Spotting “Washing” : Can Financial Experts Detect 'Say-Do Gap' in Banking Reports

Lilah Shema-Zlatokrilov (Bank of Israel)

Dissertation project. Supervisors – Dan Galai and Zvi Wiener

The Hebrew University Finance Department and Environmental Economics and Management Department, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Radboud University Nijmegen Department of Economics, and the Bank of Israel Research Department (Finance Division)

Abigail Hurwitz, PhD, Sven Nolte, PhD, Lilah Shema-Zlatokrilov (PhD Candidate)

ABSTRACT

Can financial experts detect potential greenwashing in ESG communications? Using an eye-tracking experiment with 128 professionals, we document a novel attention detection-recommendation chain. Attention to textual contradictions increases expert detection sensitivity, significantly reducing investment recommendations.

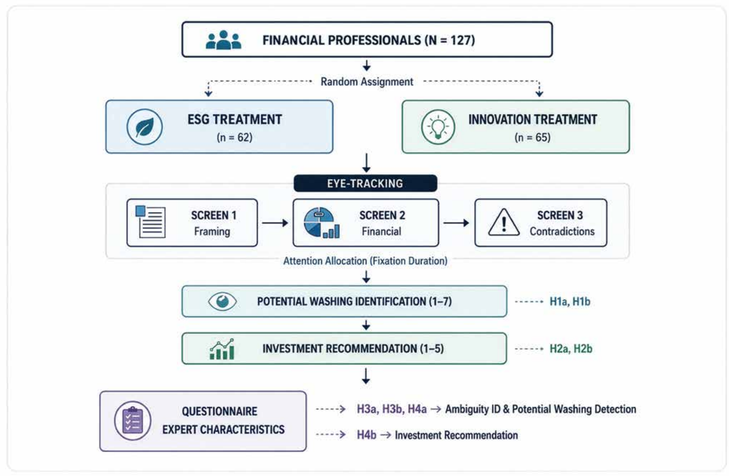
Framing impacts this chain in two distinct ways relative to a control condition: ESG exposure amplifies attention to contradictions, heightening detection—a domain-specific focusing effect; but ESG framing partially buffers the detection-to-investment penalty.

Additional results reveal a skepticism paradox: Media awareness counterintuitively reduces detection. Two mechanisms, both shaped by management communication design, allow greenwashing to persist under expert evaluation: inattention to embedded contradictions, and a warm-glow halo that attenuates the investment penalty of detection.

METHODOLOGY

First stage: Eye-Tracking Experiment : The visual attention patterns of financial experts were tracked while they reviewed banking communications that included ESG (Environmental, Social, and Governance) or innovation disclosures. Participants were randomly assigned to either the ESG version (n=62) or the Innovation version (n=66), with the presentation order counterbalanced.
Three-Screen Sequence as detailed in Figure 2.

Second stage: Participants answered detailed questions assessing “washing”, identification, their investment recommendation, and their habits, beliefs and knowledge.



PARTICIPANTS

	ESG	Innovation
N	62	66
Male	59.7%	53.8%
Age 36–60	67.7%	73.8%
MA or PhD	79.0%	76.9%

- *N* = **128** financial professionals
- Major Jerusalem–Tel Aviv financial institution
- Pre-registered: aspredicted.org/see_one.php

ESG familiarity (1–7): $M = 2.5$, $SD = 1.7$

No sig. group differences (all $p > .20$) ✓

THE ATTENTION MODEL (Gabaix, 2019)

Model

$$x^g = m \cdot x + (1 - m) \cdot x^d$$

- x = true ESG state (embedded contradiction)
- x^d = default prior (“green label”)
- $m \in [0, 1]$ = attention parameter
- **Key:** we *directly measure* m via fixation

Prior work infers m from outcomes.

Paper I (this paper): m measured at the AOI level — defined screen regions on Screens 1 & 3

Structural Equations

Detection (D_{ij})

$$D_{ij} = \beta_0 + \beta_1 A_{ij} + \beta_2 F_{ij} + \beta_3 A_{ij} F_{ij} + \beta_4' X_i + u_{ij}$$

Investment (I_{ij})

$$I_{ij} = \gamma_0 + \gamma_1 D_{ij} + \gamma_2 F_{ij} + \gamma_3 D_{ij} F_{ij} + \gamma_4' X_i + \epsilon_{ij}$$

- A_{ij} = relative fixation (our measure of m)
- F_{ij} = ESG framing indicator
- X_i = expert characteristics (PCA)
- β_3 tests H1a; β_3 tests H1b; γ_1 tests H2a; γ_3 tests H2b

FIGURE 1: EXPERIMENT FRAMEWORK

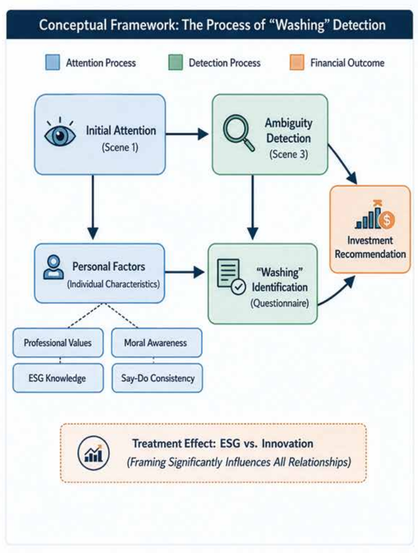
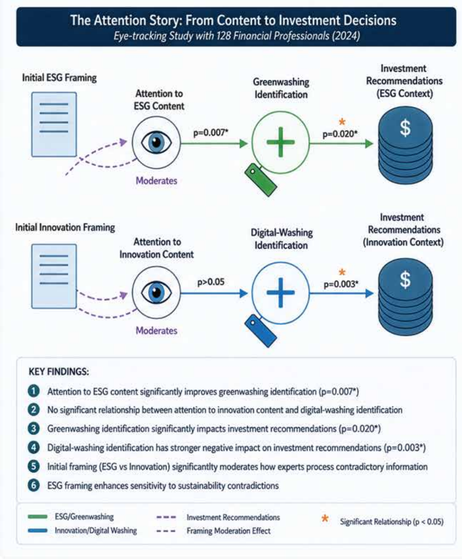


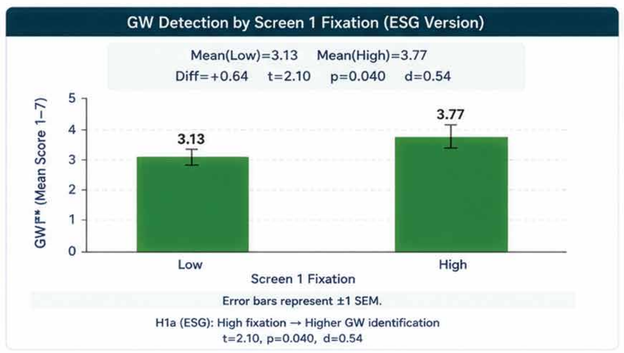
FIGURE 2: SIMPLIFIED ABSTRACT VISUALIZATION

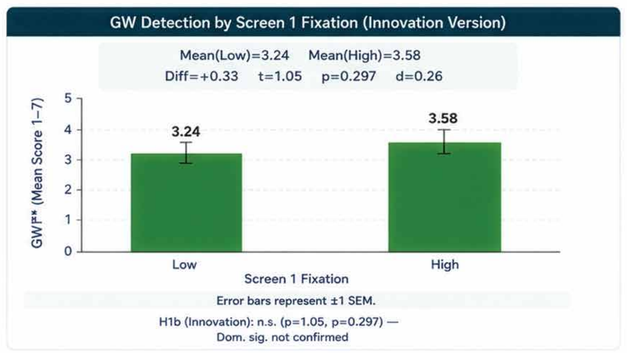


RESULTS

RESULTS — H1: ATTENTION → WASHING DETECTION

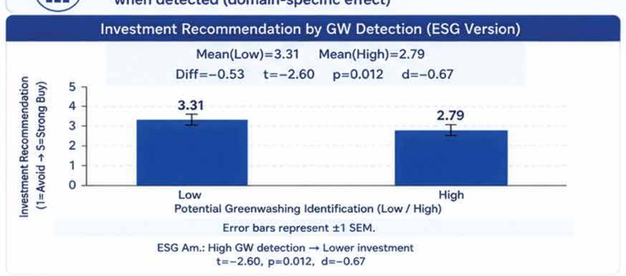
ESG contradictions attract disproportionate fixation — domain-specific attention pattern

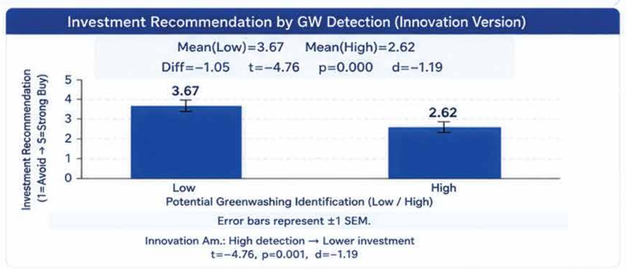




RESULTS — H2: DETECTION → INVESTMENT RECOMMENDATION

ESG contradictions reduce investment recommendation when detected (domain-specific effect)





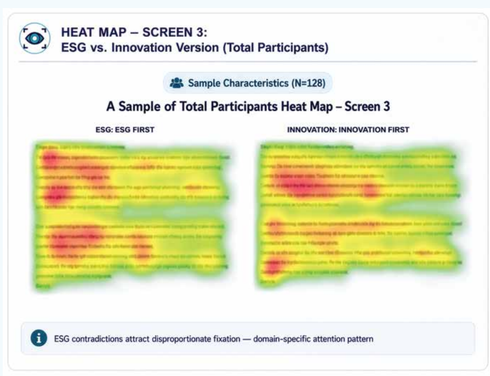
KEY TAKEAWAY
Attention to ESG contradictions (not innovation content) increases greenwashing detection, which in turn reduces investment recommendations.

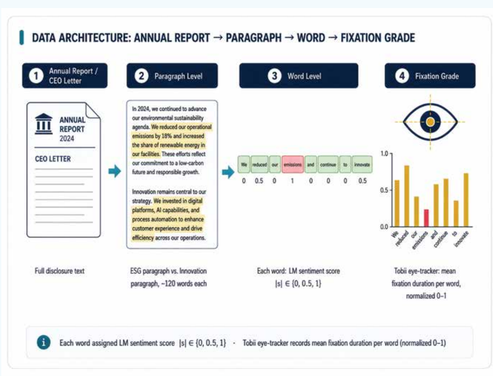
Step II - Which Words Do Experts Read? Sentiment and Attention in Financial Disclosures

Jonathan Benchimol (Bank of Israel) Lilah Shema-Zlatokrilov (Bank of Israel &Hebrew University), and Rony Zlatokrilov(Technion - Israel Institute of Technology).

ABSTRACT

We use word-by-word eye-tracking on 128 Israeli financial professionals reading paired ESG and Innovation disclosure paragraphs. Negative words attract more fixation than neutral words on average, but the relation is non-monotone in intensity. The most intensely negative words receive the lowest predicted fixation, with an identified negative-versus-non-negative gap of about 17 percentage points. Words that contradict the paragraph’s prevailing tone draw 12% more fixation, and the novelty premium reverses for negatively coded contradictions. The configuration with the lowest predicted attention is intense negative words embedded in non-negative narratives, the textual structure that the obfuscation literature analyzes on the supply side. We document a parallel demand-side pattern within the experimental stimulus. A Gabaix-style benefit-minus-cost decomposition, with a salience component and a high-intensity avoidance component, organizes the regularities and yields a closed-form break-even intensity $|s|^* = \beta 2/|\beta 3| \approx 0.42$. The pattern survives participant and participant-plus-word fixed effects and replicates across the LM and Thematic dictionaries. Coefficients are associational within a fixed stimulus pair, and the interpretation requires replication on alternative texts.





HEAT MAP → WORD-LEVEL FIXATION

The Extended Benefit Function:

$$\hat{f}_{ij} = \alpha + \beta_1 |s_j| + \beta_2 \cdot \mathbf{1}[s_j < 0] + \beta_3 |s_j| \cdot \mathbf{1}[s_j < 0] + \beta_4 \cdot \text{Contr}_j + \beta_5 \cdot \text{DoubleAvoidance}_j$$

Param.	Value	Meaning
β_1	+0.008	Intensity → fixation
β_2	+0.019**	Negativity salience
β_3	−0.085***	Avoidance
β_4	+0.079***	Contradiction
β_5	−0.189***	Double Avoid.

$|s|^* \approx 0.42$ break-even

N = 10,674 word obs.; participant + word FE

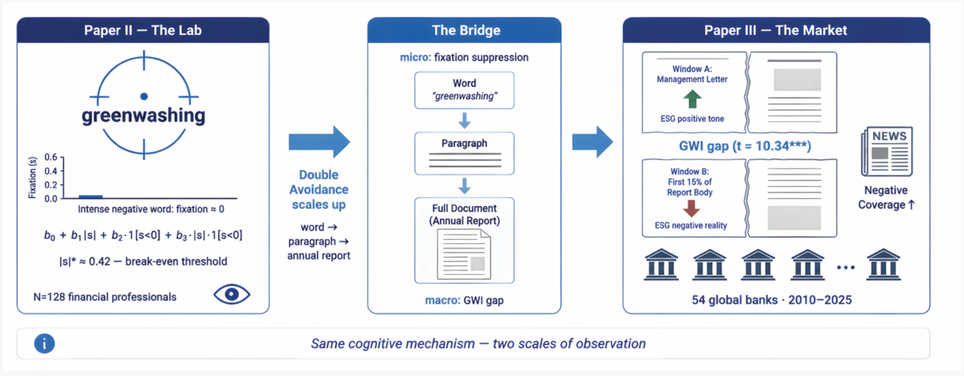
Step III - Reality Creates Words, ESG Disclosure Strategy Under Reputational Stress

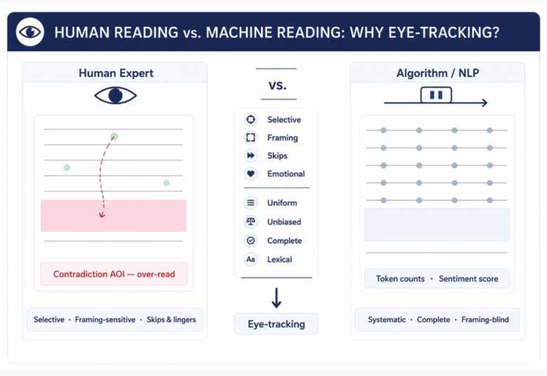
Lilah Shema-Zlatokrilov (Bank of Israel)

IMPLEMENTING PAPER I + II RESULTS AND BULDING WASHING IDENTIFICATION INDICATOR

ABSTRACT

—
"The Words Game" :This paper examines the internal rhetorical architecture of annual letters from bank CEOs. We ask whether the placement of negative disclosures within the letter—not merely their presence—shapes how readers process them. We develop a text-based measure based on Stage I and Stage II results, and examine its properties using a panel of global banks between 2010 and 2025. We analyze the correlation between the disclosure strategy and negative events.





STUDIOB