying blocks of utterances during the call and the speaker's name and affiliation. Thus, we could clearly identify the transition from the scripted initial part of the call to the less structured question-and-answer session, as well as identify when a particular analyst posed a question and how members of the management team responded.

We obtain data on the changing information environment related to firms in our sample by examining the closures of local newspapers in counties where those firms are headquartered. Prior work has used such data to examine how newspaper closures influence such outcomes as local county borrowing (Gao et al. 2020) and corporate misconduct (Choi & Valente 2023). Studies in this vein have documented a variety of ways in which the closure of local newspapers influences the information environment of firms headquartered in close proximity, especially for sell-side analysts. One reason is that even local newspapers employ financial journalists who believe that monitoring companies and holding them accountable for their actions are among their most important roles. Moreover, many rely on their personal relationships with company management and corporate communications departments to help them conduct research on firm conduct (Call, Emett, Maksymov, & Sharp 2022). Beyond financial journalists per se, Choi & Valente (2023) and Hamilton (2016) highlight the broader "accountability journalism" role of local newspapers, which produce investigative reports about the ethical conduct of firms in the community. In many cases, these local reports are then disseminated more broadly by national mainstream media.

Corroborating that a link exists between local media and the information environment about proximately headquartered firms, Allee et al. (2023) find that stock-level volatility, spreads, and illiquidity all increase as local newspaper intensity declines. These effects are amplified when analysts are more dependent on the local newspaper as an information source—for example, when they are busier or when the firm is especially important in the community. Further supporting the idea that local newspaper closures represent a negative information shock for nearby firms, Das & Zhong (2024) find that the unexpected loss of local media coverage leads to a decline in analysts' research quality: Analysts who cover such an affected firm produce less frequent, less timely, and less accurate forecasts about the firm following a newspaper closure.

In light of these studies on the informational consequences of local media loss, we begin with a subset of the newspaper closure data, covering the period from 2012 to 2014, from Gao et al. (2020). We then supplement these data with information from the University of North Carolina Hussman Database on newspapers and news deserts in the United States (Abernathy 2018). This allows us to extend the dataset to 2017. Following prior literature, we define "closure" broadly as encompassing three types of events: a newspaper fully shutting down, transitioning from being a weekly to a daily, or merging with another paper. We focus on counties with three or fewer newspapers given that a "closure" event in a county with more than three newspapers is less likely to have a material impact on the information environment about a local firm. Our resulting dataset includes closures in 103 counties across the United States between 2012 and 2017.

Our regression models include several control variables, which we obtain from the Center for Research in Security Prices (CRSP) and Compustat. We also use security-level data on abnormal stock returns and idiosyncratic volatility to help validate our empirical strategy. We gained access to these data using the beta-suite that is available to subscribers of CRSP via the Wharton Research Data Services.

Dependent Variable: Curveball Question Score

As a reminder, our theory suggests that an evaluator question is most likely to be viewed as a curveball by an evaluatee when the question is both on-topic and unexpected. To develop such a measure, we first had to make a choice about which question or questions to score along these dimensions. One important consideration is that, given the presence of multiple evaluators (i.e., sell-side analysts), the degree to which a question is on-topic or unexpected may vary based not only on what questions were previously posed and how the management team responded to them but also based on the management team's initial presentation. For example, if a person opens up an entirely new line of questioning following a string of questions on a given topic, it might be unexpected relative to the last question asked; however, it might be expected relative to the management team's original presentation. To circumvent such complications, we opted to focus on the very first question posed by an analyst following the management team's scripted presentation. The downside of this choice is that we limit the sample of questions that can be assessed on our curveball question score. Yet the advantage is that the measure is cleaner in assessing the likelihood that the management team perceived the question as a curveball. Recognizing the tradeoffs involved in the construction of this measure, we report in the Appendix an analysis showing the robustness of our measure to the inclusion of questions beyond the first one.

Our measure of curveball questions is a composite of two measures of this construct's underlying dimensions (i.e., on-topic and unexpected). We first develop separate measures of these dimensions using language models from the BiDirectional Encoder Representations from Transformers (BERT) family (Devlin, Chang, Lee, & Toutanova 2019) and then combine the two measures to develop a curveball question score. BERT is a neural network-based language model that can identify the semantic and contextual information in a given text (Devlin et al. 2019; Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, & Polosukhin 2017).

The On- Versus Off-Topic Dimension. To gauge whether a question is on- or off-topic, we use the Sentence Transformers (SBERT) library. Sentence Transformers, built on the BERT architecture to accomplish the downstream task of semantic comparison, generates embedding vectors for sentences and short pieces of text (Reimers & Gurevych 2019). Vectors corresponding to different texts can then be compared to each other to infer the semantic similarity between them. We begin with a pre-trained SBERT model "all-MiniLM-L6-v2" from the website HuggingFace. The model represents a given text as 384-dimensional embedding vector. Given that the model was trained on data from a wide range of settings (e.g., Reddit), we fine-tuned the model to the financial services context. Specifically, we fine-tuned the SBERT model on more than 20 million sentence pairs from our sample of earnings calls, under the assumption that consecutive sentences from an executive's answer to a particular question should be semantically close to one another.

SBERT has the same token limit as BERT (512 tokens), thereby limiting the amount of text that can be input at a given time. To overcome this limitation, we broke the management presentation into multiple chunks of texts and then used the fine-tuned SBERT model to measure the semantic distance between each question and each chunk of the management presentation (based on cosine distance). Larger distances imply that the question is more off-topic. Our resulting measure of "off-topicness" is provided below:

$$\text{Off-topic}_i = \frac{1}{N}\sum_{j=1}^{N} \text{CosineDistance}(\text{Mgmnt. Presentation}_j, \text{Question}_i) \tag{1}$$

The Expected Versus Unexpected Dimension. To quantify the degree to which a question was expected versus unexpected, we employed the next sentence prediction capability of FinBERT (Araci 2019). FinBERT is a pre-trained language model that has already been fine-tuned on financial text. The next sentence prediction function estimates the likelihood of a particular text following a given piece of text that is entered as an input. We utilized the publicly available language model of FinBERT from GitHub to access this function. Given BERT's token limit, we again divided the management team presentation into smaller chunks. We then computed pair-wise predictive probabilities using the different parts from the management's introduction and the first question posed by a sell-side analyst. Specifically, we calculated the likelihood of the question following each chunk of the management's presentation. Lower predictive probabilities are interpreted as markers of greater unexpectedness in the question relative to the management team's scripted presentation. We took the inverse of the result from Next Sentence Prediction, such that greater values present greater unexpectedness [1]. Our measure of unexpectedness is provided below:

$$\text{Unexpectedness}_i = \frac{1}{N}\sum_{j=1}^{N} \text{NextSentencePrediction}(\text{Mgmnt. Presentation}_j, \text{Question}_i) \tag{2}$$

The Composite Measure of Curveball Questions. Our measure of curveball questions is a composite of the two described above. Specifically, we first invert the off-topic measure such that it reflects the extent to which a question is on-topic. We then multiply this inverted measure by our measure of unexpectedness. We use percentile scores in making this calculation to ensure proportional aggregation of the two measures. The composite score is explained below:

$$\text{Curveball Question Score} = (\text{Percentile Unexpectedness}) \times (\text{Percentile On-topic})\,/100 \tag{3}$$

Recognizing that this is a novel measure, we report below a series of analyses we ran to validate it and show that it relates behaviors that prior work has linked to consequential economic outcomes. In the Appendix, we also report results of models in which we include not the composite measure but instead the two dimensions of unexpectedness and off-topicness separately, as well as their interaction. (For our main hypothesis, which focuses on the likelihood of a curveball question being asked, we had no choice but to use the composite measure as our dependent variable.)

[1]We use the term "unexpected" to refer to a question that is deemed unlikely, based on BERT's Next Sentence Prediction feature, relative to the management team presentation that preceded it. It is possible that the management team might have anticipated that such a question would arise at some point during their presentation; however, the timing of the question might still be unexpected given the conversation's narrative flow.

Independent Variables

Newspaper Closure. As noted above, we treat closures of regional newspapers as exogenous shocks to the information environment about locally headquartered firms. For each firm, we identify the zip code of its headquarters location, as recorded in Compustat. We then match each zip code to a county using the U.S. government's crosswalk between the two data. Each firm, depending on whether its county experiences a closure of a local newspaper between 2012 and 2017, is identified as being treated or not. We use a six-year treatment window, wherein treated firms are present in the sample three years prior to the change in the information environment (including the year of closure) and three years after, following a similar approach to the one used by Heese, Perez-Cavazos, & Peter (2022). Our variable of interest is *Treated * Post*, which takes the value of 1 for treated firms for the three years after the closure of the local newspaper.

Control Variables

We account for multiple variables that could potentially influence the likelihood that a management team receives surprising questions from sell-side analysts during a QEC. First, we control for firm size, as the complexity of operations in larger firms might naturally lead to more surprising questions from analysts. The information environments of larger firms may also differ from those of smaller firms. Next, we account for the firm's stock performance and trading volume for the quarter leading up to the call. This allows us to mitigate the influence of short-term market perceptions on the types of questions that analysts may ask management. Similarly, we control for the firm's financial and operational metrics, incorporating both return on assets and Tobin's Q as control variables. To further account for the firm's financial health, we include controls for leverage and the firm's cash ratio. Finally, to capture any shifts in strategic direction, we control for the firm's capital and research intensity. We winsorize the financial and operational metrics by year at the

2% and 98% levels to mitigate the impact of outliers.

In regression models that examine the impact of local newspaper closures on firm-level outcomes, we control for local economic conditions at the county level. Changes in these conditions might simultaneously affect the closure of local newspapers and firm-level outcomes. To account for this possibility, we include as controls the county-level unemployment rate, as well as county-level median income. The former is downloaded from the Economic Research Service of the U.S Department of Agriculture, while the latter is downloaded from the U.S. Census Bureau.

Analytical Strategy

As a preliminary step, we sought to validate our measure of curveball questions. As described in greater detail below, we first examined how management teams responded when they faced curveball questions. We anticipated that curveball questions would be unsettling to a management team such that: (a) they were more likely to obfuscate by giving non-answers; (b) they tended to give longer responses; (c) the CEO was more likely to respond instead of other executives who were present; and (d) conditional on the CEO giving a response, that response was more likely to be a non-answer.

Not only are these behaviors consistent with the interpretation that curveball questions surprise management teams in ways that disrupt their communication strategy, but some of these outcomes may also have material economic consequences. For example, prior work has demonstrated that non-answers are associated with lower cumulative abnormal stock returns and higher volatility (Barth et al. 2023). Given that the market is more responsive to communication from CEOs than from other executives (Goldman & Zhang 2022), non-responses by CEOs may elicit especially strong market responses.

As an alternative way of validating our curveball measure question, we also implemented an analyst-side analysis. Given that an analyst's likelihood of being able to formulate a question that the management team perceives to be a curveball is likely to be positively related to her skills and experience, we expected that such questions were more likely to be posed by star analysts relative to non-star ones.

Finally, we also conducted a call-level validation check. Although our main analyses focus on the first question posed by an analyst following the management team presentation, we also created a *call-level* curveball measure that is based on the first question posed by *every analyst* who attends the call. We then examine the association between this call-level measure and the stock's volatility on the day of the earnings call. If curveball questions are more likely to elicit surprise and unearth novel information, we expect our call-level measure to be positively correlated with the stock's volatility on the day of the call. We recognize that this analysis is only correlational; nevertheless, we think it provides suggestive evidence that curveball questions can have consequential market outcomes.

Next we sought to validate our assumption that the closure of a regional newspaper reflected a shock to the information environment about locally headquartered firms. We did so by comparing call-day stock returns (both raw returns and abnormal returns) and stock volatility (both raw and idiosyncratic volatility) in the period before a newspaper closure to the period after. We reasoned that the loss of news coverage about the firm would increase the importance of QECs as a source of novel (though not necessarily more positive or more negative) information about the firm. Given that the market is likely to be sensitive to disclosure of novel information, we therefore anticipated that a closure would be positively associated with an increase in the call-day volatility of the stocks of locally headquartered firms. Moreover, assuming the closure is a shock to the information environment that is unrelated to the focal firm's financial health or prospects, we assumed that it would not be significantly associated with call-day stock returns in the post-newspaper-closure period.

Turning to the empirical tests of our main hypothesis—that newspaper closures should produce a negative shock to the information environment that increases the likelihood of curveball questions being asked—we conducted our analyses at the firm-quarter level.

Unless otherwise specified, all models incorporate time-varying controls at the firm level. We also employ fixed effects for both the firm and the quarter to account for time-varying heterogeneity at the firm level. In models examining the impact of newspaper closures, standard errors are clustered at the county level, given that this is the level at which the treatment of newspaper closures operates.

Given that firms get treated at different points in time, we employ a staggered difference-in-differences design. The main variable of interest is the interaction term, "Treated * Post," which captures the marginal impact of newspaper closures on treated firms after the treatment has taken place (i.e., post-closure).[2]

RESULTS

Table 1 reports descriptive statistics for the variables used in our analyses. We first report results of analyses designed to validate our measure of curveball questions, then present results that help to validate that newspaper closures represent a shock to the information environment about a firm, then turn to our main results, and conclude by reporting robustness checks and supplemental analyses designed to address potential alternative explanations for our findings and to investigate the underlying mechanisms.

Insert **Table 1** here

In the Appendix (**Figure A.1**), we provide additional information on the distribution of three measures of interest: the off-topic, unexpected, and curveball scores of questions. The off-topic measure resembles a Gaussian distribution, whereas the unexpectedness score's

[2]It is important to note that our design incorporates control firms throughout the sample, eliminating the need to compare treated firms to one another based on the timing of treatment. Moreover, the issue of treatment timing is relatively inconsequential in our context. A substantial proportion (roughly 90%) of our control sample is never treated in our observation period. This mitigates concerns regarding potential biases that could arise when firms treated at different times serve as controls for each other, as highlighted in Baker, Larcker, & Wang (2022). We also focus on a relatively short time window, three years before and after the newspaper closure, thus reducing the likelihood that a firm treated early in the sample period would unduly influence treatment effects estimated for firms treated at a later time.

distribution has a long right tail (akin to a distribution emerging from a power law). The two individual measures have a correlation coefficient close to 0.35. This is intuitive as off-topic questions are more likely to be unexpected, but not always. The aggregated curveball question score, which uses the percentiles of the two individual dimensions, is slightly left-censored.[3]

To give a flavor of what curveball questions look like in our setting, we include in the Appendix (**Figure A.2**) illustrative examples of the five questions with the highest curveball score and the five questions with the lowest curveball score for two recognizable companies in data set: Apple and Netflix. Of course, the extent to which a question is a curveball depends on the management team presentation that came before it. Thus, it is difficult to interpret the extent to which a question is a curveball in isolation. Nevertheless, some informative patterns emerge from this exercise. For example, questions that are simple clarifications or requests for elaboration (e.g., Tim Cook being asked about the iPhone sales momentum) have especially low curveball scores. In contrast, questions that require the management team to envision and respond to complex but plausible scenarios appear to score higher on the curveball question score. For example, a question posed to Reed Hastings of Netflix about his potential competitive response to Amazon entering the strategic market and its implications for Netflix's pricing strategy scores especially high on our curveball measure.

Validation of Curveball Question Measure

To validate our composite measure for the degree to which a question from a sell-side analyst is curveball, we examine the responses by a firm's management team when it encounters such a question, as well as the types of analysts who are apt to ask such

[3]The censoring of a distribution from the left for the dependent variable may potentially introduce bias in estimation of a standard OLS regression; for the sake of robustness, we log transform the curveball question score, turning it into a Gaussian-like distribution, and re-run the regression models where the curveball question score is a dependent variable. We find no differences in our results.

questions. We expect management to respond to curveball questions by obfuscating, which would be reflected in both non-answers and longer responses, as well as to defer to the CEO in formulating a response. We also anticipate that analyst skill (as measured by their designation as star analysts) will be positively related to their propensity to ask curveball questions. Finally, we examine how the degree to which questions in the call are curveball correlates with stock volatility on the day of the call.

Non-answers: Our measure of non-answers comes from Barth et al. (2023). Their methodology uses sets of trigrams (i.e., three consecutive words) that were developed using a robust machine learning approach to identify obfuscating language in the context of QECs, making it an excellent fit for our context. Due to the nature of the variable and the presence of outliers, we winsorize it at the 99% level to reduce the impact of outliers.

Length of response to question: We measure response length based on the natural logarithm of the total number of words in the management's response to the question.

CEO involvement in answer: To identify the executives who responded to a curveball question, we fuzzy-matched the names in the QEC with those listed in the Execucomp database by year. Due to a significantly lower number of firms being covered in the Execucomp database relative to the Capital IQ transcripts database, we lose a part of our sample while conducting this validation check. We still retain more than half of our corpus of transcripts for this analysis. We estimate logit models of whether a CEO is likely to respond as a function of the extent to which a question is a curveball.

Non-answers from the CEO: Conditional on a question being answered by the CEO, we also separately measure the extent to which the response is a non-answer. (This variable takes the value of 0 if the CEO does not respond to the question.)

Star analysts: The magazine, *Institutional Investor*, provides rankings of star analysts in the United States. Consistent with prior literature (Vukovic, Kurbonov, Maiti, Ãzer, & Radovanovic 2024), we use this designation as an indicator of analyst reputation and human capital.

Call-level Curveball Measure: For the analysis concerning the impact of curveball questions on the firm's stock volatility, we create a call-level curveball score by calculating the median curveball-ness of the first question asked by every analyst in the call.

The results from our validation checks are included in **Table 2**. All models are run with time-varying controls at the firm level and fixed effects for the firm and quarter in which the call occurred. We cluster the standard errors at the level of the firm. In Model 1, we find that the degree to which executives provide non-answers is positively related to our measure of curveball questions ($\beta = 0.010, p < 0.01$). In Model 2, we find that executives provide much longer answers to curveball questions ($\beta = 0.076, p < 0.001$). Model 3 demonstrates that the likelihood of the CEO, rather than other executives, responding to a question is positively related to its curveball question score ($\beta = 0.142, p < 0.001$). In Model 4, we find that curveball questions also elicit greater non-answers when it is the CEO who responds ($\beta = 0.033, p < 0.001$). In Model 5, we find that our measure of curveball questions is positively related to whether the analyst posing the question is a designated star analyst ($\beta = 0.060, p < 0.01$). Finally, in Model 6, we find that our call-level curveball score is positively associated with the stock's volatility on the day of the call ($\beta = 0.005, p < 0.05$). Together, these analyses provide consistent, albeit indirect, evidence that our measure of curveball questions produces unsettling surprise in a management team.

Insert **Table 2** here

Validation of Newspaper Closures as a Shock to the Firm's Information Environment

The results of analyses designed to validate that newspaper closures represent a shock to the firm's information environment are included in **Table 3**. As a reminder, we expect that the loss of information about a firm following a newspaper closure will make subsequent QECs more important sources of novel information, which analysts will be sensitive to. This sensitivity will be reflected in greater stock volatility. However, assuming the closure is not related to fundamental shifts in the firm's financial outlook, it should have no relationship with stock returns. Consistent with the latter expectation, we find in Models 1 and 2 that the variable of interest, "Treated * Post," is not significant in predicting either call-day stock returns or call-day abnormal returns. In line with the former expectation, we find that this variable is positively and significantly related to call-day stock volatility and call-day idiosyncratic stock volatility in Models 3 ($\beta = 0.074, p < 0.01$) and 4 ($\beta = 0.062, p < 0.01$), respectively. These results suggest that newspaper closures represent an information shock yet are not a signal of either positive or negative information about the firm's future prospects.

Insert **Table 3** here

Main Results

We report tests of our main hypothesis, that a newspaper closure will increase the likelihood that the management team of a locally headquartered firm will face curveball questions from analysts, in **Table 4**. In Model 1, which is run without the inclusion of time-varying controls, we find the coefficient of the variable "Treated * Post" to be positive and significant ($\beta = 0.060, p < 0.01$). The fixed effects for the firm and the quarter in which the call happened absorb the individual coefficients of "Treated" and "Post." In Model 2, we see that the inclusion of time-varying firm controls has little impact on the size or significance of the coefficient of interest ($\beta = 0.058, p < 0.01$). Consequently, we find support for our main hypothesis.

Insert **Table 4** here

Figure 2 depicts the marginal effect of being treated from our staggered difference-in-differences estimates. The plot shows the results from the same specification as Model 2 in Table 4, except that the marginal effect is split by the number of years to treatment. It suggests that the effect of being treated is not distinguishable from zero prior to the closure of the newspaper. However, in each of the three years after the newspaper closure, we see an increase in the degree of curveball question score of the first question posed by analysts following the management team presentation.

Insert **Figure 2** here

Assessing the Theorized Mechanism

Our theory posits that local newspaper closures represent a negative information shock about proximately headquartered firms, which in turn erodes common knowledge about the firm and increases the likelihood of curveball questions being posed. Yet we do not directly observe analysts' knowledge or interpretations about firms nor what management knows that analysts know and so on. We therefore conduct three supplemental analyses that each provide indirect evidence of our theorized mechanism. First, we examine the topical breadth of management team presentations (i.e., the range of topics they speak to in their formal remarks) and, separately, the topical breadth of analyst questions (i.e., the range of topics they ask about) for the subset of treated firms in our sample before and after they experience the closure of a local newspaper. We use BERTopic (Grootendorst 2022) to classify each utterance in the management presentation and each question by sell-side analysts into a distinct topic (or set of topics). Subsequently, using the distribution of topics discussed in the call, we develop a call-level measure of topical breadth (or entropy) for both management presentations and analyst questions.

Figure 3 depicts the results of this analysis. Recall that our data on the timing of newspaper closures is coarse—at the annual level. It is therefore not possible to clearly identify the first QEC that takes place after a local paper shutters. We thus consider the year in which a newspaper closes as t=0, recognizing that some calls in this year may have taken place before the closure event. Although we might not expect to see a shift in this year, we do expect to see one in subsequent years. Such a pattern emerges across both panels of Figure 3. In Panel A, we see that topical breadth in management team presentations begins to rise in the year a local newspaper closes and continues to do so in subsequent years. The increase becomes statistically significant two years after closure. Panel B tells a similar story regarding the topical breadth of analyst questions; however, the increase becomes significant just one year after the closure event. We believe these results are consistent with the story of a negative information shock: The loss of common knowledge leads analysts to diverge in their understandings and interpretations of the firm and thus to ask a wider range of questions; moreover, because the topics analysts are interested in are harder to anticipate, management errs on the side of discussing a broader array of topics. For both reasons, the likelihood of a curveball question arising increases because it is simply harder for management to anticipate and prepare for the increased variety of questions they might face.

Insert **Figure 3** here

A second approach to assessing our theorized mechanism is to examine substitute channels of common knowledge about firms. Prior research points to the role of social connectedness in a community as a conduit through which information about local firms flows (Choi & Valente 2023). Following a similar logic, we reason that the effects of a local newspaper closure on the loss of common knowledge about a nearby firm will attenuate when the community in which it is located is highly socially connected. This is because information about a firm—for example, whether it is planning to open a new facility in anticipation of rapid growth—is more readily disseminated within a community and transmitted to mainstream media in area with high levels of social connectedness. **Table 5** reports the relevant results. We include Model 1, which is identical to our main Model 2 in **Table 4**, for ease of comparison. Using measures developed by Rupasingha, Goetz, & Freshwater (2006) (and subsequently updated in 2014), we find in **Table 5**, Model 2 that the treatment effect attenuates with the social capital of a community.

Finally, a third approach to assessing the mechanism is to consider heterogeneity by type of firm. Firms vary in the extent to which they receive coverage by local media. Moreover, Allee et al. (2023) highlight that the effects of newspaper closures on the information environment of a firm are greater for firms that are more important to their local community because analysts are more dependent on the local news source for information about such firms. In line with this view, we expect that the loss of local media coverage will lead to a greater loss of common knowledge about more prominent local firms relative to less prominent ones. Using firm size as a proxy for prominence, we expect that, all else equal, larger firms are more likely to receive curveball questions following a local newspaper closure than smaller firms. As reported in **Table 5**, Model 3, this is, indeed, what we find.

Insert **Table 5** here

Robustness Checks and Supplemental Analyses

We conducted a few additional analyses to assess the robustness of our findings and better interpret our results. In **Table 5**, Model 4, we address the previously noted limitation: that our data on newspaper closures do not indicate the specific date of the closure, just the year. Given that we are analyzing quarterly earnings calls, it is unclear if every call in our staggered difference-in-differences models represents a treated call. To increase the likelihood that our estimates are based only on calls that occurred following a newspaper closure, we report in Model 4 the results of an analysis in which we restricted the sample to only the first call of the year, which corresponds to results from the fourth quarter of the prior year. We obtain similar results in this model.

Next, recall that our measure of curveball questions is a composite of two separate measures. It was necessary to construct the composite measure because our main hypothesis pertains to the conditions under which such a question would arise, and we needed a single dependent variable to test this hypothesis. To ensure that our validation checks of this measure were not an artifact of the fact that we are using a composite measure, we first sought to test the robustness of our validation checks to disentangle the composite measure into its component parts. This was possible to do for three of our validation analyses in which the curveball question score is an independent, rather than dependent, variable. As shown in Appendix **Table A.1**, we obtain comparable results to the ones reported in **Table 2**, Models 1-3, wherein we disentangle the composite measure into separate measures of unexpectedness, off-topicness, and the interaction between the two. The marginal effects from these models are plotted in **Figure A.3** in the Appendix.

We also decomposed the treatment effect of newspaper closures on the two underlying dimensions of curveball questions: unexpectedness and off-topicness. These results are reported in Appendix **Table A.2**. For ease of comparison, Models 1 and 2 report again the results displayed in Table 4. As shown in the remaining models, the overall treatment effect appears to be driven primarily by an increase in the unexpectedness of questions that are posed to the management team.

Finally, in **Table A.3** in the Appendix, we establish that our findings pertaining to the treatment effect of newspaper closures on curveball questions are not an artifact of our empirical choice to select the first question in the earnings call. We find consistently positive and sizeable effects when we include the curveball score of the first question asked by the second and third analyst in our dependent variable. However, the results attenuate as we include more questions, highlighting the challenge of measuring curveball questions beyond the first question. This is due to the degree to which subsequent questions are curveballs depends partly partly on how unexpected or off-topic they are relative to the prior questions and the management team's response to them.

DISCUSSION

The goals of this article have been to explore why and how questions posed in evaluative contexts can sometimes surprise an actor and throw off an even carefully crafted presentation and to probe the origins of such questions. We did so by first developing a taxonomy of surprising questions that is based on the two underlying dimensions of off-topicness and unexpectedness. We first proposed that a curveball question—one that is on-topic and expected—is most likely to surprise an evaluatee in ways that unsettle her self-presentation. Next we argued that curveball questions can arise for macro-structural reasons that are unrelated to actor- or interaction-level dynamics. Specifically, we theorized that curveball questions will tend to arise when the information environment about a focal firm is or becomes impoverished. This is because, in the absence of common knowledge about evaluatees, evaluators will have divergent understandings of an evaluatee and thus formulate questions on a broader array of topics, while evaluatees will have a tougher time anticipating and preparing for this wider question range. Using the tools of computational linguistics, we then developed and validated a measure of curveball questions based on separate measures of its two component parts. Applying this measure to a corpus of corporate quarterly earnings calls and using the closure of regional newspapers as an exogenous shock to the information environment about a locally headquartered firm, we found support for our theory.

Contributions

This article makes noteworthy contributions to several literatures. First, it contributes to research on interpersonal evaluation by developing a taxonomy of surprising questions. We believe that this taxonomy can be readily extended to evaluative contexts beyond our empirical setting of QECs. For example, in developing our theory, we used the motivating example of an academic job talk. Yet the framework would appear to have application across the broad spectrum of settings that fit the profile of an "expression game" (Goffman 1970)—for example, police interrogations, venture capital pitches, political press conferences, safety inspections, financial audits, and regulatory approvals. Moreover, prior work on interpersonal evaluation has tended to emphasize actor- and interaction-level dynamics. For example, the literature on entrepreneurial venture evaluation has focused on such factors as entrepreneur minority status (Younkin & Kuppuswamy 2018), investor gut feel (Huang & Pearce 2015), and even incidental weather-related factors such as sunlight on the day of the entrepreneurial pitch (Dushnitsky & Sarkar 2022).

Similarly, in the sociological literature on labor markets and the evaluation of candidates, the focus has been on interpersonal referral networks (Fernandez, Castilla, & Moore 2000), applicant sex and race (Gorman 2005; Pager 2003), and evaluator emotions (Rivera 2015). In contrast to these approaches, we highlight how the structural property of an interaction, completely outside of the immediate context of the interaction itself, can impinge on its dynamics. This finding opens the door to understanding how other macro-structural properties—for example, shifts in the norms that prevail in a community or changes in the legal or regulatory environment—might influence the micro-dynamics of interpersonal evaluation.

Our findings also complement related research on question-asking in social psychology, political science, and criminology. Much of this work has focused on the consequences of how questions are framed—for example, whether they are directly versus indirectly stated (Bitterly & Schweitzer 2020) or open- versus close-ended (Sternberg et al. 1996). Whereas these attributes of a question are fairly static, our conceptualization of question surprise is dynamic in that the same question may or may not be surprising to an evaluatee based on when it is asked during a presentation, how it relates to what the evaluatee has previously reported, and how it relates to other potential questions that might have been previously asked. For example, if an evaluator simply rephrases a surprising question that was previously asked by a different evaluatee, it may no longer be unexpected and therefore fail to surprise the evaluatee—independent of how it is framed. This property of what makes a question surprising may have implications for question-based behavioral interventions. For example, men and women tend to get asked different types of questions when seeking entrepreneurial funding (Kanze, Huang, Conley, & Higgins 2018). More, a recent working paper suggests that a question-based intervention—encouraging evaluators to ask all evaluators, regardless of gender, about the risks and potential of their venture—can reduce discrimination in the evaluation process (Miller, Lall, Goldstein, & Montalvao 2023). Yet the efficacy of such an intervention may depend on the extent to which it is unexpected by evaluatees. If such a question becomes institutionalized within the evaluation process, evaluatees may begin to anticipate and prepare for it, thereby potentially diminishing its efficacy.

Second, we contribute to the interdisciplinary literature that analyzes discourse in quarterly earnings calls. In the accounting and finance literatures, QECs have served as a setting in which to examine executive behavior and management's decision to disclose information (Li et al. 2014; Bushee, Matsumoto, & Miller 2004; Hobson et al. 2012; Matsumoto, Pronk, & Roelofsen 2011; Li, Mai, Shen, & Yan 2021). For example, Larcker & Zakolyukina (2012) detect deceptive executives by analyzing word categories associated with deception in executive speeches. Hobson et al. (2012) uncover financial misreporting based on vocal dissonance markers in conference call audios. In more recent work, Cai et al. (2022) investigate the firm and individual career outcomes of interactions among executives in earnings calls. In the more sociological tradition, Vicinanza et al. (2023) identify the types of firms that are likely to communicate in ways that are prescient about how the field as a whole will evolve, while Gouvard et al. (2023) show how atypical forms of discourse relate to analysts' evaluations of firms.

Whereas these prior studies have focused on the speech acts of the management team, we highlight the importance of also attending to the nature and quality of questions posed by analysts. Moreover, while prior measures of discourse in analyst calls have focused on one side of the interaction or the other (i.e., the management team or analysts), we develop a construct and associated measure (i.e., the curveball question score) that is inherently dyadic in nature. The extent to which a question posed by an analyst is a curveball cannot be assessed in isolation; it can only be assessed in relationship to the prior discourse of the management team. It remains to be explored what other dyadic measures might prove informative in the QEC context. Finally, in light of prior work showing the link between non-responses and such outcomes as stock volatility (Barth et al. 2023), we highlight the role of curveball questions as an antecedent to non-responses by the management team, including the CEO.

Within the accounting and finance literature, we also add to the emerging understanding of the consequences of declines in local information availability, which is frequently examined by studying the closures of local newspapers (Abernathy 2018). The dearth of local news and information has been linked to societal-level consequences of interest such as lower voter turnout (Gentzkow, Shapiro, & Sinkinson 2011), polarized voting behavior (Darr, Hitt, & Dunaway 2018), and a rise in borrowing costs for local governments (Gao et al. 2020). Particularly to firms, research has demonstrated how newspaper closures and news deserts may potentially increase firm and facility level misconduct due to the absence of available monitors on behalf of local stakeholders (Heese et al. 2022; Jiang & Kong 2023). Our study shows a different consequence of the closure of newspapers and the subsequent decline in local information availability: the spillover effect of such an event on sell-side analysts covering firms headquartered in counties that experience such shocks. We also report evidence consistent with prior work (Choi & Valente 2023), which suggests that the degree of social connectedness in a community functions as a substitute for the local information provided by a local newspaper. In our setting too, we find that the effects of a newspaper closure attenuate in more socially connected communities.

Lastly, our paper contributes to the growing body of research that uses transformer-based machine learning models in organizational research. Whereas the prior decade saw the growing use of natural language processing methods such as word-embeddings (Mikolov, Sutskever, Chen, Corrado, & Dean 2013) and topic modeling (Blei, Ng, & Jordan 2003) to measure such constructs as firm diversification (Choi, Menon, & Tabakovic 2021), strategic differentiation (Guzman & Li 2023; Hoberg & Phillips 2016) and patent similarity/scope (Kuhn & Teodorescu 2021; Kuhn & Thompson 2019), recent years have seen transformer models gaining traction in organizational research. For example, Carlson (2023) uses sentence embeddings from the sentence transformers library (SBERT) to construct a measure that captures the degree to which a particular micro enterprise is differentiated. Vicinanza et al. (2023) train BERT models on different text corpuses, and use the decline in perplexity (or inverse-likelihood) of a given text across current and future time periods to study its prescience. Kovács et al. (2023) use BERT for a classification task, wherein books are classified into genres based on the synopses provided by the publisher. Broadly speaking, these prior studies use BERT models to measure different facets of differentiation and category alignment. Borrowing from the logic of these prior measures, we combine a form of differentiation (unexpectedness) with a form of category alignment (on-topicness) to derive a novel measure of curveball questions.

Limitations and Future Directions

Our study is not without limitations. First, our measure of curveball questions uses preceding text from a question in a QEC to classify it as being surprising or not. Although we try to validate this measure in different ways, we do not have the ground truth of which questions the management team actually found surprising and unsettling. To more thoroughly validate and potentially refine our measure, it would be useful in future research to interview management teams after a call and have them identify which questions they perceived to be curveballs.

Second, our data on newspaper closures does not indicate the exact closure date. Thus, there is some amount of measurement error in our difference-in-differences identification strategy. Although our robustness check that restricts the sample to calls that take place in the first quarter (reflecting the information environment that existed in the fourth quarter of the prior year) yields comparable results, it would be better still to have more precise closure dates.

Finally, we made the analytical choice to derive our curveball question measure based on the first question asked after the management team's presentation given that this question's potential surprise is not confounded by the surprise of other questions or the management team's response to those questions. One implication of this choice is that it is difficult to assess the economic consequences of curveball questions. Although we show that our measure is related to the tendency of the management team to give non-responses and that a call-level version of our measure is positively correlated with same-day volatility, our method would need to be adapted to more comprehensively assess whether the call as a whole represented a curveball from the perspective of management. The development of such a measure would allow for a more robust analysis of how curveball questions relate to more consequential economic outcomes.

Beyond these limitations, our study opens several pathways for future research. In this study, except for the robustness check related to star analysts, we mostly sidestep the question of how the relative status of evaluators and evaluatees might influence the likelihood of curveball questions being posed. Yet we know that status plays an important role in evaluative contexts in organizations (Russell & Fiske 2008; Bianchi, Kang, & Stewart 2012; Pearce & Xu 2012; Sharkey 2014). Future research could examine the extent to which evaluatee status might, for example, mitigate the propensity for evaluators to pose curveball questions to evaluatees or the extent to which such an effect attenuates as a function of evaluator status. In a similar vein, gender—an ascribed status characteristic—may also play a critical role in dictating discourse in the evaluative context. Recent work has called attention to such gendered patterns as lower visibility for women and a greater likelihood that they will face more aggressive questioning than men (Comprix, Lopatta, & Tideman 2022; Francis, Shohfi, & Xin 2020). We leave to future research the task of investigating how gender might play a role in the likelihood of an evaluatee being thrown curveballs.

In this study, we report evidence that star analysts are more likely to pose curveball questions than their peers who are not designated as stars. This result raises the question of what other analyst attributes might be related to the tendency to ask curveball questions. Factors that may be useful to investigate include the quality of analyst networks (e.g., the extent to which their network spans structural holes (Burt 2007)) and the extent to which they cover celebrated innovators or firms (Gouvard et al. 2023).

Finally, future research could profitably uncover the conditions under which a curveball question posed in the context of a QEC is more (or less) likely to reveal material information. The responses to some curveball questions may be benign, while the responses to others may shape such economically meaningful outcomes as stock returns and earning forecasts.

CONCLUSION

In sum, the study suggests that, in evaluative contexts, the surprises that can throw one off course have elements of both unfamiliarity and novelty (i.e., they are unexpected) but also familiarity and conventionality (i.e., they are on-topic). Moreover, when one is thrown for a curve, it can be because of macro-structural factors that are unrelated to the characteristics of actors or their interactions. Finally, we hope that the development of a novel measure of curveball questions paves the way for understanding their origins and consequences across the many types of expression games that are played in social life.

References

Abernathy PM (2018) *"The expanding news desert"* (Chapel Hill: Center for Innovation and Sustainability in local media, University of North Carolina at Chapel Hill), ISBN 978-1-4696-5324-2.

Aceves P, Evans JA (2023) Mobilizing conceptual spaces: How word embedding models can inform measurement and theory within organization science. *Organization Science* 35(3):788–814.

Alexy O, Poetz K, Puranam P, Reitzig M (2021) Adaptation or persistence? emergence and revision of organization designs in new ventures. *Organization Science* 32(6):1439–1472.

Allee KD, Cating R, Rawson C (2023) No news is bad news: local news intensity and firmsâ information environments. *Review of Accounting Studies* .

Araci D (2019) FinBERT: Financial Sentiment Analysis with Pre-trained Language Models Publisher: arXiv Version Number: 1.

Askin N, Mauskapf M (2017) What makes popular culture popular? product features and optimal differentiation in music. *American Sociological Review* 82(5):910–944.

Baker AC, Larcker DF, Wang CC (2022) How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144(2):370–395.

Barth A, Mansouri S, Woebbeking F (2023) 'let me get back to you'—a machine learning approach to measuring nonanswers. *Management Science* 69(10):6333–6348.

Bianchi AJ, Kang SM, Stewart D (2012) The organizational selection of status characteristics: Status evaluations in an open source community. *Organization Science* 23(2):341–354.

Bitterly TB, Schweitzer ME (2020) The economic and interpersonal consequences of deflecting direct questions. *Journal of Personality and Social Psychology* 118(5):945–990.

Blair-Loy M, Rogers LE, Glaser D, Wong YA, Abraham D, Cosman PC (2017) Gender in engineering departments: Are there gender differences in interruptions of academic job talks? *Social Sciences* 6(1):29.

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *The Journal of Machine Learning Research* 3:993–1022.

Botelho TL, Gertsberg M (2022) The disciplining effect of status: Evaluator status awards and observed gender bias in evaluations. *Management Science* 68(7):5311–5329.

Bourdage JS, Roulin N, Tarraf R (2018) "i (might be) just that good": Honest and deceptive impression management in employment interviews. *Personnel Psychology* 71(4):597–632.

Burt RS (2007) *Brokerage and Closure: An Introduction to Social Capital* (Oxford University Press, Oxford).

Bushee BJ, Gow ID, Taylor DJ (2018) Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research* 56(1):85–121.

Bushee BJ, Matsumoto DA, Miller GS (2004) Managerial and investor responses to disclosure regulation: The case of reg fd and conference calls. *The Accounting Review* 79(3):617–643.

Cai W, Rouen E, Zou Y (2022) Passing the mic: Career and firm outcomes of executive interactions. *Available at SSRN 3942049* .

Call AC, Emett SA, Maksymov E, Sharp NY (2022) Meet the press: Survey evidence on financial journalists as information intermediaries. *Journal of Accounting and Economics* 73(2-3):101455.

Carlson NA (2023) Differentiation in microenterprises. *Strategic Management Journal* 44(5):1141–1167.

Cattani G, Ferriani S, Allison PD (2014) Insiders, outsiders, and the struggle for consecration in cultural fields: A core-periphery perspective. *American Sociological Review* 79(2):258–281.

Chen JV, Nagar V, Schoenfeld J (2018) Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies* 23(4):1315–1354.

Choi J, Menon A, Tabakovic H (2021) Using machine learning to revisit the diversification - performance relationship. *Strategic Management Journal* 42(9):1632–1661.

Choi TJ, Valente M (2023) The crisis in local newspapers and organizational wrongdoing: The role of community social connectedness. *Organization Science* 34(5):1777–1799.

Chwe MSY (2013) *Rational ritual: Culture, coordination, and common knowledge* (Princeton University Press).

Comprix J, Lopatta K, Tideman SA (2022) The Role of Gender in the Aggressive Questioning of CEOs During Earnings Conference Calls. *The Accounting Review* 97(7):79–107.

Corritore M, Goldberg A, Srivastava SB (2020) Duality in diversity: How intrapersonal and interpersonal cultural heterogeneity relate to firm performance. *Administrative Science Quarterly* 65(2):359–394.

Darr JP, Hitt MP, Dunaway JL (2018) Newspaper Closures Polarize Voting Behavior. *Journal of Communication* 68(6):1007–1028.

Das S, Zhong R (2024) Do Analysts Learn from the News Media? Evidence from a Natural Experiment. *Working paper* .

Davis JP, Eisenhardt KM, Bingham CB (2009) Optimal structure, market dynamism, and the strategy of simple rules. *Administrative Science Quarterly* 54(3):413–452.

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186 (Minneapolis, Minnesota: Association for Computational Linguistics).

Dupas P, Modestino AS, Niederle M, Wolfers J, et al. (2021) Gender and the dynamics of economics seminars. Technical report, National Bureau of Economic Research.

Dushnitsky G, Sarkar S (2022) Here Comes the Sun: The Impact of Incidental Contextual Factors on Entrepreneurial Resource Acquisition. *Academy of Management Journal* 65(1):66–92.

Fernandez RM, Castilla EJ, Moore P (2000) Social Capital at Work: Networks and Employment at a Phone Center. *American Journal of Sociology* 105(5):1288–1356.

Francis BB, Shohfi T, Xin D (2020) Gender and Earnings Conference Calls. *SSRN Electronic Journal* .

Gao P, Lee C, Murphy D (2020) Financing dies in darkness? the impact of newspaper closures on public finance. *Journal of Financial Economics* 135(2):445–467.

Garud R, Snihur Y, Thomas LD, Phillips N (2023) The dark side of entrepreneurial framing: A process model of deception and legitimacy loss. *Academy of Management Review* .

Gentzkow M, Shapiro JM, Sinkinson M (2011) The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review* 101(7):2980–3018.

Giorgi S, Weber K (2015) Marks of distinction: Framing and audience appreciation in the context of investment advice. *Administrative Science Quarterly* 60(2):333–367.

Goffman E (1970) *Strategic Interaction*, volume 1 (University of Pennsylvania Press).

Goldberg A, Hannan MT, Kovács B (2016) What does it mean to span cultural boundaries? variety and atypicality in cultural consumption. *American Sociological Review* 81(2):215–241.

Goldman NC, Zhang Y (2022) Look whoâs talking: An analysis of market responses to executive earnings call communication. *Available at SSRN 4115713* .

Gorman EH (2005) Gender Stereotypes, Same-Gender Preferences, and Organizational Variation in the Hiring of Women: Evidence from Law Firms. *American Sociological Review* 70(4):702–728.

Gouvard P, Goldberg A, Srivastava SB (2023) Doing organizational identity: Earnings surprises and the performative atypicality premium. *Administrative Science Quarterly* 68(3):781–823.

Grootendorst M (2022) BERTopic: Neural topic modeling with a class-based TF-IDF procedure. Publisher: arXiv Version Number: 1.

Guzman J, Li A (2023) Measuring Founding Strategy. *Management Science* 69(1):101–118.

Hallen BL, Pahnke EC (2016) When do entrepreneurs accurately evaluate venture capital firmsâ track records? a bounded rationality perspective. *Academy of Management Journal* 59(5):1535–1560.

Hamilton J (2016) *Democracy's detectives: the economics of investigative journalism* (Cambridge (Mass.): Harvard university press), ISBN 978-0-674-54550-2.

Hannan MT, Freeman J (1984) Structural inertia and organizational change. *American Sociological Review* 149–164.

Heese J, Perez-Cavazos G, Peter CD (2022) When the local newspaper leaves town: The effects of local newspaper closures on corporate misconduct. *Journal of Financial Economics* 145(2):445–463.

Hoberg G, Phillips G (2016) Text-Based Network Industries and Endogenous Product Differentiation. *Journal of Political Economy* 124(5):1423–1465.

Hobson JL, Mayew WJ, Venkatachalam M (2012) Analyzing speech to detect financial misreporting. *Journal of Accounting Research* 50(2):349–392.

Hollander S, Pronk M, Roelofsen E (2010) Does Silence Speak? An Empirical Analysis of Disclosure Choices During Conference Calls. *Journal of Accounting Research* 48(3):531–563.

Hsu G (2006) Jacks of all trades and masters of none: Audiences' reactions to spanning genres in feature film production. *Administrative science quarterly* 51(3):420–450.

Huang L, Pearce JL (2015) Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. *Administrative Science Quarterly* 60(4):634–670.

Jiang JX, Kong J (2023) Green dies in darkness? environmental externalities of newspaper closures. *Review of Accounting Studies* .

Kanze D, Huang L, Conley MA, Higgins ET (2018) We Ask Men to Win and Women Not to Lose: Closing the Gender Gap in Startup Funding. *Academy of Management Journal* 61(2):586–614.

Kim M, Stice D, Stice H, White RM (2021) Stop the presses! or wait, we might need them: Firm responses to local newspaper closures and layoffs. *Journal of Corporate Finance* 69:102035.

Kovács B, Carroll GR, Lehman DW (2014) Authenticity and consumer value ratings: Empirical tests from the restaurant domain. *Organization Science* 25(2):458–478.

Kovács B, Hsu G, Sharkey A (2023) The stickiness of category labels: Audience perception and evaluation of producer repositioning in creative markets. *Management Science* .

Kovács B, Negro G, Pólos L, Pontikes E, Sharkey AJV, Le Mens G, Hsu G (2019) *Concepts and Categories: Foundations for Sociological and Cultural Analysis* (Columbia University Press).

Kovács B, Sharkey AJ (2014) The paradox of publicity: How awards can negatively affect the evaluation of quality. *Administrative Science Quarterly* 59(1):1–33.

Kuhn JM, Teodorescu MHM (2021) The Track One Pilot Program: Who benefits from prioritized patent examination? *Strategic Entrepreneurship Journal* 15(2):185–208.

Kuhn JM, Thompson NC (2019) How to Measure and Draw Causal Inferences with Patent Scope. *International Journal of the Economics of Business* 26(1):5–38.

Lamont M (2012) Toward a comparative sociology of valuation and evaluation. *Annual Review of Sociology* 38:201–221.

Larcker DF, Zakolyukina AA (2012) Detecting deceptive discussions in conference calls. *Journal of Accounting Research* 50(2):495–540.

Leung MD (2018) Learning to hire? hiring as a dynamic experiential learning process in an online market for contract labor. *Management Science* 64(12):5651–5668.

Li F, Minnis M, Nagar V, Rajan M (2014) Knowledge, compensation, and firm value: An empirical analysis of firm communication. *Journal of Accounting and Economics* 58(1):96–116.

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

Matsumoto D, Pronk M, Roelofsen E (2011) What makes conference calls useful? the information content of managers' presentations and analysts' discussion sessions. *The Accounting Review* 86(4):1383–1414.

Mayew WJ, Sethuraman M, Venkatachalam M (2020) Individual Analysts' Stock Recommendations, Earnings Forecasts, and the Informativeness of Conference Call Question and Answer Sessions. *The Accounting Review* 95(6):311–337.

Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed Representations of Words and Phrases and their Compositionality. Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, eds., *Advances in Neural Information Processing Systems*, volume 26 (Curran Associates, Inc.).

Miller A, Lall SA, Goldstein M, Montalvao J (2023) Asking Better Questions: The Effect of Changing Investment Organizations' Evaluation Practices on Gender Disparities in Funding Innovation. Technical report, The World Bank.

Minson JA, VanEpps EM, Yip JA, Schweitzer ME (2018) Eliciting the truth, the whole truth, and nothing but the truth: The effect of question phrasing on deception. *Organizational Behavior and Human Decision Processes* 147:76–93.

Negro G, Leung MD (2013) Actual and perceptual effects of category spanning. *Organization Science* 24(3):684–696.

Pager D (2003) The Mark of a Criminal Record. *American Journal of Sociology* 108(5):937–975.

Pearce JL, Xu QJ (2012) Rating performance or contesting status: Evidence against the homophily explanation for supervisor demographic skew in performance ratings. *Organization Science* 23(2):373–385.

Pontikes EG (2012) Two sides of the same coin: How ambiguous classification affects multiple audiences evaluations. *Administrative Science Quarterly* 57(1):81–118.

Reimers N, Gurevych I (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks Publisher: arXiv Version Number: 1.

Rivera LA (2012) Hiring as cultural matching: The case of elite professional service firms. *American Sociological Review* 77(6):999–1022.

Rivera LA (2015) Go with Your Gut: Emotion and Evaluation in Job Interviews. *American Journal of Sociology* 120(5):1339–1389.

Rogers T, Norton MI (2011) The artful dodger: Answering the wrong question the right way. *Journal of Experimental Psychology: Applied* 17(2):139.

Rupasingha A, Goetz SJ, Freshwater D (2006) The production of social capital in us counties. *The Journal of Socio-economics* 35(1):83–101.

Russell AMT, Fiske ST (2008) It's all relative: Competition and status drive interpersonal perception. *European Journal of Social Psychology* 38(7):1193–1201.

Sharkey AJ (2014) Categories and organizational status: The role of industry status in the response to organizational deviance. *American Journal of Sociology* 119(5):1380–1433.

Shi Y (2023) A change of tune: The democratization of market mediation and crossover production in the us commercial music industry. *Administrative Science Quarterly* 68(2):319–354.

Sternberg KJ, Lamb ME, Hershkowitz I, Esplin PW, Redlich A, Sunshine N (1996) The relation between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developmental Psychology* 17(3):439–451.

Stuart TE, Hoang H, Hybels RC (1999) Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly* 44(2):315–349.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Å, Polosukhin I (2017) Attention is All you Need. Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, eds., *Advances in Neural Information Processing Systems*, volume 30 (Curran Associates, Inc.).

Verhaal JC, Khessina OM, Dobrev SD (2015) Oppositional product names, organizational identities, and product appeal. *Organization Science* 26(5):1466–1484.

Vicinanza P, Goldberg A, Srivastava SB (2023) A deep-learning model of prescient ideas demonstrates that they emerge from the periphery. *PNAS nexus* 2(1):pgac275.

Vukovic DB, Kurbonov OOU, Maiti M, Ãzer M, Radovanovic M (2024) Outperforming the market: a comparison of Star and NonStar analystsâ investment strategies and recommendations. *Humanities and Social Sciences Communications* 11(1):16.

Younkin P, Kuppuswamy V (2018) The Colorblind Crowd? Founder Race and Performance in Crowdfunding. *Management Science* 64(7):3269–3287.

Zott C, Huy QN (2007) How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly* 52(1):70–105.

Zuckerman EW (2012) Construction, concentration, and (dis) continuities in social valuations. *Annual Review of Sociology* 38:223–245.

TABLES AND FIGURES

Table 1: Key Variables and Summary Statistics

Variable	N	Mean	S. Dev.
Question-level Measures			
Off-topic	126,910	0.817	0.089
Unexpectedness	126,910	0.923	1.424
Curveball Question Score	126,910	21.039	16.967
Answer-level Measures			
Non-Answers in Response	111,306	0.093	0.090
Answer Length	111,306	4.792	1.192
CEO Answer	72,761	0.508	0.500
Non-Answers in CEO Response	72,735	0.056	0.091
Star Analyst Question	126,665	0.064	0.245
Firm-level Variables			
Prior Q. Trading Vol	117,928	0.216	0.442
Prior Q. Stock Return	117,624	1.043	0.313
Firm Size	125,506	7.241	2.060
Capital Int.	125,503	0.038	0.048
Research Int.	125,503	0.050	0.117
Cash Ratio	125,503	0.187	0.222
Leverage	124,885	0.891	1.930
Firm RoA	125,481	-0.033	0.284
Tobin's Q	122,404	2.154	2.126
Call-day Return	116,358	0.001	0.067
Call-day Volatility	116,365	0.028	0.018
Call-day Abnormal Return	116,341	0.000	0.065
Call-day Idiosync. Volatility	116,365	0.024	0.017
County-level Variables (for DiD Sample)			
County Unemp. Rate	115,441	5.740	2.230
County Med. Income	115,252	71,420.230	20,714.94
County Social Capital	115,252	-0.426	0.627
Total Observations	126,910		

Table 2: Validation of Curveball Question Score

	(1) Non-Answers	(2) Answer Length	(3) CEO Answer	(4) CEO Non-answers	(5) Ques. Curveb. Score	(6) Stock Volatility
Quest. Curveball Score	0.010**	0.076***	0.142***	0.033***		
	(2.68)	(22.16)	(14.59)	(7.61)		
Star Analyst Question					0.060**	
					(3.18)	
Call-level Curveball						0.005*
						(2.02)
Prior Q. Trading Vol	0.001	-0.002	-0.008	-0.003	0.003	0.226***
	(0.32)	(-0.35)	(-0.45)	(-0.68)	(0.49)	(7.86)
Prior Q. Stock Ret.	0.006	0.005	0.024	0.016*	0.000	0.032***
	(1.47)	(1.44)	(1.61)	(2.06)	(0.04)	(4.47)
Firm Size	0.098***	0.110***	-0.119	0.028	0.055	-0.326***
	(4.81)	(4.74)	(-1.91)	(0.89)	(1.56)	(-10.99)
Capital Int.	0.010	0.023**	-0.037	-0.002	-0.014	-0.064***
	(1.39)	(2.84)	(-1.68)	(-0.16)	(-1.30)	(-8.19)
Research Int.	0.005	0.003	-0.011	-0.011	0.000	-0.030
	(0.54)	(0.24)	(-0.22)	(-0.40)	(0.01)	(-1.47)
Cash Ratio	0.006	0.010	-0.039	-0.023	-0.010	-0.021
	(0.58)	(0.99)	(-1.28)	(-1.49)	(-0.71)	(-1.45)
Leverage	0.002	0.007	-0.026	-0.003	0.001	0.022**
	(0.35)	(1.30)	(-1.95)	(-0.53)	(0.14)	(3.03)
Firm RoA	-0.001	0.016	-0.040	-0.025	-0.005	-0.135***
	(-0.10)	(1.74)	(-1.21)	(-1.34)	(-0.61)	(-6.88)
Tobin's Q	0.002	0.016*	0.029	0.003	0.015	-0.054***
	(0.43)	(2.53)	(1.49)	(0.36)	(1.44)	(-5.47)
Constant	0.002***	0.000		-0.012	0.017***	-0.000
	(6.10)	(0.50)		(-0.95)	(6.41)	(-0.02)
Observations	100659	100659	64393	67416	65291	108762
R-Squared	0.080	0.172		0.118	0.220	0.724
LR Chi-sq			2264.785			
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time(Quarter) F.E.	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Newspaper Closures and Stock Outcomes on the Call Day

	(1)	(2)	(3)	(4)
	Return	Ab. Return	Stock Volatility	Idio. Volatility
Treated * Post	-0.004	-0.009	0.074**	0.062**
	(-0.24)	(-0.61)	(3.12)	(3.10)
Prior Q. Trading Vol	0.005	0.006	0.236***	0.245***
	(0.70)	(1.00)	(7.05)	(7.10)
Prior Q. Stock Return	-0.017*	-0.018*	0.037***	0.042***
	(-2.40)	(-2.56)	(4.03)	(4.39)
Firm Size	-0.161***	-0.165***	-0.299***	-0.403***
	(-5.06)	(-4.97)	(-10.36)	(-14.08)
Capital Int.	-0.007	-0.009	-0.062***	-0.063***
	(-0.89)	(-1.18)	(-5.62)	(-5.88)
Research Int.	-0.003	-0.001	-0.024	-0.026
	(-0.23)	(-0.09)	(-1.13)	(-1.19)
Cash Ratio	-0.006	-0.011	-0.025	-0.035*
	(-0.67)	(-1.20)	(-1.59)	(-2.22)
Leverage	-0.004	-0.005	0.020**	0.022**
	(-0.60)	(-0.91)	(2.87)	(3.09)
Firm RoA	-0.016	-0.011	-0.125***	-0.125***
	(-1.79)	(-1.24)	(-5.61)	(-5.62)
Tobin's Q	-0.043***	-0.043***	-0.049***	-0.065***
	(-4.75)	(-4.66)	(-5.13)	(-7.05)
County Median Income	0.031	0.032	-0.103**	-0.086**
	(1.86)	(1.72)	(-2.98)	(-2.62)
County Unemployment Rate	0.026*	0.019	0.027	0.023
	(2.41)	(1.89)	(1.68)	(1.48)
Constant	0.003**	0.003**	-0.002	-0.001
	(2.59)	(2.68)	(-1.34)	(-0.44)
Observations	99260	99251	99266	99266
R-Squared	0.060	0.057	0.730	0.727
Firm, Time F.E.	Yes	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: Newspaper Closures and Curveball Question Score

	(1)	(2)
	Curveball Question Score	
Treated * Post	0.060**	0.058**
	(2.61)	(2.65)
Prior Q. Trading Vol		0.000
		(0.17)
Prior Q. Stock Return		-0.001
		(-0.17)
Firm Size		0.042
		(1.53)
Capital Int.		-0.007
		(-0.83)
Research Int.		-0.005
		(-0.42)
Cash Ratio		-0.010
		(-1.00)
Leverage		-0.002
		(-0.31)
Firm RoA		-0.006
		(-1.09)
Tobin's Q		0.015**
		(3.03)
County Median Income		-0.008
		(-0.32)
County Unemployment Rate		-0.007
		(-0.64)
Constant	-0.002*	0.001
	(-2.22)	(1.00)
Observations	115246	104783
R-Squared	0.199	0.202
Controls	No	Yes
Firm F.E.	Yes	Yes
Time(Quarter) F.E.	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Newspaper Closure & Curveballs: Mechanism and Robustness Tests

	(1)	(2)	(3)	(4)
	Curveball Question Score			
Treated * Post	0.058**	0.051*	0.062**	0.085*
	(2.65)	(2.28)	(2.71)	(2.44)
Treated * Post * Social Capital		-0.016*		
		(-1.97)		
Firm Size	0.042	0.042	0.040	0.024
	(1.53)	(1.52)	(1.44)	(0.58)
Treated * Post * Firm Size			0.069**	
			(3.07)	
Prior Q. Trading Vol	0.000	0.000	0.000	-0.007
	(0.17)	(0.17)	(0.17)	(-1.07)
Prior Q. Stock Return	-0.001	-0.001	-0.001	-0.000
	(-0.17)	(-0.16)	(-0.16)	(-0.02)
Capital Int.	-0.007	-0.007	-0.007	-0.000
	(-0.83)	(-0.85)	(-0.86)	(-0.02)
Research Int.	-0.005	-0.005	-0.005	0.012
	(-0.42)	(-0.42)	(-0.43)	(0.85)
Cash Ratio	-0.010	-0.010	-0.010	-0.002
	(-1.00)	(-1.01)	(-1.04)	(-0.13)
Leverage	-0.002	-0.002	-0.002	-0.005
	(-0.31)	(-0.31)	(-0.33)	(-0.60)
Firm RoA	-0.006	-0.006	-0.006	0.005
	(-1.09)	(-1.08)	(-1.08)	(0.35)
Tobin's Q	0.015**	0.015**	0.015**	0.012
	(3.03)	(3.04)	(3.03)	(1.25)
County Median Income	-0.008	-0.007	-0.006	0.039
	(-0.32)	(-0.28)	(-0.23)	(1.01)
County Unemployment Rate	-0.007	-0.007	-0.007	0.004
	(-0.64)	(-0.63)	(-0.68)	(0.21)
Constant	0.001	0.001	0.001	0.000
	(1.00)	(0.95)	(1.14)	(0.10)
Observations	104783	104783	104783	25889
R-Squared	0.202	0.202	0.202	0.285
Firm F.E.	Yes	Yes	Yes	Yes
Time(Quarter) F.E.	Yes	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 1: A Baseball-Inspired Framework of Surprising Questions in Expression Games

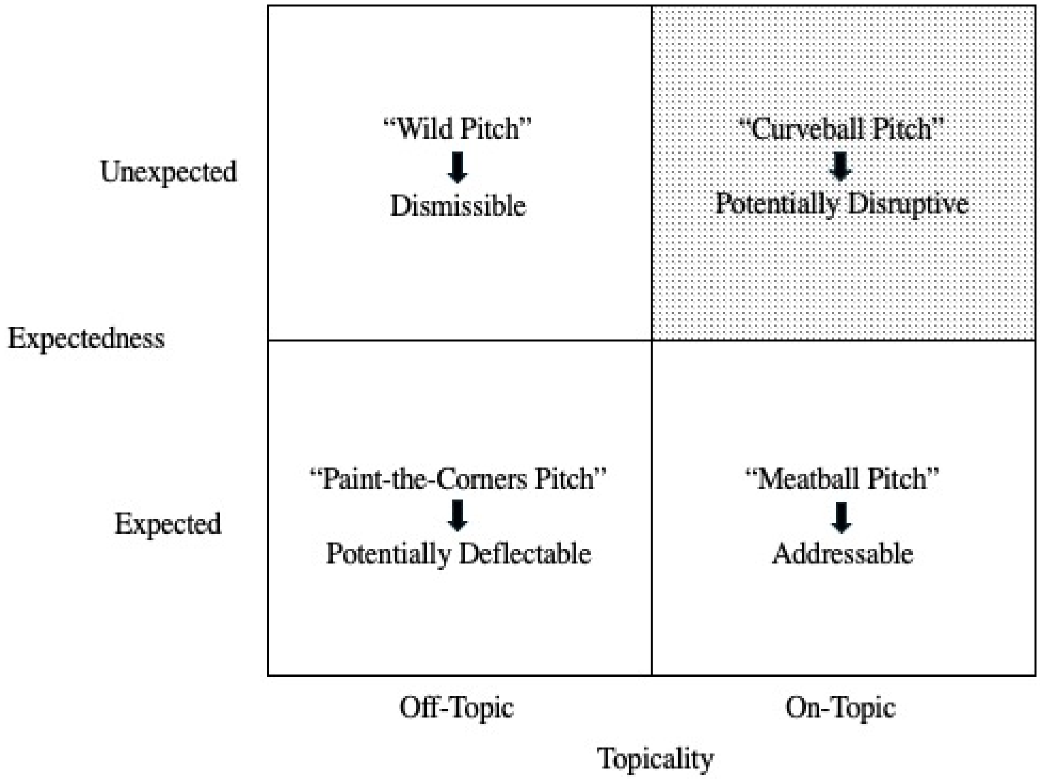


Figure 2: Difference-in-Differences Estimate: Marginal Effects by Years to Treatment

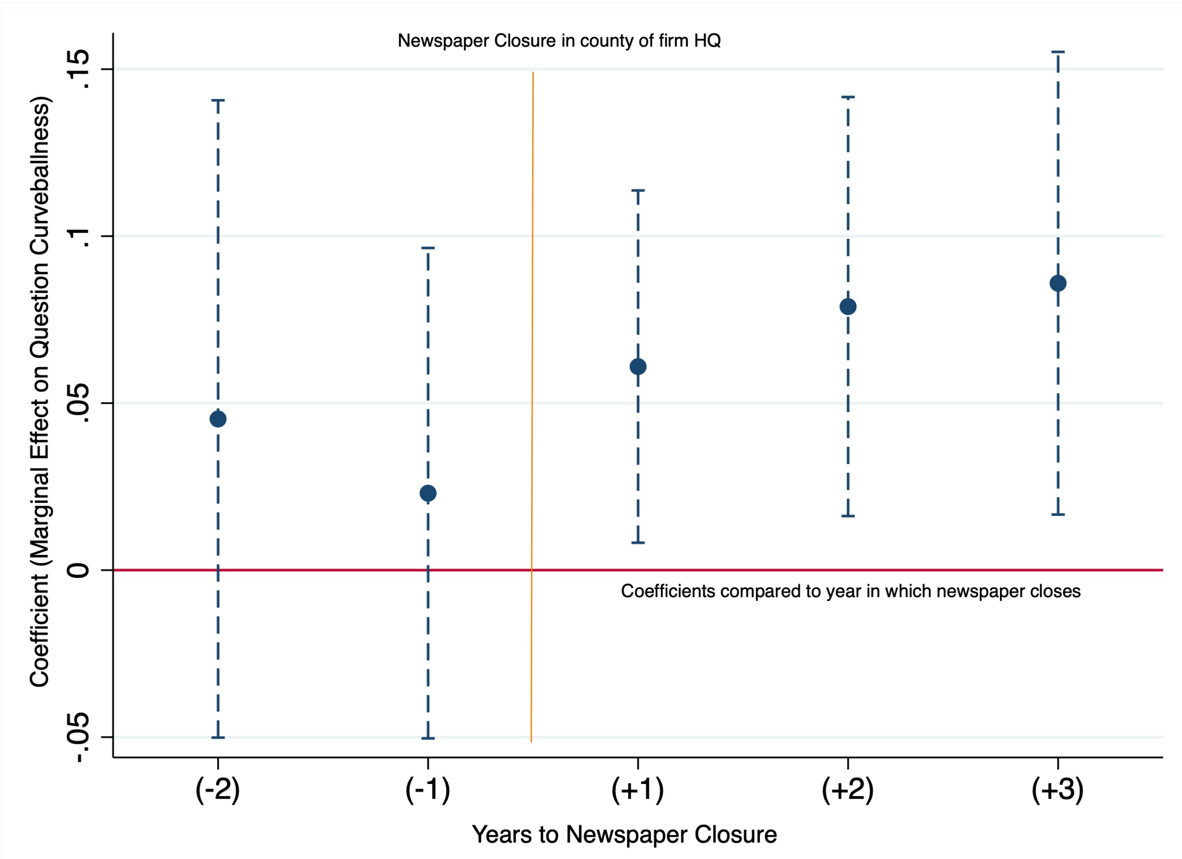
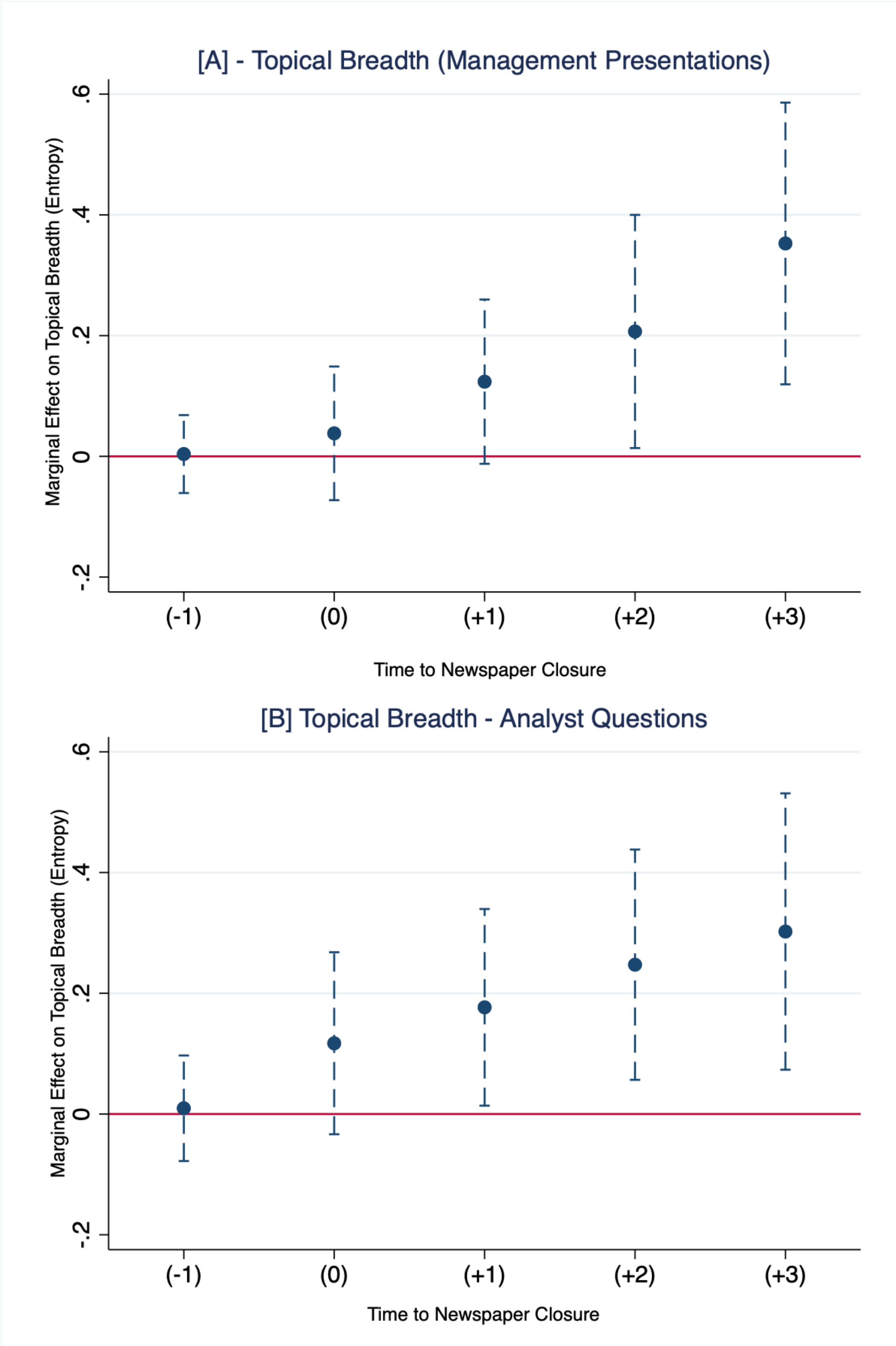


Figure 3: Topical Breadth in Treated Firms post Newspaper Closures



APPENDIX

Table A.1: Interaction Between Unexpectedness and Off-topic Scores

	(1) Non-Answers	(2) Answer Length	(3) CEO Answer
Unexpectedness	0.009	0.048***	0.097***
	(1.74)	(9.20)	(7.08)
Off-topic	-0.007	-0.117***	-0.163***
	(-1.55)	(-27.33)	(-14.77)
Unexpectedness * Off-topic	-0.014***	-0.025***	-0.035***
	(-3.75)	(-6.28)	(-3.45)
Prior Q. Trading Vol	0.001	-0.002	-0.006
	(0.33)	(-0.30)	(-0.36)
Prior Q. Stock Return	0.006	0.005	0.025
	(1.47)	(1.32)	(1.68)
Firm Size	0.098***	0.104***	-0.130*
	(4.82)	(4.50)	(-2.08)
Capital Int.	0.010	0.021**	-0.039
	(1.39)	(2.60)	(-1.78)
Research Int.	0.005	0.002	-0.021
	(0.53)	(0.16)	(-0.42)
Cash Ratio	0.006	0.010	-0.040
	(0.56)	(0.93)	(-1.32)
Leverage	0.002	0.007	-0.026
	(0.34)	(1.31)	(-1.90)
Firm RoA	-0.001	0.016	-0.039
	(-0.11)	(1.68)	(-1.20)
Tobin's Q	0.002	0.015*	0.029
	(0.44)	(2.43)	(1.46)
Constant	0.006***	0.009***	
	(4.99)	(6.74)	
Observations	100,659	100,659	64,393
R-Squared	0.080	0.176	
LR Chi-sq			2274.706
Firm F.E.	Yes	Yes	Yes
Time(Quarter) F.E.	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Newspaper Closures and the Underlying Dimensions of Curveball Questions

	(1)	(2)	(3)	(4)	(5)	(6)
	Curveball Question Score		Unexpectedness		Off-topic	
Treated * Post	0.060**	0.058**	0.037*	0.042*	0.011	0.033
	(2.61)	(2.65)	(2.47)	(2.47)	(0.48)	(1.45)
Prior Q. Trading Vol		0.000		-0.000		0.003
		(0.17)		(-0.07)		(1.02)
Prior Q. Stock Return		-0.001		-0.003		-0.003
		(-0.17)		(-1.06)		(-0.99)
Firm Size		0.042		-0.021		-0.073**
		(1.53)		(-0.85)		(-2.93)
Capital Int.		-0.007		-0.014		-0.021**
		(-0.83)		(-1.85)		(-2.87)
Research Int.		-0.005		0.001		0.002
		(-0.42)		(0.07)		(0.14)
Cash Ratio		-0.010		0.002		0.002
		(-1.00)		(0.15)		(0.19)
Leverage		-0.002		0.000		0.006
		(-0.31)		(0.07)		(1.19)
Firm RoA		-0.006		-0.002		-0.001
		(-1.09)		(-0.47)		(-0.14)
Tobin's Q		0.015**		0.001		-0.016**
		(3.03)		(0.13)		(-2.78)
County Med. Income		-0.008		-0.061**		-0.078**
		(-0.32)		(-2.68)		(-2.62)
County Unemp. Rate		-0.007		0.009		0.008
		(-0.64)		(0.75)		(0.72)
Constant	-0.002*	0.001	-0.001	-0.004***	0.004***	-0.004***
	(-2.22)	(1.00)	(-1.40)	(-4.78)	(3.58)	(-3.69)
Observations	115246	104783	115246	104783	115246	104783
R-Squared	0.199	0.202	0.213	0.210	0.300	0.300
Controls	No	Yes	No	Yes	No	Yes
Firm F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Time(Quarter) F.E.	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3: Newspaper Closures and Curveball Questions beyond First Question

	(1) First Ques. Curveball Score	(2) First 2 Ques. Curveball Score	(3) First 3 Ques. Curveball Score
Treated * Post	0.058**	0.056*	0.051
	(2.65)	(2.28)	(1.74)
Prior Q. Trading Vol	0.000	0.001	-0.000
	(0.17)	(0.32)	(-0.04)
Prior Q. Stock Return	-0.001	-0.002	-0.003
	(-0.17)	(-0.58)	(-0.97)
Firm Size	0.042	0.024	0.028
	(1.53)	(0.87)	(1.00)
Capital Int.	-0.007	-0.004	0.001
	(-0.83)	(-0.47)	(0.18)
Research Int.	-0.005	-0.008	-0.003
	(-0.42)	(-0.64)	(-0.20)
Cash Ratio	-0.010	-0.025*	-0.027**
	(-1.00)	(-2.51)	(-2.66)
Leverage	-0.002	-0.002	-0.002
	(-0.31)	(-0.42)	(-0.37)
Firm RoA	-0.006	-0.005	-0.003
	(-1.09)	(-0.70)	(-0.36)
Tobin's Q	0.015**	0.019***	0.019**
	(3.03)	(3.56)	(3.28)
County Median Income	-0.008	-0.009	-0.002
	(-0.32)	(-0.29)	(-0.06)
County Unemployment Rate	-0.007	-0.004	-0.004
	(-0.64)	(-0.31)	(-0.27)
Constant	0.001	0.002	0.003*
	(1.00)	(1.44)	(2.28)
Observations	104783	104783	104783
R-Squared	0.202	0.274	0.308
Controls	Yes	Yes	Yes
Firm F.E.	Yes	Yes	Yes
Time(Quarter) F.E.	Yes	Yes	Yes

t statistics in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A.1: Distribution of Question Scores: Off-topic, Unexpectedness, and Curveball

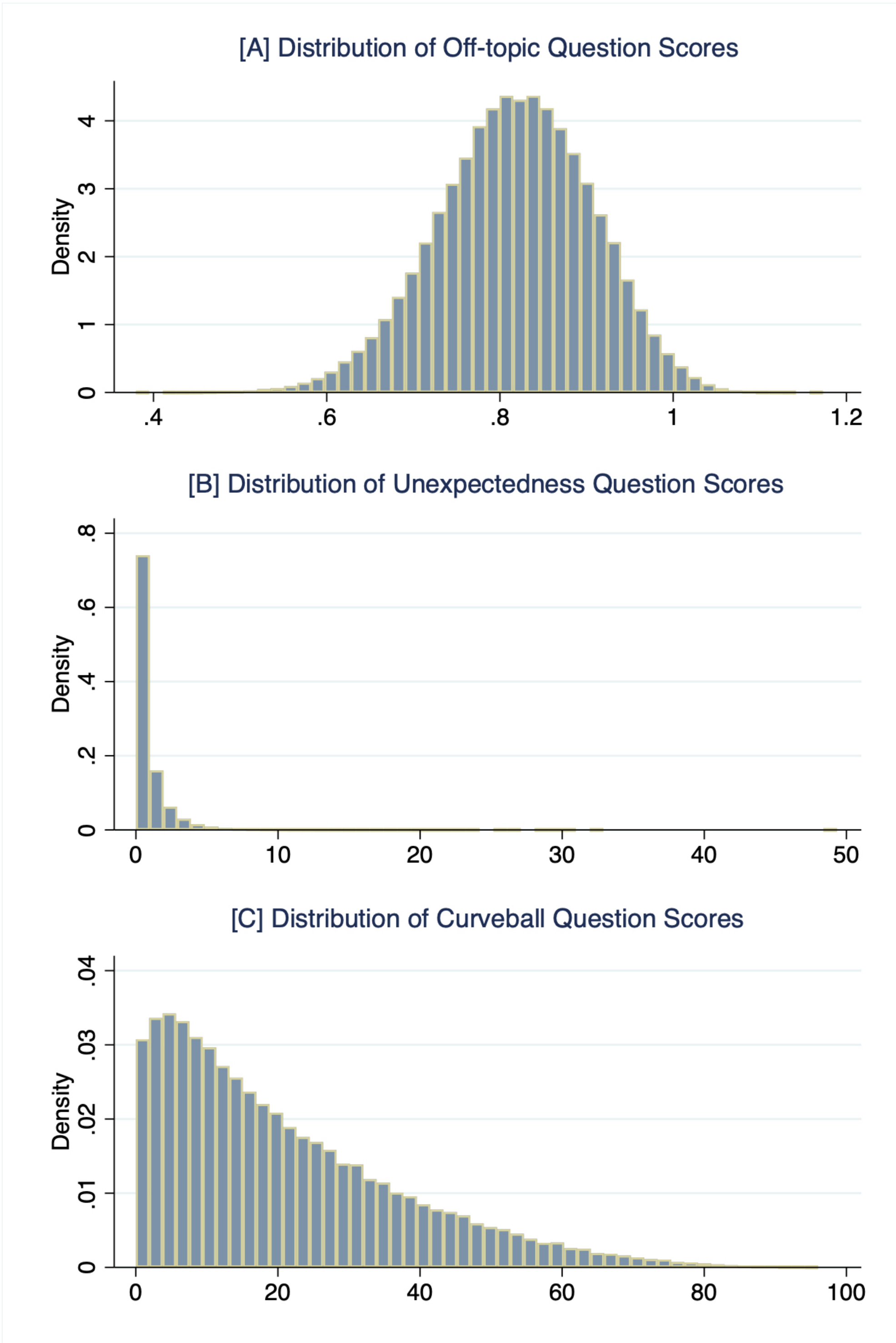


Figure A.2: Examples of Curveball Questions - Apple and Netflix

Apple – Low Curveball Questions	Tim and Peter, can you describe the iPhone momentum as you ended the quarter? And what is included for the iPhone in your guidance and your enthusiasm for the product as we head into the first half of the calendar year? Any more detail in metrics would be helpful.
	When we look at the effect of currency on your gross profit, which I think was far less than anyone anticipated, have you been able to reduce some of the pressure by adjusting terms of your suppliers? Or is that something we should expect to see more as you progress through the year?
	My first question is with regard to Apple Pay, which we actually used this afternoon. And where it was rolled out, it actually worked pretty well, which is nice. But could you provide some more color, Tim and Luca, on how you see the business model for Apple? And maybe from a high-level perspective, if you see this more sort of over time becoming a stand-alone business sort of like you look at iTunes? Or is this more just a way to sell incremental product?
	I have a question for you with regard to availability of iPhones, especially with the Verizon launch. I think you mentioned that there was a backlog that you weren't able to meet in the December quarter. So I don't know if, Tim and Peter, if you could talk a little bit about what you're doing to ensure that you have sufficient quantities of iPhones and any changes you've made to ensure more component availability?
	Okay, very helpful. Tim, the press and even some of your competitors are quite complimentary of the innovation in iPhone 5s, but it seems like investors are looking or glancing over the 64-bit ARM processor, the fingerprint sensors, the M7 processor, and part of this is that they're not exactly sure what those technologies enable, i.e. why do you need a 64-bit ARM processor in a phone? So just hoping you could enlighten us a little bit on that question.
Apple – High Curveball Questions	Fair enough. That's really helpful. And I guess if I just follow up on China, impressed to see the continued recovery you guys are seeing there despite all the headlines that are out there. Just curious, what are the few things that are driving the success in China? And how sustainable do you think those changes are for Apple as you go forward?
	Okay, great. Can you walk through how you're thinking about the supply ramp for the iPhone 5 in the holiday quarter? And how does some of the challenges you're facing relative to the strong demand compare to past iOS product launches?
	I'll ask both my questions upfront. First, for Tim, you're on track to hit your Services revenue target even earlier than planned. So how are you thinking about the next legs of Services growth as you move into the next 3 to 5 years? And then for you Luca, NAND prices are falling this year. Services mix is rising. Those should both positively influence gross margins. And yet, we're seeing gross margin sort of hang out here at 38%. What are the offsetting headwinds? And is it possible that we could see the tailwinds start to overpower those headwinds in the next couple of quarters and see gross margins drift higher?
	The 12% revenue downtick, Peter, in September is much more conservative than your typical September guidance and what you end up reporting in September, which is usually a 20% plus increase for the last 4 years. Can you help us understand why you expect this next quarter to trend softer than seasonal and maybe how you handicap the impact from some of the new software services and products that you expect to launch in the quarter?
	Congratulations on a really strong quarter. First question for Luca. The gross margin was particularly strong versus your outlook. Can you talk about whether you recognize the full impact of the weaker dollar in the December quarter given your typical currency hedges? And then how are you thinking about the headwinds and tailwinds on gross margins as you go into the March quarter? And then I have a follow-up for Tim.
Netflix – Low Curveball Questions	Makes sense. I definitely want to talk a little bit about the pricing here in a minute. But before we move on to that, you talked about broad-based basically global growth. Can I ask specifically about a couple of geographies? I understand broad-based means them all, but Asia has been one particular gigantic region, I know, a particular focus for your company. Anything you'd want to point out specifically in Asia in the quarter, or more broadly over the course of those couple quarters that are going well or not well that we should know about?
	I'll start out maybe for Reed and the team. Let's - can you reflect on the fourth quarter results for us that we're all going through right now? In particular, talk about the international strength. You mentioned you were pleased with the October -- September, October launches. So can we infer that the outperformance versus your expectation maybe came from those areas? Or any color you can give us on the international strength, to start us off?
	Actually first question for you, David, and then one for Reed. David, could you just talk a little bit about the net add results in the quarter versus expectations? And any dynamics underlying the second quarter guidance that you want investors to know about?
	So thanks for having me here. And I guess the best place to start here is Ted and Greg, congratulations on your new role, and Reed, too, I guess you can relax a little bit more. So maybe you could just start off with your priorities, Ted, and may be followed by Greg, just in terms of how you see the world evolving, what your priorities are. And of course, we are in the middle of a lot of change. So how you see the world? So maybe you could just start there.
	Okay. First question is for Reed. This was maybe an inflection point quarter in terms of the domestic streaming sub adds that came in materially ahead of your guidance. They're actually up a little bit year-over-year, which is a bit of a surprise. In the press release, you talked about both getting more subs in and retaining the better maybe than you= had expected. Could you provide a little bit more color? And is this something that you think is happening across the industry, the greater focus on streaming offerings is just helping the leader in the market, or is there something specific that you're doing on both of those ends, the gross sub adds and the churn?
Netflix – High Curveball Questions	As the usage model switches more and more towards TV viewing, could you talk about the impact that could have on the business model in terms of margins, visibility into revenue and into profits, and maybe any pricing options that, that would give you?
	Ted, takeaways for you from the third quarter. What surprised you? What was better? What was worse than expected?
	Reed, just following up on HBO GO and the competitive dynamic there, do you have a view on how it could be priced and distributed when it ultimately comes out?
	Reed, you just mentioned that one of the big reasons for churn is people not having any money. So I wonder if you would consider that if extreme picks, for example, of $4.99 and let's say Amazon comes out with a carve-out prime streaming service at something lower than $7.99, have you thought about or would you consider adjusting your price point? And then the second question would be, I think Mark asked about usage on tablets. And I would just ask, has the subscriber acquisition channel changed with the explosion in mobile? Is that something that you can comment on in terms of where you're getting gross add connects from? That would be great as well.
	But even -- because the stand-alone subscribers I think actually we're up in Q4. And so now you actually think that now that the shift from hybrid to stand-alone DVD had subsided, the net effect will be stand-alone DVD-only subscribers going down quarter after quarter?

Figure A.3: Interaction between Unexpectedness and Off-topic - Validation