Lie to Me: Video-Based Detection of ESG Washing *

Yiwei Dou
Department of Accounting
New York University
yd18@stern.nyu.edu

Binglei Duan
School of Accountancy
Central University of Finance and Economics
duanbinglei@cufe.edu.cn

Danqi Hu
Guanghua School of Management
Peking University
danqi.hu@gsm.pku.edu.cn

Changtuo (Alto) Luo
Guanghua School of Management
Peking University
luocht@stu.pku.edu.cn

May 2025

* We thank Beth Blankespoor, Shushu Jiang and seminar participants at Peking University and Tsinghua University for very helpful comments. Any errors are our sole responsibility.

Lie to Me: Video-Based Detection of ESG Washing

ABSTRACT

We examine to what extent CEOs' ESG commitment presentations reveal deception cues and thus facilitate the detection of ESG washing. Analyzing videos of bank CEOs' ESG commitment speech made available by the United Nations Principles for Responsible Banking program, we construct deception scores for 32 banks across 19 countries, representing a significant portion of total global bank assets. We find borrowers of banks that have higher deception scores in their commitment videos perform worse on various ESG outcomes including negative ESG incidents, ESG ratings, and emission intensity. The results are robust to controlling for video persuasiveness scores and available bank ESG ratings, and are mainly driven by deception cues in the visual dimension, especially in the eye area, rather than text and audio dimensions. We also find the deception score to be more powerful when bank CEOs experience greater pressure to appear ESG-friendly, are younger, or lack sales experience, consistent with more or stronger deception cues in these situations as suggested by psychology theories. Overall, our evidence indicates the usefulness of video-based deception score in the detection of ESG washing.

Keywords: ESG, commitment, deception, video disclosure

I. INTRODUCTION

There are growing concerns that companies portray their environmental, social, and governance (ESG) activities and commitment opportunistically. This practice, often referred to as "ESG washing" misleads ESG-focused shareholders and stakeholders in directing their capital and resources. Existing research documents the prevalence of ESG washing by showing disconnect between disclosures with actual performance in the ESG space (Basu et al. 2022; Reitmaier et al. 2024). However, this research does not help market participants separate truly ESG committed companies and potential washers ex ante. A few recent studies measure ESG washing in specific dimensions, where it is relatively easy to come up with benchmarks such as diversity (Baker et al. 2024). How to detect misrepresentation in broad ESG activities in a timely manner remains an open question.

In this paper, we take a different approach and explore the usefulness of video-based deception scores in detecting ESG washing. Specifically, we identify deception among CEOs/Chairmen in their ESG commitment videos using machine learning algorithms that analyze vocal, textual, and visual features.[1] Our study draws on previous studies in psychology, video-analytics, and deception detection research, which have shown that there exist significant differences in verbal and nonverbal features between truth tellers and liars. In other words, liars experience different emotional and psychological processes than truth tellers, which gives rise to behavioral deception cues that are hard to completely control or hide. When making ESG commitment, CEOs often already have knowledge through internal discussion and budgeting

[1] Most of our videos feature CEOs. There are a few exceptions where board chairmen provide the ESG commitment speech in the video. We later use "CEO" rather than "CEO/Chairman" ESG commitment videos in the paper for writing simplicity.

of whether their statements are exaggerated (or manipulated). As such, their verbal and nonverbal displays such as micro expressions contain cues that can be used to separate truly committers and potential washers.

While behavioral differences between truth tellers and liars are well-documented in psychology (Ekman 1985; Frank and Ekman 1997; DePaulo 1992), untrained individuals struggle to detect lies accurately, performing only slightly better than chance (Bond and DePaulo 2006). This could be related to suboptimal computations people engage in, relying on their own uninformative lying behaviors instead of more predictive statistical cues (Zheng, Rozenkrantz, and Sharot 2024). In addition, the inaccuracy could be due to the difficulty of identifying micromomentary facial expressions by average people without any training in doing so (Ekman and O'Sullivan 1991; Frank and Ekman 1997). Therefore, an emerging Automatic Deception Detection (ADD) literature starts to explore the usefulness of more objective tools such as machines that can analyze videos frame by frame and be trained to identify micro expressions and weigh various features more objectively. According to prior ADD literature, machine learning models for deception detection using automatic feature extractions from video data achieve an accuracy of more than 70% (e.g., Morales et al. 2017).

To train a machine learning model, we need a data set of visual and verbal features of participants whose deception labels are known. Ideally, we would train the model using CEO video data, but no such data set currently exists.[2] Thus, we have identified Real-Life Trial (RLT) data as the best available alternative. The RLT data was developed by Pérez-Rosas et al.

---

[2] Cheng and Golshan (2025) use conference call data to measure CEO depression. They also face the problem of no available CEO data with depression labels. They thus have to train their deprecation prediction model in another context than conference call CEO setting.

(2015) and it contains 121 labeled individual videos of real deception and real truth during court trials. It is the most widely used datasets of lies during real life scenarios rather than mock experiments or simulated lying experiments in the ADD literature (Morales et al. 2017; Şen et al. 2020; Constâncio et al. 2023).

With the RLT data at hand, we first extract features from the textual, audio, and visual dimensions and then concatenate the features from all three dimensions into a video-level feature vector following Morales et al. (2017). Using the video-level feature vectors as input, and the truthful and deception labels of the RLT data as outcome, we use the Random Forest (RF) model to train a deception detection algorithm that generates a continuous deception score for each video. The trained RF model achieves an accuracy of 75.83%, consistent with prior ADD literature.

To assess CEO's ESG commitment sincerity, we apply the trained deception detection algorithm on a sample of ESG commitment videos. The United Nations (UN) Principes for Responsible Banking (PRB) program, one of the leading sustainable banking frameworks that represents more than half of the global banking industry, encourages signatory banks' CEOs to produce a video disclosure upon signing up for the program. We collect these videos from UN PRB's YouTube channel and process them in a similar manner as the RLT data set to automatically extract the visual, vocal and verbal features. We then run the trained RF model on the extracted features to calculate the deception score that aims to estimate the sincerity level of bank CEOs' ESG commitment.

To validate the machine-empowered deception score of ESG commitment, we compare the ESG alignment practices of banks that are of different degrees of deceptiveness in their commitment video disclosure. Following prior literature, we use borrowers' ESG performance as a proxy for banks' ESG alignment, because numerous studies show that banks integrate ESG in their lending activities by enhancing screening and monitoring of their portfolio boroower companies (e.g., Wang 2023; Kim et al. 2023; Choy et al. 2024). To measure borrowers' ESG performance, we focus on negative ESG incidents, as they are realized outcome and thus more objective and less ambiguous than ESG ratings which are heavily affected by rating methodologies and exhibit significant disagreement across rating providers. Nonetheless, to supplement our main analyses, we also include ESG ratings and carbon emissions (CO2 Emissions) as additional metrics of the borrowers' ESG performance.

We find that the borrowers of banks with higher deception scores in their commitment videos exhibit a greater increase in the number of negative ESG incidents after the video disclosure than those with lower deception scores. In terms of economic significance, a one-standard-deviation increase in the PRB bank's video-based deception scores is associated with a 5% (=0.505×0.099) greater increase in the number of negative incidents in the borrower firms from the pre-commitment-video period to the post-period. In addition to a continuous deception score, we also examine the deception score quartile groups and find that our results are mainly driven by the differences between the top two deception quartile groups and the bottom quartile group, suggesting that our model mainly captures the differences between the extreme truthful group (i.e., bottom 25%) and the more deceptive half of the sample.

We further confirm that differential pre-existing trends before the PRB commitment video do not explain our findings. In addition, the documented patterns remain after we add bank-borrower pair fixed effects, suggesting that the results cannot be solely explained by the changes in bank-borrower pairs, for example the initiation and termination of lending relationships. Taken together, banks with higher video-based deception scores are less likely to "walk the talk" and exert less efforts in disciplining their borrowers' ESG performance than those with lower scores, consistent with them more likely being ESG washers.

We implement several cross-sectional analyses on lies detectability motivated by prior psychology literature to further strengthen our inferences. First, psychology theory suggests that deception cues become more prevalent and pronounced when individuals are highly motivated to be believed (DePaulo et al. 2003; Bond and DePaulo 2006). Based on this, we anticipate that our detection model will perform more effectively when CEOs experience greater pressure to present themselves as ESG-friendly. Consistent with this expectation, our results are primarily driven by banks in countries with heightened environmental and social (E&S) consciousness, where the pressure to appear ESG-friendly is more intense. This increased pressure in turn leads to more detectable deception cues, which are effectively captured by our detection model. Second, according to Vrij et al. (2010), more expressive people tend to be good liars, who exhibit fewer deception cues because they are either more confident and comfortable, or feel less cognitive straining when lying. Using CEOs' sales experience as a proxy for expressiveness, our findings are consistent with Vrij et al. (2010) that CEOs with sales experience exhibit fewer deception cues, making our detection model less useful. Finally, as people age, they tend to become more comfortable with lying and experience less fear or guilt associated with it, resulting in fewer deception cues. Consistent with the prediction, we find that our results only exist in the group of younger CEOs.

To unpack the driving force of our video-based deception score, we decompose the comprehensive video-based score into the scores along the textual-, audio-, and visual-dimensions. We find that our results are driven by the deception scores in the visual dimension rather than the textual and audio dimensions, consistent with visual cues being more powerful in revealing deception than text and audio features. In particular, one would expect the CEOs in our videos would be well prepared for recorded speech different from prior conference call settings. To further investigate the underlying mechanisms of visual features in deception detection, we trained six models, each using only one set of visual features as inputs. The results indicate that models trained with eye features demonstrate strong predictive power for borrowers' ESG performance, aligning with existing psychology research that highlights the eye-related features such dilated pupils as key cues to deception (Hartwig et al. 2011; Khan et al. 2021). In addition, we find that our results are weaker in the subsample of shorter videos or videos with worse facial area recognition quality, confirming that the information in the score likely comes from the videos instead of other omitted variables.

We perform several robustness checks of our results. First, our results are robust to using two alternative ESG performance measures that are relatively more input-related including borrowers' ESG ratings and carbon emission intensities. Second, we confirm that our video-based deception scores capture distinct information beyond what is revealed by the most recent ESG ratings before video disclosure if available. Third, we consider the video-based persuasiveness measure proposed by Hu and Ma (2024). While our measure is motivated to capture the truthfulness of the ESG statements in the video disclosure, the persuasiveness score is more about the emotions of the statements and the consequent impressions on the viewer.

Our results remain after we control for the video-based persuasiveness scores. Fourth, we control for the lying words identified in Larcker and Zakolyukina (2012), and our results remain robust. Finally, our results are robust to using alternative machine learning models to build the deception-detection algorithms, and to different ways of clustering standard errors.[3]

Our paper has the following contributions. First, it contributes to the literature on ESG washing. Previous research documents the prevalence of ESG washing practices, likely due to the difficulty in sanctioning manipulated ESG disclosures (Friedman and Ormazabal 2024). How to detect ESG washing ex ante, especially in the absence of historical or reliable relevant ESG performance data available, is challenging. In this paper, we propose and validate a novel ex ante detection approach based on a machine-learning empowered deception score extracted from video disclosure.

Second, it contributes to the literature on video disclosure. Prior studies on video disclosure have mostly examined the informativeness of video disclosure in capital market settings such as entrepreneurs' pitch videos (Dávila and Guasch 2022; Hu and Ma 2024), CEO interviews with respect to earnings announcements (Banker et al. 2024), and IPO roadshow videos (Blankespoor et al. 2017). These studies examine users' *perception* of visual cues presented in the videos. We contribute to this endeavor by showing the usefulness of deception scores constructed using video disclosures.

In the field of financial disclosure, Duan et al. (2024) apply a trained algorithm to IPO roadshow videos and find that the video-based deception score predicts the IPO fraud. One

[3] We also explore whether our results differ for different types of negative ESG incidents. First, the PRB program calls for the signatory banks to align their activities with the UN Sustainable Development Goals, which cover all three ESG dimensions (i.e., "Environmental", "Social", and "Government" dimensions). Indeed, we find that our results exist and are similar for negative incidents across all three dimensions.

major difference between our setting and theirs is that in the field of ESG, the costs to provide misleading disclosure seem small compared with IPO fraud, as shareholder demands for audits and successful litigations against washers are rare during our sample period. As a result, lying about ESG commitments may yield lower psychological burden and thus fewer deception cues (Ekman 1985). Therefore, ex ante, it is unclear whether the machine-learning based deception score would be effective at detecting ESG washing and we provide initial empirical evidence on this front.

Third, the paper has policy implications. Our results show that bank CEOs' videos at the time of the disclosure, combined with AI deception detection technology, may reveal how truthful/deceptive they are with their ESG statements made in the video, and thus can be used to predict their subsequent ESG efforts in the post-disclosure period. This effective ex ante measure of ESG washing is particularly important, as it may help ESG-focused investors make better decisions of the ESG capital under their management. While our research does not analyze the full costs and benefits of requiring more ESG disclosure in the video format, the evidence suggests one potential benefit of the mandatory video disclosure. That is the better separation of ESG washers ex ante and better ESG capital allocation ex post.

II. MOTIVATION AND PRIOR RESEARCH

ESG Washing

ESG washing practices are likely to be pervasive, as the costs of providing exaggerated or misleading ESG disclosures are unclear (see Friedman and Ormazabal 2024 for a review). Existing studies mostly focus on identifying ESG washing (Kim and Yoon 2021, Raghunandan and Rajgopal 2022, Giannetti et al. 2023) or measuring its intensity (Marquis et al. 2016, Baker et al. 2024) by documenting the disparities between ESG disclosures or commitments and subsequent ESG performance. For example, Giannetti et al. (2023) provide evidence of greenwashing by showing that banks with more environmental disclosures tend to extend more loans to brown borrowers. Baker et al. (2024) propose a measure of diversity-washing, defined as the difference between a firm's DEI-commitment disclosure percentile and its diversity percentile. However, measures and models that could help investors separate ESG committed companies and potential washers ex ante are rare (Briscoe-Tran 2024). In this paper, we propose and validate a novel ex ante detection approach based on a machine-learning empowered deception score extracted from video disclosure.

Psychology Foundations for Deception Detection

Various theoretical perspectives propose reasons why liars exhibit signs of deception. First the emotion and cognition theory of deception suggests that liars may experience arousal and/or an increase in emotions, cognitive load and attempt to control (Zuckerman et al. 1981; Ekman 1985; Vrij 2008). In addition, the self-presentation theory of deception proposes that deceptive presentations are often not as convincingly embraced as truthful ones, and liars typically experience a greater sense of deliberateness (DePaulo 1992). "Liars can be preoccupied with the task of reminding themselves to act the part that truth tellers are not just role-playing but living" (DePaulo et al. 2003, p78). All of these differences in the underlying psychological processes may give rise to cues that can be used to separate liars from truth tellers.[4]

Regarding whether people can hide deception cures, according to Darwin (1965) and Tomkins (1962), expressions can be seen as evolved, biologically based involuntary signs of felt emotions. Even though people, as they grow old, learn to interfere with involuntary emotional expressions to serve social functions and adapt to cultural norms, some micro expressions may still involuntarily appear, leaking the true felt emotions (Ekman and Friesen

[4] We do not aim to use the CEO ESG commitment video settings to differentiate between the different theoretical reasons of why there may exist behavioral cues to separate truthful or deceptive CEOs. From our point of view, all these theoretical reasons are plausible mechanisms of why greenwashing (deceptive) CEOs may exhibit different video cues than sincere (truthful) committers.

1982). Therefore, it is difficult for liars to perfectly control their psychological processes and eliminate all deception cues. In fact, the increased attempt to control and/or the greater sense of deliberateness that a liar experiences may be the exact reasons why they behave differently than truth tellers (DePaulo et al. 2003).

Despite the differences in why deception cues exist among different psychology theories, all the theories predict some cross-sectional differences in the strength of the cues. First, the strength and prevalence of the deception cues may differ across different lying contexts. The cues should be stronger and more prevalent in high-stake situations than in low-stake situations, such as when deception is about transgressions or identities or when senders have greater incentives to be believed (Ekman 1985; DePaulo et al. 2003).

Second, the strength and prevalence of the deception cues may also vary greatly across individuals and personalities. Manipulators are better liars, because they have no moral scruples and are thus less likely to feel fear or guilt associated with lying. Good actors or expressive people are also better at lying as they feel more confident when they lie. Being experienced in lying as people grow old will also increase the comfort level and reduces the fear or guilt felt when lying. Finally, the intelligence level or the degree of preparedness will reduce the cognitive complexity. Overall, it is more difficult to detect lies told by an individual who is more confident and experienced, feels less guilt or fear, or has a lower level of cognitive difficulty when lying.

Individuals, however, perform only marginally better than chance when it comes to lie detection (Bond and DePaulo 2006). This could be related to their overweighting of some cues based on their own lying experiences that are of no statistical significance, or their lack of ability to identify micromomentary expressions as they are brief and fleeting. Both deficiencies can be improved by using machines. In the next section, we review how machine learning and computer vision technologies are adopted in improving the lie detection performance.

Automatic Deception Detection Literature

Regarding the problem that humans may overweigh deception cues that are not statistically significant, a solution is to train a machine learning model that can take a much greater number of deception cues as input and weigh them more objectively. Perez-Rosas et al. (2015) is one such attempt. They compile and introduce RLT data for the first time, establishing the truthful and deceptive labels for speech videos in a real-life context. They then build and evaluate the performance of various machine learning models on linguistic features as well as human-annotated visual features. Their findings indicate that, compared to human performance, machine learning algorithms improve the detection performance the most for the silent video modality (i.e., only using visual features as model input). The evidence is consistent with the difficulties that untrained individuals face when identifying and incorporating fleeting micro expressions in deception detection.[5]

Later research on Automatic Deception Detection (ADD) extends Perez-Rosas et al. (2015) and builds fully automated deception detection system that eliminates the need for human annotations. For example, Morales et al. (2017) propose an open-source multimodal feature extraction tool that can automatically extract features from all the text, audio and video dimensions. They then use the features from all three dimensions to predict sentiment, deception and depression and achieve satisfying accuracy in all exercises. Because of the general effectiveness of their approach across different contexts and predicting tasks, we follow their approach and use the open-source packages provided by them to automatically extract features both in the RLT video data and in the ESG video data.

One limitation of the current automatic deception detection literature is the lack of real-life data with robust and convincing truthful and deceptive labels. As far as we know, RLT is

[5] Training may improve individuals' ability to detect micro expressions. Ekman and O'Sullivan (1991) and Frank and Ekman (1997) both find a positive correlation between the ability to identify micromomentary facial expressions of emotion and the ability to detect deception.

the only real-life data with labels available. Other studies in the literature rely on the data and labels collected from low-stake lab experiments or mocked high-stake experiments.[6] Because of the lack of real-life data with labels and the huge differences between different experiment environments, there is limited evidence on whether the deception detection algorithm trained in the RLT data can really be useful in detecting lying in other real-life contexts. Our paper is one of the first papers to explore this question.

III. SAMPLE, DATA AND VARIABLES

Data and Method for Training the Deception Detection Model

As far as we know, prior studies have not made their video-based multimodal detection models available due to proprietary concerns. Thus, we need a dataset of video recordings of individuals making truthful and deceptive statements with known labels to train a predicting model by ourselves. After carefully reviewing the recent ADD literature, the only video data set that records lying and truth telling in a real-life scenario is the RLT data, and that is why it is the most widely used data in the literature. The reasons that the RLT dataset is most suitable for our study are as follows. First, cues to deception are stronger and more prevalent in high-stake real-life situations such as court trials than in mock or simulated situations (Ekman 1985; DePaulo 1992). Second, the creators of the RLT dataset have invested substantial effort in selecting videos and assigning labels.[7] The trial scenario, characterized by multiple rounds of evidence collection and investigative efforts by police and prosecutors, offers one of the most reliable real-world contexts for determining whether an individual is lying. Finally, all court

[6] Other video data used in the ADD literature include Bag-of-lies (BgL), Box-of-lies (BxL), Mimai University deception detection database (MU3D), all of which are mock or simulated scenarios with low-stake such as lying about likes/dislikes or describing a photo or object. There are some other video data simulating high-stake lying situations such as a mock theft experiment (Tsecpenakis et al. 2005).

[7] The videos in the RLT dataset are sourced from public multimedia platforms featuring trial hearing recordings, where truthful or deceptive behavior can be reliably observed and verified. The video selection process follows a rigorous protocol with strict guidelines. These criteria ensure that the defendant or witness in each video is clearly identifiable, that their face remains visible for the majority of the clip duration, and that the visual quality is sufficient to discern facial expressions accurately.

trails data are collected from US courts and thus all individuals speak English in the video, aligning with the language primarily used in the PRB commitment video.[8]

Using well-established open-source Python packages, we extract features across textual, audio, and visual dimensions following Morales et al. (2017) to ensure replicability and methodological transparency.

For textual features, we first use the Vosk speech recognition toolkit to generate speech transcriptions of the videos. We then apply StanfordNLP to extract syntactic features. Following the ADD literature (e.g., Morales et al. 2017), we obtain 22 syntactic features for each sentence, including: (1) dependency parsing features (2 features), (2) part-of-speech (POS) tags features (18 features), and (3) sentence features (2 features). Panel B of Table OA1 in Online Appendix A details the names and descriptions of the textual features.

For audio features, we first extract the audio tracks (.wav files) from the videos. We then follow Morales et al. (2017) and use the Librosa library to derive 30 audio features for each audio segment.[9] These include: (1) chromagram features (e.g., chroma_stft, 12 features), (2) mel-frequency cepstral coefficients (MFCCs, 13 features), and (3) various spectral features such as spectral centroid, bandwidth, rolloff, root-mean-square energy, and zero-crossing rate (5 features). Panel C of Table OA1 in Online Appendix A details the names and descriptions of the audio features.

For visual features, we utilize OpenFace, a computer vision and machine learning toolkit designed for facial behavior analysis by the CMU MultiComp Lab (Baltrušaitis et al. 2018).

---

[8] In our final sample, 24 of 32 bank CEOs (75%) speak English. There are few situations in which CEOs speak the local languages, particularly those from the EU countries. We translate the transcripts of non-English videos to English before we extract the text-related features using Stanford NLP. The visual and audio features are processed the same way as English videos. Our results hold after removing the 8 CEOs who do not speak English in the videos in a reduced sample of 1,046 observations.

[9] *Librosa* is an open-source audio and music process toolkit (McFee et al. 2015) in Python. https://librosa.org/doc/main/index.html. The two timestamps of two consecutive frames marks the start and end of one audio segment. So the number of audio segments of one video closely mirror the number of frames in the video.

We follow Morales et al. (2017) to extract 709 facial features per frame and OpenFace categorizes these features into six groups: gaze-related information, eye location details, face position, face location details, face shape characteristics, and facial Action Units (AUs), with each category including 8 to 348 features. Figure OA1 provides a visual representation of OpenFace outputs and Panel A of Table OA1 in Online Appendix A details the feature names and descriptions. While all visual features are identified at the frame level, we highlight that we use dynamic modeling to identify AUs by incorporating person-specific facial movements across frames. Specifically, to account for individual differences in baseline expressions, the model calibrates these features using a person-specific neutral expression, estimated as the median of facial descriptors across a video sequence. Furthermore, the model includes correction mechanisms to mitigate both over- and under-prediction of AU activations.[10]

After extracting features from the text, audio, and visual dimensions, we apply a series of statistical functions such as maximum, minimum, mean, 25th percentile, 50th percentile, and 75th percentile to summarize the sentence-level textual features, segment-level audio features, and frame-level visual features into video-level features. This summarization step eventually provides us with one feature vector per video that includes features from all dimensions.

Then, we use the established truthful and deceptive labels in RLT data to train our machine learning models. We include four machine learning algorithms commonly used to predict deception with all visual, audio, and textual features including support vector machines (SVMs), decision trees (DT), random forest (RF) and gradient boosting decision trees (GBDT). Decision Trees model divides the feature space to build a series of partitions organized

---

[10] To corroborate the importance of the dynamic nature of facial AUs, we conduct a placebo test by randomly shuffling the original videos in units of 0.001 frames and re-generating deception scores from the resulting randomly spliced clips. We find in Online Appendix Table OA3 that the detecting power of facial AUs disappears when they are extracted from randomly spliced clips. By contrast, in Table 4 column 6, when deception scores are constructed for facial AUs identified from the original videos, they remain significant in revealing future green washing. The evidence further highlights that static frames alone are not sufficient to extract AU-related deception cues, and it is important to consider the dynamic nature and incorporate the temporal information to construct the visual deception score.

hierarchically into conditions, and Gradient Boosting Decision Trees model is a gradient-boosted version of decision trees model. Random Forest model is an example of ensemble learning (i.e., the prediction outcome is selected by the majority votes) designed to reduce the potential overfitting problem of Decision Trees.[11] We choose to present our main results using the RF model because it achieves the highest accuracy (0.7583) in predicting deception in the RLT data.[12] The accuracy level of our RF model is comparable to the RF models in existing deception detection literature that have an average accuracy of 0.7301 (Constâncio et al. 2023). We also check the robustness of our results using decision trees model.

Data and Method for Measuring Deception in ESG Commitment Videos

To detect deception in CEOs' ESG commitment, we collect bank CEOs' ESG commitment videos from UN PRB's Youtube Channel. The UN PRB aims to encourage banks to align their strategy and business practices with the vision in the UN Sustainable Development Goals (SDGs) and the Paris Climate Agreement. Figure 1 shows the growth in the number of PRB signatory banks across various regions. By 2023, the PRB has over 330 signatory banks, representing more than half of the global banking industry, and has become the world's leading sustainable banking framework.[13]

The PRB requires its signatories to make public announcements at the time of signing the commitments. However, such ex ante textual disclosures tend to be brief and boilerplate, providing little information beyond the publicly announced commitment status. Therefore, the

---

[11] Support Vector Machines (SVMs) divides the feature space into optimum hyperplanes and uses them to make decisions, and the linear kernel is the mostly used kernel in ADD literature. Each algorithm generates a continuous deception probability estimate. For SVMs model, key hyperparameters include: kernel function = linear, and regularization parameter, C = 0.5. For DT model, key hyperparameters include: max_depth = 0.1. For GBDT model, key hyperparameters include: n_estimators = 50, learning_rate=0.01. For RF model, key hyperparameters include: n_estimators = 50.

[12] Following common practices in automatic deception detection, we partition the RLT dataset into training (90%) and testing (10%) subsets using a 10-fold cross-validation approach. Specifically, we divide the sample into 10 groups, where, in each fold, one group serves as the testing subset while the remaining nine groups constitute the training subset. This process is repeated ten times, ensuring that each observation is used for both training and testing exactly once. This approach enhances model stability and mitigates overfitting.

[13] UNEP FI. "Principles for Responsible Banking" Accessed January 2024. https://www.unepfi.org/banking/banking-principles/.

PRB encourages signatory bank CEOs to produce videos explaining why they sign the Principles and what the PRB means to their business.

To help guide the content of the video disclosure, the PRB provides a list of exemplary questions that signatory CEOs can refer to. It includes topics covering motivations for joining the PRB program, views on the PRB principles, implications of the PRB for their business, etc. In addition, the PRB also provides guidelines on various aspects of video production, including the locations (i.e., where the videos should be filmed), the background of the videos, the CEO's posture, the set-up of the camera, microphone, and lighting, etc. These guidelines ensure that the videos are produced under a consistent standard and that the CEO's upper body is clearly visible and speech is audible. The list of exemplary questions and guidelines for video production can be found in Online Appendix B.

Our initial download consists of 77 videos, covering 23% of the PRB banks. The mean (median) video length is 101 (95) seconds, which is similar to the pitch videos in Hu and Ma (2024). The mean (median) number of words in the video transcripts is 204 (183). The videos are timely accessible when signatory banks sign the PRB commitment. About 89.6% of the videos were posted on the YouTube channel within the month of signing or earlier, 97.4% of the videos were posted with no more than a 4-month delay, and only two videos had a maximum delay of 7 months.[14]

To enhance the accuracy of our deception detection methodology and to eliminate potential interference from non-CEO appearances, we preprocess the bank videos by segmenting them into individual clips and retaining only those featuring the CEOs. This approach removes brief segments that may include other individuals. Following Hu and Ma (2024), we sample frames at ten frames per second to identify and compare human faces. The

[14] Online Appendix Table OA2 shows that our results are robust to deleting the observations with delayed video disclosure.

raw frames are processed using a cloud-based face recognition system (i.e., Face++), which outputs multiple face-related measures. Specifically, the face-detection API identifies all faces in the frames, while the face-comparison API determines whether two faces belong to the same individual, achieving an error rate of 0.001%. Subsequently, we employ a video editing package to compile and refine bank videos into individual CEO-exclusive clips. This preprocessing step excludes an average (median) of 9.07 seconds (5.96 seconds) per video, representing an average (median) of 10.54% (7.07%) of the total video duration. After the preprocessing, we extract the visual, audio, and textual features from the PRB videos in the same way as we extract features from the RLT data, and use the trained predictive RF model to calculate visual deception scores for the CEOs featured in the videos. The deception scores for the downloaded CEO commitment videos have a mean of 0.440 and a standard deviation of 0.158.

Other Variables and Data Sources

In line with previous studies, we use the borrowers' ESG performance to capture the banks' ESG alignment, as one significant aspect for banks to incorporate ESG factors is by enhancing their ESG involvement in lending activities, which can be achieved by carefully selecting and monitoring their borrowers (e.g., Choy et al. 2024; Wang 2023; Kim et al. 2023). In our main analyses, we first use realized ESG outcomes to capture borrowers' ESG performance. Specifically, we use negative ESG incidents from the RepRisk database, which screens over 150,000 public sources in 23 languages - such as print media, online media, social media, and regulatory filings - on a daily basis to identify any company or project associated with an ESG risk incident, covering over 250,000 public and private companies from around the world. These negative incidents represent realized outcomes, making them more objective and less ambiguous than ESG ratings, which are heavily influenced by rating methodologies and show significant disagreement across providers (Bams and Kroft 2022; Berg et al. 2022). Besides, the ESG issues identified by RepRisk align with the Ten Principles of the UN Global Compact and the 17 Sustainable Development Goals (SDGs), making them particularly relevant for examining banks' sincerity in their PRB commitments, as the PRB program is UN-led and aims to promote the SDGs.[15] Following Christensen et al. (2023), we focus on incidents classified as "severe" or "highly severe" and exclude those classified as "low severity."[16] We construct *NegIncidents* as the number of negative ESG incidents for borrower firms in a given year.

To complement our main analyses of negative ESG incidents, we also examine borrowers' ESG performance on other dimensions. We collect data on borrowers combined ESG scores, ESG reporting scores, ESG strategy scores, and carbon emissions from Refinitiv ESG. These metrics are arguably more input-based and more reflective of firms' ESG efforts, and combined together, comprehensively assess firms' overall ESG disclosures, policies, and carbon emission practices. By including these additional analyses using more input/effort-based on ESG measures, we avoid failing to capture PRB banks' efforts in developing ESG policies (Christensen et al. 2022), as improving real ESG outcomes generally takes time.

Sample Construction

We start our sample period in 2016, which is the year after the announcement of the Paris Agreement, to mitigate the confounding effect of the Paris Agreement on our results (e.g., Mueller and Sfrappini 2022). We end our sample in 2022, so that we have three years before and after the first wave of PRB commitments in 2019. Our sample begins with firms that have lending relationships with PRB banks that have videos, sourced from the Thomson Reuters Dealscan database. Thomson Reuters Dealscan provides comprehensive loan contract data, including borrower identity and loan characteristics (Bharath et al. 2011). We focus on lead

---

[15] See RepRisk's website for more introductions: https://www.RepRisk.com/research-insights/resources/methodology.

[16] The severity of incidents is assessed along three dimensions: (1) the consequences of the incident in terms of health and safety, (2) the extent of the impact, ranging from individual to a large group or population, and (3) whether the incident was caused by an accident, negligence, or intent.

banks, which play a central role in establishing and maintaining relationships with borrowers and in information collection and monitoring (Sufi 2007). We also exclude financial industry borrowers (SIC=6000-6799).

Our loan-level sample consists of 18,272 loan facilities from 2016 to 2022. For each loan initiation, we assume the bank-borrower relationship persists throughout the loan's lifecycle, following Dou and Xu (2021). This results in 22,884 bank-borrower-year observations between 2016 and 2022. We include only firms that borrow at least one loan before and after 2020 (the year of PRB commitment). We match the borrowers in DealScan with financial data from Worldscope using the link table from Beyhaghi et al. (2021). We also collect banks' accounting information from Bankscope. We exclude observations with missing borrower or bank control variables, or missing data on ESG negative incidents from the RepRisk, as well as those that are either singletons or separated by a fixed effect in the Poison regression. The final sample includes 9,260 bank-borrower-year observations, corresponding to 2,039 bank-borrower lending relationships. The details of our sample selection procedure are reported in Panel A of Table 1.

[Insert Table 1 here]

IV. EMPIRICAL RESULTS

Main Results of Video-Based Deception Scores

Research Design and Descriptive Statistics

In our main test, we investigate whether video-based deception scores can explain committed banks' ESG alignment, reflect in their borrowers' ESG performance. To empirically examine this, we estimate the following regression:

$$NegIncidents_{ijt} = \beta_0 + \beta_1 Post_{it} \times Deception\ Scores_i + Controls + FEs + \varepsilon_{ijt} \qquad (1)$$

where i denotes the bank, j the borrower, in a lending relationship (i.e., with an unmatured loan contract), and t denotes the year. *NegIncidents* is the number of negative ESG incidents of borrower j who has a lending relationship to bank i in year t. *Post* is a dummy variable that equals one after bank i joins the PRB program (i.e., year≥2020).[17] *Deception Scores* represents the deception scores of bank i' video, capturing the likelihood of deception during a bank CEO's PRB ESG commitment disclosure video.

We control for several bank characteristics, including the bank size, capital adequacy, loan loss provisions, etc., and borrower characteristics, including the firm size, profitability, leverage, investment, etc. Recent studies suggest that the ESG-related regulations may affect the ESG performance of lending relationships (Wang 2023; Ivanov et al. 2024), so we control for *Country*×*Year* fixed effects and *Industry*×*Year* fixed effects. If PRB banks are genuinely committed to improving the ESG performance of their lending relationships—either by screening out borrowers with poor ESG performance or through monitoring, such as on-site inspections and private engagement—then their connected borrowers should exhibit fewer negative incidents post-PRB.

Panel B of Table 1 presents the summary statistics of the variables used in our main analysis. We winsorize all continuous variables at the 1% and 99% levels. The mean (median) number of negative incidents is 4 (2) in our sample. 59.4% of the observations are post-PRB. The average size of PRB banks, measured as the natural logarithm of total assets, is 20.723, corresponding to approximately \$5.67 billion. Panel C of Table 1 presents the country distribution of PRB banks in our final sample, which consists of 32 PRB banks across 19 countries. Lending relationships are mainly concentrated among banks headquartered in the

[17] All banks in our analyses joined the PRB program before May 2020, with 31 joining in September 2019 and one in May 2020. Thus, for all our sample banks, *Post* is equal to one for years 2020 and onward (i.e., year≥2020). Deleting the one bank that joined in May 2020 does not affect our inferences.

United States and European countries.

Deception Scores and the Detection of ESG Washing

Table 2 presents the main results comparing the ESG performance of borrowers of PRB banks with different levels of video-based deception scores. We first estimate Equation (1) without including bank and borrower characteristics in column 1, and then add bank and borrower characteristics, *Country*×*Year* fixed effects, and *Industry*×*Year* fixed effects in columns 2-4, respectively. The coefficient estimates on *Post*×*Deception Scores* are all positive and significant at the 1% level, suggesting that PRB banks with higher video-based deception scores exhibit poorer ESG outcomes, as evidenced by more negative incidents among their borrowers. In terms of economic magnitude, the estimate in column 4 indicates that a one-standard-deviation increase in the PRB bank's video-based deception scores is associated with a 5% (=0.505×0.099) higher number of negative incidents in the borrower firms during the post-commitment-video period.[18] In column 5, we further include the *Bank*×*Borrower* fixed effects. We continue to find a positive and significant coefficient estimate on *Post*×*Deception Scores*, suggesting that the performance difference between PRB banks with different deception scores cannot be explained by the screening efforts (i.e., changes in bank-borrower pairs) alone.

[Insert Table 2 here]

Dynamics of PRB Banks' Lending Relationships

One potential concern is that the results reported in Table 2 may simply reflect pre-existing divergent trends in the ESG alignment of PRB banks with different levels of deception scores and have nothing to do with their changes in ESG practices during the post-commitment-video

[18] For comparison, the estimate in column 4 indicates that a one-standard-deviation increase in bank size is associated with a 1% (=0.823×0.011) change in the same outcome (i.e., the number of negative incidents), suggesting that the predictive power of deception scores is economically meaningful.

period. To address this, we examine whether the parallel trends assumption holds in the pre-PRB period. Specifically, we modify Equation (1) by replacing *Post* with six time indicators, each corresponding to a specific year, and interacting them with *Deception Scores*. The six time indicators are *I(year=2016)*, *I(year=2017)*, *I(year=2018)*, *I(year=2020)*, *I(year=2021)*, and *I(year=2022)*. We use *I(year=2019)* as the reference group, which is omitted from the regression. Figure 2 shows the results. Regardless of whether *Bank×Borrower* fixed effects are included, the coefficient estimates on the two-way interactions between the time indicators and *Deception Scores* are all statistically insignificant prior to 2019, and only become significant after 2019 (i.e., after the banks join the PRB program), and the significant effects remain relatively stable during the post-commitment period.

[Insert Figure 2 here]

Cross-Sectional Results

In this section, we examine the cross-sectional variation in the strength and prevalence of deception cues based on previous psychology research on deception.

[Insert Table 3 here]

First, liars' deception cues are stronger and more prevalent when they are more motivated to be believed (Ekman 1985; DePaulo et al. 2003). Prior literature suggests that E&S pressure from local stakeholders may increase firms' incentives to appear more ESG-friendly (Duchin et al. 2024). The increased pressure that managers face to appear as more ESG-friendly in the video disclosure may lead to more deception cues. Specifically, we follow Gantchev et al. (2022) and measure countries' E&S consciousness using the World Value Survey's self-expression score from the most recent survey round (2017–2022) and divide the sample into two subgroups based on the median score.[19] Panel A of Table 3 presents the results. Consistent with our prediction, the coefficient estimates on *Post*×*Deception Scores* are both positive and statistically significant in columns 2 and 4, suggesting that the detecting power of deception scores is mainly concentrated in the subgroup of countries with high E&S consciousness—where bank CEOs face greater pressure to cater to local ESG preferences.[20]

Next, the strength and prevalence of deception cues may also vary greatly across individuals and personalities. Those who experience less guilt, fear, or cognitive stress and those who are more confident and comfortable lying are better liars, leading to fewer deception cues (Ekman 1985; Vrij et al. 2010). We examine whether the performance of deception score decreases (or disappears) among the group of better liars.

First, we are motivated by Vrij et al. (2010) that good actors and salesmen tend to be good liars whose lies are more difficult to detect. Not only are these people usually expressive people, but they are trained to perform convincingly in real-time communications. Panel B of Table 3 presents the results regarding CEOs' sales experiences.[21] We find that the coefficient estimates on *Post* × *Deception Scores* for CEOs with prior salesman experience are not statistically significant and much smaller than those for CEOs without sales experience. The magnitudes of the former group are only 27%-39% of those of the latter group. The coefficient differences between the two subgroups are significant at the 5% level, consistent with CEOs with salesman experience being better liars leaving fewer deception cues.

[19] Gantchev et al. (2022) argue that attitudes toward E&S issues are effectively captured by the survival/self-expression factor, as survival values dominate in societies with low support for gender equality, human rights, and environmental protection, whereas self-expression values dominate in societies with stronger ESG awareness.
[20] Untabulated results show that deposit inflows are more sensitive to banks' PRB commitments in countries with high E&S consciousness, consistent with stronger incentives to misrepresent ESG efforts.
[21] We hand collect information on CEOs' professional background from official bank websites, LinkedIn or Facebook profiles, interview press releases, and other public sources.

Second, individuals may become more skilled at lying as they grow older, as being repeatedly experienced in lying increases their comfort level and reduces the fear or guilt felt when lying. Therefore, we expect that the strength and prevalence of deception cues may become weaker for the older CEOs. We divide the sample based on hand-collected CEO age information, with older CEOs defined as those above the median age and younger CEOs as those below the median.[22] Panel C of Table 3 shows the results. We find that the coefficient estimates on *Post* × *Deception Scores* remain significant for the younger CEO groups but become close to zero for the older CEOs, with the coefficient differences between the two groups significant at the 1% level, consistent with the accumulated experience in lying enhances their ability to lie.

Overall, the results in this section demonstrate that our deception scores perform more effectively when the deception cues in the ESG video disclosure are stronger or more prevalent as suggested by psychology theories. Specifically, the deception scores demonstrate stronger predictive power for future ESG performance when bank CEOs experience greater pressure to appear ESG-friendly, have prior sales experience, or are younger. These findings underscore the core mechanism of our machine-based deception score, which is to identify deception cues leaked during ESG commitment video presentations.

[22] The observations for the six CEOs aged 59 account for about 40% of our sample, ranging from the 35th to 75th percentile. Thus, using the median age of 59 to split the sample creates two imbalanced subgroups. CEOs with age strictly greater than 59 are categorized as the older group, and those aged 59 or below as the younger group. Our inferences remain consistent when we divide the sample into three age groups: older CEOs (top quartile), younger CEOs (bottom quartile), and middle-aged CEOs (middle two quartiles).

V. OTHER EXPLORATORY ANALYSES AND ROBUSTNESS CHECKS

Unpacking the Deception Scores

Visual, Audio, and Textual Features in Detecting Lies

In our main analyses, we focus on the deception scores trained using all video features from the visual, audio, and textual dimensions. In this subsection, we investigate which dimension of video features drives the detecting power of ESG washing. To do this, we construct the visual-based deception score (*Deception Scores_V*), the audio-based deception score (*Deception Scores_A*) and the text-based deception score (*Deception Scores_T*) by training the machine learning models using only the visual, audio, and textual features, respectively. Panel A of Table 4 shows the results. We first include the visual-based, audio-based and textual-based deception scores as independent variables separately. The coefficient estimates on the interaction terms are positive and statistically significant when we use *Deception Scores_V* as the independent variable, while they are close to zero for both *Deception Scores_A* and *Deception Scores_T*. This could be related to the well-preparedness of the ESG commitment videos, especially if the transcripts are pre-written and managers may rehearse reading the transcripts many times, leaving the textual and vocal features less useful for lie detection. By contrast, while preparedness may also reduce the power of visual features, to completely control such features including involuntary micro expressions is much harder.

In columns 4 and 8, we include all deception scores in the model to assess whether the detecting power of the visual deception score is incremental to the audio-based and text-based scores. We continue to find that only the coefficient estimates on *Deception Scores_V* are positive and statistically significant, suggesting that compared to textual and audio features, the visual features are the most useful in detecting CEO's ESG washing in ESG commitment videos.

[Insert Table 4 here]

Different Categories of Visual Features

Since the visual parts of videos are the most informative about CEO deceptions, we next compare the detecting power of different categories of visual features. Following prior literature (Baltrušaitis et al. 2018), we categorize the visual features of videos into six groups: gaze, eyes, facial pose, facial landmarks (LMK), facial shape, and facial AUs. Gaze refers to the eye gaze direction vector in world coordinates. The eye category comprises 56 eye landmarks, which capture the positions of the pupil, iris, and sclera. Facial pose describes the location and rotation of the head. Facial landmarks include the positions of 68 key points outlining the face, mouth, nose, eyes, and brows[23]. Face shape is represented by parameters of a point distribution model that delineates both rigid and non-rigid facial shapes. Facial AUs are a way to describe human facial expressions, such as upper lid raiser, cheek puffer. Using the features extracted from each category, we build classifiers to predict deception. This approach simulates scenarios that evaluate the effectiveness of a deception detection model in predicting ESG washing based on different sets of visual features, thereby identifying which categories of visual features are most important driving forces of the predictive power. Panel B of Table 4 presents the results. The visual features of eye category are incremental to all other visual categories and drive the detecting power of the visual features. The underlying physiological mechanism for the importance of eye cues may lie in the increased blood perfusion in the orbital muscles observed in individuals engaging in deceptive behavior (e.g., Tsiamyrtzis et al. 2007). The findings also align with the broader deception detection literature, which consistently identifies eye-related features including pupil size as important deception cues (Hartwig et al. 2011; Khan et al. 2021).

[23] The positions of eyes in the facial landmark category are designed to capture the relative positions of eyes to other parts of the face. By contrast, features in the eye landmark category are designed to capture eyes-related features only and are more detailed in that aspect.

Video Data Quality and the Detecting Power

We next examine the relationship between video data quality and the detecting power of video-based deception scores. First, we examine the effect of the duration of the video disclosure. The longer the ESG commitment video is, the greater detecting power it likely possesses. This is because when a CEO speaks for a longer time, the chances that she leaves deception cues and that these deception cues are captured by our algorithm should both be higher. Therefore, we expect the power of the video-based deception score in detecting ESG washing to be higher for longer videos. To empirically examine this, we divide our sample into two subgroups based on the median video length in Panel A of Table 5. The coefficient estimates on *Post×Deception Scores* are more than three times larger in the subsample of longer videos than in the subsample of shorter videos. The differences in coefficients are significant at the 1% level.

[Insert Table 5 here]

Second, we examine the effect of video quality, proxied by the degree to which facial features can be accurately measured. To measure the quality of face recognition in these CEO videos, we rely on the blurriness scores provided by Face++, which reflect the degree of clarity in the facial region of an image. The face blurriness is assessed by calculating the high-frequency information in the facial region of the image. High-frequency information refers to the finer details and textures that exhibit rapid changes in the image. Generally, blurred images contain less high-frequency information, whereas clearer images retain more of it. Higher levels of blur often result in the inability to extract clear facial features, which in turn affects the accuracy of facial recognition. We then divide our sample into two subgroups based on the median of blurriness scores. Panel B of Table 5 presents the results. The coefficient estimates on *Post×Deception Scores* are positive and statistically significant in the subsample with high face recognition quality. In contrast, the coefficients are insignificant in the subsample with low face recognition quality. The coefficient differences are significant at the 5% level.

Taken together, these patterns suggest that the detecting power of the video-based deception score varies with the length and data quality of the video disclosure. The fact that our deception score works better when video disclosure is of higher quality and greater length suggests that our results likely come from the videos instead of other omitted variables.

Robustness Tests

Other Measures of Borrower Firms' ESG Performance

In our main analyses, we capture borrower firms' ESG performance through the occurrence of negative ESG incidents, which are not subject to subjective assessments by ESG raters and reflect real ESG outcomes rather than mere "cheaptalks" (Li and Wu 2020). To further corroborate the robustness of our findings, we also test other measures of borrower firms' ESG performance such as ESG ratings and carbon emissions. Compared to negative incidents, these indicators are arguably more controllable by the borrower firms and reflect their different efforts on ESG issues. We then substitute these alternative measures for *NegIncidents* in Equation (1). Table 6 presents the results. Consistent with our baseline findings, PRB banks with higher video-based deception scores continue to exhibit poor ESG performance in their lending relationships, as evidenced by lower ESG combined scores, reporting scores, and strategy scores among their borrowers post-PRB program. Besides, the video-based deception scores also predict PRB banks' efforts in reducing the carbon footprints of their lending relationships. Overall, these results further reinforce our inferences based on negative ESG incidents as a measure of ESG performance.

[Insert Table 6 here]

The Most Recent Available ESG Ratings of PRB Banks

We investigate whether our video-based deception scores have incremental detecting power over other available information sources, in particular, the most recent available ESG ratings of PRB banks. We, however, expect the most recent available ESG ratings (prior to joining the program) of these PRB banks to be less useful in detecting ex post ESG washing. Recent studies cast doubt on commercial ESG ratings, as they mainly reflect the disclosure rather than the true actions on ESG issues (Bams and Kroft 2022). In addition, they are subject to the rewriting of data (Berg et al. 2021) and are strategically influenced by the rated firms (Cornaggia and Cornaggia 2023) and data providers (Li et al. 2024). To compare the usefulness of available ESG ratings to our deception scores, we obtain the ESG combined scores from Refinitiv ESG. Only 47% of the PRB banks in our sample have available scores, so we include a dummy variable, *I(Missing_BankESGratings)*, to indicate the missing ESG scores from Refinitiv. Columns (1) and (2) in Panel A of Table 7 report the results. After controlling for the most recent available ESG ratings of these banks, our video-based deception scores are still significantly associated with the real ESG outcomes of their lending relationships. In comparison, the detecting power using commercial ESG ratings is close to zero.

Video-Based Measures of Persuasion

Next, we examine whether the detecting power of our deception scores can be explained by other video-based measures previously examined in previous research. In particular, Hu and Ma (2024) measure entrepreneurs' persuasiveness using a set of start-up pitch videos. They find that entrepreneurs with positive (e.g., passionate and warm) pitches have a higher funding probability but underperform after receiving funding. We do not expect the persuasion scores to be particularly useful in detecting ESG washing as they mainly focus on the tone of the emotions conveyed by the video and are not designed to capture deception cues. Nonetheless, to measure persuasion, we follow Hu and Ma (2024) and employ principal component analysis (PCA) to extract the first principal component from four dimensions: visual emotion, audio emotion, text emotion, and facial beauty. Online Appendix C describes how we construct this persuasion measure in detail. We then include the persuasion measure (*Persuasiveness_PCA*) and its interaction terms with *Post* in our model. Columns (3) and (4) in Panel A of Table 7 present the results. The video-based deception scores continue to predict the ESG performance of PRB banks' lending relationships, even after controlling for the persuasion measure. Moreover, the persuasion measure contains little information about potential ESG washing, as it captures persuasiveness skills shown during the videos rather than deception cues.

Alternative Deception Scores in the Literature

We examine whether the detecting power of our video-based deception scores can be explained by alternative deception score proposed in prior literature (e.g., Larcker and Zakolyukina 2012).[24] As reported in Panel A of Table 4, we find that our machine-learning based deception scores constructed solely from textual features have limited predictive power for ESG washing. To further corroborate the inferences, we construct the textual-based deception scores using lying and truthful words proposed by Larcker and Zakolyukina (2012)in the conference call setting.[25] Columns (5) and (6) in Panel A of Table 7 show that the detecting power of video-based deception scores remains statistically significant after controlling for the lying words in the videos.

Quartile Ranks of Deception Scores

We use continuous deception scores in our main analyses. In Panel B of Table 7, we also use quartile rank indicators of deception scores to explore the differences from the least deceptive to the most deceptive CEOs. The results suggest that the detecting power of potential ESG washings is primarily driven by the differences between the top two deception quartile

[24] We do not consider the alternative deception score based on vocal dissonance markers as documented in Hobson et al. (2011) because the proprietary algorithms used in their study are not publicly available. Nevertheless, as shown in Panel A of Table 4, our machine-learning-based deception scores constructed solely from audio features (i.e., chromagram features, mel-frequency cepstral coefficients, and various spectral features; Online Appendix Table OA4 describes these audio features in detail) also exhibit limited power in detecting ESG washing.s

[25] The lying and truth word categories are listed in Table 6 of Larcker and Zakolyukina (2012). Specifically, LieWords = (References to general knowledge + Extreme positive emotion + Negations + Certainty + Tentative − Anxiety words − Shareholder value − 3rd person plural pronouns − Impersonal pronouns) / Total word count.

and the bottom quartile. In terms of economic scale, borrower firms of PRB banks in the top deception quartile exhibit 25.4%-26.1% more negative ESG incidents in the post-PRB period, while borrower firms of banks in the second-top deception quartile have 18.3%-19.0% more incidents, both compared to the bottom quartile. The results indicate that our results hold not only with continuous deception scores but also with discrete deception groups.[26]

Alternative Specifications

We test the robustness of our findings to various alternative specifications. First, we construct our video-based deception scores using an alternative machine learning model - Gradient Boosted Decision Trees (GBDT). Online Appendix Table OA4 shows the results. The deception scores trained by the GBDT model exhibit very similar detecting power to our baseline deception scores. Second, we explore alternative clustering of standard errors in Online Appendix Table OA5. Our findings remain robust to different clustering schemes. Finally, we examine whether our results are driven by specific dimensions of negative ESG incidents. RepRisk categorizes negative incidents into environmental, social, governance, and cross-cutting issues.[27] The results reported in Online Appendix Table OA6 suggest that the video-based deception scores of PRB banks are associated with an increase in their borrowers' negative incidents across all four dimensions.[28]

VI. CONCLUSION

Using banks' ESG commitment video disclosure, this paper proposes a video-based deception score as an ex ante measure of banks' potential ESG washing. To construct the score, we take advantage of the video processing technologies and advanced machine learning

[26] The means of video-based deception scores in quartiles 1-4 are 0.250, 0.388, 0.500, and 0.592, respectively.
[27] Cross-cutting issues refer to those spanning multiple dimensions of ESG.
[28] Untabulated results show that the detecting power of our deception scores remains significant after we include all interaction terms between bank characteristics and the *Post* dummy, alleviating concerns that the observed effects are driven by banks with specific characteristics changing their lending behavior in the post-commitment-video period.

algorithms trained in widely adopted real-life court trial settings to incorporate features in all visual, audio, and textual dimensions to predict the level of deception of the video disclosure. To empirically assess the usefulness of the video-based score in detecting ESG washing, we examine the ESG performance of banks with different levels of deception during the post-video-disclosure period. Following prior literature, we measure banks' ESG performance using their borrowers' ESG performance (Wang 2023; Kim et al. 2023; Choy et al. 2024). Across various ESG performance measures including outcome-based negative ESG incidents as well as relatively more input-based carbon emissions and ESG ratings, we find that banks with higher deception scores exhibit worse ESG performance during the post-disclosure period. In addition, we find that our results are more pronounced when the video disclosure is longer and only present when the facial recognition quality is high, suggesting that video quality and length is crucial for the video-based deception score to be useful. Furthermore, we investigate how banks' incentives to misrepresent their ESG commitments influence the detectability of deception. The findings reveal that deception cues are more pronounced in banks from countries with high E&S consciousness where the pressure to appear ESG-compliant is greater. This aligns with psychology research showing that heightened motivation to deceive strengthens observable deception cues. In addition, consistent with the psychology theory that good liars such as those with prior sales experience or more lying experiences may exhibit fewer deception cues, we find that the detecting power of video-based deception scores becomes significantly weaker for the salesmen-experienced and older CEOs.

To unpack the video-based deception score, we first separately examine the detecting power of the visual, textual, and audio dimensions of the videos alone. The results show that only the deception score based on the visual dimension is powerful at detecting ESG washing. To further explore which subset of the visual features explains our results, we train six distinct models, with each one solely using one set of visual features as inputs. The outcomes reveal that the model trained with eye features is the most powerful in predicting banks' ESG performance. This is consistent with psychology research that emphasizes the significance of eye movement as a primary indicator in unmasking deception.

Our results are robust to controlling for available bank ESG rating, other video-based measures in prior literature such as persuasion score, and lying words identified from conference calls, and to the use of quartile groups of the deception scores instead of the continuous version used in the main analyses. All our results hold with or without bank-borrower pair fixed effects. This means that our results still exist even after we fix the lending relationship, suggesting that our results cannot be explained solely by the screening effects (i.e., changes in the bank-borrower pair such as initiation or termination of lending relationships) and could also be related to the monitoring effects. We nonetheless do not focus on the separation of these mechanisms of how banks affect borrowers' ESG performance, which has already been widely examined in prior literature. In addition, as theories provide no clear guidance on how bank ESG commitment sincerity affect these mechanisms differently, we leave these interesting questions of exploratory nature to future research.

REFERENCES

Baker, A. C., D. F. Larcker, C. G. McClure, D. Saraph, and E. M. Watts. 2024. Diversity washing. *Journal of Accounting Research* 62 (5): 1661-1709.

Baltrušaitis, T., Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. *IEEE International Conference on Automatic Face and Gesture Recognition.*

Bams, D., and B. van der Kroft. 2022. Tilting the wrong firms? How inflated ESG ratings negate socially responsible investing under information asymmetries. MIT Center for Real Estate Research Paper, 22/12.

Banker, R. D., Ding, H., Huang, R., and Li, X. 2024. Market reaction to CEOs' dynamic hemifacial asymmetry of expressions. *Management Science*, 70(7), 4847-4874.

Basu, S., Vitanza, J., Wang, W., and Zhu, X. R. 2022. Walking the walk? Bank ESG disclosures and home mortgage lending. *Review of Accounting Studies*, 27(3), 779-821.

Berg, F., J. F. Koelbel, and R. Rigobon. 2022. Aggregate confusion: The divergence of ESG ratings. *Review of Finance* 26(6): 1315-1344.

Berg, F., K. Fabisik, and Z. Sautner. 2021. Is history repeating itself? The (un) predictable past of ESG ratings. Working paper.

Beyhaghi, M., R. Dai, A. Saunders, and J. Wald. 2021. International lending: The role of lender's home country. *Journal of Money, Credit and Banking* 53(6): 1373-1416.

Bharath, S. T., S. Dahiya, A. Saunders, and A. Srinivasan. 2011. Lending relationships and loan contract terms. *The Review of Financial Studies* 24(4): 1141-1203.

Blankespoor, E., B. E. Hendricks, and G. S. Miller. 2017. Perceptions and price: Evidence from CEO presentations at IPO roadshows. *Journal of Accounting Research* 55(2): 275-327.

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

Briscoe-Tran, H. (2024). Do employees have useful information about firms' Esg practices?. Fisher College of Business Working Paper, (2021-03), 21.

Cheng, S. Y., and Golshan, N. M. 2025. Silent Suffering: Using Machine Learning to Measure CEO Depression. *Journal of Accounting Research*.

Choy, S., S. Jiang, S. Liao, and E. Wang, 2024. Public environmental enforcement and private lender monitoring: Evidence from environmental covenants. *Journal of Accounting and Economics* 77(2-3): 101621.

Christensen, H. B., De George, E. T., Joffre, A., and Macciocchi, D. 2023. Consumer Responses to the Revelation of Corporate Social Irresponsibility. University of Miami Business School Research Paper, (4496599).

Christensen, D. M., G. Serafeim, and A. Sikochi. 2022. Why is corporate virtue in the eye of the beholder? The case of ESG ratings. *The Accounting Review* 97(1): 147-175.

Cohn, J. B., Z. Liu, and M. I. Wardlaw. 2022. Count (and count-like) data in finance. *Journal of Financial Economics* 146(2): 529-551.

Constâncio, A. S., Tsunoda, D. F., Silva, H. D. F. N., Silveira, J. M. D., and Carvalho, D. R. 2023. Deception detection with machine learning: A systematic review and statistical analysis. *Plos One*, 18(2), e0281323.

Cornaggia, J., and K. Cornaggia. 2023. ESG ratings management. Working paper.

Darwin, C. 1965. The expression of the emotions in man and animals. Chicago: University of Chicago Press. (Original work published 1872).

Dávila, A., and M. Guasch. 2022. Managers' body expansiveness, investor perceptions, and firm forecast errors and valuation. *Journal of Accounting Research* 60(2): 517-563.

DePaulo, B. M. 1992. Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2), 203.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. 2003. Cues to deception. *Psychological Bulletin*, 129(1), 74.

Dou Y, Z. Xu. 2021. Bank lending and corporate innovation: Evidence from SFAS 166/167. *Contemporary Accounting Research* 38(4): 3017-3052.

Duan, B., D. Hu, and H. Lu. 2024. Video-Based Deception Detection and Financial Fraud. Working paper.

Duchin, R., Gao, J., and Xu, Q. 2024. Sustainability or greenwashing: Evidence from the asset market for industrial pollution. *The Journal of Finance*.

Ekman, P. 1985. Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (revised edition). New York: Norton.

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

Ekman, P., and W. V. Friesen. 1982. Felt, false, and miserable smiles. *Journal of Nonverbal Behavior* 6(4): 238–258.

Ekman, P., AND O'Sullivan, M. (1991). Who can catch a liar?. *American Psychologist*, 46(9), 913.

Frank, M. G., and Ekman, P. 1997. The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72(6), 1429.

Friedman, H. L., and Ormazabal, G. 2024. The role of information in building a more sustainable economy: A supply and demand perspective. *Journal of Accounting Research*, 62(5), 1575-1609.

Gantchev, N., Giannetti, M., and Li, R. 2022. Does money talk? Divestitures and corporate environmental and social policies. *Review of Finance*, 26(6), 1469-1508.

Giannetti, M., Jasova, M., Loumioti, M., and Mendicino, C. 2023. "Glossy green" banks: the disconnect between environmental disclosures and lending activities. Banks: The Disconnect between Environmental Disclosures and Lending Activities. ECB Working Paper, (2023/2882).

Hartwig, M., and C. F. Bond Jr. 2011. Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin* 137(4): 643-659.

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

Hu, A., and S. Ma. 2024. Persuading Investors: A Video-Based Study. *Journal of Finance*, forthcoming.

Ivanov, I. T., M. S. Kruttli, and S. W. Watugala. 2024. Banking on carbon: Corporate lending and cap-and-trade policy. *The Review of Financial Studies* 37(5): 1640-1684.

Khan, W., Crockett, K., O'Shea, J., Hussain, A., and Khan, B. M. 2021. Deception in the eyes of deceiver: A computer vision and machine learning based automated deception detection. *Expert Systems with Applications*, 169, 114341.

Kim, S., Kumar, N., Lee, J., and Oh, J. 2022. ESG lending. In Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC, European Corporate Governance Institute – Finance Working Paper (No. 817).

Kim, S., and A. Yoon. 2023. Analyzing active fund managers' commitment to ESG: Evidence from the United Nations Principles for Responsible Investment. *Management Science* 69(2): 741-758.

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

Li, J., and D. Wu. 2020. Do corporate social responsibility engagements lead to real environmental, social, and governance impact?. *Management Science* 66(6): 2564-2588.

Li, X., Y. Lou, and L. Zhang. 2024. Do commercial ties influence ESG ratings? Evidence from

Moody's and S&P. *Journal of Accounting Research* 62(5): 1901-1940.

Marquis, C., Toffel, M. W., and Zhou, Y. 2016. Scrutiny, norms, and selective disclosure: A global study of greenwashing. *Organization Science*, 27(2), 483-504.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., and Nieto, O. 2015. librosa: Audio and music signal analysis in python. *SciPy*, 2015, 18-24.

Morales, M. R., S. Scherer, and R. Levitan. 2017. OpenMM: An Open-Source Multimodal Feature Extraction Tool. In: *Conference of the International Speech Communication Association (INTERSPEECH)*, 3354–3358.

Mueller, I., and E. Sfrappini. 2022. Climate change-related regulatory risks and bank lending. Working paper.

Pérez-Rosas, V., M. Abouelenien, R. Mihalcea, and M. Burzo. 2015. Deception detection using real-life trial data. In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 59–66.

Raghunandan, A., and Rajgopal, S. 2022. Do ESG funds make stakeholder-friendly investments?. *Review of Accounting Studies*, 27(3), 822-863.

Reitmaier, C., W. Schultze, and J. Vollmer. 2024. Corporate responsibility and corporate misbehavior: Are CSR reporting firms indeed responsible? *Review of Accounting Studies*: 1-69.

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

Sufi, A. 2007. Information asymmetry and financing arrangements: Evidence from syndicated loans. *The Journal of Finance* 62(2): 629-668.

Tsechpenakis, G., Metaxas, D., Adkins, M., Kruse, J., Burgoon, J. K., Jensen, M. L., ... and Nunamaker, J. F. 2005, July. HMM-based deception recognition from visual cues. In 2005 IEEE International Conference on Multimedia and Expo (pp. 824-827). IEEE.

Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. T., Frank, M. G., and Ekman, P. 2007. Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*, 71, 197-214.

Tomkins, S. 1962. Affect imagery consciousness: Volume I: The positive affects. Springer publishing company.

Tsiamyrtzis, P., J. Dowdall, D. Shastri, I. T. Pavlidis, M. G. Frank, and P. Ekman. 2007. Imaging facial physiology for the detection of deceit. *International Journal of Computer Vision*

71(2): 197-214.

Vrij, A. 2008. Detecting lies and deceit: Pitfalls and opportunities. John Wiley & Sons.

Vrij, A., Granhag, P. A., and Mann, S. 2010. Good liars. *The Journal of Psychiatry & Law*, 38(1-2), 77-98.

Walczyk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. 2003. Cognitive mechanisms underlying lying to questions: Response time as a cue to deception. Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 17(7), 755-774.

Wang, L. L. 2023. Transmission effects of ESG disclosure regulations through bank lending networks. *Journal of Accounting Research* 61(3): 935-978.

Zheng, S. Y., Rozenkrantz, L., and Sharot, T. 2024. Poor lie detection related to an under-reliance on statistical cues and overreliance on own behaviour. *Communications Psychology*, 2(1), 21.

Zuckerman, M., B. M. DePaulo, and R. Rosenthal. 1981. *Verbal and nonverbal communication of deception.* In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 14, pp. 1–59). New York: Academic Press.

Figure 1. The growth of PRB banks around the world

This figure shows the growth of PRB signatories around the world from 2019 to 2023. The information is available on the PRB's website: https://www.unepfi.org/banking/prbsignatories/.

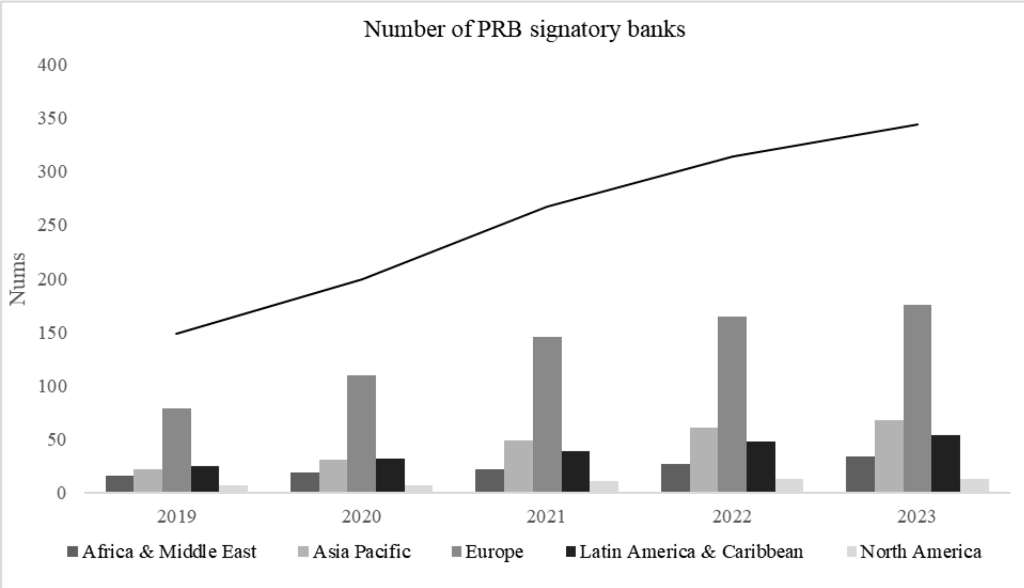
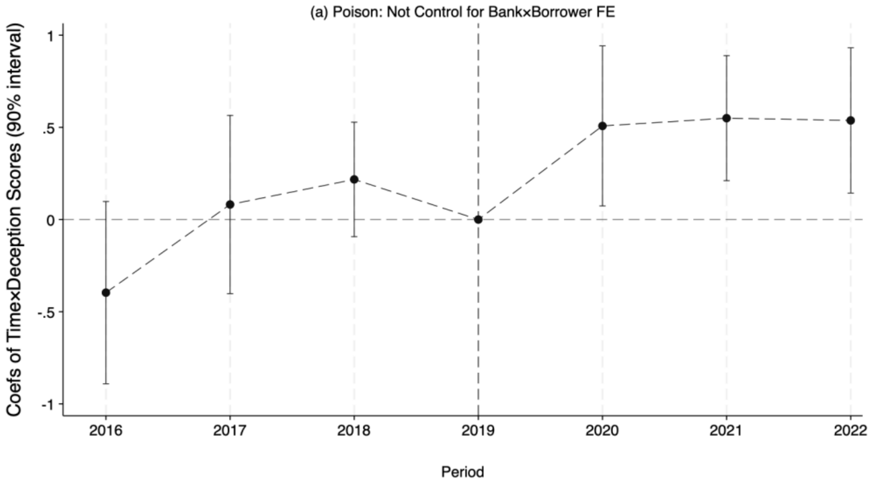


Figure 2. Dynamic effects of the video-based deception scores of PRB banks

This figure shows the Poison regression coefficients and the 10% confidence intervals. We include the full set of control variables that are consistent with Table 2. We include six time indicator variables to substitute *Post*: *I(year=2016)*, *I(year=2017)*, *I(year=2018)*, *I(year=2020)*, *I(year=2021)*, and *I(year=2022)*. We use *I(year=2019)* as the reference group and omit it from the regression.



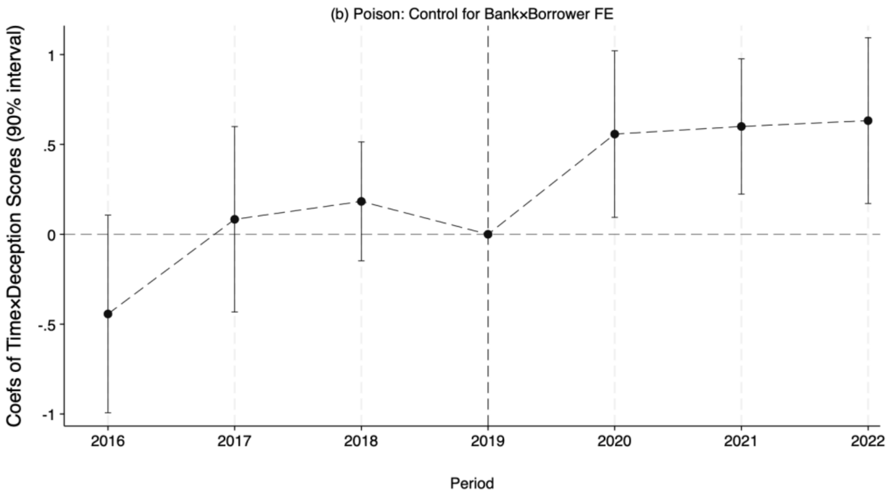


Table 1. Sample selection and descriptive statistics

This table summarizes the sample selection and descriptive statistics of our main analyses. Panel A shows the sample selection procedures. Panel B provides the descriptive statistics of the variables used in our main tests. Panel C reports the country distribution of PRB banks in our final sample. "No. of Bank-Borrower Pairs" refers to the number of borrowers that have lending relationships with PRB banks headquartered in each country. The sample consists of all lending relationships of PRB banks with videos in the period of 2016-2022. All variable definitions are listed in Appendix A.

Panel A. Sample selection

	Obs.
Initial Sample: loans data from DealScan between 2016 to 2022 (PRB banks with videos)	25,180
Less: Non-lead banks in the loan contracts	4,008
Less: Borrowers are financial firms (6000-6999)	2,900
Final loan-level observations for analysis	18,272
Initial Sample: Bank-Borrower-year level observations between 2016 to 2022 (PRB banks with videos)	22,884
Less: Borrowers that do not borrow any loans before or after the PRB program	565
Less: Observations with missing bank-level or borrower-level control variables	3,280
Less: Borrowers without ESG data	8,376
Less: Observations that are either singletons or separated by a fixed effect	1,403
Final Bank-Borrower-Year level observations for analysis	9,260
Bank-Borrower pairs in the final sample	2,039

Panel B. Descriptive statistics

	Obs.	Mean	SD	P25	Median	P75
NegIncidents	9,260	4.148	7.011	1.000	2.000	5.000
Deception Scores	9,260	0.498	0.099	0.412	0.530	0.588
Post	9,260	0.594	0.491	0.000	1.000	1.000
BankSize	9,260	20.723	0.823	20.385	21.004	21.134
LoanGr	9,260	2.332	5.486	-0.640	2.610	5.690
BankROE	9,260	7.053	4.748	4.480	8.060	11.060
Tier1	9,260	15.313	2.363	13.490	15.040	17.030
LoanRatio	9,260	39.989	13.206	32.000	37.670	50.220
NII	9,260	43.288	15.551	28.770	42.110	53.470
LLP	9,260	0.019	0.010	0.011	0.019	0.025
FirmSize	9,260	24.096	1.802	22.905	23.986	24.975
Lev	9,260	0.350	0.167	0.236	0.332	0.448
ROA	9,260	0.031	0.072	0.007	0.033	0.062
Current	9,260	0.331	0.176	0.203	0.306	0.429
InteCover	9,260	9.467	21.286	1.688	4.778	10.290
SGA	9,260	0.120	0.130	0.036	0.088	0.160

RD	9,260	0.011	0.020	0.000	0.001	0.015
CAPX	9,260	0.037	0.028	0.017	0.031	0.049

Panel C. Country distribution of PRB banks in our final sample

Bank Headquarters Country	No. Banks	No. of Bank-Borrower Pairs	No. Observations
Belgium	1	19	67
Brazil	2	10	49
Canada	1	1	4
Denmark	1	50	194
Finland	1	13	44
France	3	197	907
Germany	2	272	1,154
Ireland	1	7	22
Italy	1	4	20
Mauritius	1	1	4
Netherlands	1	62	264
Norway	2	48	182
Portugal	1	1	3
South Africa	2	20	80
Spain	5	318	1,299
Sweden	1	73	300
Switzerland	2	166	671
United Kingdom	2	379	1,771
United States	2	398	2,225
Total	32	2,039	9,260

Table 2. The video-based deception scores and the real ESG outcomes

This table reports the results of using the video-based deception scores of PRB banks to evaluate the real ESG outcomes of their lending relationships during the post-video-disclosure period. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of a bank's borrower. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)
Dep. var. =			*NegIncidents*		
Post×Deception Scores	0.544***	0.542***	0.516***	0.505***	0.573***
	(0.209)	(0.206)	(0.188)	(0.167)	(0.201)
BankSize	-0.011	-0.012	-0.017	-0.011	-0.010
	(0.020)	(0.020)	(0.018)	(0.016)	(0.018)
LoanGrowth	0.002	0.002	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
BankROE	0.003	0.003	0.003	0.003	0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Tier1	0.010	0.009	0.010	0.010	0.011
	(0.011)	(0.011)	(0.010)	(0.009)	(0.010)
LoanRatio	0.002	0.003	0.002	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
NII	0.000	0.000	0.000	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
LLP	-2.054	-2.155	-1.852	-1.957	-2.498
	(2.122)	(2.142)	(1.750)	(1.671)	(1.931)
FirmSize		0.181***	0.275***	0.231***	0.196**
		(0.061)	(0.071)	(0.072)	(0.078)
Lev		0.262	0.210	0.229	0.237
		(0.177)	(0.196)	(0.191)	(0.209)
ROA		0.767***	0.471**	0.509**	0.537**
		(0.178)	(0.209)	(0.220)	(0.226)
Current		-0.410**	-0.578**	-0.493*	-0.344
		(0.198)	(0.250)	(0.258)	(0.263)
InteCover		-0.001*	-0.001	-0.001	-0.001
		(0.001)	(0.001)	(0.001)	(0.001)
SGA		-0.444	-0.454	-0.234	-0.152
		(0.337)	(0.352)	(0.367)	(0.412)
RD		7.970***	8.420***	6.838***	6.240**
		(2.503)	(2.486)	(2.567)	(2.586)
CAPX		0.906	-0.005	0.967	1.159
		(0.811)	(0.818)	(0.841)	(0.900)

Bank FE	Yes	Yes	Yes	Yes	-
Borrower FE	Yes	Yes	Yes	Yes	-
Bank×Borrower FE	-	-	-	-	Yes
Year FE	Yes	Yes	-	-	-
Country×Year FE	-	-	Yes	Yes	Yes
Industry×Year FE	-	-	-	Yes	Yes
N	9,260	9,260	9,260	9,260	8,843
Adj. R^2	0.638	0.639	0.658	0.662	0.663

Table 3. Cross-sectional analyses: the strength and prevalence of the deception cues

This table reports the cross-sectional results based on the local E&S consciousness, CEOs' salesman experience, and CEOs' age. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of a bank's borrower. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

Panel A. Local E&S consciousness and the detecting power

	(1)	(2)	(3)	(4)
Sample =	High E&S consciousness	Low E&S consciousness	High E&S consciousness	Low E&S consciousness
Dep. var. =	*NegIncidents*			
Post×Deception Scores	0.644**	0.295	0.634*	0.299
	(0.286)	(0.350)	(0.333)	(0.423)
Dif =	0.358***		0.334**	
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	-	-
Borrower FE	Yes	Yes	-	-
Bank×Borrower FE	-	-	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	3,702	5,418	3,620	5,169
Adj. R^2	0.672	0.657	0.675	0.657

Panel B. Salesman experience and the detecting power

	(1)	(2)	(3)	(4)
Sample =	I (Salesman Experience)=1	I (Salesman Experience)=0	I (Salesman Experience)=1	I (Salesman Experience)=0
Dep. var. =	*NegIncidents*			
Post×Deception Scores	0.174	0.446**	0.125	0.469**
	(1.081)	(0.180)	(1.170)	(0.218)
Dif =	-0.272**		-0.344**	
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	-	-
Borrower FE	Yes	Yes	-	-

Bank×Borrower FE	-	-	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	3,373	5,743	3,278	5,505
Adj. R^2	0.652	0.670	0.656	0.670

Panel C. Ages and the detecting power

	(1)	(2)	(3)	(4)
Sample =	Older CEOs	Younger CEOs	Older CEOs	Younger CEOs
Dep. var. =	*NegIncidents*			
Post×Deception Scores	-0.170	0.503***	-0.450	0.583***
	(0.693)	(0.180)	(0.796)	(0.215)
Dif =	-0.674***		-1.032***	
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	-	-
Borrower FE	Yes	Yes	-	-
Bank×Borrower FE	-	-	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	2,167	6,964	2,100	6,686
Adj. R^2	0.639	0.670	0.637	0.672

Table 4. The usefulness of different features in constructing the deception scores

This table reports the results of comparing the usefulness of different features in training the deception scores. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. In Panel A, we compare the deception scores that are trained using only the visual, audio, and textual features, respectively. In Panel B, we compare the deception scores that are trained using specific categories of visual features in the videos. We divide the visual features of videos into six categories: gaze, eye, face pose, face LMK, face shape, and facial AUs. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

Panel A. Deception scores based on visual, audio, and textual features in the videos

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var. =				*NegIncidents*				
Post×Deception Scores_V	0.323**			0.339**	0.346**			0.361*
	(0.129)			(0.159)	(0.155)			(0.188)
Post×Deception Scores_A		0.252		-0.027		0.283		-0.010
		(0.193)		(0.237)		(0.229)		(0.276)
Post×Deception Scores_T			-0.016	-0.054			-0.069	-0.104
			(0.134)	(0.134)			(0.161)	(0.161)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes				
Borrower FE	Yes	Yes	Yes	Yes				
Bank×Borrower FE					Yes	Yes	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9,260	9,260	9,260	9,260	8,843	8,843	8,843	8,843
Adj. R^2	0.662	0.662	0.662	0.662	0.663	0.663	0.663	0.663

Panel B. Deception scores based on specific categories of visual features

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var. =				*NegIncidents*			
Post×Deception Scores_Gaze	0.031						-0.003
	(0.066)						(0.126)
Post×Deception Scores_Eye		0.380***					0.711*
		(0.133)					(0.416)
Post×Deception Scores_FacePose			0.118				-0.014
			(0.073)				(0.216)
Post×Deception Scores_FaceLMK				0.188			-0.322
				(0.115)			(0.347)
Post×Deception Scores_FaceShape					-0.223*		-0.187
					(0.124)		(0.277)
Post×Deception Scores_FacialAU						0.483**	0.039
						(0.200)	(0.404)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9,260	9,260	9,260	9,260	9,260	9,260	9,260
Adj. R^2	0.662	0.662	0.662	0.662	0.662	0.662	0.663
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Dep. var. =				*NegIncidents*			
Post×Deception Scores_Gaze	0.059						0.050
	(0.080)						(0.153)
Post×Deception Scores_Eye		0.411**					0.811*

		(0.161)					(0.488)
Post×Deception Scores_FacePose			0.145*				-0.049
			(0.087)				(0.256)
Post×Deception Scores_FaceLMK				0.174			-0.383
				(0.140)			(0.410)
Post×Deception Scores_FaceShape					-0.314**		-0.285
					(0.155)		(0.328)
Post×Deception Scores_FacialAU						0.609**	0.043
						(0.242)	(0.471)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank×Borrower FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	8,843	8,843	8,843	8,843	8,843	8,843	8,843
Adj. R^2	0.663	0.663	0.663	0.663	0.663	0.663	0.663

Table 5. Video data quality and the detecting power of video-based deception scores

This table reports the cross-sectional results based on the video duration and face recognition quality. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of a bank's borrower. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

Panel A. Video duration and the detecting power

	(1)	(2)	(3)	(4)
Sample =	High Video Duration	Low Video Duration	High Video Duration	Low Video Duration
Dep. var. =		*NegIncidents*		
Post×Deception Scores	1.251***	0.366*	1.338**	0.375*
	(0.481)	(0.189)	(0.542)	(0.226)
Dif =	0.885***		0.963***	
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	-	-
Borrower FE	Yes	Yes	-	-
Bank×Borrower FE	-	-	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	3,591	5,513	3,474	5,314
Adj. R^2	0.636	0.679	0.642	0.678

Panel B. The quality of face recognition and the detecting power

	(1)	(2)	(3)	(4)
Sample =	High Recognition Quality	Low Recognition Quality	High Recognition Quality	Low Recognition Quality
Dep. var. =		*NegIncidents*		
Post×Deception Scores	0.430**	0.241	0.543**	0.217
	(0.202)	(0.436)	(0.246)	(0.482)
Dif =	0.188**		0.326**	
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	-	-
Borrower FE	Yes	Yes	-	-
Bank×Borrower FE	-	-	Yes	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	5,744	3,389	5,507	3,267
Adj. R^2	0.668	0.655	0.668	0.659

Table 6. Other ESG performance of lending relationships

This table reports the results of using video-based deception scores of PRB banks to evaluate the other ESG performance of their ex post lending relationships. The dependent variables are *Borrower ESG Combined Ratings*, *Borrower ESG reporting Ratings*, *Borrower ESG strategy Ratings*, and *Borrower Co2 Intensity*. *Borrower Combined ESG Ratings* is the borrower firm's ESG combined score in the year, which captures the overall ESG performance of the borrower firms. *Borrower ESG Reporting Ratings* is the borrower firm's ESG disclosure score in the year, which captures the borrower firms' ESG disclosure performance. *Borrower ESG Strategy Ratings* is the borrower firm's ESG strategy score in the year, which reflects borrower firms' practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes. *Borrower Co2 Intensity* is the borrower firms' total Co2 and Co2-equivalent emissions (in thousands of tons), scaled by sales (in millions) in the year. We run these regressions using the OLS model. A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var. =	*Borrower Combined ESG Ratings*		*Borrower ESG Reporting Ratings*		*Borrower ESG Strategy Ratings*		*Borrower Co2 Intensity*	
Post×Deception Scores	-6.703***	-7.297**	-10.424**	-11.798***	-9.474**	-9.475**	0.485**	0.561**
	(2.575)	(2.840)	(4.133)	(4.567)	(3.826)	(4.208)	(0.238)	(0.268)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	-	Yes	-	Yes	-	Yes	-
Borrower FE	Yes	-	Yes	-	Yes	-	Yes	-
Bank×Borrower FE	-	Yes	-	Yes	-	Yes	-	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	10,795	10,354	10,795	10,354	10,795	10,354	10,795	10,354
Adj. R^2	0.800	0.758	0.688	0.622	0.872	0.847	0.853	0.828

Table 7. Robustness tests

This table reports the results of various robustness checks. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. In Panel A, we re-evaluate the video-based deception scores after controlling for the most recent available ESG ratings of PRB banks, the video-based persuasiveness scores suggested by Hu and Ma (2024), and the use of lying words as categorized by Larcker and Zakolyukina (2012), respectively. In Panel B, we use the quartile ranks of the video-based deception scores to substitute the continuous deception scores. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

Panel A. Controlling for banks' ESG ratings, video-based persuasion, and lying words

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var. =			*NegIncidents*			
Post×Deception Scores	0.575***	0.664***	0.481***	0.535***	0.539***	0.621***
	(0.180)	(0.218)	(0.174)	(0.207)	(0.172)	(0.207)
Post×Avail_BankESGratings	-0.002	-0.003				
	(0.002)	(0.003)				
Post×I(Missing_BankESGratings)	-0.123	-0.133				
	(0.128)	(0.157)				
Post×Persuasiveness_PCA			-0.006	-0.010		
			(0.013)	(0.015)		
Post×LieWords					-0.756	-1.049
					(0.857)	(1.001)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	-	Yes	-	Yes	-
Borrower FE	Yes	-	Yes	-	Yes	-
Bank×Borrower FE	-	Yes	-	Yes	-	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes
N	9,260	8,843	9,260	8,843	9,260	8,843
Adj. R^2	0.662	0.663	0.662	0.663	0.662	0.663

Panel B. Quartile rank transformation of the video-based deception scores

	(1)	(2)
Dep. var. =	*NegIncidents*	
Post×I(Deception Scores: [25%, 50%])	0.132	0.111
	(0.081)	(0.098)
Post×I(Deception Scores: [50%, 75%])	0.174**	0.168*
	(0.079)	(0.094)

Post×I(Deception Scores: [75%, 100%])	0.226***	0.232**
	(0.084)	(0.100)
Controls	Yes	Yes
Bank FE	Yes	-
Borrower FE	Yes	-
Bank×Borrower FE	-	Yes
Country×Year FE	Yes	Yes
Industry×Year FE	Yes	Yes
N	9,260	8,843
Adj. R^2	0.662	0.663

Appendix A. Variable definition

Variable	Definition
NegIncidents	The number of negative ESG incidents of borrower firms in the year. Datasource: RepRisk.
Post	A dummy variable that takes the value of one after the bank joins the PRB program (i.e., year≥2020), and zero otherwise.
Deception Scores	The deception scores of PRB banks' videos, which capture the possibility of deception. We obtain the PRB banks' videos from UNEP FI's YouTube account, in which the CEOs from PRB banks talk about why their bank signs the principles and what it means for their business. Datasource: YouTube.
BankSize	The natural logarithm of bank's total assets. Datasource: Bankscope.
LoanGr	Bank's annual growth ratio of net loans. Datasource: Bankscope.
BankROE	Bank's return on equity, defined as the ratio of net income to total equity. Datasource: Bankscope.
Tier1	Bank's tier1 risk-based capital ratio. Datasource: Bankscope.
LoanRatio	Bank's ratio of net loans to total assets. Datasource: Bankscope.
NII	Bank's non-interest income over total income. Datasource: Bankscope.
LLP	Bank's loan loss provisions over net loans. Datasource: Bankscope.
FirmSize	The natural logarithm of borrower firm's total assets. Datasource: Worldscope.
Lev	Borrower firm's ratio of total debt to total assets. Datasource: Worldscope.
ROA	Borrower firm's return on assets, defined as the ratio of net income to total assets. Datasource: Worldscope.
Current	Borrower firm's ratio of current assets to total assets. Datasource: Worldscope.
InteCover	Borrower firm's ratio of earnings before interest and tax to the interest expense. Datasource: Worldscope.
SGA	Borrower firm's selling, general, and administrative expense scaled by total assets. Datasource: Worldscope.
RD	Borrower firm's research and development expense scaled by total assets. Datasource: Worldscope.
CAPX	Borrower firm's capital expenditures scaled by total assets. Datasource: Worldscope.
Deception Scores_V (A, T)	The visual- (audio-, texual-) based deception scores of PRB banks' videos, trained only on the visual (audio, textual) features of videos.
Deception Scores_Features Category	The deception scores of PRB banks' videos, trained by specific categories of visual features. We divide the visual features of videos into six categories: gaze, eye, face pose, face LMK, face shape, and facial AUs.
Borrower ESG Ratings	Borrower firm's ESG ombined score in the year, which captures the overall ESG performance of the borrower firms. Datasource: Refinitiv ASSET4.
Borrower ESG Reporting Ratings	Borrower firm's ESG disclosure score in the year, which captures the borrower firms' ESG disclosure performance. Datasource: Refinitiv ASSET4.

Borrower ESG Strategy Ratings	Borrower firm's ESG strategy score in the year, which reflects borrower firms' practices to communicate that it integrates the economic (financial), social and environmental dimensions into its day-to-day decision-making processes. Datasource: Refinitiv ASSET4.
Borrower Co2 Intensity	Total CO2 and CO2-equivalent emissions (in thousands of tons), scaled by revenues (in millions) in the year. Datasource: Refinitiv ASSET4.
Avail_BankESGratings	The most recently available ESG ratings of PRB banks, zero if missing. Datasource: Refinitiv ASSET4.
I(Missing_BankESGratings)	A dummy variable that takes the value of one if the bank is not covered by Refinitiv ESG ratings, and zero otherwise. Datasource: Refinitiv ASSET4.
Persuasiveness_PCA	The factor with the highest eigenvalue using the Principal Component Method to estimate from visual, vocal, verbal emotions and visual beauty.

Online Appendix

Figure OA1. Illustration of Face Features Output using OpenFace

This figure illustrates the facial feature outputs generated by OpenFace. For presentation purposes, we display a subset of facial features in a graphical user interface (GUI) format. The complete set of extracted facial features is stored in CSV files during the data processing phase.

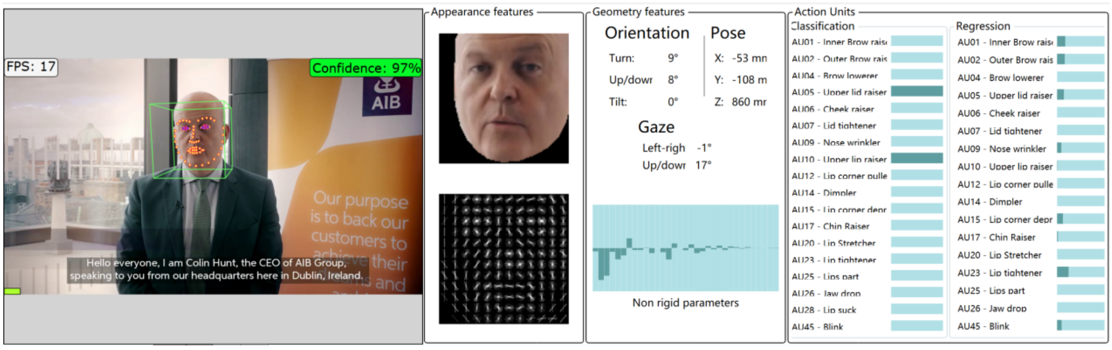


Figure OA2. Illustration of the Dynamic Encoding Process of an Action Unit Using OpenFace

Figure OA2 illustrates the dynamic encoding process of a blink action (AU45) detected in a video clip using OpenFace. The x-axis represents the timestamps of specific frames, while the y-axis indicates the values of AU45_r and AU45_c. In OpenFace, Action Units (AUs) are described in two ways: (1) Presence, indicating whether an AU is visible (e.g., AU45_c), and (2) Intensity, measuring its strength on a five-point scale from minimal to maximal. OpenFace provides both metrics. Specifically, the presence of AU45 is recorded in the AU45_c column, where "0" signifies absence and "1" indicates presence. The intensity of AU45 is captured in the AU45_r column, ranging from 0 (not present) to 5 (maximum intensity), with continuous values representing varying intensity levels.

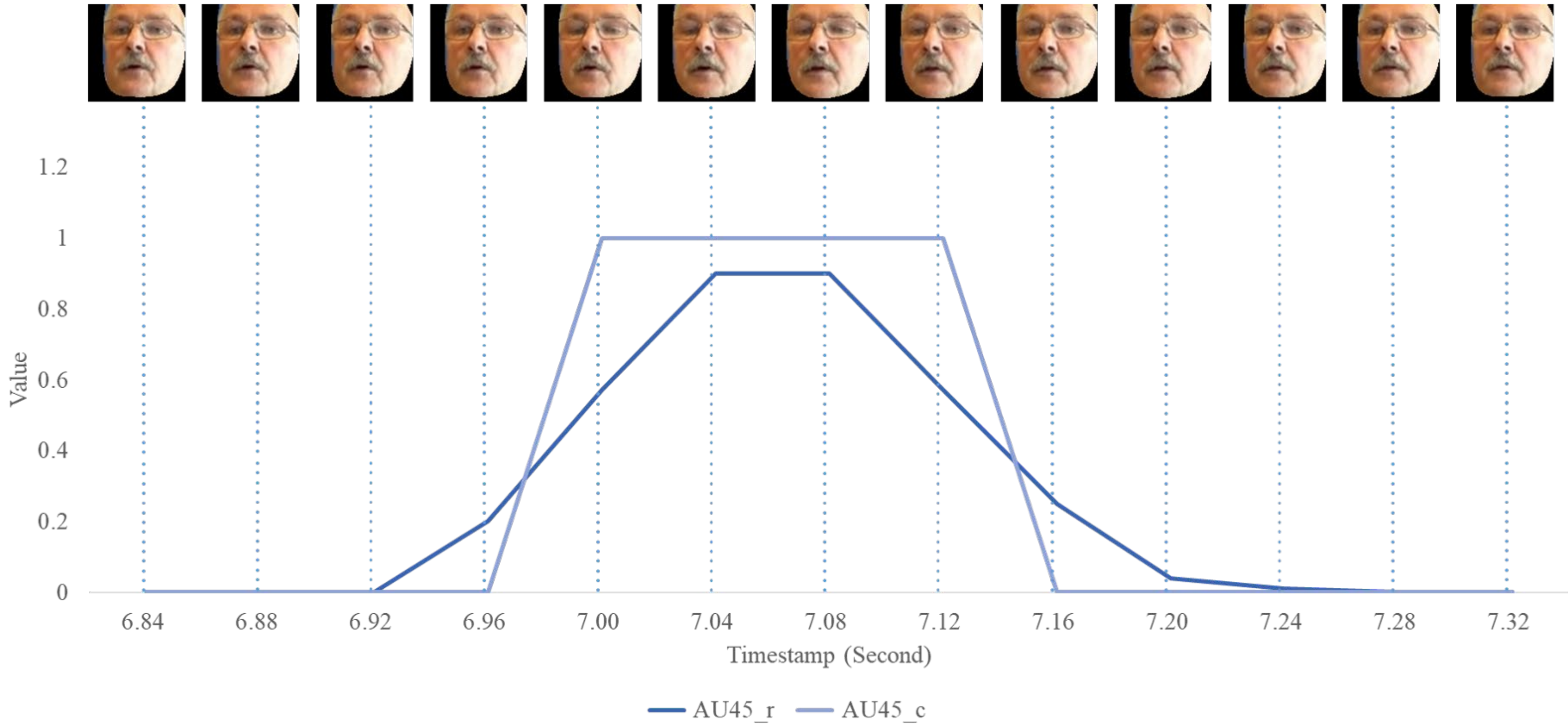
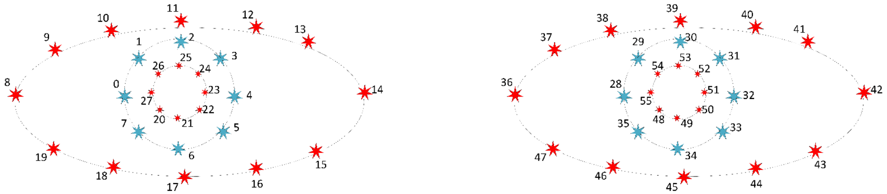


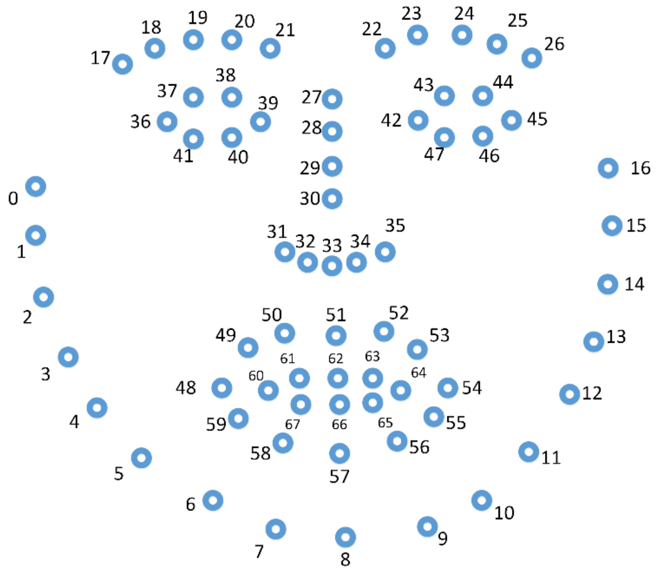
Table OA1. Feature Description

Panel A. Visual Feature Description

This panel provides the specific names and descriptions of the visual features extracted by OpenFace, organized by category.

Visual feature category	**Visual feature name**	**Visual features**	**Description**	**No. of features**	**Example Cues**
Gaze-related	Eye gaze direction vector in world coordinate	gaze_0_x, gaze_0_y, gaze_0_z; gaze_1_x, gaze_1_y, gaze_1_z	Eye 0 is the leftmost eye in the image, eye 1 is the rightmost eye in the image. Take eye 0 as an example. gaze_0_x, gaze_0_y, and gaze_0_z refer to eye gaze direction vector in world coordinates for eye landmark 0. Think of it as a ray going from the left eye in the image in the direction of the eye gaze.	6	Gaze shifting, gaze aversion.
	Eye gaze direction in radians in world coordinates averaged for both eyes and converted into more easy to use format than gaze vectors	gaze_angle_x, gaze_angle_y	If a person is looking left-right this will results in the change of gaze_angle_x (from positive to negative) and, if a person is looking up-down this will result in change of gaze_angle_y (from negative to positive), if a person is looking straight ahead both of the angles will be close to 0 (within measurement error).	2	
	Total			**8**	
Eye location detail	location of 2D eye region landmarks in pixels	eye_lmk_x_0, eye_lmk_x_1,... eye_lmk_x55; eye_lmk_y_1,... eye_lmk_y_55	There are a total 56 eye landmarks, leading to a total 112(56*2) features of 2D eye region landmark. The landmark index can be found below.	112	Pupil size changes, eyes closed, eye flutters
	location of 3D eye region landmarks in millimeters	eye_lmk_X_0, eye_lmk_X_1,... eye_lmk_X55; eye_lmk_Y_0,... eye_lmk_Z_55	There are a total 56 eye landmarks, leading to a total 168(56*3) features of 3D eye region landmark. The landmark index can be found below.	168	

	 **Eye landmark index** ** The figure illustrates the eye landmark indices used by OpenFace. The landmarks are plotted around the outline of each eye, with different indices corresponding to specific points on the eye's contour and within the eye region.*			-	
	Total			**280**	
Face Pose	The location of the head	pose_Tx, pose_Ty, pose_Tz	The location of the head with respect to camera in millimeters	3	Head shakes, head nods, head orientation
	The rotation of head	pose_Rx, pose_Ry, pose_Rz	Rotation is in radians around X,Y,Z axes with the convention R = Rx * Ry * Rz, left-handed positive sign. This can be seen as pitch (Rx), yaw (Ry), and roll (Rz). The rotation is in world coordinates with camera being the origin.	3	
	Total			6	
Face location detail (Face LMK)	Face landmarks locations in 2D	x_0, x_1, ... x_66, x_67, y_0,...,y_67	Face location of 2D landmarks in pixels. There are a total 68 eye landmarks, leading to a total 136(68*2) features of 2D eye region landmark. The landmark index can be seen below.	136	Face changes, mouth asymmetry
	Face landmarks locations in 3D	X_0,...,X_67, Y_0,...,Y_67, Z_0,...,Z_67	Face location of 3D landmarks in millimetres. There are a total 68 eye landmarks, leading to a total 204(68*3) features of 3D eye region landmark. The landmark index can be seen below.	204	

	 **Face landmark index** ** The figure illustrates the facial landmark indices used by OpenFace. The landmarks are plotted around the outline of face, with different indices corresponding to specific points on the face.*			-	
	Total			**340**	
Face shape characteristics	Rigid face shape parameters	p_scale, p_rx, p_ry, p_rz, p_tx, p_ty	Parameters of a point distribution model (PDM) that describe the rigid face shape (location, scale and rotation)	6	Relaxed face, head postural adjustments
	Non-rigid shape parameters	p_0, p_1, ... p_33	Parameters of a point distribution model (PDM) that describe the non-rigid face shape (deformation due to expression and identity).	34	
	Total			**40**	
Facial Action Units (AUs)	AU intensities	AU01_r, AU02_r, AU04_r, AU05_r, AU06_r, AU07_r, AU09_r, AU10_r, AU12_r,	The intensity (from 0 to 5) of each facial AU. Facial Action Units (AUs) are a way to describe human facial expression, more details on Action Units can be found	17	Blinking (AU 45), brow lowering (AU 4), lip stretch (AU

		AU14_r, AU15_r, AU17_r, AU20_r, AU23_r, AU25_r, AU26_r, AU45_r	“https://www.cs.cmu.edu/~face/facs.htm”		20)
	AU occurrences	AU01_c, AU02_c, AU04_c, AU05_c, AU06_c, AU07_c, AU09_c, AU10_c, AU12_c, AU14_c, AU15_c, AU17_c, AU20_c, AU23_c, AU25_c, AU26_c, AU28_c, AU45_c	The presence (0 absent, 1 present) of each facial AU. Facial Action Units (AUs) are a way to describe human facial expression, more details on Action Units can be found “https://www.cs.cmu.edu/~face/facs.htm”	18	
	Total			**35**	
Visual Total				**709**	

Panel B. Textual Feature Description

This panel provides the specific names and descriptions of the textual features extracted by StanfordNLP, organized by category.

Textual feature category	Textual feature name	Description	No. of features
Dependency parsing	The depth of the dependency tree	The dependency parsing module builds a tree structure of words from the input sentence, which represents the syntactic dependency relations between words. The depth of the dependency tree suggests how complex the syntactic structure of a sentence is.	1
	The distance of the dependency tree	The total sum of all individual dependency distances in a sentence's dependency tree, providing a compact way to evaluate how complex the syntactic structure of a sentence is.	1
Part-of-speech (POS) tags	The number of unique universal POS tags	The number of unique universal POS tags for each sentence.	1
	The frequency of each POS tag	There are a total of 17 POS tags, including ADJ, ADP, ADV, AUX, CCONJ, DET, INTJ, NOUN, NUM, PART, PRON, PROPN, PUNCT, SCONJ, SYM, VERB, X.	17
Sentence features	Total word count	Total word count for each sentence.	1
	Average word length	Average word length for each sentence.	1
Textual Total			22

Panel C. Audio Feature Description

This panel provides the specific names and descriptions of the audio features extracted by Librosa, organized by category.

Audio feature category	Audio feature name	Description	No. of features
Chromagram features	chroma_stft_1, chroma_stft_2, … chroma_stft_12	Compute a chromagram from a waveform or power spectrogram.	12
Mel-frequency cepstral coefficients (MFCCs)	mfcc_1, mfcc_2, …mfcc_13	The mel frequency cepstral coefficients (MFCCs) of an audio signal are a set of features which describe the overall shape of the spectral envelope.	13
Various spectral	spectral centroid	Compute the spectral centroid.	1
	bandwidth	Compute p'th-order spectral bandwidth.	1

features	rolloff	Compute roll-off frequency.	1
	root-mean-square energy	Compute root-mean-square (RMS) value for each frame, either from the audio samples.	1
	zero-crossing rate	Compute the zero-crossing rate of an audio time series.	1
Auidio Total			30

Table OA2. Excluding the observations with delayed video disclosure

This table reports the results of the video-based deception scores after excluding the observations with delayed video disclosure. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)
Dep. var. =	*NegIncidents*	
Post×Deception Scores	0.490***	0.631***
	(0.181)	(0.219)
Controls	Yes	Yes
Bank FE	Yes	-
Borrower FE	Yes	-
Bank×Borrower FE	-	Yes
Country×Year FE	Yes	Yes
Industry×Year FE	Yes	Yes
N	9,050	8,645
Adj. R^2	0.663	0.664

Table OA3. The dynamic nature of facial action units

This table reassesses the power of facial AUs using deception scores constructed from randomly spliced videos. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)
Dep. var. =	*NegIncidents*	
Post×Deception Scores_FacialAU_Fake	0.430	0.525
	(0.270)	(0.334)
Controls	Yes	Yes
Bank FE	Yes	-
Borrower FE	Yes	-
Bank×Borrower FE	-	Yes
Country×Year FE	Yes	Yes
Industry×Year FE	Yes	Yes
N	9260	8843
Adj. R^2	0.662	0.663

Table OA4. Deception scores trained by alternative model - Gradient Boosted Decision Trees (GBDT)

This table reports the results of using video-based deception scores trained by an alternative machine learning model (i.e., GBDT) to evaluate the real ESG outcomes of their ex post lending relationships. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)
Dep. var. =		*NegIncidents*		
Post×Deception Scores_GBDT	1.072***	1.254***		
	(0.353)	(0.422)		
Post×Deception Scores_GBDT_V			0.951***	1.092***
			(0.354)	(0.424)
Post×Deception Scores_GBDT_A			0.229	0.192
			(0.525)	(0.612)
Post×Deception Scores_GBDT_T			-0.108	-0.254
			(0.310)	(0.374)
Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	-	Yes	-
Borrower FE	Yes	-	Yes	-
Bank×Borrower FE	-	Yes	-	Yes
Country×Year FE	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes
N	9,260	8,843	9,260	8,843
Adj. R^2	0.662	0.663	0.662	0.663

Table OA5. Alternative clustering of standard errors

This table reports the results of using alternative clustering of standard errors. The dependent variable is *NegIcidents*. *NegIcidents* is the number of negative ESG incidents of borrower firms in the year. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Clustering:	Bank, Borrower		Bank		Borrower		Industry, Country	
Dep. var. =	*NegIncidents*							
Post×Deception Scores	0.505***	0.573***	0.505***	0.573***	0.505***	0.573***	0.505***	0.573***
	(0.040)	(0.057)	(0.105)	(0.120)	(0.123)	(0.152)	(0.127)	(0.183)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	-	Yes	-	Yes	-	Yes	-
Borrower FE	Yes	-	Yes	-	Yes	-	Yes	-
Bank×Borrower FE	-	Yes	-	Yes	-	Yes	-	Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	9,260	8,843	9,260	8,843	9,260	8,843	9,260	8,843
Adj. R^2	0.662	0.663	0.662	0.663	0.662	0.663	0.662	0.663

Table OA6. The dimensions of negative ESG incidents

This table reports the results of comparing the ESG negative incidents in different dimensions. We run these regressions using the Poison model as suggested by Cohn et al. (2022). A set of firm characteristics and bank characteristics are controlled for, as in the Table 2. All variable definitions are listed in Appendix A. Standard errors, adjusted for heteroskedasticity and clustered by bank-borrower pair, are reported in parentheses. ***, **, and * indicate a two-tailed test significance level of less than 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. var. =	*NegIncidents_Env*		*NegIncidents_Social*		*NegIncidents_Gov*		*NegIncidents_CrossCutting*	
Post×Deception Scores	0.485**	0.544**	0.572**	0.645**	0.542***	0.637**	0.458**	0.534**
	(0.214)	(0.263)	(0.244)	(0.298)	(0.204)	(0.250)	(0.181)	(0.215)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes		Yes		Yes		Yes	
Borrower FE	Yes		Yes		Yes		Yes	
Bank×Borrower FE		Yes		Yes		Yes		Yes
Country×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry×Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	6,372	5,913	7,125	6,751	8,084	7,627	8,592	8,201
Adj. R^2	0.705	0.704	0.655	0.655	0.555	0.552	0.663	0.664

Online Appendix B. PRB's Guide to Producing Commitment Videos

1. The exemplary questions provided by the PRB are as follows:

- #1 Why is [xx bank] involved in establishing these Principles for Responsible Banking?
- #2 Why is there a need for global Principles for Responsible Banking? What is different about them from existing frameworks? Why are they needed now?
- #3 Why do you see alignment with societal goals - as expressed in the Sustainable Development Goals and the Paris Climate Agreement - as important to strategically position your bank for future success? What value do these Principles bring to your bank, your shareholders and customers?
- #4 The Principles also call on banks to publicly set targets and report back on their progress. Why do you think that's an important feature of the Principles?
- #5 What changes in your bank do you see these Principles guiding or accelerating?
- #6 How do you see these Principles helping your bank to identify and seize emerging opportunities?

2. The guidelines for video production provided by the PRB are as follows:

Location/setting

- Maybe the CEO is sitting in a meeting room
- There should be something on the table or behind him/her that identifies it as your bank (e.g., logo, banner, etc.)
- The background should not be too distracting

Set-up

- We are looking for a tight close-up – head and shoulders – of your CEO in the frame
- Film landscape (horizontally) and place the camera level with the CEO
- Fix the camera on a tripod
- Use an external microphone (e.g., a lapel microphone) on the CEO
- If you can set the lighting, make it in front of the CEO, but to one side, not head on
- Have the CEO speak just to one side of the camera, i.e. at a hidden interviewer

Filming

- Test the focus and film and sound quality before conducting the whole interview
- Record both the questions and the answers

Let the film run on between question and answers so there is “white space” we can cut into to make the editing easier.

Online Appendix C. The Construction of Video-Based Persuasion Measure

In Online Appendix C, we describe the details of how we construct the video-based persuasion measure, following Hu and Ma (2024).

First, to capture visual emotion, we represent the PRB commitment videos as images sampled at ten frames per second. Using Face++, a face-detection machine learning algorithm, we identify human faces in these frames and generate a visual emotion measure. The *Face++* platform provides APIs through which we feed the raw images into the cloud computing system and receive a host of face-related measures constructed by *Face++*'s machine learning algorithms. Those measures include visual emotions, beauty, age, gender, etc. The *Face++* emotion recognition algorithm API classifies visual emotion into seven categories: happiness, neutral, sadness, surprise, anger, disgust, and fear. Specifically, for each frame, the API gives each category a predicted score between 0 to 1, indicating the likelihood that the frame's emotion belongs to that category. The scores of seven categories sum up to one. For each frame, the emotion category that has the highest predicted score from *Face++* is used to label the emotion of the frame. Following Hu and Ma (2024), we classify a frame as positive if its emotion label is "happiness", as negative if its emotion label is "sadness", "anger", "disgust", or "fear".[29] The visual positive tone during a PRB commitment videos is calculated as the number of positive frames scaled by the number of total frames, and the textual negative visual emotion is calculated as the number of negative frames scaled by the number of total frames.

Second, as to audio emotion, we use the deep neural networks (CNNs) model trained and provided by Pinto et al. (2020) to classify emotions from audio files extracted from PRB

[29] The "surprise" category is not classified as either positive or negative following prior literature (Curti and Kazinnik 2023; Hu and Ma 2024).

commitment videos. The Pinto et al. (2020) model classifies the audio of each word into eight different emotion categories (neutral, calm, happy, sad, angry, fearful, disgust, surprise). In line with the face emotion classification, we classify “sad”, “anger”, “disgust”, and “fearful” as negative emotions, and “happy” as a positive emotion. The audio positive emotion is calculated as the number of words with positive audio emotion scaled by the number of words, and the audio negative emotion is calculated as the number of words with negative audio emotion scaled by the number of words.

Third, we construct textual emotion by extracting speech transcriptions and applying the Loughran and McDonald (2011) dictionary. Specifically, we use *Vosk*, a speech recognition toolkit, to transcribe the PRB commitment videos. The transcriptions include a list of words, timestamps (onsets and offsets), and punctuation. The textual positive tone during a PRB commitment videos is calculated as the number of positive words scaled by the number of words, and the textual negative tone is calculated as the number of negative words scaled by the number of words. To determine positive and negative words, we rely on the 2020 version of the Loughran and McDonald (2011) dictionary.

Fourth, we measure the CEO’s facial beauty using Face++’s face detection API. The API provides two predicted beauty scores for each detected face: a male beauty score and a female beauty score, both ranging from 0 to 100, indicating the perceived beauty level of the face from male and female perspectives, respectively. To calculate the CEO’s beauty score, we take the mean of the male and female beauty scores across all frames.