Accounting, Organizations and Society 116 (2026) 101632

Contents lists available at ScienceDirect

Accounting, Organizations and Society

journal homepage: www.elsevier.com/locate/aos

Performance evaluation criteria and information processing in complex decision making☆

Xiqiong He [a], Xi (Jason) Kuang [b], Ziyang Li [c], Jordan Samet [d,*]

[a] Southwestern University of Finance and Economics, China
[b] Georgia Institute of Technology, USA
[c] Sichuan University, China
[d] Indiana University, USA

ARTICLE INFO

Keywords:
Performance evaluations
Evaluation criteria
Goal hierarchy
Numerosity heuristic
Earnings fixation
Capital budgeting
Attentional scope
Experiment

ABSTRACT

We use two experiments to examine how the design of performance evaluation criteria influences agents' information processing in complex decision-making tasks. Drawing on goal hierarchy theory, we predict that, when agents are evaluated on the actions they take to perform their task (low-level evaluation criteria), they will focus primarily on salient cues in the task setting, whereas evaluating how agents' work impacts organizational objectives (high-level evaluation criteria) will increase their attention to less salient yet relevant information cues. As predicted, in the first experiment, we find that agents exhibit the numerosity heuristic under low-level evaluation criteria, indicative of over-attention to salient task cues, but not under high-level criteria. In the second experiment, we find that, when making performance predictions, agents under low-level evaluation criteria fixate on summary earnings measures while not attending adequately to less salient cues, and high-level evaluation criteria mitigate this fixation. Taken together, these findings provide convergent support for our theory. We discuss the implications of our findings for accounting research and practice.

1. Introduction

Performance evaluation criteria can range from *low-level*, whereby agents are assessed specifically on whether the actions they take to perform their assigned task are appropriate, to *high-level*, whereby agents are assessed on how their work contributes to the higher-order objectives of the organization. For example, project managers could be evaluated based on how they organize and coordinate subordinates' work, whether they adhere to project budgets and timelines, or whether they resolve problems and conflicts in a timely and effective manner (low-level criteria). Alternatively, they can be evaluated based on the extent to which their project helps improve the organization such as how their decisions help enhance organizational efficiency or how their management approach better prepares the organization for future market changes (high-level criteria). While low-level evaluation criteria are commonplace, goal-setting frameworks, such as the SMART (Specific, Measurable, Attainable, Relevant, and Time-bound) principles, promote improving employees' motivation by connecting local goals to organizational objectives (i.e., high-level criteria) (Bipp & Kleingeld, 2011; Oracle Corporation, 2012). Departing from these motivational considerations, we examine how evaluation criteria influence agents' information processing.

A growing literature examines how performance measurement systems facilitate agents' processing of accounting information, that is, their decision-facilitating role (van Veen-Dirks, 2010). Farrell et al. (2008) show that seeding forward-looking information within performance metrics improves agents' discovery of better production strategies. Zureich (2023) finds that including uncontrollable performance measures in incentive contracts increases agents' willingness to extract useful information from these measures thereby improving performance. We build on this approach by investigating how performance evaluation criteria affect agents' information processing and judgments in complex decision-making tasks, defined as tasks that require the acquisition, retrieval, and/or integration of information, cues, and data

☆ We appreciate the helpful comments from Nicole Donahoo, Christie Hayne, Gary Hecht, Michael Majerczyk, Evelien Reusen, Nate Waddoups, Laura Wang, Michael Williamson, Li Yang, Dan Zhou, Jake Zureich, and workshop participants at the University of Illinois – Urbana Champaign, the 2021 AAA ABO Conference, and 2022 AAA Management Accounting Midyear Meeting.

* Corresponding author.
E-mail address: josamet@iu.edu (J. Samet).

https://doi.org/10.1016/j.aos.2026.101632
Received 12 September 2023; Received in revised form 19 January 2026; Accepted 20 January 2026
Available online 27 January 2026

to reach a conclusion or judgment. These tasks are both important and prevalent in practice, particularly in environments where input-output relationships are ambiguous and no universally applicable solution is readily available (Bradley, 2009; Cardinaels et al., 2024; Earley, 2002).

According to goal hierarchy theory, a primary *superordinate* goal is often achieved through the attainment of a series of intermediate *subordinate* goals (Höchli et al., 2018). Low-level evaluation criteria prompt agents to focus on the *subordinate* goal of taking actions they perceive as appropriate for accomplishing the task at hand. While subordinate goals can enhance motivation (Locke & Latham, 2002), we argue that an excessive focus on subordinate goals, induced by low-level evaluation criteria, can affect agents' information processing by narrowing their attentional scope (i.e., the breadth of contextual cues that agents attend to). This, in turn, leads agents to attend primarily to salient cues that stand out in the task context while losing sight of less salient but relevant cues. As a result, agents may be more sensitive to these salient cues in their decision making. On the other hand, high-level evaluation criteria highlight the connection between agents' actions (*subordinate* goals) and the organization's *superordinate* goals (Hammersley & Ricci, 2021; Pringle et al., 2001; Rowe et al., 2007). We argue that priming this link broadens agents' attentional scope. Consequently, agents are less likely to fixate narrowly on salient cues, which reduces the influence of these cues on their information processing and subsequent decision making.

To test our predictions, we examine how the nature of evaluation criteria (low-level versus high-level) influences agents' decision making in a capital budgeting setting.[1] Within this setting, we examine whether low-level evaluation criteria lead agents to attend more to salient contextual cues and whether high-level criteria broaden agents' attention thereby reducing the impact of those salient cues. We focus on two different kinds of contextual cues inherent in the capital budgeting setting: (1) salient *presentational attributes* of the information and (2) salient *performance indicators* within the information. Triangulating across these different types of contextual cues provides a stronger overall test of our theory. We conducted two experiments to test these two effects, respectively.

In experiment one, participants are tasked with making a profit estimation and capital investment decision for a subsidiary company. Holding constant the financial information provided, we assess the downstream effects of participants' attentional scope by examining their proclivity to engage in the numerosity heuristic, by which individuals overweight (underweight) the number (unit) used to represent quantitative information, as the number is a more salient feature than the unit in the underlying information (Raghubir & Srivastava, 2002).[2] We expect that individuals with a narrower attentional scope are more likely to overly focus on the salient presentational feature (i.e., the number) and, therefore, infer greater magnitude when financial information is presented in big numbers relative to small numbers. We manipulate (1) whether agents are provided historical financial information presented in big numbers (small units) or in small numbers (big units), and (2) whether agents are evaluated upon their written justification of the process they use to perform the task (low-level evaluation criteria) or upon their written justification for why their profit estimation and investment recommendation helps achieve organizational objectives (high-level evaluation criteria).

Consistent with our predictions, we find that agents under low-level evaluation criteria engage in the numerosity heuristic, making greater profit estimates and investment recommendations when the financial information is presented in big numbers than when the same information is presented in small numbers. By comparison, agents' profit estimates and investment recommendations do not statistically differ across numerical presentations under the high-level criteria. In addition, we find that participants under low-level criteria are more likely to refer to the salient historical financial performance information in their written justifications, whereas those under high-level criteria refer to a more balanced range of information. We also address several alternative explanations via the post-experimental questionnaire as well as through two supplemental experiments.

In experiment two, we use a similar capital budgeting setting, where participants review a subsidiary's prior performance information and then assess its future prospects. We manipulate whether agents are evaluated on low-level or high-level evaluation criteria the same way as in experiment one. Additionally, we manipulate the content of the historical financial performance while holding constant the bottom-line net income. Specifically, in one condition net income is primarily driven by persistent earning streams, whereas in the other condition it is propped up by transitory income streams such that the subsidiary's future prospects are greater in the former scenario than in the latter (Elliott et al., 2011). Our theory suggests that individuals with a narrower attentional scope are more likely to fixate on the salient net income metric (i.e., earnings fixation) and, therefore, less likely to distinguish between persistent and transitory income stream scenarios. By comparison, individuals with a broader attentional scope are less fixated on net income and more likely to make differentiated performance assessments across these scenarios. As predicted, we find that participants assess the subsidiary's performance prospects as better with persistent relative to transitory income streams under high-level evaluation criteria, and this difference is attenuated under low-level evaluation criteria. These results provide further support for our theory.

This paper makes several important contributions to theory and practice. First, our study adds to the growing literature on the decision-facilitating role of performance metrics. While prior research examines how variations in the content and structure of performance measures convey different informational signals that influence agents' task strategies (e.g., Choi et al., 2012; Farrell et al., 2008; Zureich, 2023), we hold signaling opportunities constant and investigate how incorporating the firm's higher-order objectives into evaluation criteria affects agents' judgments. We find that such evaluation criteria can dilate agents' attentional scope thereby reducing the impact of salient informational cues in complex decision-making tasks. To the extent that the level of salience naturally varies across different cues in information sets (e.g., Harris et al., 2016; Lipe & Salterio, 2000), our study informs managers as to how the design of performance evaluation criteria can influence agents' information processing approach. Our study also responds to prior calls for more research on managerial judgments in complex decision settings (Humphreys et al., 2016).

In addition, we contribute to the literature examining the detriments of agents focusing on low-level evaluation criteria. For example, Choi et al. (2012) find that agents engage in surrogation with a single, task-specific, performance metric whereas Bentley (2019) shows that written justifications can mitigate this behavior. Building on this line of work, our study finds that requiring justifications based on low-level evaluation criteria can cause agents to engage in cognitive processing that reduce the consideration of potentially relevant, yet less salient, information.

Furthermore, we extend the understanding of how cognitive

[1] Capital budgeting is a complex decision-making task because the task is ill-structured, and outcomes can be significantly influenced by whether a wide variety of information is appropriately processed and incorporated. In addition, because of the significant amount of environmental uncertainty surrounding such decisions, principals may choose to not rely on *ex post* objective performance outcomes in evaluations (Baker et al., 1994).

[2] We focus on the numerosity heuristic for several reasons. First, variations in financial information's numerosity format are commonplace in companies' internal information transmission. Second, this heuristic is driven by individuals' (in)attention to more (less) salient presentational attributes of the information. Third, to test it requires only subtle and unobtrusive experimental manipulation, minimizing the variation needed in the experimental setting across treatments. Finally, the numerosity heuristic provides an unambiguous rational benchmark in that the content and relevance of quantitative information are held constant (Cardinaels, 2008), with any difference attributable to one's (in) attention to (less) salient attributes.

approaches influence decision makers' use of accounting information. Prior research shows the benefits of deliberative, as compared to intuitive, motivational approaches by attenuating affective responses in investment decisions (Farrell et al., 2014) and the use of irrelevant information in performance evaluations (Fehrenbacher et al., 2018). We provide evidence that evaluation criteria can influence agents' attentional scope, an informational-processing cognitive state, thereby affecting how agents' attention is drawn across information cues, such that broadening one's attentional scope can help mitigate the numerosity heuristic and reduce earnings fixation. These findings are particularly useful in accounting practices, where the effective processing of quantitative financial information plays an important role in decision making.

2. Background and theory development

2.1. The decision-facilitating role of performance evaluation criteria

The establishment of performance metrics (i.e., the indicators or criteria used to assess performance) is an important part of firms' management control systems (Merchant & Van der Stede, 2017). When incorporated in compensation contracts, performance metrics can have a strong impact on agents' behavior, aligning their actions with organizational goals (Lambert, 2001). Traditionally, accounting research has focused on this decision-influencing role of performance metrics, examining how different types of measures (e.g., financial, nonfinancial, and relative performance) and evaluation principles affect agents' effort (Banker et al., 2000; Datar et al., 2001; Ittner et al., 2003; Lambert & Larcker, 1987; Lau et al., 1995; Matsumura & Shin, 2006).

An emerging literature focuses on the decision-facilitating role of performance metrics by examining their effects on agents' information processing. For example, Farrell et al. (2008) show that including product quality (a forward-looking measure) in incentive contracts directs agents to better integrate feedback information, helping them identify optimal task strategies. Farrell et al. (2012) find that accompanying performance metrics with additional information on causal linkages between these measures improves agents' effort-allocation decisions. Zureich (2023) finds that compensating agents based on both controllable (their own performance) and uncontrollable (others' performance) factors can prompt them to extract useful information from uncontrollable factors and, in turn, improve decision quality. Our study extends this literature by examining how key attributes of performance evaluation criteria influence agents' processing of different information cues in complex decision-making tasks, a common practice when utilizing accounting information to make decisions.

When designing evaluation criteria, principals need to specify which dimension of performance the evaluation will be based upon (Folger et al., 1992). A key feature of evaluation criteria is the extent to which broader organizational objectives are relevant in the agents' evaluation. Evaluation criteria may focus on whether the specific actions and steps agents take to carry out their assigned task are appropriate, without explicit reference to organizational objectives (i.e., low-level evaluation criteria), or they may be based on how agents' efforts and decisions help achieve organizational objectives (i.e., high-level evaluation criteria).[3]

Agency theory and expectancy theory suggest that low-level evaluation criteria can increase motivation (Merchant & Van der Stede, 2017; Shields & Waller, 1988), and this may explain why low-level criteria are common in corporate practices (David, 2013; Grubb, 2007).[4] On the other hand, making salient the higher-order purposes of the task (e.g., SMART goals) can improve goal commitment and thus sustain goal-pursuit effort (Harackiewicz & Elliot, 1998; Ishak et al., 2019; Weintraub et al., 2021).[5] While understanding how evaluation criteria motivate agents' effort (i.e., decision-influencing effects) is important in its own right, we examine how they differentially influence agents' information processing in complex decision-making tasks (i.e., decision-facilitating effects). This question is important because in these ill-structured tasks, the manner in which agents search, process and integrate information can play a larger role in the quality of judgments than that of effort exertion.

2.2. Goal hierarchy theory and attentional scope

Goal hierarchy theory proposes that a plan to achieve the superordinate (i.e., primary) goal often consists of a series of intermediate milestones, referred to as subordinate goals (Höchli et al., 2018; Houser-Marko & Sheldon, 2008). Assigning subordinate goals provides achievable mile-markers that individuals can complete piece-meal, which can improve self-efficacy, maintain motivation, reduce perceived outcome ambiguity, and mitigate performance fluctuations (Berger, 2019; Heath et al., 1999; Locke et al., 1989; Locke & Latham, 1990, 2002; Sun & Frese, 2013; Wright & Kacmar, 1994). However, overly emphasizing subordinate goals can cause the neglect of higher-order goals, negatively affecting performance (Kirschenbaum et al., 1981; Masicampo & Baumeister, 2012; Staw & Boettger, 1990; Tenbrunsel et al., 2000).

During information processing, agents encounter a wide array of cues, and their goals can influence how their attention is drawn toward these cues (Juola et al., 2000). Building on goal hierarchy theory, we argue that low-level evaluation criteria fixate agents on a single goal hierarchical level, that is, the subordinate goal of executing appropriate actions to fulfill the task at hand. In turn, agents' cognitive approach and neural activation adapt to best maintain this singular focus when processing information (Aiello et al., 2018; Heinze & Münte, 1993; Itthipuripat et al., 2014; Liu et al., 2014). As part of this adaptation, agents' attentional scope (i.e., the range of contextual cues they attend to during information processing) may center around information cues perceived to be most directly related to the task thereby avoiding information overload (Croft et al., 2001). As a result, agents may focus primarily on salient cues within the task because these cues are more noticeable and easier to connect to the specific task (Albers, 2007; Alcocer & Torres, 2024; Bordalo et al., 2022). These salient cues, in turn, may play a larger role in agents' information processing. Thus, we contend that increased fixation on task-specific actions, driven by low-level evaluation criteria, prompts agents to focus on salient task cues and lose sight of less salient but relevant cues.

On the other hand, high-level evaluation criteria that emphasize the degree to which agents' work helps achieve superordinate goals may attenuate this effect. High-level evaluation criteria draw agents' attention to the link between their subordinate task and the higher-order reasons for completing that task. We argue that agents' cognitive

[3] High-level evaluations could be conducted either before or after task outcome becomes quantitatively measurable. In this study, we focus on settings where objective outcome measures are unavailable or where outcomes are not a reliable indicator of agents' input. Such situations are common in complex decision-making tasks (i.e., our setting of interest) due to high levels of external uncertainty inherent in these tasks. We leave to future research a systematic investigation of whether and how our theory and predictions apply when outcomes can reliably be used as performance measures.

[4] For example, software companies often evaluate development teams based on how the project is executed, including the quality of methods used, team communication, and overall project management (Wohlin & Andrews, 2001). Similarly, in industries characterized by rapidly changing demands and technologies, performance evaluations place strong emphasis on ensuring task approaches and procedures adhere to standards (Chen & Lee, 2007; Milosevic & Patanakul, 2005).

[5] The "Relevant" principle of the SMART framework emphasizes the importance of aligning local goals with organizational goals (Oracle Corporation, 2012; Weintraub et al., 2021).

approach and neural activation will adapt to allow them to maintain both goal hierarchical levels when processing information, leading to a broader attentional scope (Aiello et al., 2018; Heinze & Münte, 1993; Itthipuripat et al., 2014). Thus, agents are likely to attend to, process, and encode a more balanced set of information cues, which attenuates the relative influence of any individual salient cues (Cave & Chen, 2016; Wang et al., 2024).[6]

In summary, we predict that low-level evaluation criteria will lead agents to overly attend to salient cues such that these cues will play a more influential role in their assessments and decision making, whereas high-level evaluation criteria will broaden their attentional focus during information processing thereby reducing the impact of salient cues on their decision making. We test our theory in a capital budgeting setting in which agents make assessments of a subsidiary's future performance and recommend investments in the subsidiary. Within this setting, we conduct two experiments, both examining how low-level versus high-level evaluation criteria influence agents' relative attention to more salient contextual cues. In the first experiment we test agents' susceptibility to the numerosity heuristic (i.e., whether they overly focus on salient presentational features), whereas in the second experiment we examine agents' susceptibility to earnings fixation (i.e., whether they overly focus on salient performance indicators). Across these two experiments, we triangulate behavioral outcomes from the processing of different types of contextual cues, which provides a stronger test of our theory and helps enhance the robustness of the test results. We develop a testable hypothesis for each experiment.[7]

3. Experiment one

3.1. Hypothesis development and the numerosity heuristic

In this experiment, we hold constant the financial information provided for a capital budgeting task and examine whether agents under low-level, relative to high-level, evaluation criteria are more likely to engage in the numerosity heuristic, a decision-making heuristic that is precipitated by the relative (in)attention to more (less) salient attributes of information. Specifically, quantitative information is typically presented as a face value (i.e., the number) paired with a measurement unit. The same informational content can be presented as a bigger number in smaller units (e.g., 1,630,000 dollars) or a smaller number in bigger units (e.g., 1.63 million dollars). From a standard economic perspective, the specific number and unit used to present quantitative information should have no impact. However, research has consistently found that the face value, relative to the measurement unit, is a more salient attribute of the quantitative information (Raghubir & Srivastava, 2002). As a result, individuals often focus on the face value without adequately considering its measurement unit (Pandelaere et al., 2011). For example, individuals may base purchasing decisions on the face value of a currency rather than its real value (Shafir et al., 1997; Wertenbroch et al., 2007).

As discussed earlier, we argue that low-level evaluation criteria heighten the attention to the subordinate goal of taking appropriate actions to carry out the specific task, leading agents to focus more on salient cues (i.e., the face value) while inadequately attending to other relevant but less salient cues (i.e., the measurement unit). Thus, agents under low-level evaluation criteria may be particularly susceptible to the numerosity heuristic. In this case, because bigger face values are associated with greater perceived magnitude of the underlying financial information, we expect these agents to infer greater magnitude from information presented in larger face values with smaller units than from the same information presented in smaller face values with larger units.

By comparison, high-level evaluation criteria draw attention to the link between the task and the superordinate goal for performing the task, leading to a relatively broad attentional scope. With a broadened attentional scope, agents are more likely to process information cues in a more holistic and comprehensive manner, which will attenuate the impact of salient attributes, thus reducing the numerosity heuristic. Consistent with prior accounting literature showing that variations across economically equivalent measurement units can influence agents' judgment and decision making (Church, Kuang, & Liu, 2019; Church, Jiang, et al., 2019), we posit that such influences can be mitigated through a high-level evaluation approach emphasizing the link between subordinate task goals and superordinate organizational goals.

In summary, we predict an interaction such that, holding financial performance constant, agents' performance assessments and investment recommendations will be affected by the numerical presentation of financial information under low-level evaluation criteria and this effect will be attenuated under high-level criteria. To provide a clean test of our theory, we hold constant the valence of all financial information such that prior financial performance is unambiguously positive. Consequently, we have a straightforward directional prediction that under low-level evaluation criteria agents will provide higher profit estimates and investment recommendations when financial information is presented in big numbers than when presented in small numbers, and this directional difference will diminish under high-level evaluation criteria.[8]

We formally state the above prediction as our first hypothesis:

H1. Agents under low-level performance evaluation criteria will make greater profit estimates and capital investment recommendations when performance information is in larger face values (smaller units) than when the same information is in smaller face values (larger units), and this effect will be attenuated for agents under high-level performance evaluation criteria.

3.2. Experimental setting and design

In the experiment, participants are assigned the role of a manager in ABC Corporation, a manufacturer of medical supplies and equipment. Senior management of ABC Corporation is considering whether to make further investment in one of its subsidiaries, the Pharmaceutical Company. Participants are given selected financial information about the subsidiary for Year *t-2* and Year *t-1*, including: (1) 11 major financial measures for Year *t-2* and Year *t-1*, (2) the value change in each measure from Year *t-2* to Year *t-1*, and (3) the percentage change in each measure from Year *t-2* to Year *t-1*. We include percentage changes because prior research found that, even with ratio information, individuals still consider the magnitude of the absolute values (Darke & Freedman, 1993). Since this measure is constant between big and small number manipulations, it provides a more conservative test of our hypothesis. Participants are asked to estimate the pharmaceutical subsidiary's net income in Year *t* and provide a recommendation of how much of two million dollars the parent corporation should invest in the subsidiary. In addition, participants are asked to provide a written response justifying aspects of their estimate and investment recommendation (described in more detail below).

The experiment uses a 2 × 2 between-participants design, with two

[6] Prior research suggests that a broader attentional scope can facilitate the integration and absorption of multiple pieces of information (Lin, 2009; Macrae & Lewis, 2002; Pringle et al., 2001).

[7] The experiments reported in this paper have been approved by the applicable universities.

[8] Note that we do not predict whether low-level criteria are less likely to lead to the "optimal" investment level or whether low-level criteria will increase or decrease investment levels, on average. Instead, we follow Lipe and Salterio's (2000) approach to test for cognitive heuristics by examining whether low-level criteria lead to different investment levels across big and small number conditions even though rational decision models predict no such difference.

manipulated factors. The first factor is evaluation *CRITERIA* via the justification report. In the LOW-LEVEL condition, participants are asked to justify "the process you used for determining your net income estimation and investment" and they will be evaluated on the quality of this justification. This evaluation criterion refers only to task-specific actions without making reference to the superordinate objectives of the task. In the HIGH-LEVEL condition, participants are asked to justify "why your net income estimation and investment recommendation help ABC Corporation achieve its strategic objective (i.e., to maximize profits)" and they will be evaluated on the quality of this justification. This evaluation criterion links participants' task-specific actions (subordinate goal) to their decision's impact on the broader organizational objectives (superordinate goal). Importantly, all participants learn the corporation's strategic objective in the case materials, but only in the HIGH-LEVEL condition is this strategic objective included in the justification requirement description and thus part of the evaluation criteria.

The second factor is whether the subsidiary's performance information, *NUMBER*, is presented in big (i.e., dollars) or small (i.e., million dollars) numbers, except for percentage changes, hereafter referred to as the BIG and SMALL conditions (presented in Fig. 1, Panels A and B, respectively). To be consistent with our manipulation, we also elicited the primary dependent measures in the same unit that the performance information was provided in.

3.3. Participants and procedures

Eighty-two business professionals were recruited from a part-time Master of Accountancy program at a major university in China to participate in the experiment. Participants' average age was 25 years and 29 percent were male. On average, participants had 1.9 years of full-time work experience. In all conditions, participants are first provided with background materials about ABC Corporation and the subsidiary as well as the subsidiary's performance information, as described above.[9] After reviewing the performance information, participants read instructions regarding their decision task to estimate the subsidiary's net income and recommend an amount, out of two million dollars, to invest in the subsidiary. Next, participants read instructions regarding the justification report as described in the previous section.

To ensure the evaluation criteria manipulation was strongly attended to, we provided participants an economic incentive contingent on the quality of their written justification. Specifically, in both conditions, participants were told that their justification report will be evaluated by a panel of experts and those rated in the top 10 percent will receive a bonus of RMB100.[10] After reading the instructions, participants were asked to provide a profit estimation and then make an investment recommendation. We adopt this sequence from prior accounting studies (Elliott, 2006; Guiral et al., 2020; Tang & Venkataraman, 2018) as it mirrors the process that financial information users typically follow when making judgments and decisions in practice (Brown et al., 2015; Sharma, 2006). Finally, participants provided their written justification and responded to a series of post-experimental questions.

3.4. Tests of H1

We have two main independent variables. For evaluation criteria, we create a dummy variable, labeled *CRITERIA*, which equals one for the HIGH-LEVEL condition and zero otherwise. For numerical presentation, we create a dummy variable, labeled *NUMBER*, which equals one for the SMALL condition and zero otherwise. We have two main dependent variables: participants' estimated profit and investment recommendation amount. Table 1 and Fig. 2 present the descriptive results of estimated profits (Panel A) and investment recommendation amounts (Panel B) in each of the four experimental conditions.[11]

Our theory argues that low-level criteria focus agents' attention on salient information cues whereas high-level criteria diffuse their attention across both salient and less salient information cues. However, prior literature finds that, if the influence of salient cues on evaluative judgments arises from intuitive "System 1" processing, increased effort can mitigate this influence by shifting individuals toward more deliberative "System 2" processing (Farrell et al., 2014; Fehrenbacher et al., 2018; Kennedy, 1993). From this perspective, even with a greater focus on salient information, increased effort might still lead to more processing of less salient information. Thus, two cognitive processes could independently influence individuals' susceptibility to the numerosity heuristic in our setting: the dispersion of attention across information cues (as predicted by our theory) and the level of overall effort in information processing.

We test our theory after controlling for this alternative—but not mutually exclusive—mechanism. In the post-experimental questionnaire, participants responded to, "How much attention and effort did you exert while performing this profit estimation task?" on an 11-point scale (1 = "no effort at all" and 11 = "extreme effort"), labeled *EFFORT*. To confirm that it is appropriate to include *EFFORT* in our hypothesis testing models, we first examine whether *EFFORT* differs across conditions: untabulated results show that *EFFORT* does not significantly differ either across BIG and SMALL conditions (4.78 versus 4.68, $t_{80} = 0.240$, $p = 0.811$, two-tailed) or across LOW-LEVEL and HIGH-LEVEL conditions (4.86 versus 4.60, $t_{80} = 0.633$, $p = 0.529$, two-tailed).[12]

To test H1, we conduct two OLS regressions, with estimated profit and investment recommendation, respectively, as the dependent variables. The independent variables include *NUMBER*, *CRITERIA*, and their interaction. We also include *EFFORT* and the interaction of *EFFORT* and *NUMBER* to account for effort's potential moderating effect (Piercey, 2023). First, we examine participants' profit estimates. In Panel A of Table 2, we find that the effect of *NUMBER* is significantly moderated by *EFFORT* ($p = 0.040$, two-tailed). Further analyses (untabulated) utilizing the Johnson-Neyman technique (Preacher et al., 2006) suggest that information presented in big numbers leads to higher profit estimates than information presented in small numbers when $EFFORT \leq 3$ (the lower 35.4th percentile) but this effect diminishes as effort increases (i.e., $EFFORT \geq 4$). Thus, consistent with prior research, increased effort attenuates differences in profit estimates across *NUMBER* conditions. More importantly, after controlling for the effect of overall effort, the results show a significant interaction between *NUMBER* and *CRITERIA* ($p = 0.032$, one-tailed). Simple effect tests further reveal that in the LOW-LEVEL condition, estimated profit is significantly higher ($p =$

[9] The experimental instrument was translated from English to Chinese using the back-translation method (Brislin, 1986).

[10] When the experiment was conducted, the exchange rate between the Chinese RMB and US dollars was approximately \$1 = RMB 6.96. After the experiment was completed, participants' justification reports were reviewed by a selected group of scholars and practitioners, and those who won the bonus were paid in cash approximately one week after the study was completed.

[11] While we ask participants to determine an investment amount out of two million dollars, two participants entered an investment amount higher than two million dollars. In our tables and analyses, we replace these two participants' investments with two million. Statistical inferences remain unchanged if their original investment amounts are used.

[12] In addition, participants indicated their perceived task difficulty and confidence in performing the task on an 11-point scale anchored at 1 (not confident/difficult at all) and 11 (extremely confident/difficult), respectively. Untabulated results indicate no statistically significant differences between the LOW-LEVEL and HIGH-LEVEL conditions for perceived task difficulty (6.57 versus 5.93, $p = 0.197$, two-tailed) or confidence in performing the task (4.24 versus 4.78, $p = 0.203$, two-tailed). These results are collectively consistent with our theory, which posits that the key difference between LOW-LEVEL and HIGH-LEVEL conditions lies in how participants' attention is drawn toward different information cues rather than in the total amount of effort or cognitive resources they dedicate to the task.

Panel A: The BIG condition

Measure	Two years prior	Three years prior	Difference	Difference in percentage
Sales (*dollars*)	24,340,000	18,170,000	+6,170,000	33.96%
Income from operations (*dollars*)	1,630,000	1,110,000	+520,000	46.85%
Net income (*dollars*)	1,450,000	990,000	+460,000	46.46%
Cash flow from operations (*dollars*)	1,400,000	910,000	+490,000	53.85%
Total assets (*dollars*)	16,300,000	14,160,000	+2,140,000	15.11%
Current assets (*dollars*)	11,440,000	9,790,000	+1,650,000	16.85%
Property, plant and equipment (*dollars*)	3,010,000	2,660,000	+350,000	13.16%
Intangible assets (*dollars*)	250,000	240,000	+10,000	4.17%
Total liabilities (*dollars*)	7,120,000	5,960,000	+1,160,000	19.46%
Current liabilities (*dollars*)	6,940,000	5,760,000	+1,180,000	20.49%
Owners' equity (*dollars*)	9,190,000	8,200,000	+990,000	12.07%

Panel B: The SMALL condition

Measure	Two years prior	Three years prior	Difference	Difference in percentage
Sales (*million dollars*)	24.34	18.17	+6.17	33.96%
Income from operations (*million dollars*)	1.63	1.11	+0.52	46.85%
Net income (*million dollars*)	1.45	0.99	+0.46	46.46%
Cash flow from operations (*million dollars*)	1.40	0.91	+0.49	53.85%
Total assets (*million dollars*)	16.30	14.16	+2.14	15.11%
Current assets (*million dollars*)	11.44	9.79	+1.65	16.85%
Property, plant and equipment (*million dollars*)	3.01	2.66	+0.35	13.16%
Intangible assets (*million dollars*)	0.25	0.24	+0.01	4.17%
Total liabilities (*million dollars*)	7.12	5.96	+1.16	19.46%
Current liabilities (*million dollars*)	6.94	5.76	+1.18	20.49%
Owners' equity (*million dollars*)	9.19	8.20	+0.99	12.07%

Fig. 1. Experiment 1: Performance information.

Table 1
Experiment 1 descriptive statistics.

Panel A: *ESTIMATED PROFIT* mean (SD) [*N*]		
NUMBER	CRITERIA	
	LOW-LEVEL	HIGH-LEVEL
BIG	2,173,437	2,005,043
	(504,247)	(152,385)
	[21]	[20]
SMALL	2,021,667	2,044,125
	(294,836)	(194,475)
	[21]	[20]
Panel B: *INVESTMENT AMOUNT* mean (SD) [*N*]		
NUMBER	*CRITERIA*	
	LOW-LEVEL	HIGH-LEVEL
BIG	1,397,143	1,206,100
	(400,714)	(442,537)
	[21]	[20]
SMALL	1,174,762	1,278,000
	(398,091)	(521,330)
	[21]	[20]

Notes.
NUMBER = a dummy variable which equals one for the SMALL condition, in which performance information is presented in small numbers (big units), and zero for the BIG condition, in which performance information is presented in big numbers (small units).
CRITERIA = a dummy variable which equals one for the HIGH-LEVEL condition, in which participants' justification report asks them to justify why their decision helps achieve the corporation's strategic objective, and zero for the LOW-LEVEL condition, in which participants' justification report asks them to justify their decision process.
ESTIMATED PROFIT = participants' estimate of The Pharmaceutical Company's subsequent year net income, in dollars.
INVESTMENT AMOUNT = participants' recommendation to The ABC Corporation how much to invest in The Pharmaceutical Company, in dollars. The investment level ranges from $0 to $2,000,000.

0.023, one-tailed) when financial information is presented in big numbers than when the same information is presented in small numbers, suggesting that these participants engage in the numerosity heuristic. By comparison, *NUMBER* does not significantly influence estimated profit in the HIGH-LEVEL condition ($p = 0.509$, two-tailed), suggesting that high-level evaluation criteria mitigate the numerosity heuristic. Taken together, these results are consistent with our theory.

Next, we examine participants' investment recommendations, reported in Panel B of Table 2. In this test, we find that *EFFORT* does not significantly influence the effect of numerical presentation on participant's investment recommendations ($p = 0.601$, two-tailed). Notably, similar to estimated profit, we find a significant interaction between *NUMBER* and *CRITERIA* ($p = 0.050$, one-tailed). Simple effect tests show that under the LOW-LEVEL condition participants recommended larger investments ($p = 0.040$, one-tailed) when financial information is presented in big numbers rather than small numbers. By comparison, investment recommendations do not significantly differ between big and small numbers in the HIGH-LEVEL condition ($p = 0.548$, two-tailed). Overall, these results are consistent with H1, suggesting that low-level performance evaluation criteria lead to a narrower attentional scope demonstrated by individuals engaging in the numerosity heuristic, whereas high-level evaluation criteria lead to a broader attentional scope by mitigating this heuristic.

3.5. Additional analyses – qualitative responses

Consistent with our theory, our main results show that participants in the LOW-LEVEL condition are more influenced by presentational formats than those in the HIGH-LEVEL condition. To further test our theory through a different lens, we examine participants' written justification reports. We expect participants in the LOW-LEVEL condition to focus on the salient numeric data (i.e., historical financial information) whereas we expect those in the HIGH-LEVEL condition to reference a more balanced range of factors.[13]

We employed two Ph.D. students, blind to condition, to code participants' written justification reports. The coders counted every instance in which a response explicitly referred to the subsidiary's historical performance information and every instance in which a response referred to forward-looking or external information.[14] The length of these reports does not significantly differ between LOW- and HIGH-LEVEL conditions ($p = 0.477$, two-tailed). However, participants in the LOW-LEVEL condition made more total historical references (7.23 vs. 4.33, $t_{72} = 2.94$, $p = 0.004$, two-tailed) and fewer total forward-looking/external references (2.57 vs. 3.82, $t_{72} = -1.83$, $p = 0.071$, two-tailed) than participants in the HIGH-LEVEL condition. We then created a measure, *REFERENCE,* calculated as the ratio of historical references over the total number of historical and forward-looking/external references, with values closer to 0.50 indicating a more balanced range of references. Untabulated t-tests reveal that *REFERENCE* is significantly greater than 0.50 (i.e., historical references accounted for more than half of the total references) in the LOW-LEVEL condition (0.72, $p < 0.001$, two-tailed) but not in the HIGH-LEVEL condition (0.53, $p = 0.614$, two-tailed). In addition, *REFERENCE* is significantly greater ($t_{72} = 2.84$, $p = 0.006$, two-tailed) in the LOW-LEVEL condition than in the HIGH-LEVEL condition.[15] These results provide further evidence that agents are more likely to focus on salient information under low-level, relative to high-level, evaluation criteria.

3.6. Additional analyses – construal levels

An alternative explanation for the results could be that highlighting the superordinate goal pushes individuals into a more abstract (versus concrete) construal level which enables them to grasp the defining features of the task (Burgoon et al., 2013), reducing their susceptibility to presentational effects. To address this, we measure participants' construal level in the post-experimental questionnaire.[16] Untabulated OLS results reveal that participants' construal level is higher ($p = 0.039$, two-tailed) in the HIGH-LEVEL condition (36.10) than in the LOW-LEVEL condition (31.38) and is also higher ($p = 0.028$, two-tailed) in the BIG condition (35.76) than in the SMALL condition (31.61), with no interaction effect ($p = 0.940$, two-tailed). These results suggest that evaluation criteria and numerical presentation additively influence participants' construal level, yet the pattern of construal level results across conditions does not match the predicted or observed pattern of

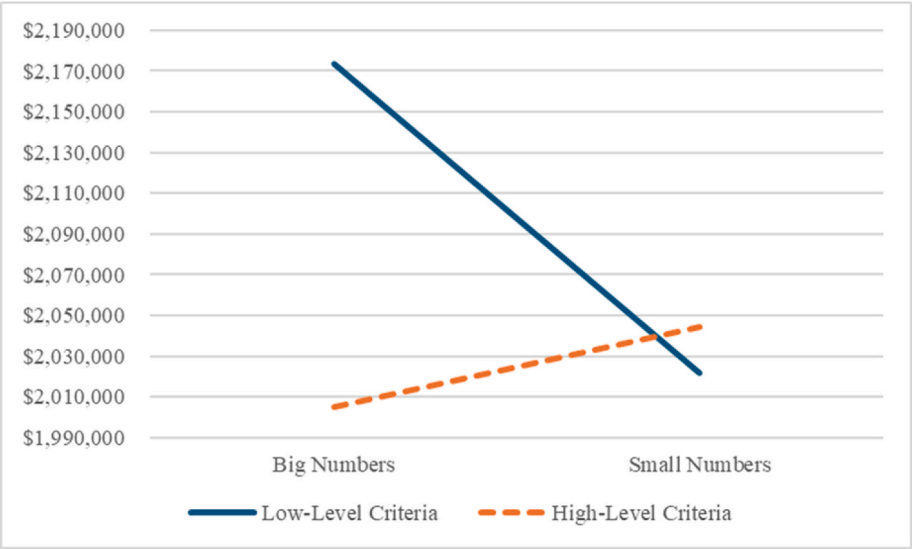
[13] The numerosity heuristic captured by our main analysis may not directly manifest in these reports because, given the nature of the justification requirement, participants are not likely to explicitly reference presentational attributes (e.g., units).

[14] The two coders' initial ratings were highly correlated (0.97 for historical information and 0.85 for forward-looking/external information; $p < 0.001$), with minor discrepancies resolved through discussion.

[15] The results reported include the 74 participants who provided a response in their justification report (eight participants did not provide written justification).

[16] We measure participants' construal level via the Behavioral Identification Form (BIF) (Vallacher & Wegner, 1987, 1989, 2014), a widely used and empirically validated instrument for assessing individuals' construal level. Each of the 25 items present a specific action (e.g., "locking a door") with two alternative restatements of this action, one at a concrete construal ("putting a key in the lock") and the other at an abstract construal ("securing the house"). Participants choose the statement that they thought best describes the given action. Each item is scored as two if the participant chose the abstract construal description and zero otherwise (i.e., range from zero to 50). The total score over all 25 items is used as a proxy for participants' construal level: the higher the score, the more abstract the construal level.

Panel A: *ESTIMATED PROFIT*



Panel B: *INVESTMENT AMOUNT*

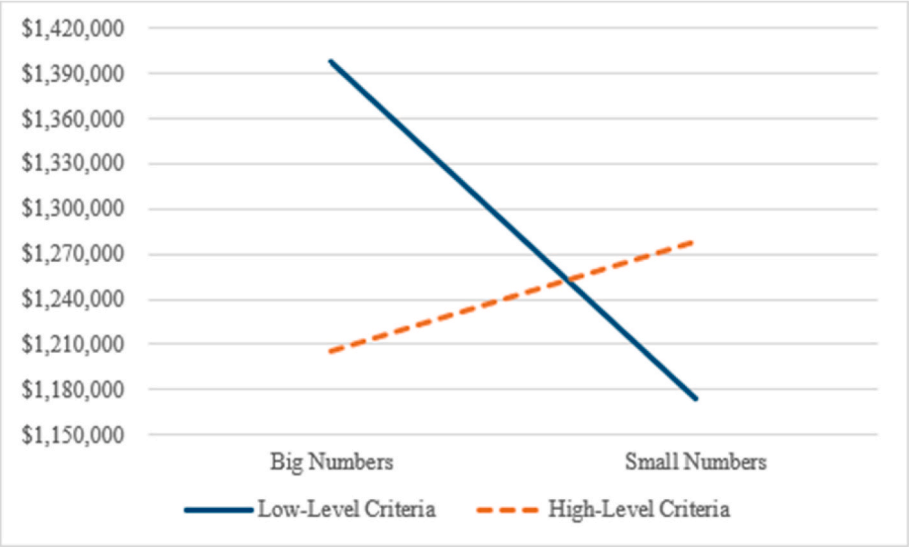


Fig. 2. Experiment 1 – mean profit estimate and investment.

the profit estimate or investment recommendation. Further, we replicate our hypothesis tests while controlling for participants' construal level and the results remain unchanged. Thus, we find no evidence that construal level drives our results.[17]

3.7. Supplemental experiment 1 – construct validity

In experiment one, while we manipulate low-level versus high-level evaluation criteria by varying what participants were asked to provide in their justification report, the description of the evaluation itself did not vary across conditions (i.e., *reports will be reviewed by a panel of experts in corporate financial analysis. The experts panel will evaluate and rate the quality of the reports*). It is possible that participants might perceive the manipulation as a difference in how the task should be performed, rather than a difference in evaluation criteria, which is our intended focus.

To evaluate the construct validity of our manipulation, we conducted a supplemental experiment recruiting 121 participants from Prolific. The design and procedures largely mirror those of experiment one, except that participants were presented with *both* HIGH-LEVEL and LOW-LEVEL evaluation criteria. They were then asked to indicate the extent to which four items differ across these two evaluation criteria: (1) performing the net income estimation, labeled *Estimation*, (2) performing the investment recommendation, labeled *Investment*, (3) the criterion by which management evaluates the employee, labeled *Criterion*, and (4) the content of the justification report, labeled *Content*.

We conduct a principal components analysis of these four items using a varimax rotation. By definition, we expect *Criterion* to naturally load on one component (*EVAL_CRITERIA*) and *Estimation* and *Investment* to load on a second component (*TASK_REQUIREMENT*). We then examine which of these two components *Content* more strongly loads on: a stronger loading on *EVAL_CRITERIA* would indicate that participants

[17] We also conducted several mediated moderation analyses (Hayes, 2018) and under no specification did participants' construal level mediate the effect of our primary results.

Table 2
Experiment 1 tests of H1.

Panel A: OLS REGRESSION (Dependent variable = *ESTIMATED PROFIT)*				
Source	Coefficient	SE	t-stat	p-value
NUMBER	−186,989.74	92,545.42	−2.02	0.023[a]
CRITERIA	−226,412.60	94,635.90	−2.39	0.019
NUMBER× CRITERIA	249,767.56	132,816.22	1.88	0.032[a]
EFFORT	−105,948.06	30,135.30	−3.52	0.001
NUMBER × EFFORT	79,049.40	37,798.70	2.09	0.040
Constant	2,206,906.90	65,776.46	33.55	<0.001
Simple effects tests				
Effect of *NUMBER* given LOW-LEVEL	−186,989.74	92,545.42	−2.02	0.023[a]
Effect of *NUMBER* given HIGH-LEVEL	62,777.82	94,576.38	0.66	0.509
Panel B: OLS REGRESSION (Dependent variable = *INVESTMENT AMOUNT)*				
Source	Coefficient	SE	*t*-stat	*p*-value
NUMBER	−242,544.47	136,918.04	−1.77	0.040[a]
CRITERIA	−222,776.28	140,010.84	−1.59	0.116
NUMBER× CRITERIA	326,966.12	196,497.41	1.66	0.050[a]
EFFORT	−57,947.98	44,584.23	−1.30	0.198
NUMBER × EFFORT	29,395.50	55,921.99	0.53	0.601
Constant	1,415,499.30	97,314.20	14.55	<0.001
Simple effects tests				
Effect of *NUMBER* given LOW-LEVEL	−242,544.47	136,918.04	−1.77	0.040[a]
Effect of *NUMBER* given HIGH-LEVEL	84,421.65	139,922.77	0.60	0.548

Notes.
NUMBER = a dummy variable which equals one for the SMALL condition, in which performance information is presented in small numbers (big units), and zero for the BIG condition, in which performance information is presented in big numbers (small units).
CRITERIA = a dummy variable which equals one for the HIGH-LEVEL condition, in which participants' justification report asks them to justify why their decision helps achieve the corporation's strategic objective, and zero for the LOW-LEVEL condition, in which participants' justification report asks them to justify their decision process.
EFFORT = participant responses to the following question, "How much attention and effort did you exert while performing this profit estimation task?" on a scale from 1 (no effort at all) to 11 (extreme effort), mean centered.
ESTIMATED PROFIT = participants' estimate of The Pharmaceutical Company's subsequent year net income, in dollars.
INVESTMENT AMOUNT = participants' recommendation to The ABC Corporation how much to invest in The Pharmaceutical Company, in dollars. The investment level ranges from $0 to $2,000,000.
[a] An asterisk indicates a one-tailed p-value for a test of directional prediction.

perceive the difference in the justification report as lying in the evaluation criteria (as our manipulation is intended to capture), whereas a stronger loading on *TASK_REQUIREMENT* would suggest that participants attribute the difference in the justification report to task requirements.

Consistent with our expectations, untabulated results find two components emerging from the data, with eigenvalues of 1.713 and 1.692, respectively, and 85.13 % of the total variation explained. The first component, *EVAL_CRITERIA*, is primarily a function of *Criterion* and, most importantly, *Content*, with component loadings of 0.918 and 0.812, respectively. The second component, *TASK_REQUIREMENT*, is primarily a function of *Estimation* and *Investment*, with component loadings of 0.922 and 0.803, respectively. Notably, *Content* loads very weakly (0.396) on the second component, *TASK_REQUIREMENT*. This analysis demonstrates that perceived differences in the justification report across conditions are more strongly correlated with evaluation criteria than with differences in the task itself, suggesting that our manipulation of the justification report captures our theoretical construct of interest.

3.8. Supplemental experiment 2 – mindset and accountability

We conduct a second supplemental experiment to address two alternative explanations for our main results. First, it is possible that low-level evaluation criteria induce an implemental mindset, focusing individuals on how to implement a task strategy, whereas high-level evaluation criteria induce a deliberative mindset, focusing them on the viability of their estimate and investment recommendation. Prior research found that a deliberative, relative to an implemental, mindset can improve decision quality in complex tasks (Griffith et al., 2015). Second, individuals tend to be more stressed under an outcome accountability regime than under a process accountability regime because their evaluation is more dependent on uncontrollable factors (Siegel-Jacobs & Yates, 1996). If our participants perceive their evaluation to be more dependent on the outcome of their profit estimation (i.e., accuracy) under high-level evaluation criteria, they might process the information more carefully to overcome the uncertainty associated with their evaluation, thus becoming less vulnerable to the numerosity heuristic.

This supplemental experiment was conducted on Prolific with 141 participants. We provide participants with the same background information and decision setting as in experiment one. We manipulate between-participants the evaluation criteria. However, we do not manipulate the numerical presentation of the financial information, nor do we have participants perform the net income estimation and investment recommendation. After describing the decision scenario, we measure participants' mindset using an established ambivalence scale (Griffith et al., 2015; Henderson et al., 2008) with more (less) ambivalence being indicative of a more deliberative (implemental) mindset. Untabulated results indicate that evaluation criteria do not influence participants' mindsets: on a scale ranging from 1 (not at all ambivalent) to 6 (extremely ambivalent), participants' ambivalence was not statistically different across low-level and high-level evaluation criteria (2.67 vs. 2.59, $t_{139} = -0.571$, $p = 0.569$, two-tailed).

To examine whether high-level evaluation criteria lead to greater perceived importance of the task *outcome* in their evaluation, participants are asked, "To what extent do you agree with the following statement, 'My performance evaluation depends on the accuracy of my net income estimation'" on an 11-point scale ranging from "Strongly Disagree" to "Strongly Agree." There is no significant difference between low-level versus high-level evaluation criteria (9.00 vs. 8.70, $t_{139} = -0.785$, $p = 0.434$, two-tailed). Overall, we find no evidence that these alternative theoretical mechanisms drive our main results.

4. Experiment two

4.1. Hypothesis development and earnings fixation

The purpose of experiment two is to test the robustness of our theory with a different salient information cue from that of experiment one.

Specifically, when making financial predictions and decisions, managers and investors may excessively rely on salient summary earnings metrics (e.g., net income) without adequately incorporating other relevant information, referred to as *earnings fixation* (Elliott et al., 2011; Harris et al., 2016). Prior research further notes that individuals' information-processing approach can influence the extent of earnings fixation (Black et al., 2024; Cikurel, 2024; Harris et al., 2016). Thus, similar to the numerosity heuristic (i.e., overly focusing on the salient nominal values of performance information), in experiment two we test our theory by examining agents' susceptibility to earnings fixation (i.e., overly focusing on a salient piece of performance information) under different evaluation criteria.

To experimentally identify earnings fixation, following prior studies (Cikurel, 2024; Harris et al., 2016) we hold the subsidiary company's net income constant but vary its favorability. To do so, we manipulate the extent to which reported earnings were driven by transitory transactions (e.g., gain on disposal of assets) such that absent these transactions, net income would have been lower in one scenario versus the other. Earnings fixation is inferred when performance assessments do not differ across these two scenarios, whereas an absence of earnings fixation is inferred when performance assessments are lower in the scenario with a greater proportion of transitory earnings.

Our theory suggests that high-level evaluation criteria draw agents' attention to the link between the subordinate and superordinate goals, leading to a broad attentional scope over both salient and less salient but relevant information cues. Thus, we expect agents under high-level evaluation criteria to attend not only to net income but also to disaggregated performance information, allowing them to better discern performance across more favorable (persistent earnings) and less favorable (transitory earnings) scenarios. On the other hand, under low-level evaluation criteria, only the subordinate goal of taking appropriate actions to complete the task at hand is prominent, narrowing agents' attentional scope to salient information cues. Thus, agents may focus their attention more on net income (i.e., a highly salient information cue), relative to other line items in the income statement. With net income playing a larger role in information processing, we expect agents to make less differentiated performance assessments across more favorable (persistent earnings) and less favorable (transitory earnings) scenarios. In summary, we predict that, relative to high-level evaluation criteria, differences in performance assessment across more and less favorable scenarios will be attenuated under low-level evaluation criteria.[18]

As such, we formally state our second hypothesis:

H2. Agents under high-level performance evaluation criteria will assess greater performance prospects for more favorably performing firms (persistent earnings) compared to less favorably performing firms (transitory earnings), and this effect will be attenuated for agents under low-level performance evaluation criteria.

4.2. Experimental design and procedures

We adopt a similar capital budgeting context as in experiment one. Participants are provided with background information about ABC Corporation (parent) and the Pharmaceutical Company (subsidiary) and selected financial data for the Pharmaceutical Company. Participants are assigned the role of financial analyst and tasked with making a judgment about the Pharmaceutical Company's future performance prospects as well as providing a brief report justifying aspects of their judgment.

We use a 2 × 2 between-participants design, with two manipulated factors. The first factor, *CRITERIA,* is manipulated the same way as in experiment one: in the LOW-LEVEL condition participants are asked to justify "the process you used to determine your performance judgment." In the HIGH-LEVEL condition, they are asked to justify "why your performance judgment helps the ABC Corporation achieve its strategic objective (i.e., to maximize firm profit)." The second factor, *FAVORABILITY*, is embedded in the subsidiary's financial performance data. Participants are presented with the Pharmaceutical Company's income statement for the past three years with net income trending upwards over this period. We hold annual net income constant but vary two line items in the most recent year. In the MORE FAVORABLE condition, Selling, General, and Administrative (SGA) expenses in the most recent year increased proportionally with gross profit relative to the prior two years (26.6 percent of gross profit), with only a modest Gain on Sale of Assets (2.6 percent of gross profit). In the LESS FAVORABLE condition, SGA expenses increased at a greater rate in the most recent year (43.6 percent of gross profit) relative to the prior two years (26.6 percent of gross profit), with a sizable Gain on Sale of Assets (19.4 percent of gross profit). Because the subsidiary's net income is propped up by transitory income streams in the LESS FAVORABLE condition, we expect that, when both SGA expenses and Gain on Sale of Assets are adequately attended to, performance assessments should be greater in the MORE FAVORABLE condition than in the LESS FAVORABLE condition. Fig. 3 presents the income statements for each condition.

After reviewing the financial performance information, participants indicated how well they believed the Pharmaceutical Company would perform in the subsequent year on a 101-point scale, anchored as zero (very poorly) and 100 (very strong). This rating serves as our primary dependent measure, labeled *PERFORMANCE JUDGMENT*. Lastly, participants provided their justification report and responded to a post-experimental questionnaire which included providing demographic information.

4.3. Participants

We recruited participants through CloudResearch Connect (Hartman et al., 2023). To ensure sufficient financial literacy for the task, our sample consisted of 391 participants with professional and/or educational accounting experience (Libby et al., 2002).[19,20] Participants completed the task online through a link provided via CloudResearch and received $4 for completing the study, which took approximately 13 min.[21] On average, participants' age was 42 years, with 19 years of

[18] Similar to H1, we do not predict whether low-level evaluation criteria will lead, on average, high or lower performance prospect judgments. Instead, we predict that differences in performance judgements across favorability conditions will be attenuated with low-level criteria.

[19] Only CloudResearch workers with at least one of the following accounting experiences were able to enroll in the study: "I am a certified public accountant (CPA) or hold an equivalent accounting certification"; "I have a bachelor's degree or higher in accounting"; "I have completed some accounting coursework or have partial accounting education"; "I have worked in an accounting role for at least 1 year (without formal accounting education)"; "I am currently studying accounting at a college/university level."

[20] We recruited a total of 400 participants but excluded three observations for providing identical written justification reports, indicative of one individual taking the study multiple times and thereby observing multiple conditions simultaneously. We excluded an additional six observations because their written justification reports suggest that they were not aware of what the experimental task entailed. Specifically, three participants provided incoherent responses ("Nnnnnn …," "Quality of performance about the work," and "I would be able to provide the following information. Financial Health Pipeline and Product Portfolio Market Position Strategic Moves"), one misinterpreted the task as requiring a prediction of a percentage growth rate, and two provided generic statements unrelated to the task, such as claims that the judgment facilitates goal setting, monitoring, and feedback, and general description of the benefits of a well-designed evaluation system.

[21] The average hourly rate was over $15 per hour, above the recommended rate of $6 to $8 per hour for online participants (Denison, 2023; Moss, 2020).

Panel A: The MORE FAVORABLE condition

Income Statement (all figures in millions) Fiscal Year Ending December 31			
	2024	**2023**	**2022**
Revenue	$84.40	$73.39	$64.38
Cost of Sales	24.99	21.73	19.06
Gross Profit	59.41	51.66	45.32
Selling, General, and Administrative Expenses	15.89	13.70	12.02
Research and Development Expenses	4.13	4.00	3.83
Gain on Sale of Assets	1.53	0.00	0.00
Earnings Before Income Taxes	40.92	33.96	29.47
Income Tax Expense	2.83	2.11	1.32
Net Income	38.09	31.85	28.15

Panel A: The LESS FAVORABLE condition

Income Statement (all figures in millions) Fiscal Year Ending December 31			
	2024	**2023**	**2022**
Revenue	$84.40	$73.39	$64.38
Cost of Sales	24.99	21.73	19.06
Gross Profit	59.41	51.66	45.32
Selling, General, and Administrative Expenses	25.91	13.70	12.02
Research and Development Expenses	4.13	4.00	3.83
Gain on Sale of Assets	11.55	0.00	0.00
Earnings Before Income Taxes	40.92	33.96	29.47
Income Tax Expense	2.83	2.11	1.32
Net Income	38.09	31.85	28.15

Fig. 3. Experiment 2: Performance information.

full-time work experience. Forty-two percent of participants indicated that their current job requires them to make projections/estimations.

4.4. Tests of H2

We have two main independent variables. For *CRITERIA* we create a dummy variable which equals one (zero) for the HIGH-LEVEL (LOW-LEVEL) condition. For *FAVORABILITY* we create a dummy variable which equals one (zero) for the MORE FAVORABLE (LESS FAVORABLE) condition. Table 3 and Fig. 4 present the descriptive results for *PERFORMANCE JUDGMENT* across the four conditions.

To test H2, we conduct an OLS regression with *PERFORMANCE JUDGMENT* as the dependent measure and *CRITERIA*, *FAVORABILITY*, and their interaction as independent variables.[22] As discussed in experiment one, greater effort can increase the processing of all information, which may independently lead to greater differentiation in performance judgments across the MORE and LESS FAVORABLE conditions, potentially mitigating earnings fixation (Farrell et al., 2014; Fehrenbacher et al., 2018; Kennedy, 1993). Thus, similar to our test of H1, we include *EFFORT*, measured as the total minutes spent completing the study, and the interaction of *EFFORT* and *FAVORABILITY* as independent variables in the OLS regression (Bonner & Sprinkle, 2002).[23] This allows us to assess the relative reliance on salient cues (H2) after accounting for participants' overall effort levels.[24]

[22] We utilize robust standard errors as there is significant heteroskedasticity in *PEFORMANCE JUDGMENT* across conditions (Breusch-Pagan Test: $\chi^2 = 4.725$, $p = 0.030$).

[23] It is appropriate to include *EFFORT* in the model as it did not significantly differ between the LOW-LEVEL (16.43) and HIGH-LEVEL (15.38) conditions ($t_{389} = -1.012$, $p = 0.312$, two-tailed) or between the LESS FAVORABLE (15.96) and MORE FAVORABLE (15.83) conditions ($t_{389} = -0.130$, $p = 0.897$, two-tailed) (Piercey, 2023).

[24] Utilizing time spent completing an online experiment as a proxy for effort is subject to limitations. For example, without direct oversight, participants could temporarily attend to something outside of the experimental instrument creating noise in the measure. Consistent with this limitation, fourteen participants took longer than three standard deviations above the mean to complete the study. Since total time is likely not indicative of total effort expended by these participants, we winsorize their *EFFORT* to three standard deviations above the mean. Results are inferentially consistent without windsorizing and when dropping these observations.

Table 3
Experiment 2 descriptive statistics.

Panel A: *PERFORMANCE JUDGMENT* mean (SD) [*N*]		
FAVORABILITY	*CRITERIA*	
	LOW-LEVEL	HIGH-LEVEL
MORE FAVORABLE	74.28 (18.168) [97]	78.52 (14.531) [98]
LESS FAVORABLE	76.22 (16.225) [94]	75.45 (16.893) [102]

Notes.
CRITERIA = a dummy variable which equals one for the HIGH-LEVEL condition, in which participants' justification report asks them to justify whether their decision helps achieve the corporation's strategic objective, and zero for the LOW-LEVEL condition, in which participants' justification report asks them to justify their decision process.
FAVORABILITY = a dummy variable which equals one for the MORE FAVORABLE condition, in which The Pharmaceutical Company's net income was primarily driven by persistent earnings, and zero for the LESS FAVORABLE condition, in which a significant portion of The Pharmaceutical Company's net income was driven by transitory gain account (i.e., sale on investment).
PERFORMANCE JUDGMENT = participant responses to the question, "please indicate on the scale below how well you believe The Pharmaceutical Company will perform in the next year" on a 101-point scale from 0 (very poorly) to 100 (very strong).

As reported in Table 4, *EFFORT* does not significantly influence the effect of *FAVORABILITY* ($p = 0.817$, two-tailed). Furthermore, we find a marginally significant interaction between *CRITERIA* and *FAVORABILITY* ($p = 0.063$, one-tailed). Within the HIGH-LEVEL condition, participants judged the subsidiary to have stronger performance prospects in the MORE FAVORABLE compared to the LESS FAVORABLE condition ($p = 0.081$, one-tailed). Conversely, in the LOW-LEVEL condition, *PERFORMANCE JUDGMENT* does not significantly differ across *FAVORABILITY* conditions ($p = 0.426$, two-tailed). Thus, consistent with H2, participants appropriately differentiate performance across *FAVORABILITY* conditions in the HIGH-LEVEL condition, and this difference is attenuated in the LOW-LEVEL condition.

Integrating these findings with those from experiment one, we provide triangulating support for our theory with a different salient information cue and a different pattern of results. Notably, results from experiment one support our theory via a difference in judgments in the LOW-LEVEL condition that disappears in the HIGH-LEVEL condition; by contrast, the results of experiment two support our theory with no difference in judgments in the LOW-LEVEL condition, and a difference emerging in the HIGH-LEVEL condition. The consistency of inferences across these opposite patterns of results lends further credence to our theory.

4.5. Additional analyses – controllability and evaluation breadth

In our setting, agents under low-level criteria are evaluated based on their execution of task-specific actions, whereas those under high-level criteria are evaluated on how their actions contribute to organizational objectives. An alternative explanation for our main results is that agents under high-level criteria may perceive their evaluation outcome to be less within their personal control. Zureich (2023) finds that, when individuals are held accountable for factors outside their span of control (e.g., others' performance), they are more willing to seek out relevant information associated with those factors.

In the post-experimental questionnaire, participants indicated where their evaluation criterion falls along the spectrum of "general and broad (i.e., encompassing many aspects of an individual's actions and impacts)" to "specific and precise (i.e., focusing only on narrowly defined aspects)," on a 101-point scale anchored at 0 (very broad and general) and 100 (very specific and precise). Untabulated results show that responses do not significantly differ across HIGH-LEVEL (61.03) versus LOW-LEVEL (60.07) conditions ($t_{389} = -0.452$, $p = 0.651$, two-tailed). We also elicit participants' perceived controllability of their performance evaluation by asking them to indicate, "to what extent do you feel you have control over the outcome of your performance evaluation" on a 101-point scale anchored at 0 (no control) to 100 (full control). Again, untabulated results show that responses do not significantly differ between HIGH-LEVEL (65.33) and LOW-LEVEL (68.10) conditions ($t_{389} = -1.259$, $p = 0.209$, two-tailed). Finally, we replicate our test of H2 while controlling for evaluation breadth and perceived controllability and the interaction of *CRITERIA* and *FAVORABILITY* remain inferentially consistent. Thus, we find no evidence that participants perceive high-level evaluation as broader or less controllable.

5. Discussion and conclusion

We extend the performance evaluation literature by investigating how evaluation criteria impact agents' information processing in ill-structured complex decision-making tasks. Drawing on goal hierarchy

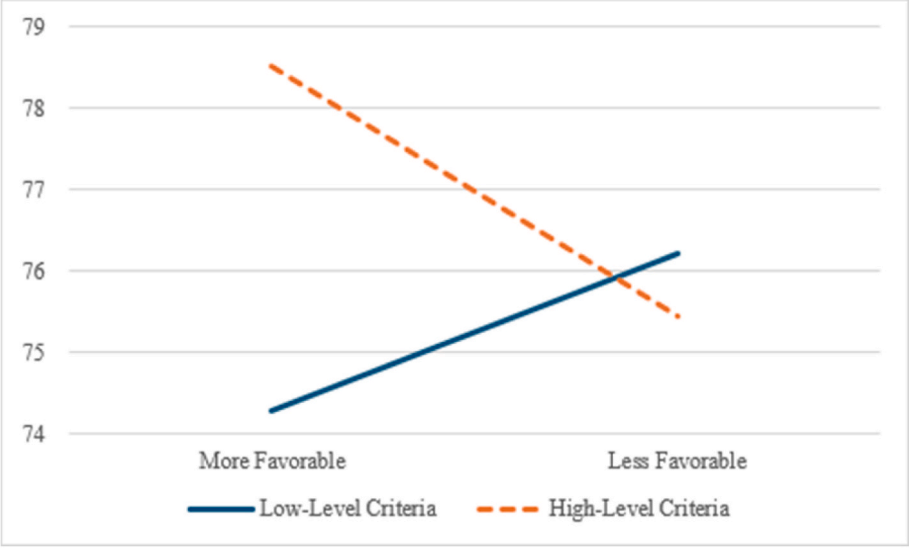


Fig. 4. Experiment 2 – mean performance judgment.

Table 4
Experiment 2 test of H2.

Panel A: OLS REGRESSION (Dependent variable = *PERFORMANCE JUDGMENT)*				
Source	Coefficient	Robust SE	*t*-stat	*p*-value
FAVORABILITY	−1.97	2.47	−0.80	0.426
CRITERIA	−0.80	2.33	−0.34	0.731
FAVORABILITY × *CRITERIA*	5.05	3.30	1.53	0.063[a]
EFFORT	−0.03	0.12	−0.22	0.827
FAVORABILITY × *EFFORT*	0.04	0.15	0.23	0.817
Constant	76.24	1.65	46.32	<0.001
Simple effects tests				
Effect of *FAVORABILITY* given LOW-LEVEL	−1.97	2.47	−0.80	0.426
Effect of *FAVORABILITY* given HIGH-LEVEL	3.09	2.20	1.40	0.081[a]

Notes.
CRITERIA = a dummy variable which equals one for the HIGH-LEVEL condition, in which participants' justification report asks them to justify whether their decision helps achieve the corporation's strategic objective, and zero for the LOW-LEVEL condition, in which participants' justification report asks them to justify their decision process.
FAVORABILITY = a dummy variable which equals one for the MORE FAVORABLE condition, in which The Pharmaceutical Company's net income was primarily driven by persistent earnings, and zero for the LESS FAVORABLE condition, in which a significant portion of The Pharmaceutical Company's net income was driven by transitory gain account (i.e., sale on investment).
PERFORMANCE JUDGMENT = participant responses to the question, "please indicate on the scale below how well you believe The Pharmaceutical Company will perform in the next year" on a 101-point scale from 0 (very poorly) to 100 (very strong).
EFFORT = participant total time, in minutes, spent completing the study, mean centered, and winsorized.
[a] An asterisk indicates a one-tailed p-value for a test of directional prediction.

theory, we contend that low-level evaluation criteria lead agents to fixate on task-specific actions and, in turn, be more sensitive to salient attributes in their information processing and decision making. We also propose that emphasizing the link between their task decisions and efforts with the higher-order impact of the task (superordinate goals) within the evaluation criteria can broaden agents' attentional scope, making them more likely to attend to both salient and less salient yet relevant task features.

Our experimental results support these predictions. First, we show that agents under low-level performance evaluation criteria engage in the numerosity heuristic, a decision-making heuristic where individuals' judgments are influenced by the numerical presentation of financial information, whereas high-level evaluation criteria mitigate this effect. Second, we demonstrate that agents under low-level evaluation criteria are susceptible to earnings fixation, a decision-making heuristic where individuals overly focus on summary earnings measures while not attending adequately to other relevant information in making financial predictions and, again, high-level evaluation criteria mitigate this effect. Moreover, we address several possible alternative explanations for our main findings through additional analyses and supplemental data. Overall, the findings provide converging evidence for our theory, suggesting that the nature of performance criteria used to evaluate agents influences how they process information cues, ultimately impacting their decision making.

We extend the accounting literature on the decision-facilitating role of performance metrics and criteria. Prior studies find that using performance metrics that capture aspects beyond agents' immediate task can influence their judgment and decision making (Farrell et al., 2008; Zureich, 2023). We extend this stream of research by examining whether the link between goal hierarchical levels is emphasized in evaluation criteria affects agents' information processing in complex decision tasks. Additionally, prior accounting studies find that attending to salient task cues can sometimes be beneficial: for instance, drawing employees' attention to the performance-contingent nature of compensation schemes enhances their subsequent performance (Farrell et al., 2008; Kachelmeier et al., 2019). Other research suggests that increasing the salience of the "right" information cue can facilitate task strategy selection (e.g., Bryant, 2014). Our study helps depict a more complete picture about the salience effect by showing that over-attention to salient cues can lead to unintended consequences. Our study also provides useful insights for firms to improve the effectiveness of management control systems. To the extent that performance evaluations can influence agents' cognitive approach and thus have unintended effects on their information processing, firms need to be aware of such effects in the design of performance evaluation systems (Booth & Shultz, 2004).

Our study is subject to several limitations, which provide ample avenues for future research. We test the effect of low-level versus high-level performance evaluation criteria in one decision-making scenario – capital budgeting. Future research could test the robustness of these effects on other decision-making settings. While our study shows significant findings in an ill-structured complex decision-making task, it is possible that variations in the task procedures could limit the implications of the study. For example, when making profit estimation and investing decisions across multiple divisions or subsidiaries for the purpose of distributing investment funds from a fixed pool, judgments associated with the absolute magnitude of the evaluation (e.g., the numerosity heuristic) might be mitigated by the relative comparisons across units. In both experiments, individuals make decisions only once, yet in naturally occurring settings, individuals might perform such procedures on a recurring basis. Repeated decision making could lead to different behavior, due to variations in unit performance, changes in incentives, or learning through experience. In addition, we study agents' relative attention to salient information cues indirectly by examining downstream judgments that are influenced by the relative use of salient and less salient information cues. Future research can investigate this relationship more directly, for instance, by examining agents' attention across information cues with eye-tracking software.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aos.2026.101632.

Data availability

Please contact the authors.

References

Aiello, M., Merola, S., Lasaponara, S., Pinto, M., Tomaiuolo, F., & Doricchi, F. (2018). The influence of visual and phonological features on the hemispheric processing of hierarchical Navon letters. *Neuropsychologia, 109*, 75–85.
Albers, M. J. (2007). Information salience and interpreting information. In *Proceedings of the 25th annual ACM international conference on design of communication* (pp. 80–86). October.
Alcocer, C. D., & Torres, E. R. T. (2024). Salience bias: A framework about the importance of prices and budget constraints perceptions. *Journal of Behavioral and Experimental Economics, 110*, Article 102212.
Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics, 109*(4), 1125–1156.
Banker, R. D., Potter, G., & Srinivasan, D. (2000). An empirical investigation of an incentive plan that includes nonfinancial performance measures. *The Accounting Review, 75*(1), 65–92.
Bentley, J. W. (2019). Decreasing operational distortion and surrogation through narrative reporting. *The Accounting Review, 94*(3), 27–55.

Berger, L. (2019). The importance of task complexity when rewarding nonfinancial performance. *Journal of Management Accounting Research, 31*(2), 1–17.

Bipp, T., & Kleingeld, A. (2011). Goal-setting in practice: The effects of personality and perceptions of the goal-setting process on job satisfaction and goal commitment. *Personnel Review, 40*(3), 306–323.

Black, P. W., Jackson, K. E., Rowe, S. P., & Zimbelman, A. F. (2024). The effect of personalized communication on investor earnings fixation. *Journal of Financial Reporting, 9*(1), 1–19.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: Theories, evidence, and a framework for research. *Accounting, Organizations and Society, 4–5*, 303–345.

Booth, P., & Shultz, A. K.-D. (2004). The impact of ethical environment on managers' project evaluation judgments under agency problem conditions. *Accounting, Organizations and Society, 29*, 473–488.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2022). Salience. *Annual Review of Economics, 14*(1), 521–544.

Bradley, W. E. (2009). Ability and performance on ill-structured problems: The substitution effect of inductive reasoning ability. *Behavioral Research in Accounting, 21*(1), 19–35.

Brislin, R. W. (1986). The wording and translation of research instruments. In W. J. Lonner, & J. W. Berry (Eds.), *Field methods in cross-cultural research* (pp. 137–164). New York, NY: Sage Publications.

Brown, L. D., Call, A. C., Clement, M. B., & Sharp, N. Y. (2015). Inside the "black box" of sell-side financial analysts. *Journal of Accounting Research, 53*(1), 1–47.

Bryant, D. J. (2014). Strategy selection in cue-based decision making. *Canadian Journal of Experimental Psychology, 68*(2), 97–110.

Burgoon, E. M., Henderson, M. D., & Markman, A. B. (2013). There are many ways to see the forest for the trees: A tour guide for abstraction. *Perspectives on Psychological Science, 8*(5), 501–520.

Cardinaels, E. (2008). The interplay between cost accounting knowledge and presentation formats in cost-based decision-making. *Accounting, Organizations and Society, 33*(6), 582–602.

Cardinaels, E., Kramer, S., & Maas, V. S. (2024). Navigating through the noise: The effect of color-coded performance feedback on decision-making. *Contemporary Accounting Research, 41*(2), 1031–1057.

Cave, K. R., & Chen, Z. (2016). Identifying visual targets amongst interfering distractors: Sorting out the roles of perceptual load, dilution, and attentional zoom. *Attention, Perception, & Psychophysics, 78*(7), 1822–1838.

Chen, S. H., & Lee, H. T. (2007). Performance evaluation model for project managers using managerial practices. *International Journal of Project Management, 25*(6), 543–551.

Choi, J.(W.), Hecht, G. W., & Tayler, W. B. (2012). Lost in translation: The effects of incentive compensation on strategy surrogation. *The Accounting Review, 87*(4), 1135–1163.

Church, B. K., Jiang, W., Kuang, X., & Vitalis, A. (2019). A dollar for a tree or a tree for a dollar? The behavioral effects of measurement basis on managers' CSR investment decision. *The Accounting Review, 94*(5), 117–137.

Church, B. K., Kuang, X., & Liu, Y. S. (2019). The effects of measurement basis and slack benefits on honesty in budget reporting. *Accounting, Organizations and Society, 72*, 74–84.

Cikurel, D. (2024). How does perceived ease of information access affect investors' judgments? *Contemporary Accounting Research, 41*(4), 2325–2353.

Croft, R. J., Lee, A., Bertolot, J., & Gruzelier, J. H. (2001). Associations of P50 suppression and desensitization with perceptual and cognitive features of "unreality" in schizotypy. *Biological Psychiatry, 50*(6), 441–446.

Darke, P. R., & Freedman, J. L. (1993). Deciding whether to seek a bargain: Effects of both amount and percentage off. *Journal of Applied Psychology, 78*(6), 960–965.

Datar, S., Kulp, S. C., & Lambert, R. A. (2001). Balancing performance measures. *Journal of Accounting Research, 39*(1), 75–92.

David, E. M. (2013). Examining the role of narrative performance appraisal comments on performance. *Human Performance, 26*(5), 430–450.

Denison, G. (2023). How much should you pay research participants? Prolific. https://www.prolific.com/blog/how-much-should-you-pay-research-participants.

Earley, C. E. (2002). The differential use of information by experienced and novice auditors in the performance of ill-structured audit tasks. *Contemporary Accounting Research, 19*(4), 595–614.

Elliott, W. B. (2006). Are investors influenced by pro forma emphasis and reconciliations in earnings announcements? *The Accounting Review, 81*(1), 113–133.

Elliott, W. B., Hobson, J. L., & Jackson, K. E. (2011). Disaggregating management forecasts to reduce investors' susceptibility to earnings fixation. *The Accounting Review, 86*(1), 185–208.

Farrell, A. M., Goh, J. O., & White, B. J. (2014). The effect of performance-based incentive contracts on system 1 and system 2 processing in affective decision contexts: fMRI and behavioral evidence. *The Accounting Review, 89*(6), 1979–2010.

Farrell, A. M., Kadous, K., & Towry, K. L. (2008). Contracting on contemporaneous versus forward-looking measures: An experimental investigation. *Contemporary Accounting Research, 25*(3), 773–802.

Farrell, A. M., Kadous, K., & Towry, K. L. (2012). Does the communication of causal linkages improve employee effort allocations and firm performance? An experimental investigation. *Journal of Management Accounting Research, 24*(1), 77–102.

Fehrenbacher, D. D., Schulz, A. K. D., & Rotaru, K. (2018). The moderating role of decision mode in subjective performance evaluation. *Management Accounting Research, 41*, 1–10.

Folger, R., Konovsky, M. A., & Cropanzano, R. (1992). A due process metaphor for performance appraisal. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 14, pp. 129–177). Greenwich, CT: JAI.

Griffith, E. E., Hammersley, J. S., Kadous, K., & Young, D. (2015). Auditor mindsets and audits of complex estimates. *Journal of Accounting Research, 53*(1), 49–77.

Grubb, T. (2007). Performance appraisal reappraised: It's not all positive. *Journal of Human Resources Education, 1*(1), 1–22.

Guiral, A., Moon, D., Tan, H. T., & Yu, Y. (2020). What drives investor response to CSR performance reports? *Contemporary Accounting Research, 37*(1), 101–130.

Hammersley, J. S., & Ricci, M. A. (2021). Using audit programs to improve auditor evidence collection. *The Accounting Review, 96*(1), 251–272.

Harackiewicz, J. M., & Elliot, A. J. (1998). The joint effects of target and purpose goals on intrinsic motivation: A mediational analysis. *Personality and Social Psychology Bulletin, 24*(7), 675–689.

Harris, L. L., Hobson, J. L., & Jackson, K. E. (2016). The effect of investor status on investors' susceptibility to earnings fixation. *Contemporary Accounting Research, 33*(1), 152–171.

Hartman, R., Moss, A. J., Jaffe, S. N., Rosenzweig, C., Litman, L., & Robinson, J. (2023). Introducing connect by CloudResearch: Advancing online participant recruitment in the digital age. https://doi.org/10.31234/osf.io/ksgyr. Retrieved October 2025.

Hayes, A. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York, NY: The Guildford Press.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38*, 79–109.

Heinze, H.-J., & Münte, T. F. (1993). Electrophysiological correlates of hierarchical stimulus processing: Dissociation between onset and later stages of global and local target processing. *Neuropsychologia, 31*(8), 841–852.

Henderson, M. D., de Liver, Y., & Gollwitzer, P. M. (2008). The effects of an implemental mind-set on attitude strength. *Journal of Personality and Social Psychology, 94*(3), 396–411.

Höchli, B., Brügger, A., & Messner, C. (2018). How focusing on superordinate goals motivates broad, long-term goal pursuit: A theoretical perspective. *Frontiers in Psychology, 9*, 1–14.

Houser-Marko, L., & Sheldon, K. M. (2008). Eyes on the prize or nose to the grindstone? The effects of level of goal evaluation on mood and motivation. *Personality and Social Psychology Bulletin, 34*(11), 1556–1569.

Humphreys, K. A., Gary, M. S., & Trotman, K. T. (2016). Dynamic decision making using the balanced scorecard framework. *The Accounting Review, 91*(5), 1441–1465.

Ishak, Z., Fong, S. L., & Shin, S. C. (2019). SMART KPI management system framework. In *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)* (pp. 172–177).

Itthipuripat, S., Garcia, J. O., Rungratsameetaweemana, N., Sprague, T. C., & Serences, J. T. (2014). Changing the spatial scope of attention alters patterns of neural gain in human cortex. *Journal of Neuroscience, 34*(1), 112–123.

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review, 78*(3), 725–758.

Juola, J. F., Duvuru, P., & Peterson, M. S. (2000). Priming effects in attentional gating. *Memory & Cognition, 28*(2), 224–235.

Kachelmeier, S. J., Wang, L. W., & Williamson, M. G. (2019). Incentivizing the creative process: From initial quantity to eventual creativity. *The Accounting Review, 94*(2), 249–266.

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research, 31*(2), 231–245.

Kirschenbaum, D. S., Humphrey, L. L., & Malett, S. D. (1981). Specificity of planning in adult self-control: An applied investigation. *Journal of Personality and Social Psychology, 40*(5), 941–950.

Lambert, R. A. (2001). Contracting theory and accounting. *Journal of Accounting and Economics, 32*(1–3), 3–87.

Lambert, R. A., & Larcker, D. F. (1987). An analysis of the use of accounting and market measures of performance in executive compensation contracts. *Journal of Accounting Research, 25*, 85–125.

Lau, C. M., Low, L. C., & Eggleton, I. R. C. (1995). The impact of reliance on accounting performance measures on job-related tension and managerial performance: Additional evidence. *Accounting, Organizations and Society, 20*(5), 359–381.

Libby, R., Bloomfield, R., & Nelson, M. W. (2002). Experimental research in financial accounting. *Accounting, Organizations and Society, 27*(8), 775–810.

Lin, L. (2009). Breadth-biased versus focused cognitive control in media multitasking behaviors. *The Proceedings of the National Academy of Sciences, 106*(37), 15521–15522.

Lipe, M. G., & Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review, 75*(3), 283–298.

Liu, L., Zhang, G., Zhou, R., & Wang, Z. (2014). Motivational intensity modulates attentional scope: Evidence from behavioral and ERP studies. *Experimental Brain Research, 232*, 3291–3300.

Locke, E. A., Chah, D.-O., Harrison, S., & Lustgarten, N. (1989). Separating the effects of goal specificity from goal level. *Organizational Behavior and Human Decision Processes, 43*, 270–287.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting & task performance*. Prentice-Hall, Inc.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705–717.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13*(2), 194–196.

Masicampo, E. J., & Baumeister, R. F. (2012). Committed but closed-minded: When making a specific plan for a goal hinders success. *Social Cognition, 30*(1), 37–55.

Matsumura, E. M., & Shin, J. Y. (2006). An empirical analysis of an incentive plan with relative performance measures: Evidence from a postal service. *The Accounting Review, 81*(3), 533–566.
Merchant, K. A., & Van der Stede, W. A. (2017). *Management control systems: Performance measurement, evaluation and incentives* (4th ed.). Harlow, UK: Pearson Education Limited.
Milosevic, D., & Patanakul, P. (2005). Standardized project management may increase development projects success. *International Journal of Project Management, 23*(3), 181–192.
Moss, A. (2020). https://www.cloudresearch.com/resources/blog/five-things-you-should-be-doing-in-online-data-collection/?_hstc=206527749.ee17fb44469 8f768f0c2ab8618bd057b.1602862718922.1607629663657.16079 65484304.38&_hssc=206527749.1.1607965484304&_hsfp=2686683400.
Oracle Corporation. (2012). *Goal setting: A fresh perspective*. White Paper. https://www.oracle.com/us/%20media1/goal-setting-fresh-perspective-ee-1679275.pdf.
Pandelaere, M., Briers, B., & Lembregts, C. (2011). How to make 29% increase look bigger: The unit effect in option comparisons. *Journal of Consumer Research, 38*, 308–322.
Piercey, M. D. (2023). "Throw it in as a covariate?" common problems using measured control variables in experimental research. *Auditing: A Journal of Practice & Theory, 42*(2), 183–205.
Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics, 31*(3), 437–448.
Pringle, H. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin & Review, 8*(1), 89–95.
Raghubir, P., & Srivastava, J. (2002). Effect of face value on product valuation in foreign currencies. *Journal of Consumer Research, 29*, 335–347.
Rowe, G., Hirsh, J. B., & Anderson, A. K. (2007). Positive affect increases the breadth of attentional selection. *The Proceedings of the National Academy of Sciences, 104*(1), 383–388.
Shafir, E., Diamond, P., & Tversky, A. (1997). Money illusion. *Quarterly Journal of Economics, 112*(2), 341–374.
Sharma, D. S. (2006). Effects of professional and non-professional investors' perceptions of board effectiveness on their judgments: An experimental study. *Journal of Accounting and Public Policy, 25*(1), 91–115.
Shields, M. D., & Waller, W. S. (1988). A behavioral study of accounting variables in performance—incentive contracts. *Accounting, Organizations and Society, 13*, 581–594.
Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes, 65*(1), 1–17.
Staw, B. M., & Boettger, R. D. (1990). Task revision: A neglected form of work performance. *Academy of Management Journal, 33*(3), 534–559.
Sun, S. H., & Frese, M. (2013). In E. A. Locke, & G. P. Latham (Eds.), *Multiple goal pursuit in new developments in goal setting and task performance* (pp. 177–194). New York, NY: Routledge.
Tang, M., & Venkataraman, S. (2018). How patterns of past guidance provision affect investor judgments: The joint effect of guidance frequency and guidance pattern consistency. *The Accounting Review, 93*(3), 327–348.
Tenbrunsel, A. E., Wade-Benzoni, K. A., Messick, D. M., & Bazerman, M. H. (2000). Understanding the influence of environmental standards on judgments and choices. *Academy of Management Journal, 43*(5), 854–866.
Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review, 94*, 3–15.
Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology, 57*(4), 660–671.
Vallacher, R. R., & Wegner, D. M. (2014). *A theory of action identification*. New York, NY: Psychology Press.
van Veen-Dirks, P. (2010). Different uses of performance measures: The evaluation versus reward of production managers. *Accounting, Organizations and Society, 35*(2), 141–164.
Wang, S., Karabay, A., & Akyürek, E. G. (2024). Attentional blur and blink: Effects of adaptive attentional scaling on visual awareness. *Consciousness and Cognition, 117*, Article 103627.
Weintraub, J., Cassell, D., & DePatie, T. P. (2021). Nudging flow through 'SMART' goal setting to decrease stress, increase engagement, and increase performance at work. *Journal of Occupational and Organizational Psychology, 94*(2), 230–258.
Wertenbroch, K., Soman, D., & Chattopadhyay, A. (2007). On the perceived value of money: The reference dependence of currency numerosity effects. *Journal of Consumer Research, 34*(1), 1–10.
Wohlin, C., & Andrews, A. A. (2001). Assessing project success using subjective evaluation factors. *Software Quality Journal, 9*(1), 43–70.
Wright, P. M., & Kacmar, K. M. (1994). Goal specificity as a determinant of goal commitment and goal change. *Organizational Behavior and Human Decision Processes, 59*, 242–260.
Zureich, J. T. (2023). The (un)controllability principle: The benefits of holding employees accountable for uncontrollable factors. *Journal of Accounting Research, 61*(2), 653–690.

Update

Accounting, Organizations and Society
Volume 116, Issue , June 2026, Page

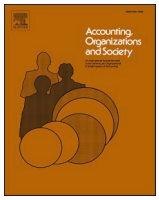
DOI: https://doi.org/10.1016/j.aos.2026.101635

Accounting, Organizations and Society 116 (2026) 101635

Contents lists available at ScienceDirect

Accounting, Organizations and Society

journal homepage: www.elsevier.com/locate/aos



Corrigendum to "Performance evaluation criteria and information processing in complex decision making" [Accounting, Organizations and Society, 116 (2026), 101632]

Xiqiong He [a], Xi (Jason) Kuang [b], Ziyang Li [c,*], Jordan Samet [d]

[a] Southwestern University of Finance and Economics, China
[b] Georgia Institute of Technology, USA
[c] Sichuan University, China
[d] Indiana University, USA

The corresponding author was incorrectly labelled in the original article. The correct corresponding author is: Ziyang Li (lzy_feng@scu.edu.cn).

The authors would like to apologise for any inconvenience caused.

DOI of original article: https://doi.org/10.1016/j.aos.2026.101632.
* Corresponding author.
E-mail address: lzy_feng@scu.edu.cn (Z. Li).

https://doi.org/10.1016/j.aos.2026.101635

Available online 12 February 2026