Table of Contents

Can Randomization Reveal the Truth?

--Exploring Potentials for Behavioral Approach in Accounting Research

1. Introduction

The increasing stream of research in accounting and finance is focused on causal inference. Many scholars try to find as-if random conditions to identify the quasi-experiments since they increase the credibility of research (Gow et al., 2016). This trend shows us randomization can be the state-of-the-art if you want your research's reasoning to be convincing and to gain recognition from the audience. However, Thaler (2016) discusses the view from behavioral science that raises the importance of individuals' heterogeneity in their judgement and decision-making. So, in this essay, we review the approach (view and technique) commonly used by behavioral science to see if we can benefit from it.

We revisit analyst attention to the annual report and proxy statement that related to psychological factors in the accounting research. We realize experiment's evidence may be clearer than archival ones because a well-structured experiment can manipulate the variable of our interest. However, this setting may not reflect the context of the real world. As a result, generalizability will be the major threat. Limitations of time consuming and sample size also pose challenges that we are concerned with.

In addition to analyst attention, we explore biological-based evidence that reveal the process of cognitive and emotional reasoning. This provides us a solid foundation for learning the impact of CEOs characteristics on firm decision-making. But relying on these biological factors may also raise questions on whether individual differences can be extended to whole management styles.

Above all, behavioral science approached give us new potentials in accounting research compared to the current mainstream of economic view that use archival data as empirical evidence to generalize to the homo economic beings.

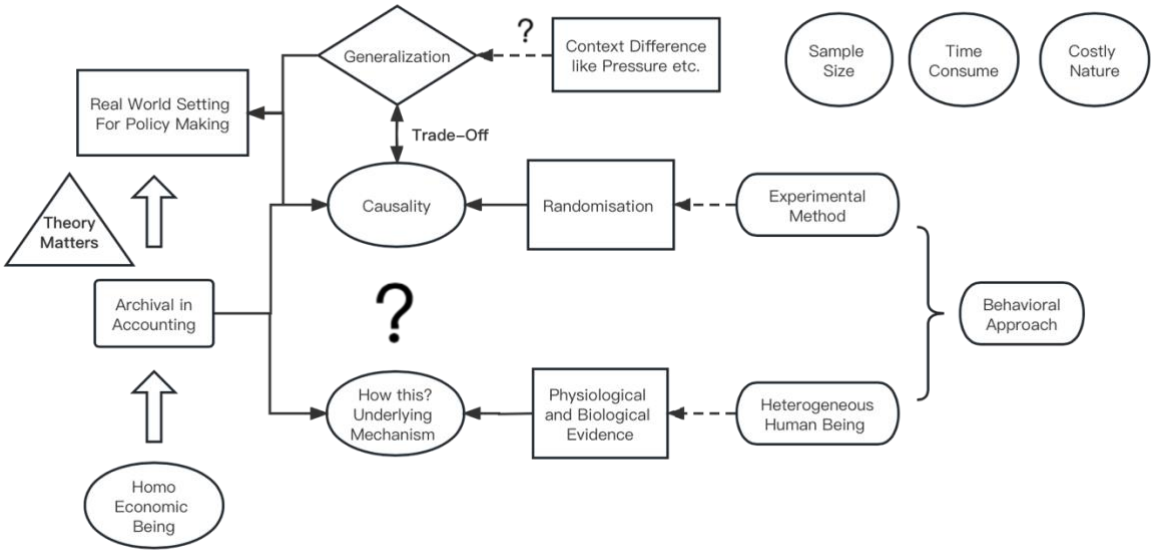

Figure 1 Potentials and Concerns About Behavioral Accounting

2. Potentials for Experiments in Causal Inference

2.1 Causality and as-if random concerns in accounting research

Recently, establishing causal relationships are important in accounting research. As Armstrong et al. (2022) discussed, scholars seek for providing high standard evidence as they want to support their theories convincingly. Many accounting theories require rigorous empirical evidence which may enable researchers to produce more credible findings. Consequently, a growing number of research are motived by finding the as-if random conditions to articulate causal effect (Christensen, 2019). But these settings for quasi-experiment cannot reach to true randomization, which is crucial for causality.

Therefore, Random Controlled Trials (RCT), usually employed in clinical and behavioral research, may offer benefits when we apply it in accounting research. In fact, our goals of using quasi-experiment techniques like Difference in Differences (DID), Regression Discontinuity Design (RDD) is to create an as-if RCT setting (Angrist and Pischke, 2008), which can address the influence of omitted variables. Nevertheless, accounting phenomena related to decision-making is still trapped into endogenous problem. For example, we cannot manipulate analyst attention and behavior to test if variation of attention may lead to different choice for analyst. Analysts may foresee the potential attention's outcome and choose optimum option. Also, in real world, we cannot find a natural setting that two groups are randomly assigned.

To be specific, if we want to check whether readability of annual report and proxy statement can influence the analysts' or investor's information judgment, and then influence their forecast accuracy, we need to exclude the possibility that higher quality analysts whose forecasts are more accurate will endogenously choose more readable reports to analyze. Indeed, we may find an exogenous mandate from U.S. Securities and Exchange Commission (SEC) that requires a part of firms' disclosure format as the quasi-experiment setting. But exogenous shock does not fully represent randomization. Also, we cannot ensure the perfect compliance for participants in treatment groups and control groups. Prominent example is like the electronic filing of SEC forms (Samuels et al., 2021), this mandate just changes the subset of firms that does not voluntarily adopt this action. It creates classic selection bias when we want to draw the real effect of forms readability on analyst accuracy. Subsequently, we hope we can establish a setting like a clinical trial to have real randomization. So, we call for experimental research to mitigate the endogenous issue in real world.

2.2 Experimental method and related concerns

Lab experiments used in behavioral science can reveal the "true effect" compared to quasi-experiments. Fuzzy RDD and DID may be problematic in quasi-experiments as shown by Angrist et al. (1996). However, real experimental research can address those problematic issues and lead to three advantages: 1. Randomization; 2. Double-Blinded; 3. Perfect Compliance. By using experimental method, we can avoid placebo effect, always-takers (compliance issue), no parallel problems and knowing treatment intention issues that discussed in prior research (Armstrong et al., 2021).

In the following section, we will mention our protentional topics about analysts' attention. If we conduct experiment, we can draw on the real randomization. While if we want our research have impacts on policy making, we need to persuade that our sample and context can be applied into the real capital market. Indeed, we can use survey experiment to get large representative sample from the society, but our concern is that participants may not be all equipped the investment skills. What we are curious is investor's behavior, not others. Also, we may think lab experiment can restrict participants backgrounds like we can recruit our accounting and finance students from the university. But recent research (Haushofer et al., 2022) in prosocial behavior finds that there is a significant difference between the pressure from lab judgement and the pressure from real context, which may disturb the outcome's robustness. Both two experimental methods are facing threats for external validity. All of these told us that we need to trade-off between generalization and randomization.

More detailed information for the experimental application on this topic will be discussed in Section 4.

3. Heterogeneity Perspective in Decision-Making

3.1 Psychological and biological factors can be more convincing in behavioral economics

We are also interested with the topics of management decision-making in corporate governance. Randomization technique can be an effective tool as we discussed above since corporate governance topics are hard to escape from questions of endogeneity (Guest, 2021). Moreover, behavioral science may give us hints that some psychology theories might be more reasonable to explain investors processing of financial information. Adopting behavioral economics views followed by Thaler (2016), researchers treat individuals as human beings with heterogeneity rather than as purely rational economics beings which have well-defined unbiased preference, maximization of self-interest and so on. So, the behavioral theory recognize that individual judgment can be influenced by informal institutions and non-economic incentives like early experience (Hanlon et al., 2022).

3.2 Recent behavioral views applications and potential directions for future research

In CEO's and analyst's environment research in accounting, imprinting theory from biology has provided convincing evidence that early experience can shape individuals' behavior persistently. For example, CEOs who experienced the Great Depression may be less likely to publish the forecast (Bamber et al. 2010). But recent evidence-based literature disagrees with imprinting's long-term effect. Bourveau and Law (2021) shows analysts who are living in extreme weather affected states tend to have pessimistic forecasts for no extreme weather impacted firms. This phenomenon, contrary to the explanation of imprinting theory, disappear after the second fiscal year. To better understand this mood-effect, we may call for exploring more biological-level factors like serotonin and melatonin level in relation to weather shocks and early experiences to distinguish long term imprinting effect from short term adjustment in human body when facing different weather conditions. We will enquiry for further advice from behavioral science group and collaborate with experts in medical school if it is available. This could help reveal the underlying mechanism behind accounting scholars' observations.

As for analyst appearance and attention research, scholars have used testosterone level to justify the analysts' beauty effect on the accuracy of forecast (He et al., 2019). Indeed, they find male analysts' performance are better than women in Chinese setting. However, Kumar (2010) discovered the opposite outcomes in US setting. We may keep eyes on the possibility that biological factors impact on individual behavior cannot be generalized across different cultures. This inconsistency highlights culture limitations related to individual effect. Scholars in accounting should consider the influence of cultural heterogeneity when investigating the role of biological factors in decision-making process.

4. Applying Behavioral Approach to Analyst Attention and Management Style Accounting Research

4.1 The impact of repetitive and interaction terms on analyst attentions

Regulators like SEC have been working with disclosures quality's improvement for many years to make the annual report and proxy statement more effective for investors. As financial intermediaries, analysts also need more useful information from these disclosures. For our interest topic, repetitive terms in the statements may both benefit and disturb individual's judgements. Cazier and Pfeiffer (2017) evidence that managers intentionally repeat some valuable information. On the other hand, repeated terms may also include information that is not valuable. Further, repeated information disturb and distract analysts and investors' attention for non-repeated terms when reading through the reports. So, analysts may oversee important information that is not repeated. Therefore, regulators should tradeoff between the benefits of emphasizing valuable information through repeated terms and cost of

obscuring important but non-repeated information. As a result, how to regulate the repetition in disclosure incur debates given the opposite outcomes. So, it's important to find a way to mitigate repetition side effect.

Since the existing literature discuss interactive terms that can navigate the investors and analysts more easily between different sections may enhance the readability and improve the exposure of useful key information (Fleming, 2015), we hope our study can provide evidence that interactive terms can increase analysts' attention on the important non-repeated information. Through our study, regulator might design more effective disclosure-format mandates that address potential distractions while retaining repeated terms benefits.

To isolate the interaction term's effect, we need randomization as above sections discussed. So, experiments are needed. When comparing the survey experiment with lab experiment, we find our setting is more suitable to conduct lab experiment in the university. Although there is self-selection bias, we want our sample to represent the future investors and analysts' behaviors working in the financial sectors. This selection bias, in turn, can help us limit our sample within students who already have accounting and finance knowledge to read the annual and proxy statements (Libby et al., 2002). As for survey experiment, it will introduce noises like costs for filtering the participant with accounting and finance skills and full-time survey participants that their intentions are to finish the survey rather than processing information to find high-value entities to invest (Dennis et al., 2020; Eyal et al., 2022). Motivation deviations will influence our inference. So, we will choose senior and MBA students in the university rather than using more complicated sample group that is unnecessary and costly.

To convince the audience why interaction can draw analyst attention on finding more non-repeated important information, we need a theory to have a sound reasoning. That's why we need behavioral science approach: we need to admit individuals' differences. Based on psychology theory, individual's perception will be influenced by the repeated terms. This form of repetition increases the truthfulness, reliability, and importance of the information (Hasher et al., 1977; Koch and Zerback, 2013). So, analyst will focus more on the repeated terms with increasing attention of them, and this increasing attention can deepen the process of handling repeated information and reduce the process of non-repeated information. In this situation, analysts treat the sign of repetition as "a heuristic cue" that will reduce their engagements of value relevant information that is shown in non-repeated, no interactive way (Sundar et al., 2014).

We argue that interactive terms can arise individual's attention on value relevant information, even this information is non-repeated. Moreover, they will increase their engagement of processing information and drilldown the useful value behind the information. So, our experiment will set different key adjustments for revenue and expense to detect whether our participants are engaged in the specific part of the annual report. Our key manipulations are repetitive frequency and interactive level of the annual report. From our topics of interest, we are now trying to use behavioral science approach to construct the randomization and psychological reasoning for our prediction.

4.2 Eye tracking measurement from physiological field

Followed by the prior reliable explorations for behavioral accounting research (Lynch and Andiola, 2019; Rose et al., 2022), we will try to use eye tracking technique to capture analyst attention. To keep our measurement consistent, we control the screen brightness and participants' interpupillary distance. To capture individual's processing of repeated and non-repeated information, we want to know participants' different level of attentions. Therefore, we use Fixation Count and Fixation Duration to proxy them.

These two measures are often used in research related to eye tracking and are also adopted in accounting research recently (Fehrenbacher, 2018; Brink, 2020). We will set the key manipulations statements as our capturing area (CA). For Fixation Count, it calculates the number of times

individuals focus on manipulated repeated and non-repeated information. Only focusing more than 60ms will be recorded. Fixation Duration records the total time participants' eyes focus on the CA. These two can represent different attention levels among the participants.

We expect that manipulating the repetition and interaction terms will yield differences in Fixation Count and Fixation Duration. Higher repeated terms may increase the Fixation Count and Duration in the repeated CA, which will enlarge the adjustment of the specific repeated discretionary item in analyst's forecast. In turn, higher repeated terms will decrease the adjustment of the specific non-repeated discretionary item. When we add more interactive terms, we expect the negative effect on non-repeated item will be mitigated. Setting both fixation and accounting adjustment measurements, we would better proxy changes in participants' physiological state and forecast accuracy with more physiological-based evidence. With behavioral approach like above, we can handle more interesting topics that traditional accounting research haven't touched before.

4.3 How to pinpoint early experience with observed outcome (management style) in firms?

As discussed in Section 3.2, the existing literature's limitation is we don't know clearly how CEO's early experience influences on the later decisions. Our research interest is on how CEO with economic shrinking experience in their first jobs will have aggressive behavior. We want to use physiological factor—cortisol level—to represent the underlying mechanism for aggressive behavior. Biology research evidence we can use cortisol level to represent the stress (Hellhammer et al., 2009). We want to first check if economic downturn CEOs have systematic higher level of salivary cortisol level than economic upturn CEOs have. Or we want to know if economic downturn CEO's cortisol level is more sensitive when facing business stress. These will support the management style literature using imprinting theory. Second, we will investigate the cortisol effect on firm's aggressive strategy, which will reveal biological factor's impact on firm's outcome. Biological evidence may be more convincing than only using the theory itself. We can benefit from interdisciplinary behavioral science approach.

5. Limitations compared to other approaches

5.1 Concerns about differences between lab and real-world contexts and ethical problem

The different contexts between experiment-based finding and archival-based finding may cause concerns about the causal inference. Participants' responses in a lab setting may not reflect real world behaviors. One possibility is that they are under different pressure level in different settings. In the Section 4.1 experiment, if we want to stimulate the stress and pressure in the real world when analysts make forecasts, we can add one more experiment to manipulate stress level in order that we can create a more realistic decision-making context. As mentioned in Section 4.3, we can use hydrocortisone pills to manipulate cortisol levels—a biomarker that directly represent physical and mental press in participant's body. While this can strengthen our experiment's inference, it is impossible to stimulate and replicate every context to let the experiment has the exact same setting as the real world.

Additionally, ethic concerns regarding participants' consent for double-blind experiments and taking medicine in the treatment group need to be well-managed before starting the studies. These require accounting researchers understand ethical guidelines in different disciplinaries and strictly follow them.

5.2 Generalization issues

In management style research, our experimental sample may not reflect actual CEO's behaviors. Also, many behavioral economic theories are drawn from the general population. As a result, these situations and assumptions may not apply to highly selected managers. For example, preference difference based on the gender may not exist in top management team, such as CEOs and CFOs. One explanation is that the person who are appointed in that position experience extensive selection processes before promoting into one firm's core (Hambrick and Mason, 1984). So, sample's representative issue may not be addressed before we have better understanding of the difference between general population and top executive.

6. Discussion

In general, whether samples in the experiment match the real-world entities is the key question and challenge when we want to conduct experimental research. For me, heterogeneity and generalization create a paradox; as we recognize our brain's recognitive processing is different with each other, we are hard to generalize our findings. Limited sample size, time-consuming and costly nature of the experiment also hinder the external validity in our research. Therefore, if we want to contribute to public policy, we need to balance between our evidence quality of causality and generalizability potentials.

We see the potential benefits for experimental method in addressing endogenous issue. Randomization can be realized in the well-designed experiment. Moreover, incorporating physiological and biological evidence, we can make our research's reasoning of the executives and analysts more persuasive.

However, it is insufficient to rely solely on these. Although behavioral science could give us randomization and solid interdisciplinary background to support our research, we cannot give up the reasoning from the existing theories. In this term, lecturers give us different literature using behavioral science that inspire me a lot in my current accounting research. But they also show us that experiment itself cannot give us intuitive idea and cannot identify the meaningful question in accounting research. Behavioral science can give us more effective and powerful tools on the "accounting" table. But without prior accounting theory, we cannot reveal the truth and maintain the relevance of the accounting profession in real-world practice.

Reference

Angrist, J. D., G. W. Imbens, and D. B. Rubin. (1996). Identification of causal effects using instrumental variables. Journal of the American Statistical Association 91 (434): 444–455. https://doi.org/10.1080/01621459.1996.10476902

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Armstrong, C. S., Glaeser, S. A., & Huang, S. (2022). Contracting with controllable risk. The Accounting Review, 97(4), 27-50.

Armstrong, C., Kepler, J. D., Samuels, D., & Taylor, D. (2022). Causality redux: The evolution of empirical methods in accounting research and the growth of quasi-experiments. Journal of Accounting and Economics, 101-521.

Bamber, L. S., J. Jiang, and I. Y. Wang. (2010). What's my style? The influence of top managers on voluntary corporate financial disclosure. The Accounting Review 85 (4): 1131–62.

Bourveau, T., and K. K. Law. (2021). Do disruptive life events affect how analysts assess risk? The Accounting Review 96 (3): 121–40.

Brink, A., A. Gouldman, J. Rose, and K. Rotaru. (2020). Effects of superiors' compensation structures on psychophysiological responses and real earnings management decisions of subordinate managers. Management Accounting Research 48: 100691. https://doi.org/10.1016/j.mar.2020.100691

Cazier, R. A., and R. A. Pfeiffer. "10-K Disclosure Repetition and Managerial Reporting Incentives." Journal of Financial Reporting 2 (2017): 107–31.

Christensen, H. B. (2019). Broad-versus narrow-sample evidence in disclosure regulation studies: A discussion of Badia, Duro, Jorgensen, and Ormazabal (2018). Contemporary Accounting Research, Forthcoming.

Dennis, S. A., Goodson, B. M., & Pearson, C. A. (2020). Online worker fraud and evolving threats to the integrity of MTurk data: A discussion of virtual private servers and the limitations of IP-based screening procedures. Behavioral Research in Accounting, 32(1), 119-134.

Eyal, P., David, R., Andrew, G., Zak, E., & Ekaterina, D. (2021). Data quality of platforms and panels for online behavioral research. Behavior Research Methods, 1-20.

Fleming, R. (2015). Effective Disclosure for the 21st Century Investor. Remarks at the Practising Law Institute (2015). Available at https://www.sec.gov/news/speech/022015-spchraf.html

Fehrenbacher, D., A. Schulz, and K. Rotaru. (2018). The moderating role of decision mode in subjective performance evaluation. Management Accounting Research 41: 1–10. https://doi.org/10.1016/j.mar.2018.03.001

Gow, I. D., Larcker, D. F., & Reiss, P. C. (2016). Causal inference in accounting research. Journal of Accounting Research, 54(2), 477-523

Guest, N. M. (2021). The information role of the media in earnings news. Journal of Accounting Research, 59(3), 1021-1076.

Haushofer, J., Lowes, S., Musau, A., Ndetei, D., Nunn, N., Poll, M., & Qian, N. (2022). Stress, Ethnicity, and Prosocial Behavior. NBER Working Paper

Hanlon, M., Yeung, K., & Zuo, L. (2022). Behavioral economics of accounting: A review of archival research on individual decision makers. Contemporary Accounting Research, 39(2), 1150-1214

Hasher, L.; D. Goldstein; and T. Toppino. (1977) "Frequency and the Conference of Referential Validity." Journal of Verbal Learning and Verbal Behavior, 16, 107–12.

He, X., H. Yin, Y. Zeng, H. Zhang, and H. Zhao. (2019). Facial structure and achievement drive: Evidence from financial analysts. Journal of Accounting Research 57 (4): 1013–57.

Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology, 34(2), 163-171.

Kandasamy, Narayanan, Ben Hardy, Lionel Page, Markus Schaffner, Johann Grag- gaber, Andrew S. Powlson, Paul C. Fletcher, Mark Gurnell, and John Coates, (2014). "Cortisol Shifts Financial Risk Preferences," Proceedings of the National Academy of Sciences, 111 (9), 3608–3613.

Koch, T., and T. Zerback. "Helpful or Harmful? How Frequent Repetition Affects Perceived Statement Credibility." (2013). Journal of Communication, 63, 993–1010.

Kumar, A. 2010. Self-selection and the forecasting abilities of female equity analysts. Journal of Accounting Research 48 (2): 393–435.

Libby, R.; R. Bloomfield; and M. W. Nelson. (2002). "Experimental Research in Financial Account- ing." Accounting, Organizations and Society, 27, 775–810.

Lynch, E. J., and L. M. Andiola. (2019). If eyes are the window to our soul, what role does eye-tracking play in accounting research? Behavioral Research in Accounting 31 (2): 107–133. https://doi.org/10.2308/bria-52283

Rose, A. M., Rose, J. M., Rotaru, K., Sanderson, K. A., & Thibodeau, J. C. (2022). Effects of uncertainty visualization on attention, arousal, and judgment. Behavioral Research in Accounting, 34(1), 113-139.

Samuels, D., Taylor, D. J., & Verrecchia, R. E. (2021). The economics of misreporting and the role of public scrutiny. Journal of Accounting and Economics, 71(1), 101-340.

Sundar, S. S.; S. Bellur; J. Oh; Q. Xu; and H. Jia. (2014). "User Experience of On-Screen Interac- tion Techniques: An Experimental Investigation of Clicking, Sliding, Zooming, Hovering, Dragging, and Flipping." Human-Computer Interaction, 29, 109–52.

Thaler, R. H. (2016). Behavioral economics: Past, present, and future. American economic review, 106(7), 1577-1600.