Congruent Photographs and Text in Annual Reports

Sudipta Basu

sudipta.basu@temple.edu

Ruikai Ji

ruikai.ji@temple.edu

Xueming Luo

luoxm@temple.edu

Fox School of Business

Temple University

June 2025

We appreciate helpful comments from Dmitri Byzalov, Leting Liu (discussant), Samuel Rosen, Wei Wang, Forester Wong (discussant), Shijia Wu, Enshuai Yu (discussant), workshop participants at Temple University, Fox School Brown Bag, 2024 *CJAR* Summer Research Workshop, 2024 Haskayne and Fox Accounting Conference, 2024 AAA Spark Meeting, and 2023 Conference on Artificial Intelligence, Machine Learning, and Business Analytics.

Congruent Photographs and Text in Annual Reports

Abstract

We investigate management's strategic use of photographs and text in annual reports. Communications research shows that when text and adjacent photographs share similar meanings (are congruent), they improve the understanding and memorability of the focal context. By employing the deep-learning Bootstrapping Language-Image Pre-Training (BLIP) model to measure photo-text congruence, we document that firms that are performing poorly tend to enhance the photo-text congruence in annual reports, suggesting an emphasis on favorable data. This strategy appears to mitigate the negative impact brought by the bad performance, as evidenced by higher shareholder supporting rates for director elections and say-on-pay proposals at the next annual meeting. The results are more significant when these poorly performing firms have low active institutional ownership or higher operating, especially R&D, expenses. Meanwhile, the high photo-text congruence is not correlated with better future firm performance. Our findings suggest that while enhancing photo-text congruence of favorable data can benefit management, it can convey more comprehensive yet less valuable data that distracts shareholders.

Keywords: Strategic disclosure; shareholder votes; information processing costs; unstructured data
JEL Classification: M40, M41, G14, G17, D83

1. Introduction

We investigate whether management strategically combines text and photographs in annual reports to emphasize favorable or obfuscate unfavorable data. The financial and non-financial data in annual reports can help shareholders make important performance-based decisions like voting on executive compensation and director (re)appointments. When firm performance is poor, management usually receives lower voting support from shareholders (e.g., Cai, Garner, and Walkling, 2009). We examine whether the use of a particular disclosure strategy—namely, the congruence between text and adjacent photographs in annual reports, defined as the extent to which their meanings align—is common among firms with bad performance. For these firms, we examine whether the photo-text congruence in annual reports is associated with shareholder voting outcomes at annual meetings and with these firms' future performance. Although prior papers have studied the content and tone of firms' textual disclosures in SEC filings, we study whether and how the color photographs in glossy annual reports affect shareholders' voting decisions. Compared with early quantitative papers that mostly study just one communication mode (e.g., text), we quantitatively explore the interaction between the contents of multiple communication modes, i.e., photographs and text, in firms' annual reports.

The annual report format is discretionary and ranges from a single Form 10-K (e.g., Apple Inc.) to an integrated report combining an opening section, a Form 10-K, and a proxy statement (e.g., HP Inc.). Some firms issue a standalone annual report (e.g., Hilton Worldwide Holdings Inc.) that summarizes the previous fiscal year's performance without including the Form 10-K or proxy statement. The format and content in the voluntarily disclosed section (either the whole standalone report or the opening section in an integrated report) vary a lot across firms, which makes it challenging for researchers to examine the cause and impact of firms' potential disclosure strategies in this section. Investors, with limited attention and resources, may not analyze 10-Ks in detail (e.g., Kumar and Lee, 2006), but rely on less complex corporate disclosures, such as the glossy opening section of the annual report, to make investment decisions. By creating sections beyond SEC requirements, firms may seek to convey more understandable data for better reader engagement and easier information processing;[1] or they may obfuscate information by including misleading or irrelevant contents.

In annual reports, the most frequently used communication mode after text and tables is visual data. In our sample, 63.7% of firm-year observations have photographs in their annual reports. Although the visual materials used in the voluntarily disclosed section of annual reports should not contain data materially different from that in regulatory filings such as 10-Ks according to Regulation S-T Rule 304, regulators allow some subtle differences, such as visual data in annual reports that cannot be easily perceived in the plain text 10-Ks (Deng, Gao, Hu, and Zhou, 2023). Thus, firms may reduce the processing costs of favorable data by using visual data to attract attention and emphasize certain topics (Blankespoor, deHaan, and Marinovic, 2020; Deng et al., 2023; Ben-Rephael, Ronen, Ronen, and Zhou, 2023). Ben-Rephael et al. (2023) provide evidence that analysts read annual reports to understand firms' priorities and what they want to promote. Compared to the mandatory 10-Ks and proxy statements, which are carefully prepared to deter lawsuits, the voluntarily disclosed sections use more understandable language and user-friendly formats to convey messages. Even though these sections repeat some data disclosed in 10-Ks, the information processing costs could be lower. Moreover, firms can justify negative financial data. For example, a firm's R&D expense might be high in a given year, lowering profitability. The firm can frame this spending as an investment that will benefit shareholders in the long term, so that

[1] This *Wall Street Journal* article provides examples of how European firms are trying to produce more user-friendly corporate reporting: https://www.wsj.com/articles/companies-find-ways-to-keep-their-annual-reports-from-being-a-bore-11583231402.

shareholders perceive the high R&D expense as a positive signal. However, management may have an incentive to extract resources by misleading investors (Dechow, Ge, Larson, and Sloan, 2011), and if the management is overoptimistic about the R&D investment outcome, it might garner increased shareholder support temporarily but hurt long-term shareholder interests. When firm performance is poor, management may also try to obfuscate data by disclosing complex contents in annual reports because the market reacts slowly to complicated disclosures (Bloomfield, 2002; Li, 2008). Overall, the voluntarily disclosed section can serve as an additional communication channel.

We use a deep-learning multimodal model, the Bootstrapping Language-Image Pre-Training (BLIP) model, to measure the photo-text congruence in the annual reports. Theories and empirical evidence in communications research show that high photo-text congruence can improve the understanding and memorability of a given context (e.g., Geise and Baden, 2015; Huang and Fahmy, 2013). Therefore, if management wants to highlight favorable data, enhancing the photo-text congruence in voluntary disclosures could be an effective strategy. In our empirical analyses, we first find that firms with poor performance have higher photo-text congruence in their annual reports, consistent with poorly performing firms' management using photo-text congruence as a disclosure strategy to emphasize favorable data about future performance or obfuscate bad news. As this disclosure strategy does not add new information beyond 10-Ks, our finding differs from prior evidence that firms voluntarily disclose new information when performance is poor (e.g., Ebert, Simons, and Stecher, 2017; Nagar and Schoenfeld, 2025).

We next examine the impact of high photo-text congruence on two outcomes, shareholder voting outcomes and future firm performance. Given the importance of shareholder votes, management has strong incentives to persuade shareholders to vote for its proposals and against activists' proposals at the annual meetings. When firm performance is poor, management is motivated to expend greater effort in persuading shareholders to maintain their support. We hypothesize that for poorly performing firms, using photo-text congruence in their annual reports is positively associated with shareholder votes and future performance. Our results show that for poorly performing firms, high photo-text congruence helps mitigate the negative impact of the bad performance, allowing these firms to garner stronger shareholder support as reflected by higher supporting rate for director elections and executive say-on-pay (SOP) proposals. However, this high congruence is not associated with future performance, suggesting that the data highlighted by high photo-text congruence might be misleading. We then exploit a quasi-natural experiment setting, in which we define poorly performing firms that are scrutinized by Institutional Shareholder Services (ISS) as the treated firms and assume that they are similar to observations in the control group, which are also poorly performing firms right below the threshold for scrutiny. In the regression discontinuity test, our findings hold for the treated observations, whose high photo-text congruence is more likely to be noticed by shareholders due to the more intense shareholder engagement activities triggered by ISS scrutiny. This result lets us better infer causality for our findings. In cross-sectional tests, we find that management in poorly performing firms with low active institutional ownership or high operating expenses (including R&D expenses) is more likely to benefit from high photo-text congruence in annual reports, but this high congruence is not associated with better short-term performance, suggesting that these firms may provide misleading data or overstate their operating investment to maintain shareholder support.

Our findings can inform shareholders, who should evaluate whether the topics discussed by the management could benefit the firm and should not get distracted by misleading or exaggerated perspectives from management. Since humans have limited cognitive resources, such as attention, memory, and computation (Simon, 1955), they could be distracted by irrelevant (but more comprehensible) contents in the annual reports that distort how they interpret firms' performance. Our findings can also guide firms in better preparing their disclosure to more accurately convey what they have achieved and what they are pursuing, so as to facilitate more effective shareholder engagement. We shed new light on the informativeness of the voluntarily disclosed annual reports, adding to the research on management's communication with investors (e.g., Kothari, Li, and Short, 2009).

Our study also adds to the discussion of a call for more relevant and user-friendly corporate disclosures encouraged by regulators and standard setters (e.g., SEC, 1998; SEC, 2016; FASB, 2012). Our findings reveal that while user-friendly disclosures may decrease data processing costs, they also allow firms to distract outsiders with clearer yet less valuable information, leaving room for impression management (e.g., Bowen, Davis, and Matsumoto, 2005; McVay, 2006). We are one of the first to examine the impact of the combination of photographs and text in annual reports. Recent studies in finance and accounting that explore settings containing photographs typically focus on the impact of photographs (e.g., Obaid and Pukthuanthong, 2022; Nekrasov, Teoh, and Wu, 2022; Cao, Cheng, Wang, Xia, and Yang, 2023), but often neglect the interaction between photographs and text (or other communication modes, such as audio and video) that could convey valuable data, with a few recent exceptions (e.g., Elliott, Loftus, and Winn, 2024; Ben-Rephael et al., 2023).[2] We use an advanced deep-learning model to measure photo-text congruence and demonstrate the importance of considering multiple disclosure modes.

[2] Ben-Rephael et al. (2023) examine the role of images in reinforcing text in annual reports. Our study is different in that we focus on outcomes such as shareholder votes and firms' future performance, while they are more interested in analyst forecast outcomes. Also, as we discuss in section 2.2, our congruence measure differs in that they count the number of matches between image labels and text, while we use a deep-learning model to measure the interaction between photos and text. Our model captures the semantics of words more accurately and can calculate the extent to which the meanings of an image and text are similar, leading to a more robust congruence measure.

2. Literature review and hypothesis development

2.1 The importance of shareholder votes

Shareholder participation in corporate decisions can increase firm value (Harris and Raviv, 2010). As an approach to participation, shareholder votes at annual meetings affect corporate governance. For example, Aggarwal, Dahiya, and Prabhala (2019) find that directors who face dissent in shareholder votes are more likely to leave boards. Focusing on shareholder activism, Ertimur, Ferri, and Muslu (2011) document that firms with excess CEO pay targeted by "vote-no" campaigns significantly reduce CEO pay.[3] Liu, Low, Masulis, and Zhang (2020) find that less institutional investor monitoring leads to less board effectiveness, partly due to reduced monitoring by independent directors in response to less shareholder voting pressure. They further show that poorly monitored boards make worse CEO pay decisions when their institutional investors are distracted, which is consistent with Hartzell and Starks' (2003) finding that institutions serve a monitoring role in mitigating the agency problem between shareholders and managers. The management teams thus have strong incentives to persuade shareholders to vote for the management's proposals and against activists' proposals at the annual meetings.

Prior studies have examined different aspects of the determinants of shareholders' voting decisions. For example, Cai et al. (2009) document that meeting attendance and ISS recommendations have significant positive impacts on shareholder votes. Meanwhile, Yermack (2010) raises the concern that shareholders who lack specific data about the firm could make suboptimal voting choices. We examine whether the annual reports and the photographs combined with text therein affect shareholders' decisions.

[3] Armstrong, Gow, and Larcker (2013) find evidence suggesting that shareholder votes for equity pay plans have little substantive impact on firms' incentive compensation policies. However, Chen, Hribar, and Melessa (2018) raise concerns about the use of decomposed compensation by Armstrong et al. (2013), whose findings therefore should be interpreted with caution.

2.2 Research on annual reports

After each fiscal year-end, managements use annual reports to convey performance data to shareholders before the annual meetings. In designing these reports, firms often include their Forms 10-K (and/or proxy statements) as part of or the entirety of the reports. Although managers likely do not directly design the annual reports themselves, they probably provide guidance to the teams involved, as managers need to make sure that the messages are properly conveyed to the public. Thus, they should influence the content presented in annual reports. U.S. public firms are required by the Securities and Exchange Commission (SEC) to annually file a Form 10-K that comprehensively summarizes their financial performance and operations.[4] Given the important role that Forms 10-K play, prior studies have documented different aspects of impacts from 10-K disclosures, mostly concentrating on the textual data (e.g., Li, 2008; Christensen, Floyd, Liu, and Maffett, 2017). Li (2008) documents that firms whose 10-Ks have a higher level of readability have more persistent positive earnings. Christensen et al. (2017) find that mandatorily including safety records in Forms 10-K decreases mining-related citations and injuries and reduces labor productivity. Focusing on data visualization in 10-Ks, Christensen, Fronk, Lee, and Nelson (2024) document that infographic disclosure increases over time, and firms vary in their choices of image type, data content, and infographic placement. Blankespoor et al. (2020) argue that managers strategically use various methods to reduce processing costs of favorable data and increase processing costs of unfavorable data, such as shifting expense classifications, highlighting positive non-GAAP metrics, or emphasizing prior non-recurring gains but not losses (Schrand and Walther, 2000; Bowen et al., 2005; McVay, 2006). In some cases, managers may refrain from explicitly

[4] https://www.sec.gov/files/form10-k.pdf.

mentioning favorable data, such as low effective tax rates, to avoid potential scrutiny (Chychyla, Falsetta, and Ramnath, 2022).

Unlike Forms 10-K, many firms voluntarily disclose different types of data in their annual reports, which can include textual data such as a letter to shareholders from the CEO, infographics such as graphs and figures of financial summaries, and other visual data such as photographs of products, workspaces, and directors and executives. This non-standardized section makes the unstructured data difficult to compare across firms. Only a few studies investigate this section, such as Dikolli, Keusch, Mayew, and Steffen (2020), who use textual analysis in the "a letter to shareholders from CEO" section to measure CEO characteristics and find that auditors undertake additional work in response to low CEO integrity.

Because Forms 10-K contain too much data for an average investor to digest quickly, investors find it hard to use all the data to make relevant decisions (Hawkins and Hawkins, 1986). Although Regulation S-T Rule 304(a) requires that there be no material differences between the data in annual reports and that in 10-Ks, Rule 304(b) leaves room for subtle differences, which could be visual data in annual reports but not easily perceived in the plain 10-Ks (Deng et al., 2023).[5] Thus, firms can include visual data in annual reports to better convey their messages without revising their 10-Ks. Indeed, Lee (1994) argues that firms use images in annual reports to establish corporate identity and influence corporate stakeholders. Preston, Wright, and Young (1996) draw on art theories and develop four phases to categorize pictures in annual reports, from "the reflection of a basic reality" to "constitute rather than merely represent reality." Preston and Young (2000) further use an international setting to examine how corporations use pictures to

[5] Rule 304(b)(2) states that "to the extent such descriptions, representations or transcripts represent a good faith effort to fairly and accurately describe omitted graphic, image, audio or video material, they are not subject to the civil liability and anti-fraud provisions of the federal securities laws."

construct themselves as global entities. Findings from more recent studies suggest that shareholders and other stakeholders do read firms' annual reports. For example, Townsend and Shu's (2010) experiment shows that if the first two pages of an annual report include more visuals, the focal firm is likely to receive more investment. Deng et al. (2023) find that firms' initial inclusion of graphics in annual reports is positively associated with stock market reactions and with institutional holdings. Concurrent studies focusing on visual data in annual reports use machine learning tools to conduct large-sample in-depth visual analyses (Deng et al., 2023; Ben-Rephael et al., 2023). Most closely related, Ben-Rephael et al. (2023) use machine learning to label image pages and check whether the labels match (reinforce) the text in the annual reports. They document that more label-to-text matches in an annual report is associated with higher analyst forecast accuracy in subsequent quarters. While their idea of images reinforcing text is similar to our photo-text congruence, their research question is different from ours. Also, we use a deep-learning model that considers the context of the text and thus captures the semantics of words more accurately. Our model calculates the similarity of the meanings of an image and text, leading to a more refined congruence measure.

2.3 Visual information, multimodal theory, and congruence

In the social psychology literature that explores information releases with visual and textual data, Chaiken and Eagly's (1976) classic study shows that visual data is more effective at changing opinions when the conveyed message is easy to understand, while textual data is more effective when the message is complex. Visual data attract humans much more than textual data (e.g., Larkin and Simon, 1987), as in the common adage that "a picture is worth a thousand words." In a related accounting study, Fronk (2023) finds evidence suggesting that incorporating data visualizations in earnings conference calls reduces information asymmetry and lowers investor processing costs.

Prior research continues to debate whether visuals are more effective than text. On the one hand, the inclusion and manipulation of images in an information release influences the perceptions and memorability of an issue (Zillmann, Gibson, and Sargent, 1999; Gibson and Zillmann, 2000). For example, including graphic photographs in military news articles can decrease support for war (Scharrer and Blackburn, 2015). On the other hand, the text accompanying images can also guide interpretations (Pfau, Haigh, Fifrick, Holl, Tedesco, Cope, Nunnally, Schiess, Preston, Roszkowski, and Martin, 2006). Pfau et al. (2006) document that altering the accompanying text of images can affect readers' opinions and responses towards certain issues. Domke, Perlmutter, and Spratt (2002) argue that the inclusion of images influences opinions when readers can relate the images to their existing knowledge; thus images may not drive public opinions by themselves. Zillmann, Knobloch, and Yu (2001) also show that incorporating images draws attention to the accompanying texts and enhances their recall.

This debate suggests that it might not be appropriate to focus on only one communication mode (e.g., textual data) when analyzing a context that contains multiple communication modes (e.g., textual and visual data). Following the communication literature, we refer to the communication mode as "modality" and accordingly, we use "multimodality" to describe the co-existence of different communication modes, such as text, image, and audio. In particular, multimodality focuses on the interaction of meaning embedded in different modalities in a certain context (Kress and van Leeuwen, 2001; LeVine and Scollon, 2004). In addition to complementing one another, different modalities in a context may interact with each other and generate meaning beyond that of an individual modality (O'Halloran and Smith, 2012; van Leeuwen, 2012). Geise and Baden (2015) further theorize the multimodal visual and textual framing effects, proposing that visuals are perceived quickly and are highly salient, but their interpretation can be ambiguous;

the accompanying text can guide interpretations, especially when the textual data is clear and explicit (Huang and Fahmy, 2013). Empirical studies have found that high congruence between different modalities in a context (i.e., the meanings of different modalities are very similar (Powell, Boomgaarden, De Swert, and de Vreese, 2015)) can improve the understanding and memorability of the context (e.g., Reese, 1984; Graber, 1990; Paivio, 1991; Huang and Fahmy, 2013; Powell et al., 2015). More recently, Cao, Li, and Zhang (2025) find that high or low image-text congruence of a product's description, compared to a medium level congruence, can increase consumers' preference towards the product. Thus, in our setting, management teams can enhance the photo-text congruence in the annual reports as a disclosure strategy to emphasize certain messages to their shareholders.

2.4 Hypotheses

When firm performance is poor, management teams receive lower shareholder voting support (e.g., Cai et al., 2009), which may motivate them to justify the poor performance by further communicating with the shareholders in the annual reports, such as providing more information like non-GAAP reporting to explain poor performance (Bloomfield, 2008; Leung and Veenman, 2018). They may also emphasize long-term goals by including more congruent photo-text pairs related to favorable data that indicates promising future performance, such as investment in expanding operations. Other communication channels, such as social media, can be used to moderate the negative price reactions to bad news (e.g., Lee, Hutton, and Shu, 2015).

Besides truthful communication, management teams may obfuscate information in annual reports (Bushee, Gow, and Taylor, 2018). For example, while non-GAAP reporting can improve firms' information environment (e.g., Gomez, Heflin, and Wang, 2023), management may also use non-GAAP reporting to manage investors' perception of the firm when performance is poor

(Black, Christensen, Joo, and Schmardebeck, 2017). Bloomfield (2002) hypothesizes that because the market reacts slowly to complicated disclosures, managers are incentivized to obfuscate information when firm performance is poor, potentially lessening the negative impact on their interests. Empirical studies such as Li (2008) have documented evidence supporting this hypothesis. Firms may strategically disclose incongruent photos and text pairs as a way to complicate annual reports. Besides disclosing complicated contents, another obfuscating tactic is to include attractive, congruent photo-text pairs covering data that is either exaggerated or not value-relevant to mislead shareholders from the poor performance.

In addition to disclosing favorable or distracting data, a third disclosure option could be that management teams do not differentiate their disclosure strategy based on firm performance, but just describe reasons for poor results in the same manner as for good performance. Similar to the choice between efficient and opportunistic accounting procedures—where managers choose accounting procedures to either efficiently maximize firm value or opportunistically prioritize their own interests at others' expense (Holthausen, 1990; Christie and Zimmerman, 1994)—management may opt for either efficient or opportunistic image disclosure in annual reports when firm performance is poor. Also, prior research is inconclusive on the use of infographics in annual reports conditional on firm performance: while Beattie and Jones (2008) document that firms are more likely to highlight favorable performance in infographics, Christensen et al. (2024) find that firms with extreme performance are more likely to include infographics in 10-Ks. Our first hypothesis is as follows:

H1: *Firms with poor performance have higher photo-text congruence in annual reports.*

To further investigate the impact of photo-text congruence in annual reports, we focus on two outcomes, shareholder votes and firms' future performance.[6] Firms usually attach their annual reports and proxy statements to the proposals that shareholders will vote on. Appendix B provides two examples of firms' annual meeting voting page. In both examples, the firms provide the resources including a proxy statement and an annual report for shareholders to use, which provides evidence that shareholders can directly access annual reports when making voting decisions. Because congruent photo-text pairs improve the understanding and memorability of the context (e.g., Reese, 1984; Graber, 1990; Paivio, 1991; Huang and Fahmy, 2013; Powell et al., 2015), we argue that photo-text congruence can be used to highlight favorable data in annual reports and affect the shareholders' perception of the firms and the management teams, thus affecting shareholder voting outcomes at the annual meeting. This prediction is not without tension because whether shareholders are aware of the disclosure strategy, or whether they are convinced by what management has conveyed, is not clear. We hypothesize (in the alternative) that:

H2a: *High photo-text congruence in poorly performing firms' annual reports benefits the management's favored proposals at annual meetings.*

Whether shareholders are better off in the future hinges on firms' future performance. To frame poor current performance in a positive way, management disclosures may focus more on the future (Asay, Libby, and Rennekamp, 2018). We next examine whether the use of photo-text congruence in annual reports indicates good future performance when current performance is poor. If management's emphasis on favorable data in annual reports predicts better future performance, then this disclosure strategy, i.e., increasing photo-text congruence, is truthful and benefits

[6] We refer to voting outcomes in year t (e.g., 2023 voting outcomes) and annual reports of year t-1 (e.g., 2022 Annual Report). In a given fiscal year, firms send out their annual reports for the previous fiscal year. For example, JPMorgan Chase & Co. released its 2022 Annual Report in April 2023 before their annual meeting on May 16, 2023.

shareholders. On the other hand, if the photo-text congruence highlights some distracting data, it might help the management garner stronger shareholder voting support but does not relate to future firm performance. We hypothesize that:

H2b: *High photo-text congruence in poorly performing firms' annual reports is positively associated with the firms' future performance.*

3. Sample selection and measure of photo-text congruence

3.1 Sample selection

We download firms' annual reports from AnnualReports.com and extract all images from the annual reports. Because we focus on only photographs, we train a ResNet18 model to identify and remove infographics from the images that we have extracted.[7] We then label each photograph with the number of the page it appears on. We obtain financial data from Compustat, voting results from ISS Voting Analytics, institutional ownership data from Thomson Reuters, and analyst related data from IBES. Our sample covers S&P 500 firms from 2002 to 2022. After removing observations with missing variables (except for the congruence measure), we have 4,674 firm-year observations, and 2,978 of them contain photographs in the annual reports.

3.2 Congruence measure

We rely on a deep-learning multimodal model to measure photo-text congruence. In particular, we use a neural network model called the Bootstrapping Language-Image Pre-Training (BLIP) model, which was developed by researchers (Li, Li, Xiong, and Hoi, 2022) at Salesforce AI Research, a research arm of the cloud-based software company Salesforce Inc. BLIP is "a pre-training framework for unified vision-language understanding and generation, which achieves

[7] The ResNet model can be used for image recognition. See https://pytorch.org/vision/main/models/resnet.html for more details.

state-of-the-art results on a wide range of vision-language tasks."[8] In short, BLIP decodes a large number of matched image-text pairs (e.g., an image and a caption that describes the image), analyzes the features of the image and text in each pair, and "learns" about the reason why each pair is a match. One of the original purposes of BLIP is to predict which images are paired with which texts in a large dataset, which can be extended to measure the extent to which the image is matched to the text in a pair. We take advantage of this feature of BLIP to measure the cosine similarity between photos and text, which produces the photo-text congruence scores for our analysis. The current version of the BLIP model was trained using 129 million image samples. We obtain and modify the BLIP code from Salesforce's GitHub repository.

We calculate congruence scores for pages that have photographs nearby. Because BLIP calculates a score for one photo-text pair at a time, we proceed as follows: for each page, we have BLIP calculate the congruence between each sentence and the photograph(s) on the one preceding page, the text page, and the following page, i.e., $sentence_{i,n}$ and [$photo_{n-1}$, $photo_n$, and $photo_{n+1}$], in which i refers to an individual sentence and n refers to the page number.[9] For example, when computing the congruence scores for page six of an annual report, we focus on the text on page six and all photographs from page five to page seven. If there are three sentences on page six and four photographs on these three pages, BLIP will return twelve $(3 * 4)$ congruence scores respectively for the twelve photo-sentence pairs; if there is no photograph on these three pages, we exclude that page from our analysis.[10] We then keep the highest score for each sentence; within

[8] See https://blog.salesforceairesearch.com/blip-bootstrapping-language-image-pretraining/ for a more detailed introduction. The model became publicly available in 2022 and has been widely used in computer science research; the paper that introduces the BLIP model has been cited over 5,000 times as of May 2025.
[9] We consider adjacent photos instead of all photos in the annual report, because according to bounded rationality (Simon, 1955), humans have limited cognitive attention and memory. Thus, it is more likely that only the adjacent photos affect readers' understanding of the focal page's text. We consider photos within three adjacent pages because sometimes a page has only photos or only text, or photos can run over two facing pages. We adopt the 3-page range to reduce data loss.
[10] For text on the first page of the annual report, we focus on photos from page one to page two; for text on the second page, we focus on photos from page one to page three.

each page, we keep the highest score; within each annual report, we then take the average of the scores across pages (*ConScore*). The reason why we use the highest congruence score for each page is that annual reports are usually so long that the readers, with limited attention (e.g., Hirshleifer and Teoh, 2003) and limited capability of consuming all the data (Hawkins and Hawkins, 1986), may just read the sections that they deem most important and interesting. Photo-text pairs with the highest congruence score are typically more eye-catching and memorable, thus more likely to leave a nontrivial impression on the readers.[11]

3.3 Validation of the congruence measure

We manually check a small sample with high and low photo-text congruence scores that are produced by BLIP. Appendix C provides three examples. In Example 1, page eight of the Hess Corporation 2020 Annual Report has a high congruence score 0.42. There are two photographs on pages seven to nine. The text on page eight discusses production, exploration, and development, which are highly relevant to the surrounding photographs. In Example 2, page five of Ross Stores, Inc. 2020 Annual Report has a low congruence score of 0.17. On this page, the firm mainly discusses the COVID-19 pandemic and its operational performance, which are topics that are relevant to the business environment, while the surrounding photographs seem to be more relevant to products and customers.

We also use the infographics, which are not considered in our main tests, to further evaluate the performance of the BLIP model. Because in their annual reports, firms usually describe the infographics in the surrounding text, it is likely that the infographic-text congruence scores are generally higher than the photo-text congruence scores. Following the same process in Section 3.2, we generate congruence scores for all the infographic-sentence pairs in our sample and notice that

[11] In an untabulated robustness test, for each page, we take the average of each sentence's highest score, instead of keeping the photo-sentence pair with the highest congruence score. Our results remain robust.

they do have relatively higher congruence scores (untabulated). Example 3 in Appendix C, for instance, shows that Loews Corporation is describing a graph on page 42, which contributes to a high congruence score (0.54). In an untabulated test, we create a firm-year level infographic-text congruence measure and find no significant correlation between this measure and poor firm performance, suggesting that poorly performing firms may not consider infographic-text interaction as an effective disclosure strategy. All three examples in Appendix C help illustrate the effectiveness of the BLIP model.

4. Empirical design

We employ the following empirical model to test our first hypothesis:

$$ConScore_{i,t} = \beta_0 + \beta_1 \mid Neg\ ROA \mid_{i,t} + \gamma X + \varepsilon_{i,t}, \tag{1}$$

where *ConScore* is the average of the highest photo-text congruence scores across pages in an annual report as explained in Section 3.2. Following Christensen et al. (2024), we divide ROA into absolute positive ROA, |*Pos ROA*|, and absolute negative ROA, |*Neg ROA*|. |*Pos ROA*| is equal to ROA if ROA is positive and equal to zero if ROA is negative; |*Neg ROA*| is equal to the absolute value of ROA if ROA is negative and equal to zero if ROA is positive. We use |*Neg ROA*| to proxy for firms' poor performance. *X* is a vector of other firm-level characteristics. Business complexity is shown to be positively associated with various attributes of 10-Ks such as length (e.g., Li, 2008; Cazier and Pfeiffer, 2016; Christensen et al., 2024). Firms with complex operations can use photos and high photo-text congruence that are likely to help readers better understand their message. Thus, we use firm size (*Size)* and the number of business segments (*Bus_seg*) and geographic segments (*Geo_seg*) to proxy for complexity. Other characteristics include common variables such as annual returns (*Annual_ret)* and institutional ownership *(Institution).* We follow Kim and Skinner (2012) to construct firms' litigation risk exposure (*Lit_risk*) as firms facing higher litigation risks may be more cautious in preparing disclosures and follow Koh and Reeb (2015) to create a missing-R&D dummy (*Missing_RD*) to accompany *R&D expense*. We also follow Lev, Petrovits, and Radhakrishnan (2010) to create a dummy variable *B2C*, which is equal to one if the firm belongs to one of the business-to-consumer industries, as firms in these industries may be more inclined to tailor their disclosures to individual consumers, who can be affected by the visual data in annual reports. We also include year fixed effects and 2-digit SIC industry fixed effects. Appendix A provides detailed variable definitions. All continuous variables are winsorized at their 1% and 99% values. All standard errors are clustered by firm. The coefficient of interest is β_1, which our hypothesis H1 predicts is positive.

We employ the following empirical model to test our H2a:

$$Voting\ outcome_{i,t} = \beta_0 + \beta_1 ConScore_{i,t} \times |\ Neg\ ROA\ |_{i,t} + \gamma X + \varepsilon_{i,t}, \quad (2)$$

where *Voting outcome* is one of the two outcome variables we use: *Vote_comp* and *Vote_dir*. *Vote_comp* is the median shareholder supporting rate (i.e., the percentage of votes supporting a proposal) for executive say-on-pay (SOP) plans within a firm-year observation. Since director elections affect 1) how boards make CEO turnover decisions and 2) the strength of CEOs' incentives (Fos, Li, and Tsoutsoura, 2018), we use *Vote_dir* as another dependent variable, which is the median shareholder supporting rate for director elections within a firm-year observation. We are interested in β_1, which captures the association between photo-text congruence in annual reports and the voting outcomes on these managements' favored proposals for poorly performing firms. *X* is a vector of control variables including firm-level characteristics that are the same as in equation (1) and report-level characteristics: the text sentiment in the annual report measured using the Loughran-McDonald Sentiment Word List (*Sentiment*),[12] the file size of the annual report in

[12] See https://sraf.nd.edu/loughranmcdonald-master-dictionary/ for more details about this word list.

Megabytes (*File_size*), and Loughran-McDonald Plain English Index (Loughran and McDonald, 2014; SEC, 1998) to proxy for text readability of the annual reports (*LM_PE_Index*).

We employ the following empirical model to test our H2b:

$$\Delta ROA_{i,t+1} = \beta_0 + \beta_1 ConScore_{i,t} \times \mid Neg\ ROA \mid_{i,t} + \gamma X + \varepsilon_{i,t}, \tag{3}$$

where ΔROA_{t+1} is the difference between ROA_{t+1} and ROA_t divided by ROA_t. We also consider operating ROA as an alternative measure because firms' financial outcomes could be affected by nonrecurring items in the short term. Negative nonrecurring events could decrease a firm's net income in a certain year without hurting its operational activities. If firms have low profitability due to nonrecurring items, it might be easier for the management to persuade shareholders to continue supporting the current leaders. By contrast, if firms' low profitability is caused by some persistent deficiency in their operations, their shareholders might think it too risky to continue investing in the firms. Other variables are defined the same as in equation (2).

5. Empirical results

5.1 Summary statistics

Table 1 reports the summary statistics of our variables. After requiring all variables but *ConScore* to be non-missing, we show that 63.7% of the 4,674 annual reports contain photographs. In our further analyses, we focus on the 2,978 firm-year observations that have photos in their annual reports, i.e., observations that have a non-missing *ConScore*. The mean and the median of *Vote_dir* are 77.6% and 79.0%, respectively. Because not all firms propose say-on-pay plans every year, we lost some observations for *Vote_comp*, whose mean and median are 70.6% and 73.3%, respectively. The mean and median of our congruence score measure, *ConScore,* are 0.289, and 0.286, respectively. S&P 500 firms in general perform well, as reflected in measures such as ROA and annual returns. The proportion of firms with a Big 4 auditor is also high (98.9%). The mean of the *LM_PE_Index* is -0.353, which is comparable to Loughran and McDonald's (2014) sample mean 0.363.[13] Figure 1 shows that firms in B2C industries are more likely to include photos in their annual reports (Panel A) and generally have higher photo-text congruence scores (Panel B) compared to non-B2C firms. The difference between the average congruence scores of B2C firms (0.292) and non-B2C firms (0.286) is statistically significant ($p < 0.01$).

Before our regression analysis, in Table 2 we report the comparison of firm characteristics between observations in the high *ConScore* subsample and the low *ConScore* subsample, defined as whether the *ConScore* is higher or lower than the sample median. Nine control variables have a statistically significant difference between the two subsamples: the high *ConScore* group has lower institutional ownership, more leverage, fewer current and intangible assets, more inventory, lower quick ratio, longer history of going public, fewer R&D expenses, and is more likely to report R&D.

The results suggest that the two subsamples are not similar. To ensure a more balanced comparison between them, we employ entropy balancing as the main approach for testing our H2. Specifically, we reweight the low *ConScore* subsample by imposing a three-moment constraint (the mean, variance, and skewness) on certain variables. The variables that we use include all the firm characteristics except for |*Pos ROA*| and |*Neg ROA*| (i.e., all the variables starting from *Annual_ret* in Table 2). Entropy balancing ensures that, the low and high *ConScore* subsamples have identical mean, variance, and skewness for each variable used for matching.

5.2 Do poorly performing firms have higher photo-text congruence in annual reports?

In Table 3, we examine whether the photo-text congruence in the annual report is associated with current poor firm performance. The coefficients on both |*Pos ROA*| and |*Neg ROA*|

[13] The index that we calculate is Loughran and McDonald's (2014) index multiplied by negative one. We transform the index to make it comparable to other measures in the literature, i.e., the higher the measure, the lower the readability. Loughran and McDonald's (2014) sample covers 40,251 Forms 10-K from 1998 to 2009.

are positive, consistent with Christensen et al. (2024), who document that firms with extreme performance are more likely to include infographics in 10-Ks. More importantly, the coefficients on |*Neg ROA*| are statistically significant in all columns, showing that firms with poor performance are more likely to have higher photo-text congruence. The alternative performance measure, *Annual_ret*, is also negatively correlated with *ConScore* albeit less statistically significant. The results are consistent with our H1, suggesting that poorly performing firms strategically reduce the processing costs of favorable data by enhancing photo-text congruence in their annual reports. The results also suggest that high photo-text congruence could be a disclosure strategy for these firms.

5.3 Do management-favored proposals at poorly performing firms benefit from higher photo-text congruence?

We test our H2a in Table 4. In Panel A where we use the entropy-balanced sample, the coefficients on the interaction term $ConScore \times | Neg\ ROA|$ are positive and statistically significant across all columns, showing that for poorly performing firms, higher photo-text congruence in their annual reports is associated with better shareholder voting outcomes for management's favored proposals. Panel B reports similar results using the full sample.

Further interpreting the economic magnitude, we find that high photo-text congruence can mitigate or even reverse the negative impact brought by the poor performance. Taking Panel A Column (3) as an example, assuming that *ex ante* a firm with negative ROA had the median *ConScore*, a one standard deviation increase in |*Neg ROA*| changes *Vote_dir* by -0.32%. If the firm's *ConScore* was one standard deviation higher than the median, a one standard deviation increase in |*Neg ROA*| changes *Vote_dir* by 0.14%, suggesting that the negative impact brought by the poor performance can be reversed with higher photo-text congruence in place.[14] For the same firm, the same increase of *ConScore* can mitigate the negative impact of *|Neg ROA|* on *Vote_comp* from -1.54% to -0.85% if we focus on Column (6). Overall, our results suggest that managements in poorly performing firms enjoy stronger shareholder support when enhancing the photo-text congruence in their annual reports, which is consistent with H2a.

5.4 Does poorly performing firms' photo-text congruence indicate future performance?

So far, we have documented that firms with poor performance tend to have high photo-text congruence in their annual reports, and this high congruence benefits management's favored proposals at annual meetings. These findings are consistent with management using strategic disclosure to reconcile poor performance, leading to more favorable management outcomes. However, it is not clear whether the disclosure strategy of enhancing photo-text congruence really indicates future performance or is just cheap talk. On the one hand, management may distract shareholders' attention from poor performance by emphasizing other topics to make them seem important and adding photographs to enhance shareholders' memories of those topics. These topics might not align with the firm's strategy and thus not benefit future performance (e.g., Banker, Ma, Pomare, and Zhang, 2023). On the other hand, management could mention other materially important topics such as production, operation expansions, and research and development activities, which deliver a positive message that persuades shareholders to trust management. This strategy is likely useful when management expects that the poor financial performance is temporary or wants to convince shareholders that this is the case even when untrue.

---

[14] Specifically, when a firm-year observation has negative ROA and the median *ConScore*, a one standard deviation (SD) increase in *|Neg ROA|* can change *Vote_dir* by: $\beta_{|Neg\ ROA|} \times SD_{|Neg\ ROA|} + \beta_{ConScore*|Neg\ ROA|} \times ConScore_{median} \times SD_{|Neg\ ROA|} = -1.349 \times 0.020 + 4.153 \times 0.286 \times 0.020 = -0.0032$. If this observation increased its *ConScore* by one SD, 0.056, it would further change *Vote_dir* by: $\beta_{ConScore*|Neg\ ROA|} \times SD_{ConScore} \times SD_{|Neg\ ROA|} = 4.153 \times 0.056 \times 0.020 = 0.0046$. This further change turns the previous negative impact to positive (-0.0032+0.0046 = 0.0014). A similar calculation applies to the calculation for Column (6).

Example 1 in Appendix C shows that Hess explicitly mentioned its production "was negatively impacted by a reduction in energy demand due to COVID-19." It continued to emphasize its further exploration and development and provided several photographs that were related to these topics. Its financial performance over time is consistent with our second explanation: From 2015 to 2020, Hess's ROA was negative, and then it turned positive in 2021 and 2022.

Our H2b investigates whether higher photo-text congruence is associated with better future performance for poorly performing firms. Table 5 presents the results. In both Panels A and B, the coefficients on $ConScore \times |\ Neg\ ROA|$ are positive but not statistically significant, which is inconsistent with our prediction. In an untabulated robustness test, we replace our dependent variable with ΔROA_{t+2}, ΔROA_{t+3}, ROA_{t+1}, or corresponding operating ROA variables, and the coefficients on the interaction term are still not statistically significant.[15] The results suggest that when firm performance is poor, although management is better-off by enhancing photo-text congruence in the annual reports, this disclosure strategy is not indicative of future performance, i.e., shareholders are not better-off in the future. Therefore, whereas enhancing photo-text congruence can serve as an effective communication tool to help management garner more support, shareholders should be cautious when processing the glossy annual reports.

5.5 Regression discontinuity setting

To better identify causality for our findings, we follow Dey, Starkweather, and White (2024) to exploit a quasi-natural experiment. Specifically, when a publicly listed firm's say-on-pay (SOP) voting support rate falls below 70%, the largest advisory firm—Institutional Shareholder Services (ISS)—will formally scrutinize and evaluate the firm's disclosed shareholder engagement

[15] We calculate ΔROA_{t+2} by using ROA_{t+2} minus ROA_t, scaled by ROA_t. Similarly, ΔROA_{t+3} is calculated using ROA_{t+3} minus ROA_t, scaled by ROA_t.

activities before the firm's next annual meeting. During the scrutiny, ISS requires the management to provide sufficient disclosure and take subsequent actions about the firm's engagement with its shareholders regarding the low voting support. Analyzing a sample of firms that receive an SOP supporting rate near the 70% threshold, Dey et al. (2024) document that firms receiving ISS scrutiny significantly increase shareholder engagement, thus benefiting firms' governance and information environments.

Firms use various communication tools, including annual reports, to interact with shareholders. High photo-text congruence in annual reports could facilitate firms' engagement with shareholders. Following Dey et al. (2024), we identify firm-year observations that received an SOP voting support rate above 67.5% and below 72.5% in the previous fiscal year.[16] This sample starts in 2011 because ISS started mandating this scrutiny policy in 2011. To infer causality from this research design, an important assumption is that firms cannot manipulate the SOP voting outcomes and whether they receive ISS scrutiny is random around the 70% threshold. Because shareholders are the ones who vote, firms should not be able to determine whether the support rate will be just above the 70% threshold. In Figure 2, we show the distribution of firms' SOP voting support around this threshold and we do not see an obvious discontinuity; in an untabulated test, we find that receiving ISS scrutiny is not associated with photo-text congruence in annual reports for firms with poor performance. These two pieces of evidence are consistent with Dey et al. (2024) and Bach and Metzger (2019), who find no evidence of firms' manipulation of voting outcomes around the 70% threshold, suggesting that receiving ISS scrutiny is not endogenous.

For poorly performing firms, if high photo-text congruence in annual reports predicts better shareholder voting outcomes and future firm performance, we should see stronger predictive

[16] The results are similar when we use alternative bandwidths, e.g., 67%--73%, 68%--72%.

power when these firms receive ISS scrutiny because the high congruence is more likely to be salient to shareholders. We use the following empirical models to re-examine our H2a and H2b:

$$Voting\ outcome_{i,t} = \beta_0 + \beta_1 Treat_{i,t} \times ConScore_{i,t} \times Neg\_ROA_{i,t} + \gamma X + \varepsilon_{i,t}, \quad (4)$$

$$\Delta ROA_{i,t+1} = \beta_0 + \beta_1 Treat_{i,t} \times ConScore_{i,t} \times Neg\_ROA_{i,t} + \gamma X + \varepsilon_{i,t}. \quad (5)$$

We create a dummy variable *Treat*, which is equal to one if the firm's previous fiscal year's SOP voting support rate was lower than 70%, and equal to zero otherwise. For easier interpretation, in the three-term interaction we include *Neg_ROA*, which is equal to one if ROA is negative.[17] All other variables are defined the same as in equations (2) and (3). The main outcome variable of interest is the SOP voting outcome *Vote_comp* because it is directly related to the ISS scrutiny. We also test with the other two dependent variables, *Vote_dir* and ROA_{t+1}, to see whether the management's engagement effort has spillover effects. We predict that β_1 is positive in both equations.

The results in Table 6 Columns (1) – (4) are consistent with our H2a. The coefficients in the first three lines are not statistically significant, showing that when firms perform well, having high photo-text congruence in annual reports does not help garner more shareholder voting support. The coefficients on $Treat \times ConScore \times Neg\_ROA$ are positive and statistically significant, suggesting that following ISS scrutiny regarding the previously low SOP voting support, poorly performing firms' management can garner stronger shareholder voting support when enhancing photo-text congruence in the annual reports. However, the results in Columns (5) and (6) are consistent with our results from Table 5, again showing that the high congruence does not indicate future performance. Taken together, the findings in Table 6 support what we find in Sections 5.3

[17] Our results are robust when we replace *Neg_ROA* with |*Neg ROA*| in the interaction.

and 5.4, providing some causal evidence that management can be better-off by using a disclosure strategy, i.e., high photo-text congruence in the annual reports, at the expense of shareholders.

6. Cross-sectional tests

6.1 Does ownership structure affect the impact of photo-text congruence in annual reports?

Since individual investors tend to devote less time to investment analysis (e.g., Kumar and Lee, 2006), it is likely that they rely on more comprehensible corporate disclosures to make investment decisions. To test whether our documented impact of photo-text congruence varies with different ownership structures, we divide our sample into low- and high-institutional ownership groups based on whether an observation's institutional ownership is lower than the sample median. The results in Table 7 Panel A show that for firms with low institutional ownership, enhancing photo-text congruence in annual reports is associated with better shareholder voting outcomes and better future performance when current firm performance is poor, and this is not the case for firms with high institutional ownership. In the Wald test between the coefficients of the two groups, we find weak evidence that the impact of the congruence on shareholder votes is significantly larger for firms with low institutional ownership than for those with high institutional ownership, but this is not the case when we use future ROA as the dependent variable.

Because passive institutional investors do not actively participate in corporate governance (e.g., Bushee, 2001; Schmidt and Fahlenbrach, 2017), we further differentiate active versus passive institutional ownership by following Bushee (2001) to construct the ratio of quasi-indexers as a proxy for passive institutional ownership. We then divide our sample into low- and high-active institutional ownership groups based on whether an observation's active ownership—defined as institutional ownership minus quasi-indexer ownership—is below the sample median. As shown in Table 7 Panel B, the difference between the coefficients of the two groups are statistically significant in the first four columns, but not the last two columns, consistent with the results in Panel A. In an untabulated test, we compare the high-active ownership subsample with the high-retail ownership subsample and the result is also similar. Taken together, our finding suggests that when active institutional ownership is low, management in poorly performing firms benefits more from high photo-text congruence in annual reports, although the high congruence does not indicate better future performance. This finding is also consistent with retail investors devoting less time to analyzing complex corporate filings and more likely to be attracted by the glossy annual reports when firm performance is poor.

6.2 Do intense operating activities affect the impact of photo-text congruence in annual reports?

We next explore whether management in loss-making firms with intense operating activities benefits more from photo-text congruence in annual reports. Prior studies document that firms' losses may be driven by intangibles expensing, such as R&D expenditure, which are "hidden assets" that can contribute to future profits (e.g., Darrough and Ye, 2007; Gu, Lev, and Zhu, 2023). Thus, loss firms with intense operating activities, including significant investment in intangibles, may highlight the firms' long-term potential in their message to shareholders by depicting relevant activities. Although intangibles usually cannot be easily measured, firms can envision positive outcomes using visualizations, which may help convince shareholders to maintain their confidence. Meanwhile, high operating expenses can be long-lasting and risky, which may result in continuous losses.

To test whether our documented impact of photo-text congruence varies with different levels of operating expenditure, we divide our sample into high and low-operating expense groups based on whether an observation's operating expense is higher than the sample median. Table 8

Panel A shows that the difference between the coefficients on $ConScore \times |Neg\ ROA|$ of the two groups are statistically significant in the first four columns, but not the last two columns. As a corroborating test, in Panel B we repeat the test in Panel A by comparing high and low-R&D expense groups, which are based on whether an observation's ratio of R&D expense to total assets is higher than the 75th percentile of the sample, i.e., 2.9%. The results are consistent with those in Panel A. In both panels, when we replace ΔROA_t with ΔROA_{t+2} or ΔROA_{t+3} in the last two columns, the results still hold. The findings in Table 8 together show that management in poorly performing firms benefits more from highlighting favorable data in annual reports when the firms have high operating expenses or high R&D expenses, but their highlighted data does not indicate better future performance. This finding suggests that poorly performing firms may exaggerate the potential of their "hidden assets," which the shareholders perceive as valuable but fail to deliver on expectations in the short term.

7. Conclusion

In this study, we investigate management's strategic use of text and photographs in annual reports. In annual reports, firms convey performance data to shareholders before annual meetings. Given the importance of shareholder votes at annual meetings, firms' management teams have strong incentives to persuade shareholders to vote for their proposals. When firm performance is poor, shareholders are less likely to support management's proposals, which can motivate management to push for its favored voting outcomes. A potential approach that management can use is to enhance the photo-text congruence of its favorable data in corporate disclosures, since congruent photos and text can improve readers' understanding and memorability of the emphasized content. We tackle the challenge of measuring the photo-text congruence in annual reports by using the deep-learning Bootstrapping Language-Image Pre-Training (BLIP) model to quantify the similarity between the meanings of photos and text.

Consistent with management emphasizing favorable data when firm performance is poor, we find that the photo-text congruence level in annual reports is higher when firms have poor performance. The higher congruence in these firms can mitigate the negative impact of the bad performance by garnering stronger shareholder voting support on director elections and say-on-pay proposals. However, the high congruence is not associated with these firms' future performance either on average or in our cross-sectional tests where we differentiate firms with low and high active institutional ownership or high and low operating or R&D expenses, which suggests that these firms may emphasize their operating investment to maintain shareholder support. Overall, our findings suggest that the disclosure strategy, i.e., enhancing photo-text congruence in annual reports when firm performance is poor, benefits the management but does not add value for the shareholders.

Our findings inform both management and shareholders: Management can strategically design corporate disclosures to more effectively communicate with shareholders; shareholders, in turn, should carefully assess whether the topics management discusses predict future firm performance. Additionally, shareholders should remain vigilant and not be swayed by visually appealing but potentially distracting photographs provided by management.

References:

Aggarwal, R., Dahiya, S., & Prabhala, N. R. (2019). The power of shareholder votes: Evidence from uncontested director elections. *Journal of Financial Economics*, *133*(1), 134–153. https://doi.org/10.1016/j.jfineco.2018.12.002

Armstrong, C. S., Gow, I. D., & Larcker, D. F. (2013). The efficacy of shareholder voting: Evidence from equity compensation plans. *Journal of Accounting Research*, *51*(5), 909–950. https://doi.org/10.1111/1475-679X.12023

Asay, H. S., Libby, R., & Rennekamp, K. (2018). Firm performance, reporting goals, and language choices in narrative disclosures. *Journal of Accounting and Economics*, *65*(2–3), 380–398. https://doi.org/10.1016/j.jacceco.2018.02.002

Bach, L., & Metzger, D. (2019). How close are close shareholder votes? *The Review of Financial Studies*, *32*(8), 3183–3214. https://doi.org/10.1093/rfs/hhy126

Banker, R. D., Ma, X., Pomare, C., & Zhang, Y. (2023). When doing good for society is good for shareholders: Importance of alignment between strategy and CSR performance. *Review of Accounting Studies*, *28*(2), 1074–1106. https://doi.org/10.1007/s11142-021-09664-y

Beattie, V., & Jones, M. J. (2008). Corporate reporting using graphs: A review and synthesis. *Journal of Accounting Literature*, *27*, 71–110.

Ben-Rephael, A., Ronen, J., Ronen, T., & Zhou, M. (2023). “Show me!” The informativeness of images. *SSRN Working Paper*. http://dx.doi.org/10.2139/ssrn.3954219

Black, E. L., Christensen, T. E., Taylor Joo, T., & Schmardebeck, R. (2017). The relation between earnings management and non-GAAP reporting. *Contemporary Accounting Research*, *34*(2), 750–782. https://doi.org/10.1111/1911-3846.12284

Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure processing costs, investors’ information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, *70*(2–3), 101344. https://doi.org/10.1016/j.jacceco.2020.101344

Bloomfield, R. J. (2002). The “Incomplete Revelation Hypothesis” and financial reporting. *Accounting Horizons*, *16*(3), 233–243.

Bowen, R. M., Davis, A. K., & Matsumoto, D. A. (2005). Emphasis on Pro Forma versus GAAP earnings in quarterly press releases: Determinants, SEC intervention, and market reactions. *The Accounting Review*, *75*(2), 151–177.

Bushee, B. J. (2001). Do institutional investors prefer near-term earnings over long-run value? *Contemporary Accounting Research*, *18*(2), 207–246. https://doi.org/10.1506/J4GU-BHWH-8HME-LE0X

Bushee, B. J., Gow, I. D., & Taylor, D. J. (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research*, *56*(1), 85–121. https://doi.org/10.1111/1475-679X.12179

Cai, J., Garner, J. L., & Walkling, R. A. (2009). Electing directors. *The Journal of Finance*, *64*(5), 2389–2421. https://doi.org/10.1111/j.1540-6261.2009.01504.x

Cao, J., Li, X., & Zhang, L. (2025). Is relevancy everything? A deep-learning approach to understand the effect of image-text congruence. *Management Science*, mnsc.2022.01896. https://doi.org/10.1287/mnsc.2022.01896

Cao, S. S., Cheng, Y., Wang, M., Xia, Y., & Yang, B. (2023). Visual information in the age of AI: Evidence from corporate executive presentations. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.4490834

Cazier, R. A., & Pfeiffer, R. J. (2016). Why are 10-K filings so long? *Accounting Horizons*, *30*(1), 1–21. https://doi.org/10.2308/acch-51240

Chaiken, S., & Eagly, A. H. (1976). Communication modality as a determinant of message persuasiveness and message comprehensibility. *Journal of Personality and Social Psychology*, *34*(4), 605–614.

Chen, W., Hribar, P., & Melessa, S. (2018). Incorrect inferences when using residuals as dependent variables. *Journal of Accounting Research*, *56*(3), 751–796. https://doi.org/10.1111/1475-679X.12195

Christensen, H. B., Floyd, E., Liu, L. Y., & Maffett, M. (2017). The real effects of mandated information on social responsibility in financial reports: Evidence from mine-safety records. *Journal of Accounting and Economics*, *64*(2–3), 284–304. https://doi.org/10.1016/j.jacceco.2017.08.001

Christensen, T. E., Fronk, K. E., Lee, J. A., & Nelson, K. K. (2024). Data visualization in 10-K filings. *Journal of Accounting and Economics*, *77*(2–3), 101631. https://doi.org/10.1016/j.jacceco.2023.101631

Christie, A. A., & Zimmerman, J. L. (1994). Efficient and opportunistic choices of accounting procedures: Corporate control contests. *The Accounting Review*, *69*(4), 539–566.

Chychyla, R., Falsetta, D., & Ramnath, S. (2022). Strategic choice of presentation format: The case of ETR reconciliations. *The Accounting Review*, *97*(1), 177–211. https://doi.org/10.2308/TAR-2019-0333

Darrough, M., & Ye, J. (2007). Valuation of loss firms in a knowledge-based economy. *Review of Accounting Studies*, *12*(1), 61–93. https://doi.org/10.1007/s11142-006-9022-z

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research*, *28*(1), 17–82. https://doi.org/10.1111/j.1911-3846.2010.01041.x

Deng, X., Gao, L., Hu, B., & Zhou, G. (2023). Seeing is believing: Annual report “graphicity” and stock returns. *SSRN Working Paper*. https://doi.org/10.2139/ssrn.3723126

Dey, A., Starkweather, A., & White, J. T. (2024). Proxy advisory firms and corporate shareholder engagement. *The Review of Financial Studies*, *37*(12), 3877–3931. https://doi.org/10.1093/rfs/hhae045

Dikolli, S. S., Keusch, T., Mayew, W. J., & Steffen, T. D. (2020). CEO behavioral integrity, auditor responses, and firm outcomes. *The Accounting Review*, *95*(2), 61–88. https://doi.org/10.2308/accr-52554

Domke, D., Perlmutter, D., & Spratt, M. (2002). The primes of our times?: An examination of the ‘power’ of visual images. *Journalism*, *3*(2), 131–159. https://doi.org/10.1177/146488490200300211

Ebert, M., Simons, D., & Stecher, J. D. (2017). Discretionary aggregation. *The Accounting Review*, *92*(1), 73–91. https://doi.org/10.2308/accr-51434

Elliott, W. B., Loftus, S., & Winn, A. (2024). To read or to listen? Does disclosure delivery mode impact investors’ reactions to managers’ tone language? *Contemporary Accounting Research*, *41*(1), 7–38. https://doi.org/10.1111/1911-3846.12898

Ertimur, Y., Ferri, F., & Muslu, V. (2011). Shareholder activism and CEO pay. *Review of Financial Studies*, *24*(2), 535–592. https://doi.org/10.1093/rfs/hhq113

Financial Accounting Standards Board (FASB). (2012). *Disclosure Framework. Invitation to Comment.* https://www.iasplus.com/en/news/2012/july/fasb-issues-an-invitation-to-comment-on-disclosure-framework

Fos, V., Li, K., & Tsoutsoura, M. (2018). Do director elections matter? *The Review of Financial Studies*, *31*(4), 1499–1531. https://doi.org/10.1093/rfs/hhx078

Fronk, K. (2023). Disclosure salience in earnings conference calls and investor information processing. *Ph.D. Thesis, University of Georgia*. https://esploro.libs.uga.edu/esploro/outputs/doctoral/Disclosure-salience-in-earnings-conference-calls/9949558927502959

Geise, S., & Baden, C. (2015). Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, *25*(1), 46–69. https://doi.org/10.1111/comt.12048

Gibson, R., & Zillmann, D. (2000). Reading between the photographs: The influence of incidental pictorial information on issue perception. *Journalism & Mass Communication Quarterly*, *77*(2), 355–366. https://doi.org/10.1177/107769900007700209

Gomez, E. A., Heflin, F. L., & Wang, J. (2023). Securities and Exchange Commission regulation and non-GAAP income statements. *The Accounting Review*, *98*(2), 149–175. https://doi.org/10.2308/TAR-2018-0719

Graber, D. A. (1990). Seeing is remembering: How visuals contribute to learning from television news. *Journal of Communication*, *40*(3), 134–156. https://doi.org/10.1111/j.1460-2466.1990.tb02275.x

Gu, F., Lev, B., & Zhu, C. (2023). All losses are not alike: Real versus accounting-driven reported losses. *Review of Accounting Studies*, *28*(3), 1141–1189. https://doi.org/10.1007/s11142-023-09799-0

Harris, M., & Raviv, A. (2010). Control of corporate decisions: Shareholders vs. management. *Review of Financial Studies*, *23*(11), 4115–4147. https://doi.org/10.1093/rfs/hhq081

Hartzell, J. C., & Starks, L. T. (2003). Institutional investors and executive compensation. *The Journal of Finance*, *58*(6), 2351–2374. https://doi.org/10.1046/j.1540-6261.2003.00608.x

Hawkins, B. A., & Hawkins, D. F. (1986). *The effectiveness of the annual report as a communication vehicle: A digest of the relevant literature*. Financial Executives Research Foundation.

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1–3), 337–386. https://doi.org/10.1016/j.jacceco.2003.10.002

Holthausen, R. W. (1990). Accounting method choice: Opportunistic behavior, efficient contracting, and information perspectives. *Journal of Accounting and Economics*, *12*(1–3), 207–218.

Huang, Y., & Fahmy, S. (2013). Picturing a journey of protest or a journey of harmony? Comparing the visual framing of the 2008 Olympic torch relay in the US versus the Chinese press. *Media, War & Conflict*, *6*(3), 191–206. https://doi.org/10.1177/1750635213495601

Kim, I., & Skinner, D. J. (2012). Measuring securities litigation risk. *Journal of Accounting and Economics*, *53*(1–2), 290–310. https://doi.org/10.1016/j.jacceco.2011.09.005

Koh, P.-S., & Reeb, D. M. (2015). Missing R&D. *Journal of Accounting and Economics*, *60*(1), 73–94. https://doi.org/10.1016/j.jacceco.2015.03.004

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. *The Accounting Review*, *84*(5), 1639–1670. https://doi.org/10.2308/accr.2009.84.5.1639

Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Oxford University Press.

Kumar, A., & Lee, C. M. C. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, *61*(5), 2451–2486. https://doi.org/10.1111/j.1540-6261.2006.01063.x

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *11*(1), 65–100. https://doi.org/10.1111/j.1551-6708.1987.tb00863.x

Lee, L. F., Hutton, A. P., & Shu, S. (2015). The role of social media in the capital market: Evidence from consumer product recalls. *Journal of Accounting Research*, *53*(2), 367–404. https://doi.org/10.1111/1475-679X.12074

Lee, T. (1994). The changing form of the corporate annual report. *Accounting Historians Journal*, *21*(1), 215–232.

Leung, E., & Veenman, D. (2018). Non-GAAP earnings disclosure in loss firms. *Journal of Accounting Research*, *56*(4), 1083–1137. https://doi.org/10.1111/1475-679X.12216

Lev, B., Petrovits, C., & Radhakrishnan, S. (2010). Is doing good good for you? How corporate charitable contributions enhance revenue growth. *Strategic Management Journal*, *31*(2), 182–200. https://doi.org/10.1002/smj.810

LeVine, P., & Scollon, R. (2004). *Discourse and technology: Multimodal discourse analysis*. Georgetown University Press.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics*, *45*(2–3), 221–247. https://doi.org/10.1016/j.jacceco.2008.02.003

Li, J., Li, D., Xiong, C., & Hoi, S. (2022). *BLIP: Bootstrapping Language-Image Pre-training for unified vision-language understanding and generation* (arXiv:2201.12086). arXiv. http://arxiv.org/abs/2201.12086

Liu, C., Low, A., Masulis, R. W., & Zhang, L. (2020). Monitoring the monitor: Distracted institutional investors and board governance. *The Review of Financial Studies*, *33*(10), 4489–4531. https://doi.org/10.1093/rfs/hhaa014

Loughran, T., & McDonald, B. (2014). Regulation and financial disclosure: The impact of plain English. *Journal of Regulatory Economics*, *45*(1), 94–113. https://doi.org/10.1007/s11149-013-9236-5

McVay, S. E. (2006). Earnings management using classification shifting: An examination of core earnings and special items. *The Accounting Review*, *81*(3), 501–531. https://doi.org/10.2308/accr.2006.81.3.501

Nagar, V., & Schoenfeld, J. (2025). Do weaker firms disclose more? *Journal of Financial Reporting*. https://doi.org/10.2308/JFR-2024-020

Nekrasov, A., Teoh, S. H., & Wu, S. (2022). Visuals and attention to earnings news on twitter. *Review of Accounting Studies*, *27*(4), 1233–1275. https://doi.org/10.1007/s11142-021-09630-8

Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, *144*(1), 273–297. https://doi.org/10.1016/j.jfineco.2021.06.002

O'Halloran, K., & Smith, B. (2012). *Multimodal studies: Exploring issues and domains*. Routledge.

Paivio, A. (1991). *Images in mind: The evolution of a theory*. Harvester Wheatsheaf.

Pfau, M., Haigh, M., Fifrick, A., Holl, D., Tedesco, A., Cope, J., Nunnally, D., Schiess, A., Preston, D., Roszkowski, P., & Martin, M. (2006). The effects of print news photographs of the casualties of war. *Journalism & Mass Communication Quarterly*, *83*(1), 150–168. https://doi.org/10.1177/107769900608300110

Powell, T. E., Boomgaarden, H. G., De Swert, K., & De Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, *65*(6), 997–1017. https://doi.org/10.1111/jcom.12184

Preston, A. M., Wright, C., & Young, J. J. (1996). IMag[in]ing annual reports. *Accounting, Organizations and Society*, *21*(1), 113–137. https://doi.org/10.1016/0361-3682(95)00032-5

Preston, A. M., & Young, J. J. (2000). Constructing the global corporation and corporate constructions of the global: A picture essay. *Accounting, Organizations and Society*, *25*(4–5), 427–449. https://doi.org/10.1016/S0361-3682(98)00066-X

Reese, S. D. (1984). Visual-verbal redundancy effects on television news learning. *Journal of Broadcasting*, *28*(1), 79–87.

Scharrer, E., & Blackburn, G. (2015). Images of injury: Graphic news visuals' effects on attitudes toward the use of unmanned drones. *Mass Communication and Society*, *18*(6), 799–820. https://doi.org/10.1080/15205436.2015.1045299

Schmidt, C., & Fahlenbrach, R. (2017). Do exogenous changes in passive institutional ownership affect corporate governance and firm value? *Journal of Financial Economics*.

Schrand, C. M., & Walther, B. R. (2000). Strategic benchmarks in earnings announcements: The selective disclosure of prior-period earnings components. *The Accounting Review*, *75*(2), 151–177. https://doi.org/10.2308/accr.2000.75.2.151

Securities and Exchange Commission (SEC). (1998). *A Plain English Handbook*. SEC Office of Investor Education and Assistance. https://www.sec.gov/pdf/handbook.pdf

Securities and Exchange Commission (SEC). (2016). *Business and Financial Disclosure Required by Regulation S-K*. https://www.sec.gov/files/rules/concept/2016/33-10064.pdf

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852

Townsend, C., & Shu, S. B. (2010). When and how aesthetics influences financial decisions. *Journal of Consumer Psychology*, *20*(4), 452–458. https://doi.org/10.1016/j.jcps.2010.06.013

van Leeuwen, T. (2012). The Semiotics of Decoration. In *Multimodal studies: Exploring issues and domains* (pp. 115–127). Routledge.

Yermack, D. (2010). Shareholder voting and corporate governance. *Annual Review of Financial Economics*, *2*(1), 103–125. https://doi.org/10.1146/annurev-financial-073009-104034

Zillmann, D., Gibson, R., & Sargent, S. L. (1999). Effects of photographs in news-magazine reports on issue perception. *Media Psychology*, *1*(3), 207–228. https://doi.org/10.1207/s1532785xmep0103_2

Zillmann, D., Knobloch, S., & Yu, H. (2001). Effects of photographs on the selective reading of news reports. *Media Psychology*, *3*(4), 301–324. https://doi.org/10.1207/S1532785XMEP0304_01

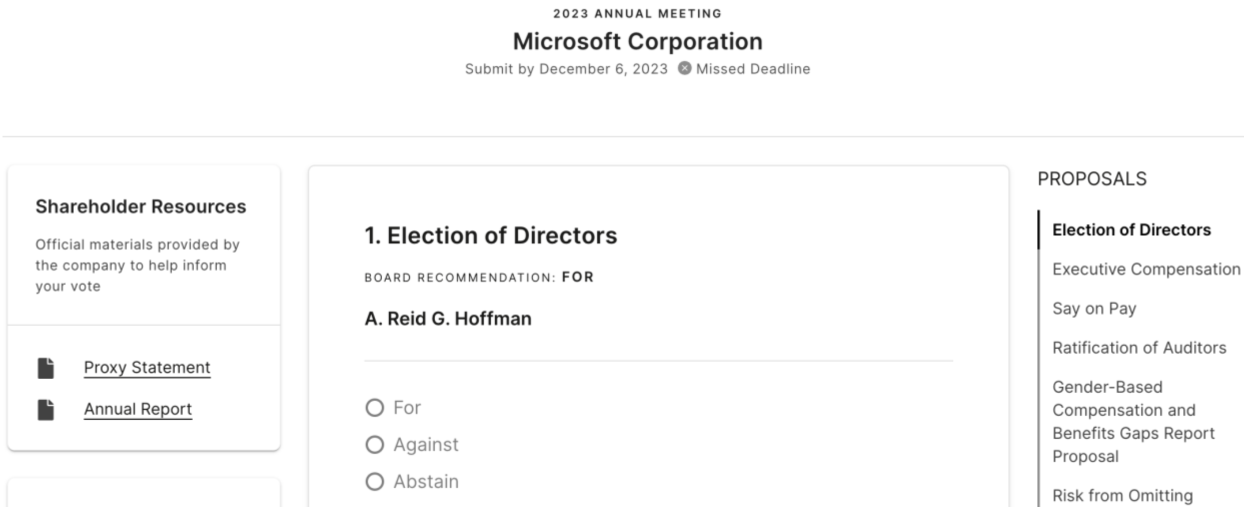
Appendix A Variable definitions

Congruence score	Definition
ConScore	Congruence score between photographs and text for a firm-year observation. To calculate this score, we adopt the following measure: for each page, we have BLIP calculate the congruence between each sentence and the photograph(s) on the one preceding page, the text page, and the following page, i.e., $sentence_{i,n}$& $(photo_{n-1}, photo_n, photo_{n+1})$, in which i refers to an individual sentence and n refers to the page number. For example, when computing the congruence scores for page six of an annual report, we focus on the text on page six and all photographs from page five to page seven. If there are three sentences on page six and four photographs on these three pages, BLIP will return twelve (3*4) congruence scores respectively for the twelve photo-sentence pairs; if there is no photograph on these three pages, we exclude that page from our analysis. We then keep the highest score for each page, and within each annual report, we take the average of the highest scores across pages for each page, which is the *ConScore*. Data source: AnnualReports.com.
Voting outcome	
Vote_dir	The median shareholder supporting rate (the percentage of votes supporting a proposal) for director elections within a firm-year observation. Specific proposals that we consider: "Elect Director", "Elect Director (Cumulative Voting or More Nominees Than Board Seats)", "Elect Director (Management)", "Elect Director and Approve Director's Remuneration", "Elect Directors (Bundled Dissident Slate)", "Elect Directors (Bundled)", "Elect Directors (Opposition Slate)", "Fix Number of and Elect Directors (Bundled)". Data source: ISS Voting Analytics.
Vote_comp	The median shareholder supporting rate for executive say-on-pay (SOP) plans within a firm-year observation. Specific proposals that we consider: "Advisory Vote to Ratify Named Executive Officers' Compensation." Data source: ISS Voting Analytics.
Report-level characteristics	
Photo	An indicator equal to one if a firm-year observation has photograph(s); equal to zero otherwise. Data source: AnnualReports.com.
Sentiment	The text sentiment score in the annual report measured using the positive and negative keywords from the Loughran-McDonald Sentiment Word List (https://sraf.nd.edu/loughranmcdonald-master-dictionary/). The equation to calculate the sentiment score in an annual report: (number of positive keywords – number of negative keywords) / (number of positive keywords + number of negative keywords). Data source: AnnualReports.com.
File_size	The file size of the annual report in Megabyte. Data source: AnnualReports.com.
LM_PE_Index	The Loughran-McDonald Plain English Index (Loughran and McDonald, 2014; SEC, 1998). To calculate this index, we first measure the average sentence length, average word length, the ratio of passive verbs to total words, the ratio of legal terms, the ratio of personal pronouns, and other items (see p.100-101 in Loughran and McDonald (2014) for a more comprehensive explanation of each item). Then we standardize each item and sum the average sentence length, average word length, the ratio of passive verbs to total words, the ratio of legal terms, and other items, and subtract the ratio of personal pronouns to create our *LM PE Index*. The higher the index, the lower the readability. The index that we calculate is equivalent to Loughran and McDonald's (2014) index multiplied by negative one. Data source: AnnualReports.com.
Firm-level characteristics	
|Neg ROA|	The absolute value of negative ROA. Calculated using income before extraordinary items divided by total assets, if negative, and 0 otherwise. Data source: Compustat.
|Pos ROA|	The absolute value of positive ROA. Calculated using income before extraordinary items divided by total assets, if positive, and 0 otherwise. Data source: Compustat.
Annual_ret	The 12-month cumulative stock returns. Data source: CRSP.
Institution	The ratio of shareholdings of institutions to total shares outstanding. Data source: Thomson Reuters.

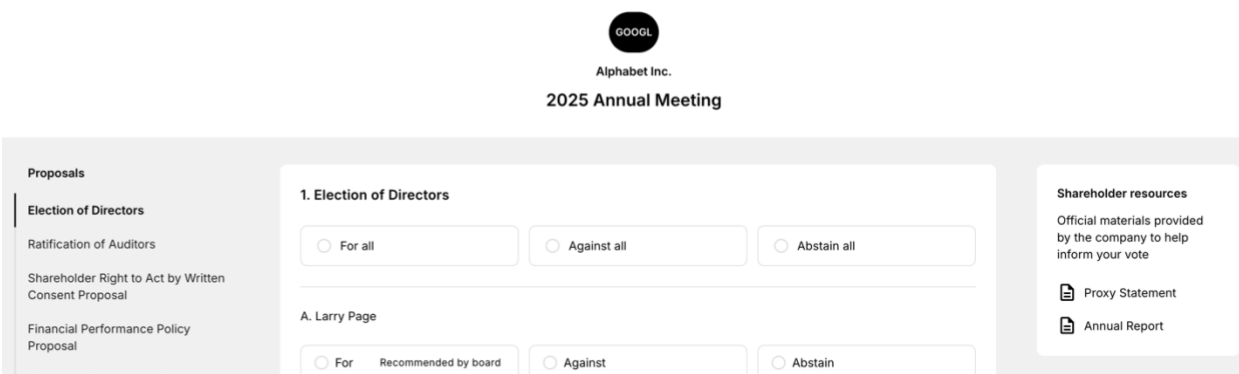
# Analyst	The number of analysts following the firm-year observation. Data source: IBES.
Size	The natural logarithm of the total assets of a firm-year observation. Data source: Compustat.
BTM	Book-to-market ratio. The ratio of book value of equity to market value of equity, where market value is calculated as fiscal year-end stock price multiplied by total shares outstanding. Data source: Compustat.
Leverage	Long-term debt scaled by total assets for a firm-year observation. Data source: Compustat.
Loss	An indicator equal to one if a firm-year observation's net income is less than zero; equal to zero otherwise. Data source: Compustat.
Current_assets	Current assets scaled by total assets for a firm-year observation. Data source: Compustat.
Intangible	Intangible assets scaled by total assets for a firm-year observation. Data source: Compustat.
Inventory	Total inventory scaled by total assets for a firm-year observation. Data source: Compustat.
Quick_ratio	Current assets less inventory and scaled by current liabilities for a firm-year observation. Data source: Compustat.
Non_opi	Nonoperating income scaled by total assets for a firm-year observation. Data source: Compustat.
Sale_grow	Sales for firm i in year t minus sales for firm i in year t-1, all scaled by sales for firm i in year t-1. Data source: Compustat.
Lit_risk	Litigation risk. Calculated using the coefficients from Model (3) in Kim and Skinner (2012).
Missing_RD	An indicator equal to one if a firm-year observation's R&D expense is missing; equal to zero otherwise. Data source: Compustat.
R&D_expense	Research and development expense scaled by total assets. For firm-year observations with missing R&D expense, we replace the missing value with two-digit SIC industry average. Data source: Compustat.
Bus_seg	The number of business segments. Data source: Compustat.
Geo_seg	The number of geographic segments. Data source: Compustat.
Big4	An indicator equal to one if a firm-year observation is audited by a Big 4 auditors; equal to zero otherwise. Data source: Compustat.
Age	Natural logarithm of the number of years that the firm-year observation has been listed in CRSP. Data source: CRSP.
B2C	An indicator equal to one if a firm belongs to one of the B2C industries; equal to zero otherwise. Following Lev et al. (2010), B2C industries are in the consumer goods and finance sectors. Data source: Compustat.
ΔROA	Change of ROA between year *t* and year *t*+1. Calculated using the difference between ROA_{t+1} and ROA_t scaled by ROA_t. Data source: Compustat.
ΔOp_ROA	Change of operating ROA between year *t* and year *t*+1. Calculated using the difference between Operating ROA_{t+1} and Operating ROA_t scaled by Operating ROA_t. Operating ROA is the operating income before depreciations scaled by total assets. Data source: Compustat.
Neg_ROA	An indicator equal to one if a firm-year observation has negative ROA; equal to zero otherwise. Data source: Compustat.
Treat	An indicator equal to one if a firm-year observation received a voting support rate for the SOP proposal larger than or equal to 67.5% and smaller than 70%; equal to zero if the voting support rate was between 70% and 72.5%. Data source: ISS Voting Analytics.

Appendix B Examples of firms' annual meeting voting page

Example 1 Microsoft Corporation 2023 annual meeting



Example 2 Alphabet Inc. 2025 annual meeting



Appendix C Examples of the congruence measure

Example 1 High photo-text congruence

(Hess Corporation 2020 Annual Report: text on page 8 and its surrounding photographs)

(*Congruence score*: 0.42)

2020 averaged 23,000 barrels of oil equivalent per day, compared with 28,000 barrels of oil equivalent per day in 2019. Production at both assets was negatively impacted by a reduction in energy demand due to COVID-19.

In the Danish North Sea, net production from the South Arne Field (61.5% working interest, operator) averaged 6,000 barrels of oil equivalent per day in 2020, compared with 7,000 barrels of oil equivalent per day in 2019, reflecting natural field declines. In the first quarter of 2021, we announced the sale of our interests in Denmark. This transaction is expected to close in the third quarter.

Net production from Libya averaged 4,000 barrels of oil equivalent per day in 2020, compared with 21,000 barrels of oil equivalent in 2019. Production operations were largely shut in during 2020 due to a force majeure caused by civil unrest.

Developments

Offshore Guyana, Hess holds a 30% interest in the 6.6 million acre Stabroek Block. Esso Exploration and Production Guyana Limited, a subsidiary of ExxonMobil, is operator and holds a 45% interest. CNOOC Petroleum Guyana Limited, a wholly owned subsidiary of CNOOC Limited, holds the remaining 25% interest. As a result of further exploration and appraisal success during 2020, the estimate of gross discovered recoverable resources on the block was increased to approximately 9 billion barrels of oil equivalent.

First oil from the second phase of the Liza development is on track for early 2022, utilizing the Liza Unity floating production storage and offloading vessel (FPSO) with production capacity of approximately 220,000 gross barrels of oil per day. A third development, at the

Appendix C Examples of the congruence measure

Example 2 Low photo-text congruence

(Ross Stores, Inc. 2020 Annual Report: text on page 5 and its surrounding photographs)

(*Congruence score*: 0.17)

To Our Stockholders

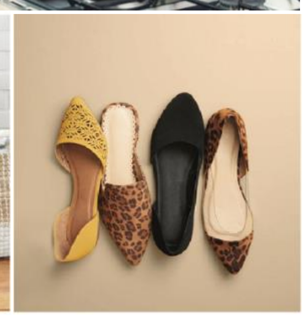
Fiscal 2020 was an extremely difficult and challenging year. Like so many other retailers and businesses, our operations and financial results reflect the negative impact of the COVID-19 pandemic.

Effects of the COVID-19 Pandemic on Our Fiscal 2020 Business

In early 2020, to prioritize the safety and well-being of our customers and associates and help slow the spread of COVID-19, we temporarily closed all store locations, our distribution centers, and buying and corporate offices. We also instituted "work from home" capabilities for many of our associates.

Given the uncertainty surrounding the duration and overall impact on consumer demand from the spread of this virus, we also took decisive actions to significantly increase our liquidity and strengthen our financial flexibility. These measures included drawing down on our $800 million revolving credit facility, issuing $2.0 billion in new senior notes, entering a new $500 million credit facility, suspending our stock repurchase and dividend programs, slowing new store growth, and aggressively cutting both ongoing expenses and capital expenditures. We have since repaid the $800 million under the revolving credit facility, terminated the undrawn $500 million credit facility, and refinanced a portion of the senior notes.

In mid-May, based on local government and health mandates, we began a phased reopening process, and by the end of June, the vast majority of our stores and all of our distribution centers were operating again.



Appendix C Examples of the congruence measure

Example 3 Infographic-text congruence

(Loews Corporation 2022 Annual Report: text on page 42 and its surrounding figure)

(*Congruence score*: 0.54)

PART II

Item 5. Market for the Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities.

Our common stock is listed on the New York Stock Exchange under the symbol "L".

The following graph compares annual total return of our Common Stock, the Standard & Poor's 500 Composite Stock Index ("S&P 500 Index") and our peer group set forth below ("Loews Peer Group") for the five years ended December 31, 2022. The graph assumes that the value of the investment in our Common Stock, the S&P 500 Index and the Loews Peer Group was $100 on December 31, 2017 and that all dividends were reinvested.

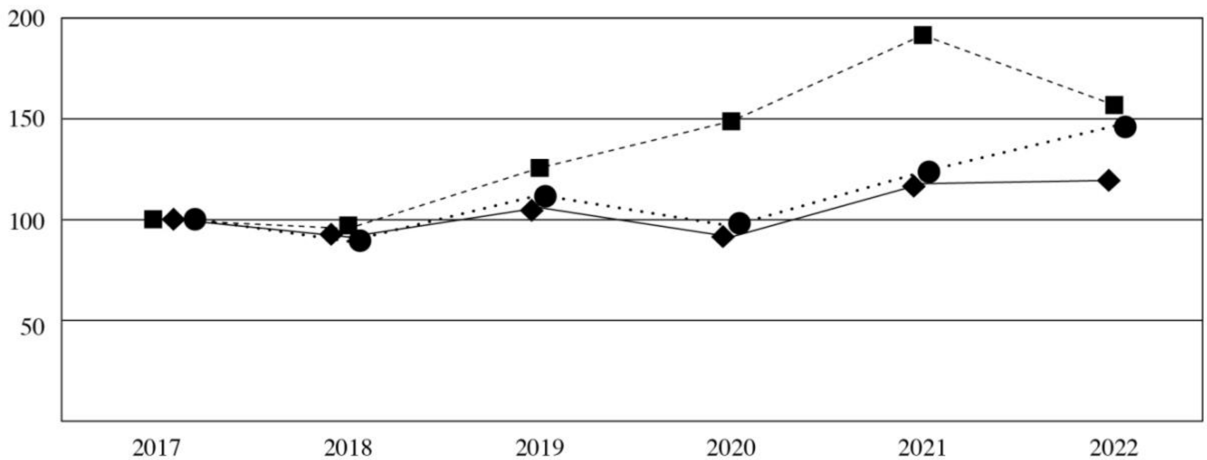
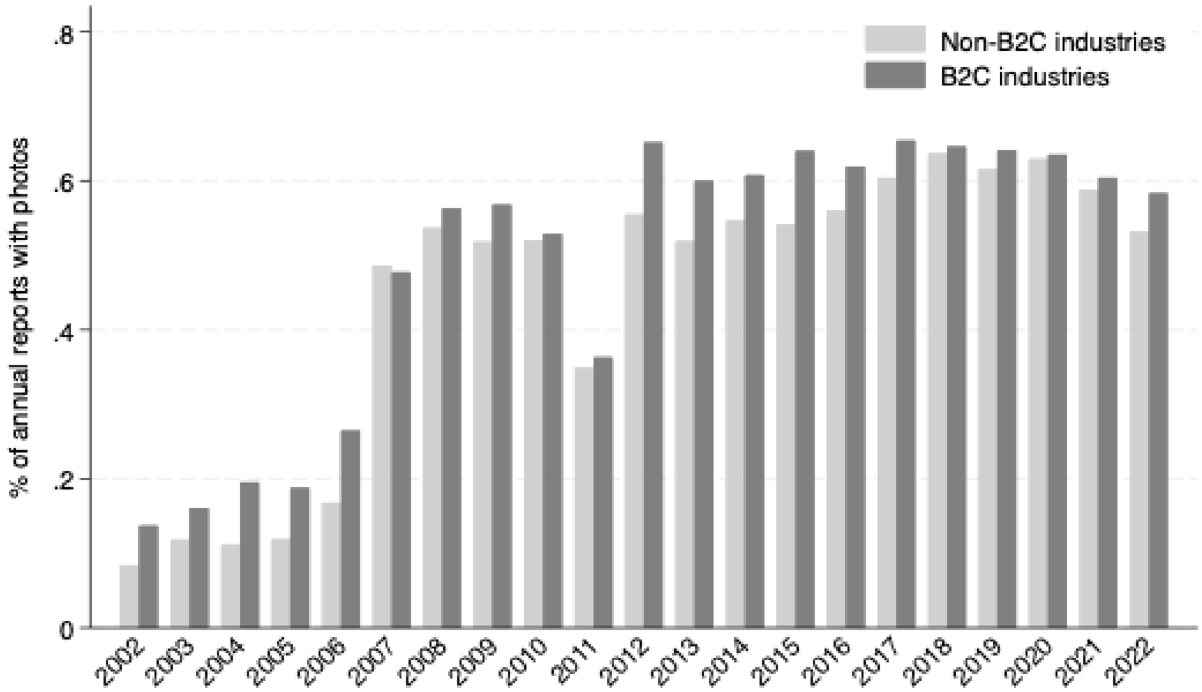


Figure 1 Comparison of B2C and non-B2C firms

Panel A Percentage of annual reports with photos over time



Panel B Average congruence score by industries

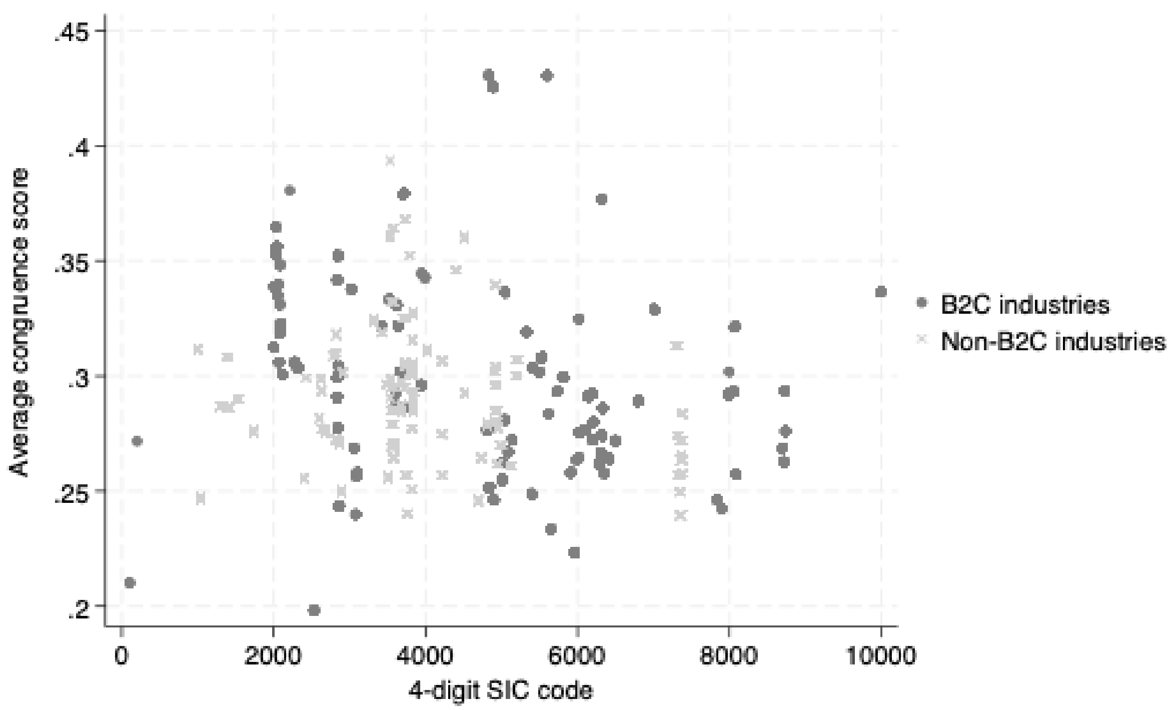


Figure 2 Distribution of say-on-pay (SOP) voting support around the 70% threshold

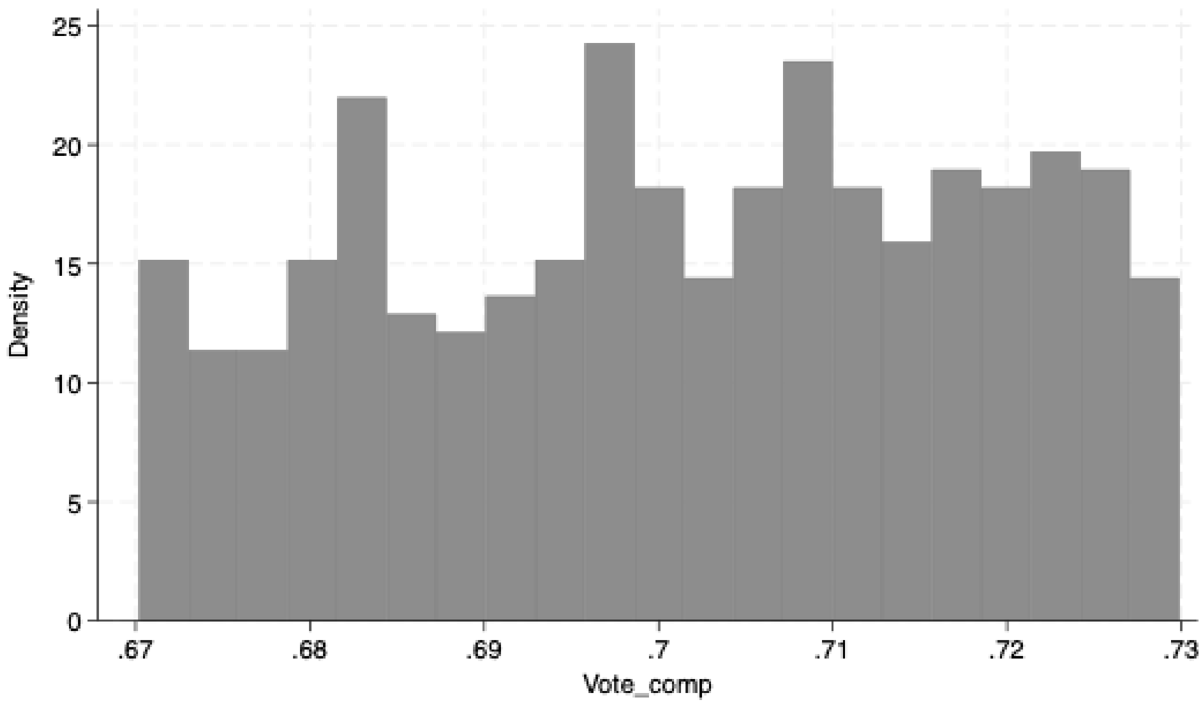


Table 1 Summary statistics

Variable	Obs.	Mean	S.D.	1st	25th	50th	75th	99th
Photo	4674	0.637	0.481	0.000	0.000	1.000	1.000	1.000
ConScore	2978	0.289	0.056	0.174	0.246	0.286	0.330	0.421
Vote_dir	2978	0.776	0.088	0.513	0.725	0.790	0.837	0.954
Vore_comp	2405	0.706	0.130	0.250	0.656	0.733	0.793	0.910
|*Neg ROA*|	2978	0.004	0.020	0.000	0.000	0.000	0.000	0.133
|*Pos ROA*|	2978	0.078	0.057	0.000	0.035	0.069	0.107	0.266
Annual_ret	2978	0.173	0.313	-0.555	-0.010	0.161	0.337	1.223
Institution	2978	0.794	0.154	0.268	0.720	0.820	0.901	1.000
#Analyst	2978	8.697	4.507	1.000	5.000	8.000	11.000	23.000
Size	2978	9.492	1.346	6.350	8.545	9.548	10.446	12.457
BTM	2978	0.329	0.256	-0.111	0.148	0.281	0.454	1.153
Leverage	2978	0.265	0.171	0.000	0.147	0.251	0.356	0.876
Loss	2978	0.071	0.257	0.000	0.000	0.000	0.000	1.000
Current_assets	2978	0.349	0.188	0.043	0.206	0.329	0.468	0.821
Intangible	2978	0.282	0.224	0.000	0.082	0.244	0.450	0.768
Inventory	2978	0.084	0.099	0.000	0.011	0.058	0.112	0.449
Quick_ratio	2978	1.343	0.948	0.177	0.729	1.109	1.651	5.283
Non_opi	2978	0.004	0.008	-0.016	0.000	0.002	0.006	0.037
Sales_grow	2978	0.082	0.172	-0.363	0.004	0.063	0.132	0.843
Lit_risk	2978	-0.786	1.093	-2.714	-1.549	-0.917	-0.154	2.645
Missing_RD	2978	0.322	0.467	0.000	0.000	0.000	1.000	1.000
R&D_expense	2978	0.025	0.042	0.000	0.001	0.010	0.029	0.253
Bus_seg	2978	2.811	1.899	1.000	1.000	3.000	4.000	7.000
Geo_seg	2978	3.444	2.751	1.000	1.000	3.000	5.000	15.000
Big4	2978	0.989	0.105	0.000	1.000	1.000	1.000	1.000
Age	2978	3.372	0.814	1.014	2.869	3.486	3.981	4.533
B2C	2978	0.420	0.494	0.000	0.000	0.000	1.000	1.000
Sentiment	2978	-0.292	0.330	-0.706	-0.508	-0.387	-0.218	0.755
File_size	2978	4.674	4.329	0.484	1.774	3.324	5.967	22.887
LM_PE_index	2978	-0.353	2.950	-7.979	-2.058	-0.230	1.557	6.636

Note: Appendix A provides detailed definitions for all variables.

Table 2 Comparison of high and low firm-year *ConScore*

	High *ConScore*	Low *ConScore*	Difference (High - Low)	
Variable	**Mean**	**Mean**	**Difference**	**t-statistics**
ConScore	0.335	0.242	**0.093**	**80.488**
Vote_dir	0.774	0.778	-0.005	-1.506
Vote_comp	0.698	0.714	**-0.016**	**-3.051**
|Neg ROA|	0.004	0.003	0.001	0.927
|Pos ROA|	0.079	0.077	0.002	1.140
Annual_ret	0.163	0.182	-0.018	-1.612
Institution	0.784	0.804	**-0.020**	**-3.502**
# Analyst	8.686	8.707	-0.021	-0.126
Size	9.528	9.457	0.071	1.449
BTM	0.337	0.321	**0.016**	**1.702**
Leverage	0.271	0.260	**0.010**	**1.672**
Loss	0.071	0.072	-0.001	-0.143
Current_assets	0.341	0.358	**-0.017**	**-2.494**
Intangible	0.273	0.290	**-0.017**	**-2.060**
Inventory	0.091	0.078	**0.013**	**3.546**
Quick_ratio	1.304	1.381	**-0.077**	**-2.211**
Non_opi	0.004	0.004	0.000	0.046
Sale_grow	0.077	0.087	-0.010	-1.582
Lit_risk	-0.783	-0.790	0.006	0.153
Missing_RD	0.297	0.348	**-0.051**	**-2.983**
R&D_expense	0.024	0.027	**-0.003**	**-2.025**
Bus_seg	2.834	2.787	0.047	0.675
Geo_seg	3.486	3.403	0.083	0.819
Big4	0.989	0.989	-0.001	-0.175
Age	3.416	3.329	**0.087**	**2.928**
B2C	0.433	0.406	0.027	1.485
Sentiment	-0.249	-0.335	**0.086**	**7.153**
File_size	4.791	4.558	0.233	1.466
LM_PE_Index	-0.458	-0.248	**-0.210**	**-1.947**

Note: Appendix A provides detailed definitions for all variables.

Table 3 Do poorly performing firms have higher photo-text congruence in annual reports?

	(1)	(2)	(3)	(4)
	ConScore	*ConScore*	*ConScore*	*ConScore*
|Neg ROA|	0.144**	0.131*	0.161**	0.152**
	(0.049)	(0.078)	(0.019)	(0.029)
|Pos ROA|	0.067*	0.069**	0.074**	0.079**
	(0.056)	(0.044)	(0.031)	(0.021)
Annual_ret	-0.004	-0.000	-0.005	-0.003
	(0.278)	(0.911)	(0.122)	(0.500)
Institution	-0.030**	-0.022	-0.012	-0.006
	(0.028)	(0.114)	(0.312)	(0.656)
#Analyst	-0.000	-0.000	-0.000	-0.000
	(0.954)	(0.720)	(0.498)	(0.283)
Size	-0.003	-0.000	-0.001	0.001
	(0.140)	(0.855)	(0.566)	(0.607)
BTM	0.015*	0.014*	0.009	0.009
	(0.055)	(0.075)	(0.228)	(0.271)
Leverage	0.010	0.017	-0.006	0.002
	(0.420)	(0.165)	(0.617)	(0.865)
Loss	-0.010*	-0.010*	-0.007	-0.007
	(0.060)	(0.074)	(0.153)	(0.169)
Current_assets	-0.033**	-0.031**	-0.013	-0.012
	(0.013)	(0.018)	(0.425)	(0.457)
Intangible	-0.017*	-0.016	-0.026*	-0.023
	(0.085)	(0.118)	(0.073)	(0.118)
Inventory	0.040*	0.041*	0.045	0.051
	(0.094)	(0.094)	(0.165)	(0.115)
Quick_ratio	0.001	0.002	-0.002	-0.000
	(0.640)	(0.359)	(0.501)	(0.854)
Non_opi	-0.206	-0.315	-0.319*	-0.410**
	(0.291)	(0.111)	(0.061)	(0.017)
Sales_grow	-0.001	-0.005	0.002	-0.001
	(0.858)	(0.539)	(0.774)	(0.920)
Lit_risk	0.001	-0.000	0.002	0.002
	(0.735)	(0.976)	(0.116)	(0.310)
Missing_RD	-0.009**	-0.008*	-0.000	-0.000
	(0.047)	(0.080)	(0.944)	(0.936)
R&D_expense	-0.064	-0.049	-0.122*	-0.106
	(0.332)	(0.446)	(0.065)	(0.103)
Bus_seg	-0.000	-0.000	-0.001	-0.001
	(0.992)	(0.929)	(0.473)	(0.479)
Geo_seg	0.001	0.001	-0.000	-0.000
	(0.497)	(0.476)	(0.689)	(0.706)
Big4	-0.011	-0.016	-0.033	-0.036
	(0.624)	(0.471)	(0.197)	(0.136)
Age	0.003	0.003	-0.002	-0.002
	(0.320)	(0.301)	(0.499)	(0.537)

B2C	0.005	0.005	0.014*	0.012
	(0.199)	(0.235)	(0.077)	(0.109)
Constant	0.341***	0.313***	0.357***	0.329***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,978	2,978	2,978	2,978
Adj. R^2	0.033	0.047	0.128	0.139
Year FE	No	Yes	No	Yes
Industry FE	No	No	Yes	Yes

Note: Appendix A provides detailed definitions for all variables. The shaded area represents the coefficient test of Hypothesis 1. Robust p-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by firm.

Table 4 Do management-favored proposals at poorly performing firms benefit from higher photo-text congruence?

Panel A: Entropy-balanced sample results

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote_dir	*Vote_dir*	*Vote_dir*	*Vote_comp*	*Vote_comp*	*Vote_comp*
ConScore	-0.008	0.006	-0.029	0.008	-0.070	-0.037
	(0.901)	(0.914)	(0.632)	(0.919)	(0.392)	(0.610)
|Neg ROA|	-1.381**	-1.146*	-1.349**	-2.659***	-2.690***	-2.537***
	(0.030)	(0.060)	(0.017)	(0.000)	(0.000)	(0.000)
*ConScore * |Neg ROA|*	4.454**	3.479*	4.153**	6.740***	6.791***	6.186***
	(0.032)	(0.077)	(0.023)	(0.007)	(0.002)	(0.005)
|Pos ROA|	-0.130	-0.178	-0.163	-0.018	-0.074	-0.008
	(0.460)	(0.299)	(0.333)	(0.936)	(0.730)	(0.969)
*ConScore * |Pos ROA|*	0.289	0.531	0.573	0.071	0.543	0.379
	(0.617)	(0.340)	(0.295)	(0.923)	(0.446)	(0.561)
Annual_ret	0.008	-0.015***	0.006	0.038***	0.039***	0.037***
	(0.207)	(0.002)	(0.331)	(0.000)	(0.000)	(0.000)
Institution	0.118***	0.107***	0.117***	0.103***	0.122***	0.111***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.001)	(0.002)
#Analyst	-0.001*	-0.001*	-0.001*	-0.001	-0.002**	-0.001
	(0.071)	(0.055)	(0.061)	(0.272)	(0.014)	(0.263)
Size	-0.014***	-0.017***	-0.016***	-0.017***	-0.015***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.007)	(0.001)
BTM	-0.040**	-0.051***	-0.042***	-0.014	-0.022	-0.012
	(0.019)	(0.001)	(0.010)	(0.459)	(0.242)	(0.510)
Leverage	-0.026	-0.039*	-0.021	-0.015	0.031	0.007
	(0.218)	(0.066)	(0.292)	(0.609)	(0.276)	(0.801)
Loss	0.001	0.001	-0.003	0.014	0.021	0.023
	(0.941)	(0.927)	(0.813)	(0.414)	(0.188)	(0.141)
Current_assets	-0.039	-0.058*	-0.053	-0.047	-0.071*	-0.061
	(0.143)	(0.079)	(0.110)	(0.216)	(0.095)	(0.143)
Intangible	0.042**	0.008	0.021	0.048**	-0.001	-0.002
	(0.024)	(0.738)	(0.378)	(0.015)	(0.973)	(0.947)
Inventory	0.062	-0.028	-0.021	0.147**	0.029	0.007
	(0.165)	(0.648)	(0.716)	(0.011)	(0.666)	(0.915)
Quick_ratio	0.002	-0.000	0.002	0.003	0.002	0.000
	(0.591)	(0.908)	(0.630)	(0.553)	(0.713)	(0.954)
Non_opi	0.058	0.444	0.018	0.235	0.093	0.356
	(0.874)	(0.152)	(0.955)	(0.576)	(0.821)	(0.355)
Sales_grow	0.002	0.017	-0.008	0.045**	0.044**	0.031*
	(0.891)	(0.197)	(0.525)	(0.021)	(0.022)	(0.086)
Lit_risk	-0.007**	-0.011***	-0.007**	-0.020***	-0.021***	-0.016***
	(0.019)	(0.000)	(0.036)	(0.000)	(0.000)	(0.000)
Missing_RD	0.004	0.011	0.010	0.013	0.025*	0.023*
	(0.674)	(0.290)	(0.325)	(0.234)	(0.085)	(0.093)
R&D_expense	-0.186	-0.108	-0.103	-0.221*	-0.152	-0.202

	(0.113)	(0.392)	(0.415)	(0.100)	(0.360)	(0.207)
Bus_seg	-0.001	0.000	0.000	-0.002	-0.001	-0.001
	(0.725)	(0.937)	(0.972)	(0.435)	(0.595)	(0.593)
Geo_seg	0.000	0.001	0.001	-0.001	0.001	0.001
	(0.951)	(0.452)	(0.372)	(0.687)	(0.552)	(0.691)
Big4	0.003	0.014	0.008	0.067	0.057	0.063
	(0.955)	(0.778)	(0.873)	(0.300)	(0.391)	(0.355)
Age	-0.010**	-0.012**	-0.009*	-0.014**	-0.020***	-0.015***
	(0.046)	(0.019)	(0.068)	(0.011)	(0.001)	(0.003)
B2C	0.010	-0.007	-0.004	0.002	-0.008	-0.005
	(0.167)	(0.653)	(0.811)	(0.809)	(0.725)	(0.827)
Sentiment	-0.001	0.022**	0.006	0.005	-0.002	0.011
	(0.926)	(0.044)	(0.584)	(0.700)	(0.872)	(0.391)
File_size	-0.001	-0.001**	-0.000	-0.000	0.000	-0.000
	(0.257)	(0.045)	(0.365)	(0.479)	(0.725)	(0.702)
LM_PE_Index	-0.001	0.002	0.001	-0.001	0.001	0.001
	(0.604)	(0.184)	(0.525)	(0.704)	(0.442)	(0.470)
Constant	0.864***	0.933***	0.894***	0.759***	0.784***	0.807***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,978	2,978	2,978	2,403	2,403	2,403
Adj. R^2	0.335	0.338	0.392	0.371	0.313	0.410
Year FE	Yes	No	Yes	Yes	No	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes

Note: Appendix A provides detailed definitions for all variables. The shaded area represents the coefficient test of Hypothesis 2a. Robust *p*-values in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered by firm.

Table 4 Do management-favored proposals at poorly performing firms benefit from higher photo-text congruence?

Panel B: Full sample results

	(1) *Vote_dir*	(2) *Vote_dir*	(3) *Vote_dir*	(4) *Vote_comp*	(5) *Vote_comp*	(6) *Vote_comp*
ConScore	-0.071	-0.038	-0.075	0.000	-0.060	-0.042
	(0.273)	(0.521)	(0.211)	(0.997)	(0.459)	(0.574)
|Neg ROA|	-1.387**	-1.074*	-1.311**	-2.343***	-2.329***	-2.277***
	(0.028)	(0.081)	(0.022)	(0.003)	(0.001)	(0.002)
*ConScore * |Neg ROA|*	4.268**	3.172	3.935**	5.730**	5.672**	5.451**
	(0.038)	(0.110)	(0.033)	(0.033)	(0.019)	(0.026)
|Pos ROA|	-0.150	-0.182	-0.180	0.082	0.047	0.060
	(0.362)	(0.239)	(0.229)	(0.707)	(0.821)	(0.764)
*ConScore * |Pos ROA|*	0.392	0.647	0.693	-0.210	0.154	0.167
	(0.481)	(0.210)	(0.167)	(0.771)	(0.828)	(0.797)
Annual_ret	0.005	-0.016***	0.003	0.041***	0.039***	0.039***
	(0.363)	(0.000)	(0.560)	(0.000)	(0.000)	(0.000)
Institution	0.116***	0.110***	0.116***	0.103***	0.124***	0.113***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)
#Analyst	-0.001	-0.001*	-0.001*	-0.001	-0.002***	-0.001*
	(0.121)	(0.057)	(0.061)	(0.127)	(0.002)	(0.079)
Size	-0.013***	-0.016***	-0.015***	-0.017***	-0.018***	-0.019***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
BTM	-0.032*	-0.043***	-0.034**	-0.018	-0.028	-0.020
	(0.064)	(0.005)	(0.035)	(0.335)	(0.146)	(0.276)
Leverage	-0.018	-0.030	-0.017	-0.011	0.019	0.004
	(0.395)	(0.122)	(0.387)	(0.707)	(0.471)	(0.868)
Loss	0.005	0.003	-0.002	0.019	0.026*	0.025*
	(0.651)	(0.814)	(0.886)	(0.183)	(0.072)	(0.073)
Current_assets	-0.027	-0.038	-0.032	-0.043	-0.043	-0.040
	(0.265)	(0.222)	(0.316)	(0.214)	(0.274)	(0.301)
Intangible	0.047**	0.025	0.036	0.049**	0.016	0.014
	(0.016)	(0.332)	(0.152)	(0.013)	(0.581)	(0.622)
Inventory	0.046	-0.028	-0.026	0.123**	-0.002	-0.021
	(0.244)	(0.616)	(0.636)	(0.017)	(0.975)	(0.739)
Quick_ratio	0.001	-0.001	0.002	0.003	0.001	-0.000
	(0.801)	(0.897)	(0.705)	(0.539)	(0.876)	(0.946)
Non_opi	0.053	0.341	-0.054	0.163	0.072	0.307
	(0.900)	(0.299)	(0.874)	(0.718)	(0.870)	(0.453)
Sales_grow	0.002	0.013	-0.010	0.043***	0.041**	0.029*
	(0.892)	(0.265)	(0.396)	(0.007)	(0.013)	(0.059)
Lit_risk	-0.008**	-0.010***	-0.007**	-0.020***	-0.020***	-0.017***
	(0.011)	(0.000)	(0.016)	(0.000)	(0.000)	(0.000)
Missing_RD	0.006	0.009	0.008	0.017*	0.025*	0.022*
	(0.519)	(0.362)	(0.409)	(0.100)	(0.081)	(0.096)
R&D_expense	-0.141	-0.076	-0.068	-0.237*	-0.206	-0.230

	(0.218)	(0.529)	(0.580)	(0.052)	(0.158)	(0.110)
Bus_seg	-0.001	0.000	0.000	-0.001	-0.001	-0.001
	(0.731)	(0.814)	(0.906)	(0.634)	(0.634)	(0.777)
Geo_seg	-0.000	0.001	0.001	0.000	0.002	0.002
	(0.845)	(0.507)	(0.420)	(0.924)	(0.181)	(0.295)
Big4	0.000	0.008	0.005	0.061	0.053	0.055
	(0.999)	(0.868)	(0.919)	(0.340)	(0.445)	(0.430)
Age	-0.010**	-0.011**	-0.009**	-0.011**	-0.016***	-0.013**
	(0.038)	(0.017)	(0.048)	(0.044)	(0.006)	(0.015)
B2C	0.010	-0.006	-0.005	0.006	-0.006	-0.003
	(0.185)	(0.719)	(0.760)	(0.502)	(0.791)	(0.878)
Sentiment	-0.000	0.017*	0.004	0.001	-0.007	0.005
	(0.998)	(0.077)	(0.712)	(0.912)	(0.604)	(0.674)
File_size	-0.000	-0.001*	-0.000	-0.000	0.000	-0.000
	(0.428)	(0.085)	(0.464)	(0.846)	(0.529)	(0.941)
LM_PE_Index	-0.000	0.001	0.001	-0.000	0.001	0.001
	(0.796)	(0.181)	(0.498)	(0.906)	(0.430)	(0.392)
Constant	0.875***	0.922***	0.891***	0.750***	0.790***	0.801***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,978	2,978	2,978	2,403	2,403	2,403
Adj. R^2	0.314	0.323	0.374	0.351	0.299	0.390
Year FE	Yes	No	Yes	Yes	No	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes

Note: Appendix A provides detailed definitions for all variables. The shaded area represents the coefficient test of Hypothesis 2a. Robust *p*-values in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered by firm.

Table 5 Does poorly performing firms' photo-text congruence indicate future performance?

Panel A: Entropy-balanced sample results

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔROA	*ΔROA*	*ΔROA*	*ΔOp_ROA*	*ΔOp_ROA*	*ΔOp_ROA*
ConScore	0.467	0.131	0.308	0.036	-0.007	-0.013
	(0.729)	(0.926)	(0.828)	(0.892)	(0.978)	(0.960)
|Neg ROA|	4.546	3.698	3.973	-6.908*	-6.684	-6.729
	(0.521)	(0.583)	(0.571)	(0.099)	(0.111)	(0.110)
*ConScore * |Neg ROA|*	15.066	17.630	16.380	1.830	0.569	0.630
	(0.426)	(0.366)	(0.421)	(0.898)	(0.968)	(0.965)
|Pos ROA|	-5.030	-7.428**	-6.881*	-1.431**	-1.817**	-1.714**
	(0.128)	(0.043)	(0.060)	(0.033)	(0.013)	(0.017)
*ConScore * |Pos ROA|*	-1.916	3.224	1.150	1.434	1.816	1.533
	(0.861)	(0.784)	(0.922)	(0.503)	(0.428)	(0.499)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,832	2,832	2,832	2,832	2,832	2,832
Adj. R^2	0.155	0.148	0.155	0.121	0.117	0.139
Year FE	Yes	No	Yes	Yes	No	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes

Panel B: Full sample results

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔROA	*ΔROA*	*ΔROA*	*ΔOp_ROA*	*ΔOp_ROA*	*ΔOp_ROA*
ConScore	0.067	-0.041	0.072	0.085	0.055	0.022
	(0.959)	(0.976)	(0.958)	(0.705)	(0.806)	(0.920)
|Neg ROA|	3.081	1.437	2.306	-5.600	-5.704	-5.753
	(0.684)	(0.843)	(0.763)	(0.179)	(0.176)	(0.175)
*ConScore * |Neg ROA|*	14.491	20.625	17.765	-1.977	-2.289	-2.191
	(0.482)	(0.332)	(0.426)	(0.892)	(0.878)	(0.883)
|Pos ROA|	-5.502*	-7.545**	-7.317**	-0.994*	-1.348**	-1.313**
	(0.083)	(0.031)	(0.037)	(0.067)	(0.030)	(0.031)
*ConScore * |Pos ROA|*	0.180	4.170	3.161	0.342	0.558	0.490
	(0.986)	(0.711)	(0.779)	(0.850)	(0.780)	(0.802)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,832	2,832	2,832	2,832	2,832	2,832
Adj. R^2	0.127	0.121	0.125	0.116	0.111	0.128
Year FE	Yes	No	Yes	Yes	No	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes

Note: Appendix A provides detailed definitions for all variables. The shaded area represents the coefficient test of Hypothesis 2b. Robust p-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by firm.

Table 6 Regression discontinuity design

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote_comp	*Vote_comp*	*Vote_dir*	*Vote_dir*	*ΔROA*	*ΔROA*
Treat	0.021	0.011	0.048	0.038	-0.681	-0.776
	(0.529)	(0.751)	(0.125)	(0.212)	(0.376)	(0.312)
ConScore	0.141	0.146*	0.097	0.098	-0.673	-0.760
	(0.136)	(0.095)	(0.191)	(0.175)	(0.740)	(0.700)
*Treat * ConScore*	-0.122	-0.079	-0.208*	-0.167	1.388	1.765
	(0.276)	(0.487)	(0.054)	(0.103)	(0.579)	(0.477)
Neg_ROA	0.163***	0.140**	0.092**	0.056	-15.038***	-14.650***
	(0.003)	(0.029)	(0.046)	(0.180)	(0.000)	(0.000)
*Treat * Neg_ROA*	-0.233***	-0.260**	-0.320***	-0.230*	-2.318	-1.321
	(0.004)	(0.019)	(0.004)	(0.074)	(0.866)	(0.927)
*ConScore * Neg_ROA*	-0.151	-0.129	-0.042	0.049	8.693	7.406
	(0.374)	(0.509)	(0.775)	(0.726)	(0.189)	(0.331)
*Treat * ConScore * Neg_ROA*	0.737***	0.841**	0.957***	0.696*	7.451	5.059
	(0.004)	(0.013)	(0.005)	(0.072)	(0.855)	(0.905)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	384	384	388	388	364	364
Adj. R^2	0.079	0.114	0.111	0.135	0.281	0.279
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes

Note: Appendix A provides detailed definitions for all variables. The shaded area represents the coefficient test of interest. Robust *p*-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by firm.

Table 7 Does ownership structure affect the impact of photo-text congruence in annual reports?

Panel A: Low and high institutional ownership

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote_comp	*Vote_comp*	*Vote_dir*	*Vote_dir*	Δ*ROA*	Δ*ROA*
			Low Institutional Ownership			
ConScore * |*Neg ROA*|	5.561**	4.169**	8.265*	6.733*	-5.544	-1.306
	(0.014)	(0.021)	(0.060)	(0.086)	(0.867)	(0.970)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,489	1,202	1,200	1,417	1,415
Adj. R^2	0.356	0.453	0.421	0.493	0.179	0.183
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
			High Institutional Ownership			
ConScore * |*Neg ROA*|	-0.465	0.068	-0.928	0.139	14.771	19.496
	(0.781)	(0.965)	(0.799)	(0.969)	(0.657)	(0.588)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,487	1,202	1,201	1,417	1,415
Adj. R^2	0.202	0.353	0.169	0.204	0.127	0.120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
Wald test			Low - High Institutional Ownership			
X^2	4.48**	3.12*	2.60	1.61	0.19	0.19
p-value	(0.034)	(0.077)	(0.107)	(0.204)	(0.662)	(0.666)

Note: Appendix A provides detailed definitions for all variables. Robust *p*-values in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors are clustered by firm.

Table 7 Does ownership structure affect the impact of photo-text congruence in annual reports?

Panel B: Low and high active institutional ownership

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote_comp	*Vote_comp*	*Vote_dir*	*Vote_dir*	ΔROA	ΔROA
			Low Active Institutional Ownership			
ConScore * |*Neg ROA*|	8.370***	7.910***	14.116***	13.435***	-10.388	1.408
	(0.001)	(0.000)	(0.000)	(0.000)	(0.679)	(0.954)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,486	1,202	1,197	1,417	1,415
Adj. R^2	0.388	0.452	0.471	0.518	0.204	0.215
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
			High Active Institutional Ownership			
ConScore * |*Neg ROA*|	1.271	1.063	-0.280	-0.758	31.647	31.385
	(0.407)	(0.456)	(0.915)	(0.787)	(0.294)	(0.325)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,487	1,484	1,203	1,201	1,415	1,413
Adj. R^2	0.401	0.453	0.241	0.292	0.123	0.120
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
Wald test			Low - High Active Institutional Ownership			
X^2	6.40**	8.49***	11.67***	13.21***	1.31	0.68
p-value	(0.011)	(0.004)	(0.001)	(0.000)	(0.252)	(0.409)

Note: Appendix A provides detailed definitions for all variables. Robust *p*-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by firm.

Table 8 Do intense operating activities affect the impact of photo-text congruence in annual reports?

Panel A: High and low operating expenses

	(1) *Vote_comp*	(2) *Vote_comp*	(3) *Vote_dir*	(4) *Vote_dir*	(5) *ΔROA*	(6) *ΔROA*
			High Operating Expenses			
ConScore * |*Neg ROA*|	6.002***	5.542***	9.542***	8.788***	-1.473	-9.952
	(0.000)	(0.001)	(0.000)	(0.001)	(0.934)	(0.690)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,488	1,202	1,198	1,417	1,416
Adj. R^2	0.295	0.376	0.403	0.452	0.156	0.218
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
			Low Operating Expenses			
ConScore * |*Neg ROA*|	1.235	-0.515	-3.531	-2.328	19.878	27.448
	(0.529)	(0.764)	(0.273)	(0.495)	(0.574)	(0.446)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,489	1,487	1,201	1,198	1,417	1,414
Adj. R^2	0.411	0.458	0.348	0.379	0.166	0.161
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
Wald test			High - Low Operating Expenses			
X^2	3.84**	6.97***	10.49***	6.83***	0.29	0.75
p-value	(0.050)	(0.008)	(0.001)	(0.009)	(0.592)	(0.388)

Note: Appendix A provides detailed definitions for all variables. Robust *p*-values in parentheses: *** *p*<0.01, ** *p*<0.05, * *p*<0.1. Standard errors are clustered by firm.

Table 8 Do intense operating activities affect the impact of photo-text congruence in annual reports?

Panel B: High and low R&D expenses

	(1)	(2)	(3)	(4)	(5)	(6)
	Vote_comp	*Vote_comp*	*Vote_dir*	*Vote_dir*	ΔROA	ΔROA
			High R&D Expenses			
ConScore * |*Neg ROA*|	7.787***	6.471***	10.062***	9.655***	-1.004	4.903
	(0.000)	(0.001)	(0.000)	(0.000)	(0.980)	(0.907)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	744	743	566	563	711	710
Adj. R^2	0.470	0.520	0.358	0.369	0.238	0.242
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
			Low R&D Expenses			
ConScore * |*Neg ROA*|	1.387	1.033	2.921	3.241	18.302	18.788
	(0.449)	(0.548)	(0.391)	(0.289)	(0.488)	(0.489)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,233	2,233	1,839	1,838	2,122	2,120
Adj. R^2	0.339	0.401	0.382	0.437	0.148	0.147
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes	No	Yes
Wald test			High - Low R&D Expenses			
X^2	6.14**	4.90**	3.28*	2.81*	0.17	0.08
p-value	(0.013)	(0.027)	(0.070)	(0.094)	(0.679)	(0.772)

Note: Appendix A provides detailed definitions for all variables. Robust *p*-values in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Standard errors are clustered by firm.