Responding to Surprises in a Complex World

Thomas Graeber, Christopher Roth, Marco Sammon

Discussion

Eben Lazarus
UC Berkeley Haas

NBER Behavioral Finance

November 2024

Outline

1. Placing the paper

2. Interpreting the empirics

The paper's context

Rich literature related to categorical thinking

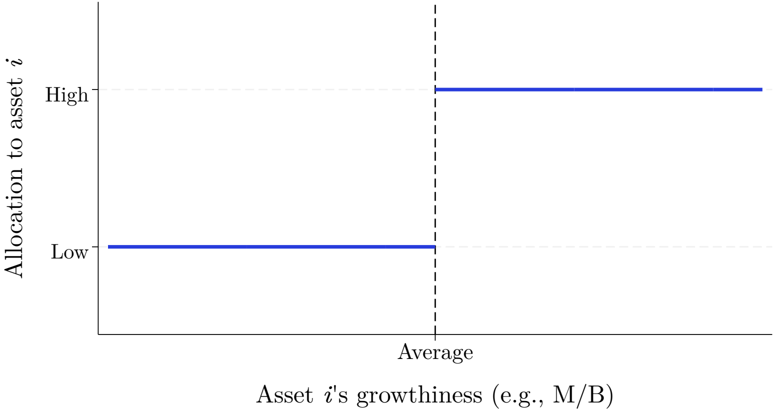
- Useful to place the paper within this literature
- Will paint with a very broad brush (sorry!)
- Highlight some key commonalities & distinctions between small # of seemingly unrelated papers

Important questions:

1. Are the categorization criteria objectively goal-relevant?
2. How does behavior change *within* a category?

Style investing [Barberis & Shleifer 2003]

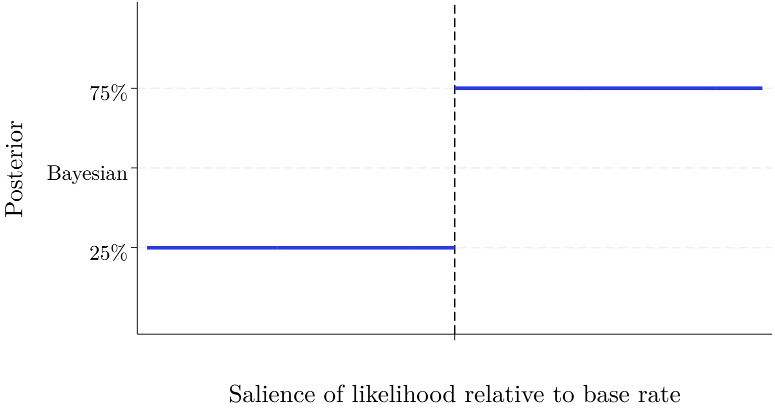
Portfolio choice when growth stocks have recently done well relative to value



1. Are categorization criteria objectively goal-relevant? **Probably not.** Allocations based on past returns.
2. How does behavior change *within* a category? **Not at all** *(n.b. style allocations do change with past returns)*

Selective attention [BCGKS 2024, BGLS 2024]: diff. foundations, some broad-brush similarities

Posterior belief given base rate of 25%, signal likelihood 75%



1. Are categorization criteria objectively goal-relevant? **No.** Bayesian line would be flat.
2. How does behavior change *within* a category? **Not at all** *(can be relaxed)*

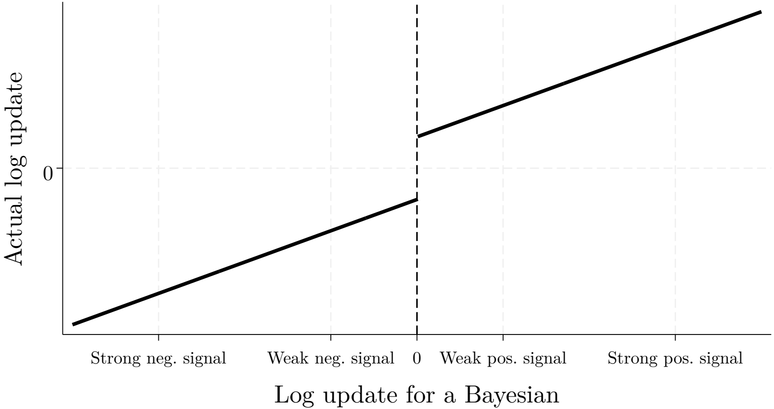
Overinference from weak signals, . . . [Augenblick, Lazarus, Thaler 2024]

The basic premise:

- In many settings, easy to categorize a piece of news as “good” or “bad”
- Harder to figure out **exactly how good or bad**
  - Good poll numbers: $\nearrow$ Pr(win)...but how much?
  - Basket: $\nearrow$ Pr(win)...but how much?
  - Earnings beat expectations: $\nearrow$ fundamental value...but how much?
- Can understand if info is good/bad, but must generate imperfect estimate of strength
- Shrinkage to moderate strength: after good news, shade toward “average” good news

Overinference from weak signals, . . . [Augenblick, Lazarus, Thaler 2024]

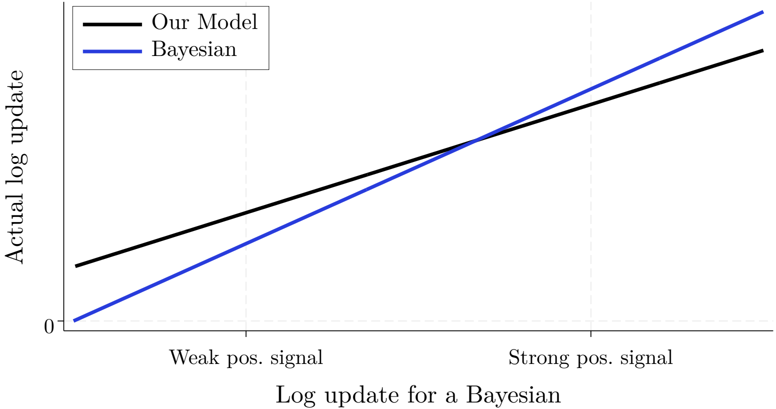
Updating behavior given a signal where direction (good/bad) is clear, strength less clear



1. Are categorization criteria objectively goal-relevant? **Yes**

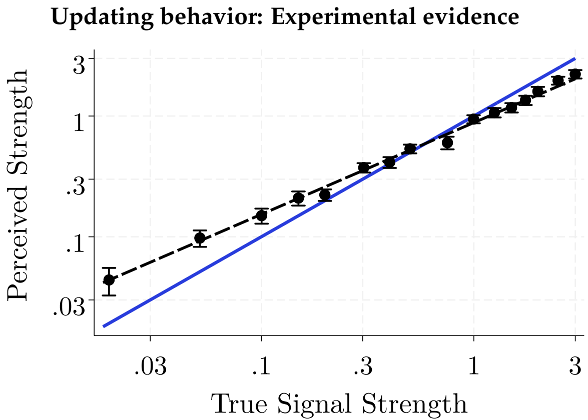
Overinference from weak signals, . . . [Augenblick, Lazarus, Thaler 2024]

Updating behavior given a signal where direction (good/bad) is clear, strength less clear



1. Are categorization criteria objectively goal-relevant? **Yes**
2. How does behavior change *within* a category? **Attenuated, but not fully flat**

Overinference from weak signals, . . . [Augenblick, Lazarus, Thaler 2024]



1. Are categorization criteria objectively goal-relevant? **Yes**
2. How does behavior change *within* a category? **Attenuated, but not fully flat**

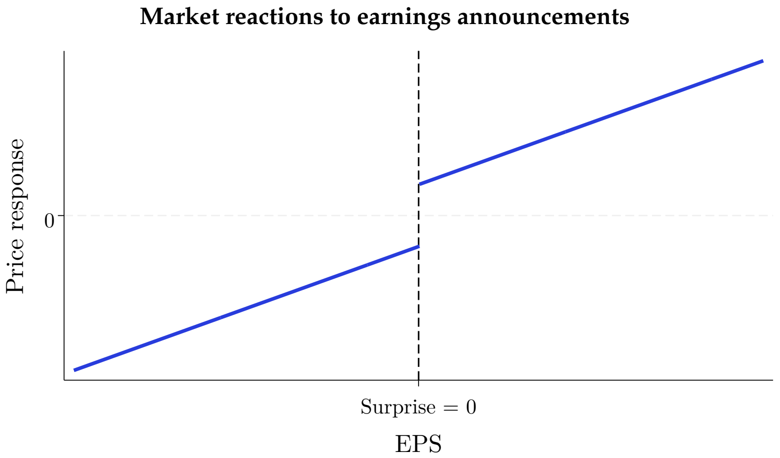
This paper [finally!]

The basic premise:

1. Easy to classify news as “high” or “low,” possibly compared to multiple categorical default levels
2. Harder to incorporate precise numerical information
   - Earnings (a) beat expectations and (b) were positive
   - . . .and (c) EPS = $3.25, with various one-time items

- #1 generalizes our setup, while #2 slightly constrains to numerical processing

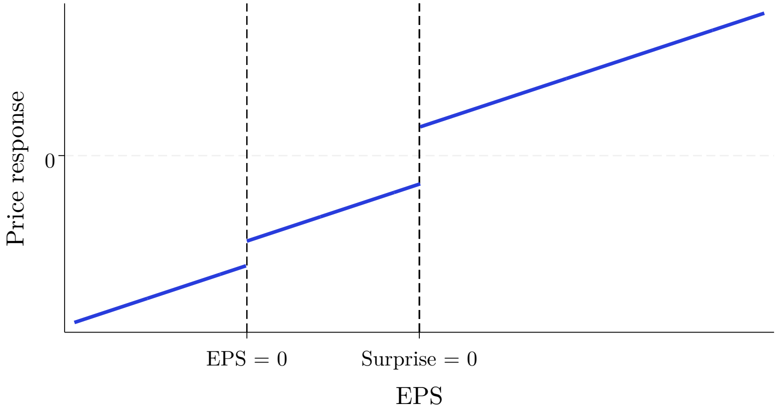
This paper



1. Are categorization criteria objectively goal-relevant? **Yes**
2. How does behavior change *within* a category? **Attenuated, but not fully flat**

This paper

Market reactions to earnings announcements with multiple comparison points

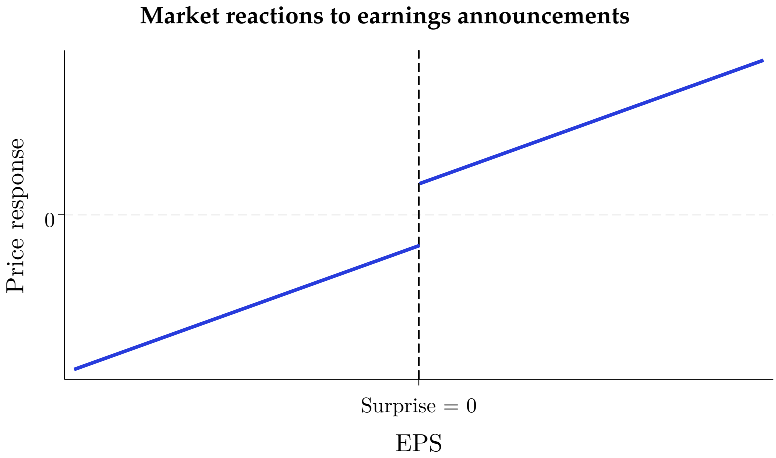


- This is a substantive generalization: in our case, wouldn't work to be switching overinference $\rightarrow$ underinference $\rightarrow$ overinference
- They show evidence for mult. comparison points, but won't discuss

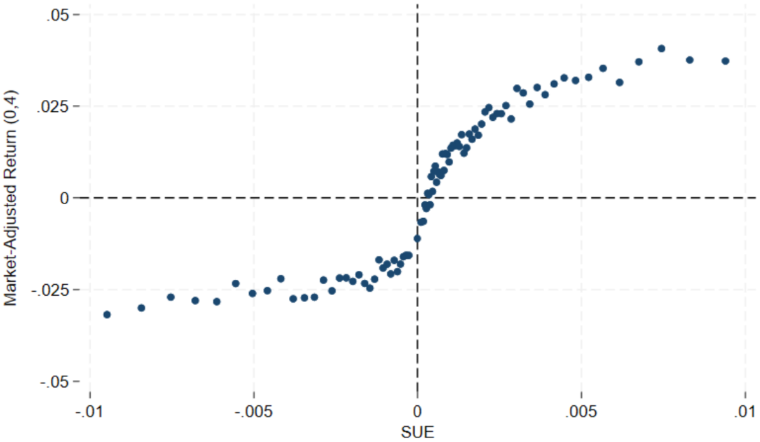
Outline

1. Placing the paper
2. Interpreting the empirics
   - How they relate to the theory
   - My comments

My basic version of the theory



Problem: Basic theory $\neq$ data



1. No clear break at category boundary (pos. vs. neg. surprise)
2. Diminishing sensitivity away from boundary

Where they take things

They're upfront about these issues, and theory is geared to address them

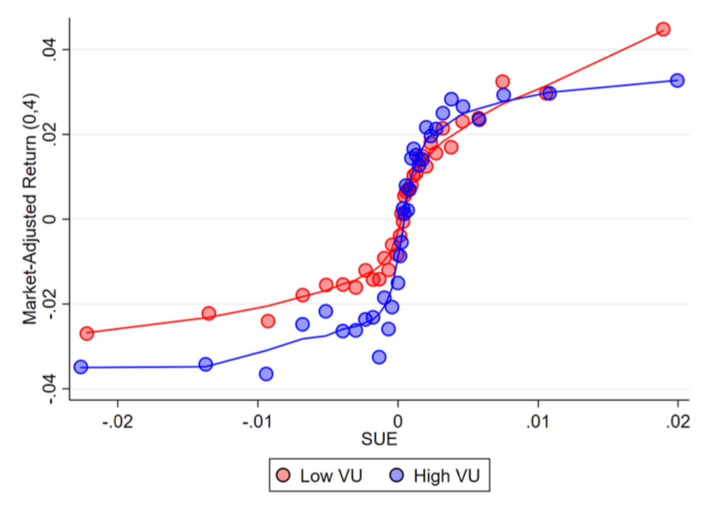
- Recall question #2 from before: How does behavior change *within* a category?
- Style investing, some selective attention models: **Not at all**
- Over-/underinference model: **Attenuated**, possibly for good reasons (e.g., constrained optimal response to qualitative signal), possibly not (e.g., selective attention to only a subset of the signal)
- **Their model:** Response to signal s is $r(s) = \lambda(s) r_{\text{cognitive signal}}(s) + (1 - \lambda(s)) r_{\text{category default}}(s)$
  - Standard parameterization: $\lambda = \frac{\sigma^2_{\text{prior}}}{\sigma^2_{\text{signal}} + \sigma^2_{\text{prior}}} \implies$ piecewise linear response function $\neq$ data
  - They assume: $\sigma^2_{\text{signal}}(s)$ increases in the distance of s from default

    $\implies$ S shape, diminishing sensitivity: $\lambda(s)$ close to 1 near $s = s_d$, and shrinks to 0 far away
  - Why? Processing noise $\nearrow$ in distance to simple defaults (Enke et al. 2024), or $\searrow$ in empirical mass of signal due to efficient coding (Frydman & Jin 2022)
  - Further prediction: $\lambda(s) \searrow$ everywhere — leading to sharper changes at boundaries, and less sensitivity away from them — for hard-to-value firms for which $\sigma^2_{\text{signal}}(s)$ is high everywhere

Theory now appears to match the data

Returns for firms with high vs. low valuation uncertainty

(where VU is dispersion in fundamental value across different accounting-based methods)



Theory now appears to match the data. . .and the experiment

Participants' predicted returns across complexity treatments

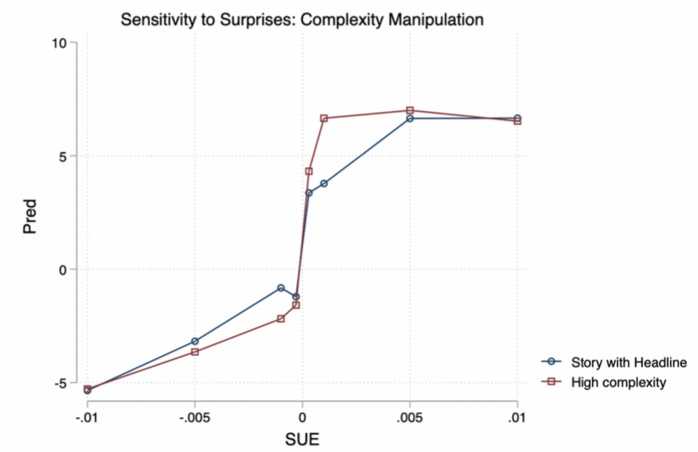
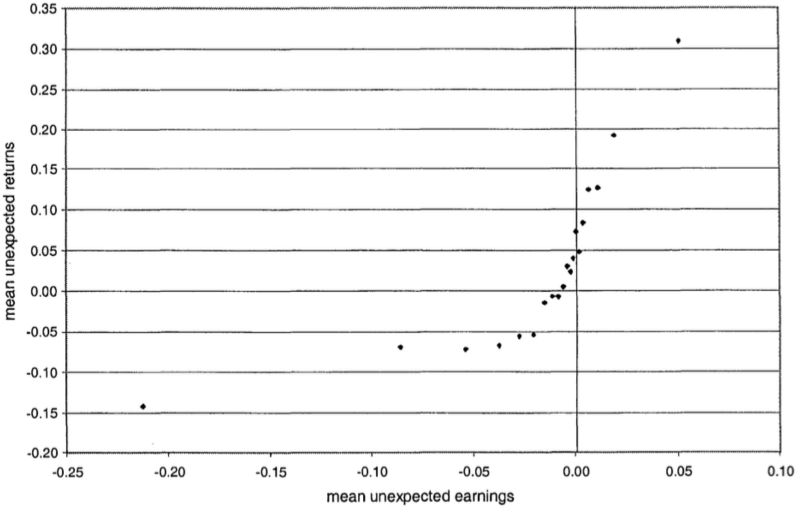


Figure 5: Processing Noise and the S-shape

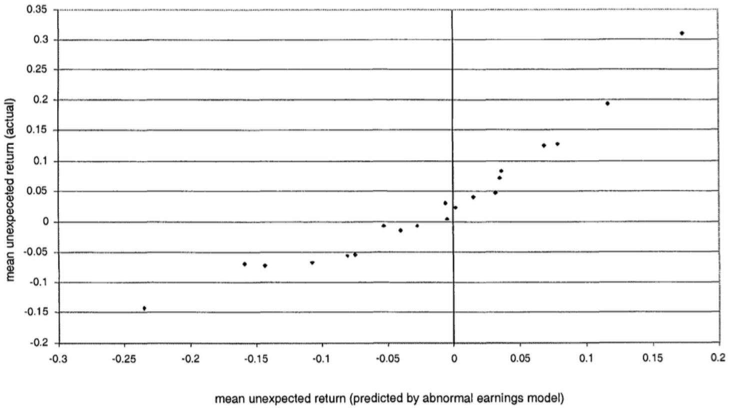
My thoughts

- A lot of steps to get to diminishing sensitivity within category…
- …but the basic idea is surely right in many contexts
- I think the experiment is one such context: Hard to predict the market response to good vs. bad news, especially with extraneous info $\Rightarrow$ jump at clear good/bad threshold and attenuation elsewhere
- The empirical application is trickier: Small return diffs for low vs. high VU, and **not clear what the benchmark (correct) response is**
- Old idea in accounting lit: EPS surprises contain permanent and transitory news about future earnings, and the transitory component tends to be larger for big surprises
  - May matter more for high-VU firms: lower earnings quality (Golubov & Konstantinidi 2024)
  - They try to rule out by showing SUE_t doesn't predict SUE_{t+4} differently for high vs. low VU
  - But predicting the future earnings *surprise* (SUE_{t+4}) doesn't measure earnings persistence: it just shows predictability of forecast errors. Want to measure change in expected future earnings *level*.
  - Luckily for me, Liu & Thomas (2000) do this: they look at IBES forecast revisions after earnings responses to measure expected return response given constant discount rates.

Liu & Thomas: S-shaped return response vs. SUE. . .



. . .attenuates when accounting for future EPS forecast revisions



- Of course, EPS forecast revisions might not be rational!
- But suggests that diminishing sensitivity may be less important over and above whatever is happening to those forecasts

Slight counterpoint 1: Small but significant PEAD for large |SUE|

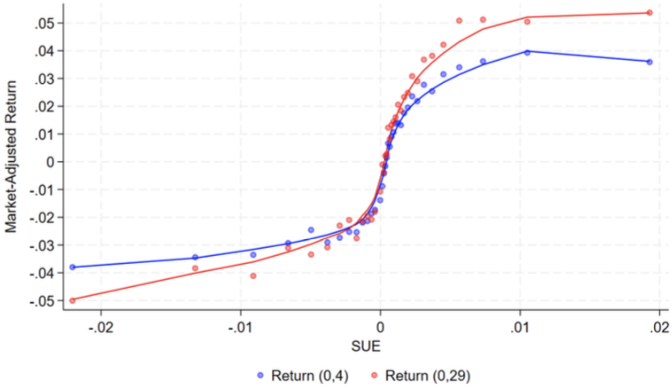


Figure 7: Post-Earnings Announcement Drift

- But overall S shape in 30-day return further suggests that this shape may be a reasonable benchmark response function

Counterpoint 2: Overreaction for small $|\text{SUE}|$

- In appendix, they show that in small window around $|\text{SUE}| = 0$, the big 5-day return responses for high-VU firms tend to revert slightly by day 30
  - I predict I will have negative time left by now, so won't bother showing the table
- Suggests that there *is* meaningful overreaction for weak signals
- Very hard to generate this if the observed S shape is correct long-run benchmark

Final Notes

- Very nice, thought-provoking paper
- Their motivation in terms of categorical defaults spurred me to see connections between previous papers' frameworks that I hadn't realized were there
- Basic idea seems correct in lots of settings
- But some caution warranted around the specific empirical application considered here

Thank you!