UvA-DARE (Digital Academic Repository)

Selective Attention as a Determinant of Escalation Bias in Subjective Performance Evaluation Judgments

Kramer, S.; Maas, V.S.

DOI

[10.2308/bria-18-021](https://doi.org/10.2308/bria-18-021)

Publication date

2020

Document Version

Author accepted manuscript

Published in

Behavioral Research in Accounting

[Link to publication](#)

Citation for published version (APA):

Kramer, S., & Maas, V. S. (2020). Selective Attention as a Determinant of Escalation Bias in Subjective Performance Evaluation Judgments. *Behavioral Research in Accounting*, 32(1), 87-100. <https://doi.org/10.2308/bria-18-021>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Selective Attention as a Determinant of Escalation Bias in Subjective Performance

Evaluation Judgments

Stephan Kramer

Rotterdam School of Management

Erasmus University Rotterdam

Burgemeester Oudlaan 50

3062PA Rotterdam

skramer@rsm.nl

Victor S. Maas

Amsterdam Business School

University of Amsterdam

Plantage Muidersgracht 12

1018TV Amsterdam

vmaas@uva.nl

Published as:

Kramer, S. and V. S. Maas (2020) Selective Attention as a Determinant of Escalation Bias in Subjective Performance Evaluation Judgments. *Behavioral Research in Accounting* 32(1), 87-100.

We thank two anonymous reviewers and Steve Salterio (editor) for guidance and suggestions during the review process and Markus Arnold, Steve Kachelmeier, Weiming Liu, Marcel van Rinsum, Mike Shields, and Michael Williamson for comments on previous versions of this manuscript. The manuscript has also benefited from comments and discussion at the 2015 AAA MAS midyear meeting, the 2015 ACMAR conference, a research seminar at the University of Strathclyde, and brownbag seminars at Michigan State University, the University of Texas at Austin and the University of Amsterdam. We thank Christiaan Tieman and Marcel Boom for research assistance with the eye-tracker.

Selective Attention as a Determinant of Escalation Bias in Subjective Performance

Evaluation Judgments

Abstract

We use an experiment to examine escalation bias in subjective performance evaluations. Participants assume the role of manager and evaluate the performance of an employee based on a balanced scorecard-type performance report. We manipulate whether managers recommended positively or negatively about the evaluated employee's promotion to his current position. Consistent with the presence of escalation bias, we find that managers give higher performance ratings to employees about whom they advised positively than to employees about whom they advised negatively. We then examine the role of selective attention in the emergence of this escalation bias. Using eye-tracking data, we investigate whether escalation bias arises because managers with different prior commitments towards the evaluated employee pay attention to different items in the scorecard. We find no evidence that this is the case. Specifically, the eye-tracking data indicate that evaluators' prior recommendation did not affect what proportion of their visual attention was given to favorable (versus unfavorable) performance measures, and that the relative attention paid to favorable measures is not associated with the performance rating.

Keywords: Attention, Balanced scorecard, Escalation of commitment, Eye-tracking, Subjective performance evaluation.

INTRODUCTION

Management accounting information is frequently used as input to subjective performance evaluation processes (Luft and Shields 2009). For example, managers often have discretion over the specific weights they attach to performance measures when determining an overall performance assessment or bonus payment (e.g., Prendergast 1999; Lipe and Salterio 2000; Bol, Kramer, and Maas 2016). While, in principle, such discretion allows managers to provide a more accurate rating of their subordinates' performance (Bol 2008), it also opens the door to performance evaluation bias (Luft and Shields 2009; Hoque 2014). One specific bias that has often been documented is escalation of commitment, which is the tendency of managers to give higher performance ratings to employees whom they hired or promoted themselves - or whom they recommended for hiring or promotion in a previous period - compared to equally well performing employees who were selected by others (Bazerman, Beekun, and Schoorman 1982; Schoorman 1988; Slaughter and Greguras 2008; Müller and Weinschenk 2015; Angelovski, Brandts, and Sola 2016). While there is general consensus that the main psychological determinant of escalation behavior is self-justification (Staw 1997; Sleesman, Conlon, McNamara, and Miles 2012), relatively little is known about the specific process through which decision makers arrive at evaluation judgments that are consistent with their prior decisions. In this study, we examine whether escalation bias in subjective performance evaluation should at least partly be attributed to biased attention during the information acquisition phase, i.e. the phase preceding the actual evaluation decision during which evaluators visually scan the information on which they need to base their decision.

We performed an experiment in which we asked participants to assume the role of manager and evaluate the performance of a subordinate in a specific year, based on a performance report in

the form of a balanced scorecard containing target values and actual values for eight performance measures. The report was deliberately constructed to provide an ambiguous picture of the subordinate's overall performance, and contained both measures on which the subordinate scored above target and measures on which he scored below target. We manipulated the context such that some participants read that they had previously recommended the employee for promotion, whereas others read that they had previously advised against the promotion. We tracked the participants' eye-movements while they studied the information in the scorecard and provided their performance rating, which allows us to observe which specific items they looked at - and for how long.

Consistent with our reasoning and with prior literature, we find that evaluators who advised in favor of the promotion of the subordinate subsequently rate the subordinate's performance significantly higher than evaluators who advised against the promotion of the subordinate. Using eye-tracking data, we then test whether managers who advised in favor of the subordinate's promotion pay more attention to measures that reflect positively on the subordinate's performance, and less attention to measures that reflect negatively on the subordinate's performance, compared to managers who advised against the subordinate's promotion. In addition, we examine whether the amount of attention a manager pays to favorable versus unfavorable measures predicts that manager's overall evaluation judgment. We find no evidence that biased attention mediates escalation of commitment in subjective performance evaluation. Specifically, our data indicate that evaluators' prior recommendation does not affect what proportion of their visual attention is given to favorable (versus unfavorable) performance measures. Moreover, the relative attention paid to favorable and unfavorable measures is not associated with the final performance rating. This result is in line with recent research in psychology that suggests that the role of attention bias

as a mediator of escalation of commitment may be limited (Schultze, Pfeiffer, and Schulz-Hardt 2012).

Our paper makes several contributions to the literature. First, we complement the existing literature on subjective performance evaluations based on comprehensive performance reports such as the balanced scorecard (Lipe and Salterio 2000; Dilla and Steinbart 2005; Liedtka, Church, and Ray 2008) by investigating whether escalation bias exists in such settings, and find that it does. This result is important because it indicates that providing evaluators with objective performance reports does not fully mitigate escalation. In addition, it confirms that allowing managers discretion in attaching weights to performance signals can compromise the accuracy of evaluation outcomes (Ittner, Larcker, and Meyer 2003; Bol et al. 2016).

Next, by making use of eye-tracking technology, we directly measure attention in judgment and decision making processes based on accounting information. While it has been suggested that attention is a crucial mediator between accounting information and judgments and decisions (e.g., Birnberg and Shields 1984), our current understanding of the role of attention in accounting-based decision making is still limited. Investigating whether escalation bias is mediated by attention bias is important for management accounting research and practice, because if prior decisions affect evaluation judgments by steering evaluators' attention in certain directions, organizations might be able to influence evaluation outcomes through policies and procedures that shape managers' cognitive information acquisition processes. For example, organizations could change the layout of performance reports (Cardinaels and van Veen-Dirks 2010), develop procedures or 'nudges' that make it less likely that information items are ignored or overlooked (Emby and Finley 1997; Thaler and Sunstein 2008), or make managers accountable for processes rather than outcomes (Libby, Salterio, and Webb 2004; Dalla Via, Perego, and van Rinsum 2019).

At a more fundamental level, our study also contributes to the debate in cognitive psychology on the roles of information acquisition and information processing as mediators of escalation of commitment and similar judgment biases such as confirmation bias and assimilation effects (e.g., Conlon and McLean Parks 1987; Schultze et al. 2012; Schultze and Schulz-Hardt 2015; Chaxel, Russo, and Kerimi 2013). Specifically, we are the first to use eye-tracking to determine whether prior decisions affect current decisions by influencing the information to which decision makers attend. As we find no evidence that this is the case, our results confirm evidence from Schultze et al. (2012) that suggests that escalation processes are not due to decision makers selectively exposing themselves to information that justifies their prior decision.

HYPOTHESIS DEVELOPMENT

Escalation of commitment in general refers to the tendency of decision-makers who have committed themselves to a particular course of action to bias subsequent decisions in order to justify their previous commitment (Staw 1976, 1981, Brockner 1992). Escalation bias has been documented in a wide variety of settings (Staw 1997; Sleesman et al. 2012), including settings within the domain of accounting such as investment decision making (Cheng, Schulz, Luckett, and Booth 2003; Kadous and Sedor 2004; Denison 2009) and internal auditing (Brody and Kaplan 1996). Bazerman et al. (1982) were the first to examine escalation bias in a subjective performance evaluation context. Their experiment provided evidence that managers who evaluate the performance of an employee rate this performance higher if they had previously hired that employee themselves versus when the employee was hired by someone else. In a follow-up study, using field data, Schoorman (1988) confirmed that managers who were asked for their input in a hiring or promotion decision process and agreed with the decision to hire or promote an employee

subsequently rated that employee's performance higher than managers who did not participate in the decision process. In addition, Schoorman (1988) found that managers who participated in the decision process but *disagreed* with the outcome, subsequently rated an employee's performance lower than managers who did not participate in the decision. Slaughter and Greguras (2008) replicated the first finding of Schoorman (1988) on positive escalation bias but not the second on negative escalation bias. Finally, in a recent lab experiment, Angelovski et al. (2016) found evidence for both positive and negative escalation bias.

We investigate a setting in which a senior manager needs to evaluate the performance of a junior manager based on the information contained in a balanced scorecard. The junior manager was promoted to his current position after consultation of the senior manager, who either advised in favor or against the promotion. In line with existing research in organizational psychology and economics (Bazerman et al. 1982; Schoorman 1988; Slaughter and Greguras 2008; Angelovski et al. 2016), we hypothesize that in this setting the performance rating will be affected by the evaluator's prior recommendation. Specifically, we argue that managers who in the past advised favorably about that subordinate's promotion, will provide a higher rating of the subordinate's current period performance than managers who advised against the subordinate's promotion.

H1. Evaluators who in a previous period advised in favor of an employee's promotion to his current position will provide higher ratings of that employee's current period performance than evaluators who advised against the employee's promotion.

This first hypothesis predicts that escalation bias in subjective performance evaluation will be observed in a setting in which evaluators rate employees based on information contained in a

balanced scorecard. The main objective of our study, however, is to shed new light on the cause-effect relationship through which prior commitments affect subjective evaluations, and especially to establish whether escalation bias is at least partly explained by biased attention to available performance measures.¹ While there is general consensus that escalation bias reflects decision-makers' need to justify their prior decision, there is much less clarity about the specific manner in which the need to justify oneself results in biased evaluation judgments. In their seminal study, Bazerman et al. (1982) propose three possible mechanisms. First, it could be that prior commitments affect how raters look at current period performance information, "[...] that is the rater may pay greater attention to information that supports his or her promotion decision than to information that discredits the earlier decision" (Bazerman et al. 1982, p. 874). Second, it could be that "[...] given the set of information that the rater perceives, he or she may make biased judgments due to the prior decision. For example the rater may give greater weight to confirmatory than to disconfirmatory information" (Bazerman et al. 1982, p. 874). Finally, raters may deliberately bias their public evaluation judgment to be consistent with their prior decision. That is, while their private judgment of the subordinate's performance is unbiased, their publicly expressed evaluation judgment is more or less favorable than their private assessment, depending on the manager's prior commitment. In summary, according to Bazerman et al. (1982), escalation bias can arise because of biased attention to information, biased weighing of information, or deliberate actions.

¹ Note that as our interest is less in establishing *whether* or *when* prior recommendations affect subjective evaluations, and more in *how* this happens, we do not distinguish between positive and negative escalation. Instead, we follow the setup of Bazerman et al. (1982) and compare the performance ratings of evaluators who made a positive prior recommendation with the ratings of evaluators who made a negative prior recommendation and we do not have a control condition with evaluators who did not make a prior recommendation.

Only recently have researchers begun to explicitly disentangle these different mechanisms (Schultze et al. 2012). A central question in this emerging literature is whether biased attention to information indeed mediates escalation bias, as suggested by Bazerman et al. (1982) and other authors (e.g. Colon and Parks 1987; Dietz-Uhler 1996; Ku 2008). Attention has generally been defined as the process of focusing mental capacity on a selection of sensory input while disregarding other input from the sensory realm such that the mind can process the stimulus of interest (Kahneman 1973; Duchowski 2007). Attention plays a gatekeeping role between environmental stimuli and the working memory, because it filters and directs the encoding of information (Cowan 1988; Awh, Vogel, and Oh 2006). The working memory is a brain area in which information is temporarily stored and made accessible to other cognitive processes (Baddeley 1992; Fournie 2009). Consequently, informational cues that are not being attended to, do not enter the working memory, and are ignored in judgment and decision making. In this paper we are particularly interested in visual attention, i.e., what people look at (Orquin and Mueller Loose 2013).

Visual attention can be measured through eye-tracking because visual attention and eye-movements are tightly coupled (Wedel and Pieters 2007). Eye-movements consist of fixations and saccades. Fixations are periods of time in which the eye is relatively still and a scene is projected on the fovea for detailed visual processing. Saccades are very rapid (30-80 milliseconds) jumps of the eye from one fixation to another, during which vision is suppressed and no new information is acquired (Rayner 2009; Holmqvist et al. 2011). Eye-trackers record patterns of fixations and saccades. The underlying idea is that a fixation on a position (e.g., a word or a number on a screen) indicates that attention is paid to that position. More specifically, most eye-tracking studies in the judgment and decision making literature measure the number of fixations on a certain position and

the total time spent fixated at that position as indicators of attention (Sütterlin, Brunner, and Opwis 2008; Kuo, Hsu, and Day 2009; Glöckner and Herbold 2011; Ashby, Dickert, and Glöckner 2012).²

That biased attention mediates escalation bias seems reasonable because there is a plethora of research that suggests that selective attention to information is an important driver of closely related cognitive biases such as confirmation bias and assimilation effects.³ This research suggests that prior beliefs can bias judgments through a process of biased information acquisition, or *selective exposure* (Frey 1986; Hart, Albarracín, Eagly, Brechan, Lindberg, and Merrill 2009; Fischer and Greitemeyer 2010). Selective exposure is said to exist when individuals, consciously or unconsciously, expose themselves disproportionately to a subset of all the information that is available to them (e.g., Hart et al. 2009). Most literature on selective exposure specifically focuses on the idea that individuals have a tendency to selectively expose themselves to information that is consistent with their existing beliefs. This idea has its roots in Festinger's (1957) theory of cognitive dissonance. The basic tenet of cognitive dissonance theory is that individuals experience discomfort if information or behavior is not in line with their existing beliefs. Individuals will

² Three recent studies use eye-tracking technology in a management accounting context. Chen, Jermias, and Panggabean (2016) show that managers who look more at strategically linked performance measures, are more likely to make decisions consistent with the achievement of their subordinates' strategic objectives. Moreover, when they are aware of the strategy, managers focus more on strategically linked performance measures than on non-linked measures. Fehrenbacher, Schulz, and Rotaru (2018) use eye-tracking to measure decision-makers' decision modes (intuitive versus deliberative) and find that decision mode moderates spillover effects in subjective performance evaluation (Bol and Smith 2011). Finally, Dalla Via et al. (2019) study the effects of accountability on a balanced scorecard-based investment decision. These authors use eye-tracking to measure the intensity of information search effort and show process accountability increases search effort compared to outcome accountability and that search effort – in turn – increases performance. Lynch and Andiola (2019) provide a more detailed description of eye-tracking technology and review eye-tracking studies in the broader accounting literature.

³ Confirmation bias is the general tendency to seek or interpret information in ways that are partial to one's existing beliefs, expectations, or hypotheses in hand (Nickerson 1998). Assimilation effects similarly refer to a tendency to bias judgments based on current information in the direction of previous information or beliefs (Murphy, Balzer, Lockhart, and Eisenman 1985; Sumer and Knight 1996).

strive for internal consistency and will try to avoid information or behavior that is likely to increase the dissonance (Festinger 1957).

Most empirical work in the selective exposure domain consists of experiments in which participants, after making an initial choice or somehow acquiring an initial attitude, opinion or belief, have to select the information based upon which they will make a follow-up judgment or decision. Typically, participants can choose a limited number of information items from a larger set, and they are informed in advance whether a specific item is consistent or inconsistent with their initial choice or belief. Hart et al. (2009) present a meta-analysis of the selective exposure literature, which shows that while most individuals are motivated by both a desire to make accurate judgments and a desire to defend their original position, this latter ‘defense motivation’ is often stronger such that there is indeed a bias in information acquisition. Specifically, Hart et al.’s (2009) results indicate that individuals are more than twice as likely to select information items that are consistent with their initial beliefs, attitudes or behavior than information items that are inconsistent.

In the present context, selective exposure implies that prior commitments (i.e. a favorable or unfavorable recommendation about the employee’s promotion to his current position) affect how managers attend to information in a current period performance report. For example, if managers need to evaluate the performance of a subordinate whom they recommended for promotion in an earlier period, they may pay greater attention to information that justifies their prior recommendation than to information that discredits the earlier decision (Bazerman et al.

1982, p. 874).⁴ This reasoning implies that analysis of evaluators' eye-movements would reveal that managers who advised positively about the subordinate spend more time looking at measures that are above target, and less time looking at measures that are below target, than managers who advised negatively about the subordinate. It furthermore suggests that gaze patterns will be associated with ratings, such that managers who pay more attention to above-target measures and less attention to below-target measures, provide higher ratings of the overall performance of the subordinate.

While the evidence that decision makers selectively expose themselves to information that confirms their existing beliefs is generally strong (Hart et al. 2009), we do not yet know whether selective exposure also contributes to the explanation of escalation bias in subjective performance evaluation. As proposed by Bazerman et al. (1982), there are at least two other possible mechanisms through which escalation bias might emerge: biased *evaluation* of information and deliberate action, and it is unclear to what extent these mechanisms should be considered complements or substitutes.

In summary, while selective exposure theory suggests escalation bias in a subjective performance measure weighting context might at least be partly due to selective attention, the literature also points towards other mechanisms through which escalation can occur. To examine whether attention to performance measures mediates the relation between prior commitments and subjective performance ratings, we test the following hypothesis:

⁴ Consistent with prior literature (e.g., Loftus and Mackworth 1978; Rayner 1998), we assume a two-stage serial process of perception. During the first stage, basic categorical information, such as the favorability of a performance indicator is perceived by initial brief viewing. In the second stage, the information content is attended to in more detail and transcribed to the individual's memory. While selective attention does not affect the first stage, it does affect the second stage.

H2. Evaluators who, in a previous period, advised in favor of (against) an employee's promotion to his current position, will pay relatively more attention to favorable (unfavorable) performance indicators, which in turn will increase (decrease) their rating of the employee's current period performance.

RESEARCH METHOD

Experimental design

We test our hypotheses using an experiment with a 1×2 between-subjects design. Participants assumed the role of regional manager of a clothing retailer and received a balanced scorecard-format performance report about an employee who was recently promoted to store manager by the board of the company. We manipulated the evaluators' prior commitment by varying the description of the manager's recommendation regarding the employee's promotion to his current position (in favor or against). Half of the performance measures in the scorecard are above target and the other half is below target, such that the overall picture of the performance of the employee is ambiguous. Participants visually scan the scorecard and subsequently provide an overall performance rating on a 0 - 100 scale. While the participants are looking at the scorecard, we use an eye-tracking device to follow their eye-movements across the screen and register the number and duration of their fixations on each individual information item in the scorecard.

Participants and procedures

The participants in our experiment were 97 students from a West European business school, who were enrolled in the subject pool of the school's behavioral laboratory.⁵ The participants self-registered for participation in the experiment after an announcement on the laboratory website and were compensated with a fixed fee of €5. An initial inspection of the data revealed that the eye-tracker had failed to capture the gaze data of two participants. Those two participants were excluded from all further analyses and thus the final sample size is 95. Of the 95 participants in the sample, 45 (47.4%) are male and 50 (52.6%) are female. The participants are on average 21.79 years old and have 3.05 years of work experience.

Each participant was assigned an individual 30 minute time slot. The participants were first given a brief introduction to data recording with the eye-tracker and were informed about their rights as a participant and about the basic experimental procedures, to which they gave their verbal informed consent. Next, participants were positioned in front of the eye-tracker, which was then adjusted to their position and calibrated.⁶ After the calibration was successful, the experimenter started the real study which was programmed in E-Prime 2.0. From this point on, the participants could move through the study at their own pace using the mouse or keyboard. They first read some basic instructions and were introduced to their role and to the task. Next, they read the case scenario, performed the task and then moved to an exit questionnaire. After filling out the questionnaire, the participants received their payment and were dismissed. Participants were

⁵ The use of human subjects in this study was approved by the research institute where the experiment was conducted.

⁶ Calibration is necessary because variations in shape and radius of the eyeball or the use of glasses could otherwise distort the geometrical values that are used to calculate the gaze positions (Tobii Technology AB 2010; Holmqvist et al. 2011).

debriefed at a later stage, after we analyzed the data. They received an email explaining the basic purpose of the study and describing some preliminary results.

Task

Participants read a hypothetical case scenario about a company named Women's Apparel International (WAI). WAI was described as a chain that operates 96 clothing stores throughout the Netherlands and Europe. The participants were asked to assume the role of one of WAI's twelve regional managers. Their task was to review and rate the overall performance of a store manager in their region in the past year. To determine the rating, the participants received a performance report of the store manager containing target values and actual values of eight performance measures (described as "Key Performance Indicators"). The set of measures was organized in a balanced scorecard format with four different dimensions, i.e. a financial perspective, a customer perspective, an internal business perspective and a learning and growth perspective. The performance measures reported in the financial perspective were *return on sales* (a percentage) and *sales revenue growth* (a percentage). In the customer perspective, *repeat sales* (a percentage) and *average customer satisfaction rating* (on a 1-100 scale) were listed. The internal business perspective was comprised of *return to suppliers* (a percentage) and *average markdowns* (a percentage). Finally, the learning and growth perspective contained *hours of employee training per employee* (an absolute number) and *employee suggestions per employee* (an absolute number). The scorecard was constructed such that in each category one performance measure fell short of the given target and one performance measure exceeded the given target. The relative deviation from the target was kept constant between the performance measures on which the employee over- and underachieved to provide an ambiguous picture of the store manager's overall performance.

Manipulation and dependent variable

Following existing research (Bazerman et al. 1982; Slaughter and Greguras 2008), we manipulated the prior commitment of the rater by varying the descriptions in the case scenario. Participants in all conditions read that the store manager whose performance had to be rated (“Richard Berg”) had been promoted to his current position a year ago, and that prior to the promotion decision WAI’s corporate board had contacted the participant, in their role as regional manager, for advice. In the positive prior recommendation condition the participants furthermore read: *Based on your positive impression of Richard in his previous position as assistant store manager, you were very confident that he would perform well as a store manager. For this reason, you strongly recommended WAI’s corporate management board to promote Richard Berg to the position of store manager of the Rotterdam store.* In the negative prior recommendation condition the participants instead read *Based on your negative impression of Richard in his previous position as assistant store manager, you were very confident that he would perform badly as a store manager. For this reason, you strongly recommended WAI’s corporate management board NOT to promote Richard Berg to the position of store manager of the Rotterdam store.*

The dependent variable in our experiment is the overall performance rating that the participants assigned to the store manager based on their assessment and weighting of the different KPI’s provided in the performance report. Participants were asked to “[...] rate Richard Berg’s overall performance as a store manager in 2012 on scale from 0 (exceptionally bad performance) to 100 (exceptionally good performance).” The instructions furthermore explicitly specified that they should base their evaluation of Richard’s performance on the results in the provided balanced scorecard.

As described in more detail below, we used eye-tracking to measure the total duration and the frequency of participations' visual fixations at the target values and actual values of each of the eight performance measures. Our main mediator variable is the *relative time spent on favorable measures*. This was calculated as the total time that participants spent looking at the target values and actual values of favorable performance measures divided by the total time spent looking at the target values and actual values of unfavorable and favorable performance measures.⁷

Eye-tracking data

The university laboratory where the experiment was run has a specific eye-tracking section which consists of a sound-attenuated and dimly lit room with an eye-tracker of the type Tobii® 2150. The Tobii® 2150 looks like a normal 21-inch LCD monitor and it records gazing patterns in a non-intrusive way. It samples at 50 Hz, which means that it registers the user's eye position on the screen every 20 milliseconds. Each time the gaze position is registered, a new row is created in a database. The raw data for each participant thus consists of thousands of rows (one row, or 'sample', for each 20 milliseconds that the scorecard was projected on the participant's screen). Important columns in the database are a validity code that indicates whether the eye signal was captured without problems, the x-y coordinate on the screen for the left eye and for the right eye, and indicators of whether the x-y coordinates fall within a previously defined "area of interest" (AOI).

AOIs define screen regions of particular interest (e.g. Lynch and Andiola 2019). When designing the experiment, we defined several AOIs. As a result, each sample of the eye-tracking

⁷ In our study, the favorable performance measures were return on sales, average customer satisfaction rating, return to suppliers and number of employee suggestions. The remaining unfavorable measures were sales revenue growth, repeat sales, average markdowns and hours of employee training per employee.

raw data contained a variable that indicated at which AOI, if any, the user was looking at a given moment in time. The predefined AOIs included the scorecard's caption, the names of the performance measures and performance measure categories and the area in which the participants entered their performance rating. Most importantly, AOIs were defined around the target levels and actual levels of the performance indicators that were presented in the scorecard. To test our hypotheses we only focus on these latter AOIs. To make the time spent looking at these AOIs comparable, we ensured that they were all of equal size. Figure 1 shows the scorecard and the AOIs for the target and actual values of the measures as they were used in the study.⁸

--- Insert Figure 1 about here ---

Following existing literature (e.g., Sütterlin et al. 2008), we calculate two complementary indicators of visual attention: the number of fixations on each AOI and the total time spent fixated on each AOI, i.e. the *total dwell time* (Lynch and Andiola 2019). To distill these measures from the raw data, we used an algorithm that summed up the total time spent on an AOI if the participant viewed it for a period of at least ten consecutive samples, which corresponds to an attention threshold of 200 milliseconds. In other words, in line with recommendations in the literature (e.g. Holmqvist et al. 2011), we filter out periods where the eyes are fixed within an AOI for a period that is so short that no visual information can be extracted.⁹

⁸ Other AOIs are not depicted in Figure 1 for parsimony reasons.

⁹ While there is some debate in the literature on the appropriate algorithm to filter eye-tracking data (e.g., Holmqvist et al. 2011), a 200 millisecond threshold for fixations is common.

RESULTS

Preliminary analyses

In the exit questionnaire we asked participants whether they felt the eye-tracking device interfered with their judgment and decision making process. Specifically, we asked them to indicate (on seven-point Likert scales anchored 1 = strongly disagree and 7 = strongly agree), to what extent they agreed that they felt controlled by the eye-tracking device and to what extent they agreed that the fact that their eyes were tracked had affected performance rating. Neither participants in the positive recommendation condition (mean = 3.89, SD = 2.03) nor participants in the negative recommendation condition (mean = 3.34, SD = 1.94) felt controlled by the eye-tracker. The feeling of being controlled did not differ between conditions ($t = 1.324$; $p = 0.189$).¹⁰ Participants in both the positive recommendation condition (mean = 2.16, SD = 1.32) and the negative recommendation condition (mean = 2.07, SD = 1.24) furthermore disagreed with the statement that the eye-tracker had affected their rating. Again, there was no significant difference between the two conditions ($t = 0.335$; $p = 0.738$).

Next, to confirm that our setting provided ambiguous signals about the performance of the evaluated employee, we asked participants to what extent they agreed that “*the information in the Balanced Scorecard provided ambiguous signals about Richard Berg’s performance.*” As expected, we find that participants in both conditions tend to agree with this statement. In the positive recommendation condition the mean score on this item is 4.608 (SD = 1.313), which is significantly higher than the scale midpoint of 4 ($t = 3.307$; $p = 0.002$). In the negative

¹⁰ All p -values reported in this study are two-tailed. All questionnaire items discussed in this section were formulated as statements and were scored on a seven point Likert scale anchored 1 = strongly disagree and 7 = strongly agree.

recommendation condition the mean score on this item is 4.386 (SD = 1.104), which is also significantly higher than the scale midpoint of 4 ($t = 2.321$; $p = 0.025$). The mean item score is not significantly different between the two conditions ($t = 0.882$; $p = 0.380$).

We also included two items in the questionnaire to check the effectiveness of the manipulation. For the item “*my impression of Richard at the time of his promotion to store manager was very positive*” the mean score is significantly higher in the positive prior recommendation condition than in the negative prior recommendation condition (mean = 6.28, SD = 0.94 vs. mean = 1.82, SD = 1.15; $t = 20.814$; $p < 0.001$). For the item, “*I had strongly recommended Richard for promotion to store manager*”, the mean score is also significantly higher in the positive prior recommendation condition than in the negative prior recommendation condition (mean = 6.28, SD = 0.94 vs. mean = 1.46, SD = 0.93; $t = 25.091$; $p < 0.001$). These results indicate that the manipulation was successful.

Table 1 provides descriptive statistics for the performance rating (the dependent variable) and several variables related to the viewing patterns of the participants. The table shows that, on average, participants assigned a performance rating of 64.59 to the employee. The table further shows that the average participant spent 3.19 seconds looking at favorable performance measures and 2.86 seconds looking at unfavorable measures.¹¹ The average relative time spent on favorable measures is 52.4 percent. Participants fixate on average 10.94 times on favorable measure values and 9.80 times on unfavorable measure values. The mean favorable-to-total fixation ratio is 52.5 percent.

¹¹ The average (median) time that participants took for the performance evaluation decision, i.e. the time they spent on the screen on which the scorecard was projected and on which they had to fill out their rating is 49.40 (50.34) seconds.

--- Insert Table 1 about here ---

Table 2 contains bivariate correlations of our main variables and some participant demographics. From this table it is clear that the manipulation is not correlated with any demographic variable (all $p > 0.10$), consistent with a random allocation of the participants over the two experimental conditions. Age, gender and work experience are also not correlated with the performance rating or any of the eye-tracking measures. The manipulation is correlated with the performance rating ($r = 0.337$; $p < 0.001$), but neither the manipulation nor the rating is correlated with any of the eye-tracking measures (all $p > 0.10$). There are several correlations between the different eye-tracking measures, as should be expected. The absolute time spent on favorable measures has a strong positive correlation with the absolute time spent on unfavorable measures ($r = 0.867$; $p < 0.001$). In addition, for both favorable measures ($r = 0.990$; $p < 0.001$) and unfavorable measures ($r = 0.987$; $p < 0.001$) the number of fixations is almost perfectly correlated with the absolute time spent looking at the measures.

--- Insert Table 2 about here ---

Hypothesis tests

Our baseline H1 predicts that the performance ratings provided by evaluators who in a previous period advised in favor of an employee's promotion will be higher than the performance ratings provided by evaluators who advised against an employee's promotion in a previous period. An independent sample t -test, reported in Table 1, indicates that this indeed the case. From this table it is clear that the mean performance rating in the positive recommendation condition is 69.25 (SD = 13.66), whereas in the negative recommendation condition it is 59.18 (SD = 14.77). The

mean difference of 10.07 between the positive and negative recommendation conditions is significant ($t = 3.451$; $p < 0.001$) and thus H1 is supported by the data.

H2 predicts that attention mediates the effect of managers' prior recommendation on their performance rating. We test this hypothesis using ordinary least squares regression analysis, following the standard approach suggested by Baron and Kenny (1986). The Baron and Kenny (1986) approach to mediation analysis involves a three step procedure. In the first step we assess the unmediated model by testing whether the effect of the independent variable on the dependent variable is significant. As is also clear from the t -test, the effect of the prior recommendation on the performance rating is significant ($\beta = 0.337$; $p < 0.001$). In the second step, we test whether the independent variable affects the mediating variable. The results show that the recommendation does not significantly affect the relative time spent looking at favorable performance measures ($\beta = 0.075$; $p = 0.471$). In the third step, we test whether the mediating variable affects the dependent variable and whether this effect also influences the relationship between the independent and the dependent variable. When we test this joint effect we find that the proportion of time spent looking at favorable performance measures has no significant effect on the performance rating ($\beta = -0.090$; $p = 0.363$) while the effect of the recommendation on the performance rating is highly significant ($\beta = 0.344$; $p < 0.001$). Because both mediating paths are insignificant while the direct path stays almost identical we conclude that there is no support for H2. Thus, we find no evidence that the effect of prior recommendations on performance ratings is mediated by biased attention to favorable (versus unfavorable) information items. The results are summarized in Table 3.

--- Insert Table 3 about here ---

Robustness tests and questionnaire data

To increase our understanding of the results of the experiment we run several robustness checks and supplemental tests, and we also examine data collected using the post-experimental questionnaire. The results of these additional analyses are reported in detail in Online Appendix A. The robustness checks show that the results are not qualitatively different if we exclude participants who spent very little time looking at the performance report or participants who provided exceptionally low performance ratings. We also confirm that the findings do not critically depend on the specific algorithm that we use to transform the raw eye-tracking data or on our specific measure of visual attention. The questionnaire data are consistent with our main analyses and support our conclusions that prior recommendations influence subjective performance evaluations, but not by biasing evaluators' attention.

DISCUSSION AND CONCLUSION

While it has often been assumed that biases in accounting-based decisions are caused by differences in the extent to which decision makers pay attention to the various items in accounting reports, there is little evidence that this is actually the case. This paper uses an eye-tracking experiment to shed new light on the role of attention as a possible determinant of escalation bias in subjective performance evaluation judgments. We find no evidence that, in our setting, escalation bias emerges because evaluators with different prior commitments pay attention to different items in the performance report.

The results of our study reconfirm the need to better understand the role of attention in accounting-based decision making processes (Birnberg and Shields 1984). Accounting information can only affect decisions if it is attended to. However, why decision makers attend to

specific pieces of accounting information and how – if at all - attention patterns are context-dependent is unknown. Similarly, the current literature does not provide many clues as to whether the amount of attention that individuals pay to a piece of information in an accounting report affects the decisions that they make. Nevertheless, both researchers and practitioners need to make assumptions about when, why, and how individuals using accounting reports attend to the information in those reports. We believe that explicating those assumptions, firmly grounding them in psychological theory and putting them to an empirical (eye-tracking) test will improve our understanding of the role of visual attention in accounting-based decision tasks.

Our findings are likely of interest to researchers examining the determinants of escalation of commitment and biases in subjective performance evaluation, but also contribute to the more general debate in the psychology literature on attention and selective exposure as drivers of judgment and decision-making biases (Hart et al. 2009; Fischer and Greitemeyer 2010; Schultze et al. 2012; Chaxel et al. 2013). Of course, when drawing inferences based on our findings it is important to consider the study's limitations. First, participants in our experiment were aware that their eyes were being tracked while they progressed through the study. Even though participants stated in the post-experimental questionnaire that they did not feel controlled by the eye-tracker and that it did not affect their rating decision, we cannot rule out that the use of the eye-tracker affected participants' gazing patterns or rating decisions. Second, participants in our experiment had relatively little experience in evaluating subordinates, which may have made it more difficult for them to put themselves into the position of an evaluating manager. We cannot rule out that the limited experience of our participants may have exacerbated escalation bias in our experiment. Third, participants in our experiment made their rating decisions on the basis of a case-based scenario. No financial consequences or reputational costs were associated with providing an

inaccurate rating, which may have attenuated or mitigated the extent to which participants biased their performance rating. Fourth, one specific design choice regarding our case may have contributed to the lack of support for H2. This design choice is that rather than identifying the performance measures in the balanced scorecard as favorable or unfavorable, we left it to the participants to figure this out for themselves. Had we, instead, grouped the positive and negative measures together, or otherwise highlighted which measures were more or less favorable, it would have been easier for the evaluators to identify which measures were and which were not consistent with their prior decision. Future researchers may want to identify how – if at all - explicitly highlighting performance indicators as positive or negative (e.g., using ‘traffic light’ systems) affects judgments and decision making processes of managers, including performance evaluations. Fifth, we examined a setting in which evaluators rated the performance of a subordinate who was promoted to his current position, and manipulated whether or not this promotion was consistent with the evaluator’s prior recommendation. In another possible scenario, evaluators have to rate the performance of a subordinate who – consistent or inconsistent with the evaluator’s recommendation – did not receive the promotion. While we have no reason to believe that we would find different results in a such a scenario, we acknowledge that this is ultimately an empirical question that can be addressed in future research. Finally, while our experiment allowed us to examine the possible role of selective attention as a determinant of escalation bias, it does not tell us much about which alternative mechanisms might have been at work. This is something that clearly requires further research. In addition, future researchers may want to investigate whether attention bias contributes to other biases in subjective performance evaluation based on accounting information, e.g. common measure bias (Lipe and Salterio 2000) or outcome effects (Ghosh and Lusch 2000; Ittner et al. 2003). Another potential avenue for future research would be

to further explore the effects of performance report layout on evaluation decisions (e.g., Maas and Verdoorn 2018). Finally, researchers could extend the investigation of the role of visual attention to other types of performance reports (e.g., annual reports), other types of judgments and decisions (e.g., auditor opinions, investor decisions), and information search and learning patterns (Lynch and Andiola 2019).

REFERENCES

- Angelovski, A., J. Brandts, and C. Sola. 2016. Hiring and escalation bias in subjective performance evaluations: A laboratory experiment. *Journal of Economic Behavior and Organization* 121: 114-129.
- Ashby, N.J.S., S. Dickert, and A. Glöckner. 2012. Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making* 7 (3): 254-267.
- Awh, E., E. Vogel, and S.-H., Oh. 2006. Interactions between attention and working memory. *Neuroscience* 139 (1): 201-208.
- Baddeley, A. 1992. Working memory. *Science* 255 (5044): 556-559.
- Baron, R.M., and D.A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51 (6): 1173-1182.
- Bazerman, M.H., R.I. Beekun, and F.D. Schoorman. 1982. Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee. *Journal of Applied Psychology* 67 (6): 873-876.
- Birnberg, J.G., and M.D. Shields. 1984. The role of attention and memory in accounting decisions. *Accounting, Organizations and Society* 9 (3-4): 365-382.
- Bol, J.C. 2008. Subjectivity in compensation contracting. *Journal of Accounting Literature* 27: 1-32.
- Bol, J.C., S. Kramer, and V.S. Maas. 2016. How control system design affects performance evaluation compression: The role of information accuracy and outcome transparency. *Accounting, Organizations and Society* 51: 64-73.
- Bol, J.C., and S.D. Smith. 2011. Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review* 86 (4): 1213-1230.
- Brockner, J. 1992. The escalation of commitment to a failing course of action: Toward theoretical progress. *Academy of Management Review* 17 (1): 39-61.
- Brody, R. G., and S.E. Kaplan. 1996. Escalation of commitment among internal auditors. *Auditing: A Journal of Practice & Theory* 15 (1): 1-15.
- Brownstein, A.L. 2003. Biased predecision processing. *Psychological Bulletin* 129 (4): 545-568.
- Cardinaels, E., and P.M.G. van Veen-Dirks, 2010. Financial versus non-financial information: The impact of information organization and presentation in a Balanced Scorecard. *Accounting, Organizations and Society* 35 (6): 565-578.
- Chaxel, A-S, J.E. Russo, and N. Kerimi 2013. Preference-driven biases in decision makers' information search and evaluation. *Judgment and Decision Making* 8 (5): 561-576.
- Chen, Y., J. Jermias, and T. Panggabean. 2016. The role of visual attention in the managerial judgment of balanced-scorecard performance evaluation: Insights from using an eye-tracking device. *Journal of Accounting Research* 54 (1): 113-146.
- Cheng, M.M., A.K. Schulz, P.F. Lockett, and P. Booth. 2003. The effects of hurdle rates on the level of escalation of commitment in capital budgeting. *Behavioral Research in Accounting* 15 (1): 63-85.
- Conlon, E. J., and J.M. Parks. 1987. Information requests in the context of escalation. *Journal of Applied Psychology* 72 (3): 344-350.
- Cowan, N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin* 104 (2): 163-191.

- Dalla Via, N., P. Perego, and M. van Rinsum. 2019. How accountability type influences information search processes and decision quality. *Accounting, Organizations and Society* 75: 79-91.
- Denison, C. A. 2009. Real options and escalation of commitment: A behavioral analysis of capital investment decisions. *The Accounting Review* 84 (1): 133-155.
- Dietz-Uhler, B.E.T.H. 1996. The escalation of commitment in political decision-making groups: A social identity approach. *European Journal of Social Psychology* 26 (4): 611-629.
- Dilla, W. N., and P.J. Steinbart. 2005. Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting* 17 (1): 43-53.
- Duchowski, A. 2007. *Eye-tracking Methodology: Theory and Practice*. Secaucus NJ: Springer.
- Emby, C., and D. Finley, D. 1997. Debiasing framing effects in auditors' internal control judgments and testing decisions. *Contemporary Accounting Research* 14 (2): 55-77.
- Fehrenbacher, D.D., A.K.D. Schulz, and K. Rotaru. 2018. The moderating role of decision mode in subjective performance evaluation. *Management Accounting Research* 41: 1-10.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
- Fischer, P., and T. Greitemeyer. 2010. A new look at selective-exposure effects: An integrative model. *Current Directions in Psychological Science* 19 (6): 384-389.
- Fougnie, D. 2009. The relationship between attention and working memory. In *New research on short-term memory*, edited by N.B. Johansen. New York, NY: Nova Science Publishers.
- Frey, D. 1986. Recent research on selective exposure to information. In *Advances in Experimental Social Psychology - Volume 19*, edited by L. Berkowitz, 41-80. New York, NY: Academic Press.
- Ghosh, D., and R.F. Lusch. 2000. Outcome effect, controllability and performance evaluation of managers: Some field evidence from multi-outlet businesses. *Accounting, Organizations and Society* 25 (4-5): 411-425.
- Glöckner, A., and A.K. Herbold. 2011. An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making* 24 (1): 71-98.
- Hart, W., D. Albarracín, A.H. Eagly, I. Brechan, M.J. Lindberg, and L. Merrill. 2009. Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological Bulletin* 135 (4): 555-588.
- Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. 2011. *Eye-tracking: A Comprehensive Guide to Methods and Measures*. Oxford, U.K: Oxford University Press.
- Hoque, Z. 2014. 20 years of studies on the balanced scorecard: Trends, accomplishments, gaps and opportunities for future research. *The British Accounting Review* 46 (1): 33-59.
- Ittner, C.D., D.F. Larcker, and M.W. Meyer. 2003. Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*. 78 (3): 725-758.
- Kadous, K., and L.M. Sedor. 2004. The efficacy of third-party consultation in preventing managerial escalation of commitment: The role of mental representations. *Contemporary Accounting Research* 21 (1): 55-82.
- Kahneman, D. 1973. *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Ku, G. 2008. Learning to de-escalate: The effects of regret in escalation of commitment. *Organizational Behavior and Human Decision Processes* 105 (2): 221-232.

- Kuo, F.Y., C.W. Hsu, and R.F. Day. 2009. An exploratory study of cognitive effort involved in decision under framing: An application of the eye-tracking technology. *Decision Support Systems* 48 (1): 81-91.
- Libby, T., S.E. Salterio, and A. Webb. 2004. The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review* 79 (4): 1075-1094.
- Liedtka, S.L., B.K. Church, and M.M Ray. 2008. Performance variability, ambiguity intolerance, and Balanced Scorecard-based performance assessments. *Behavioral Research in Accounting* 20 (2): 73-88.
- Lipe, M.G., and S.E. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75 (3): 283-298.
- Loftus, G.R., and N.H. Mackworth. 1978. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human perception and Performance* 4: 565-572.
- Luft, J., and M.D. Shields. 2009. Psychology models of management accounting. *Foundations and Trends in Accounting* 4 (3-4): 199-345.
- Lynch, E.J., and L.M. Andiola. 2019. If eyes are the window to our soul, what role does eye tracking play in accounting research? Forthcoming in: *Behavioral Research in Accounting*.
- Maas, V.S. and N. Verdoorn. 2017. The effects of performance report layout on managers' subjective evaluation judgments. *Accounting and Business Research* 47 (7): 731-751.
- Müller, D., and P. Weinschenk. 2015. Rater bias and incentive provision. *Journal of Economics and Management Strategy* 24 (4): 833-862.
- Murphy, K.R., W.K. Balzer, M.C. Lockhart, and E.J. Eisenman. 1985. Effects of previous performance on evaluations of present performance. *Journal of Applied Psychology* 70 (1): 72-84.
- Nickerson, R.S. 1998. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology* 2 (2): 175-220.
- Orquin, J.L., and S. Mueller Loose. 2013. Attention and choice: A review on eye-movements in decision making. *Acta Psychologica* 144 (1): 190-206.
- Prendergast, C. 1999. The provision of incentives in firms. *Journal of Economic Literature* 37 (1): 7-63.
- Rayner, K. 1998. Eye-movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124 (3): 372-422.
- Rayner, K. 2009. Eye-movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology* 62 (8): 1457-1506.
- Schoorman, F.D. 1988. Escalation bias in performance appraisals: An unintended consequence of supervisor participation in hiring decisions. *Journal of Applied Psychology* 73 (1): 58-62.
- Schultze, T., F. Pfeiffer, and S. Schulz-Hardt. 2012. Biased information processing in the escalation paradigm: information search and information evaluation as potential mediators of escalating commitment. *Journal of Applied Psychology* 97 (1): 16-32.
- Schultze, T., and S. Schulz-Hardt. 2015. The impact of biased information and corresponding meta-information on escalating commitment. *Journal of Economic Psychology* 49: 108-119.
- Slaughter, J.E., G.J. Greguras. 2008. Bias in performance ratings: Clarifying the role of positive versus negative escalation. *Human Performance* 21 (4): 414-426.

- Sleesman, D.J., D.E. Conlon, G. McNamara, and J.E. Miles. 2012. Cleaning up the big muddy: A meta-analytic review of the determinants of escalation of commitment. *Academy of Management Journal* 55 (3): 541-562.
- Staw, B.M. 1976. Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance* 16 (1): 27-44.
- Staw, B.M. 1981. The escalation of commitment to a course of action. *Academy of Management Review* 6 (4): 577-587.
- Staw, B.M. 1997. The escalation of commitment: An update and appraisal. In *Organizational Decision Making* edited by Z. Shapira, 191 – 215. New York: Cambridge University Press.
- Sumer, H.C., and P.A. Knight. 1996. Assimilation and contrast effects in performance ratings: Effects of rating the previous performance on rating subsequent performance. *Journal of Applied Psychology* 81 (4): 436-442.
- Sütterlin, B., T.A. Brunner, and K. Opwis, 2008. Eye-tracking the cancellation and focus model for preference judgments. *Journal of Experimental Social Psychology* 44 (3): 904-911.
- Thaler, R.H., and C.S. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
- Tobii Technology AB. 2010. *User Manual Tobii Studio*.
- Wedel, M., and R. Pieters. 2007. A review of eye-tracking research in marketing. In *Review of Marketing Research Volume 4* edited by N. K. Malhotra, 123 – 147. New York, NY: M.E. Sharpe Inc.
- Woods, A. 2012. Subjective adjustments to objective performance measures: The influence of prior performance. *Accounting, Organizations and Society* 37 (6): 403-425.

TABLE 1
Descriptive Statistics and Mean Comparisons Across Conditions

		Negative recommendation	Positive recommendation	Overall	<i>t</i> -statistic	<i>p</i> -value
Performance rating (0-100)	Mean	59.18	69.25	64.59	3.451	<0.001
	SD	14.77	13.66	14.99		
	Min	6	7	6		
	Max	85	90	90		
	n	44	51	95		
Absolute time spent on favorable measures (seconds)	Mean	3.00	3.35	3.19	0.561	0.576
	SD	2.54	3.38	3.01		
	Min	0.00	0.20	0.00		
	Max	9.63	14.21	14.21		
	n	44	51	95		
Absolute time spent on unfavorable measures (seconds)	Mean	2.67	3.03	2.86	0.597	0.552
	SD	2.42	3.25	2.88		
	Min	0.00	0.00	0.00		
	Max	10.33	16.62	16.62		
	n	44	51	95		
Relative time spent on favorable measures (0.00-1.00)	Mean	0.509	0.537	0.524	0.723	0.471
	SD	0.214	0.166	0.189		
	Min	0.00	0.15	0.00		
	Max	1.00	1.00	1.00		
	n	44	51	95		
Number of fixations on favorable measures	Mean	10.39	11.41	10.94	0.513	0.609
	SD	8.01	10.97	9.67		
	Min	0	1	0		
	Max	29	46	46		
	n	44	51	95		
Number of fixations on unfavorable measures	Mean	9.09	10.41	9.80	0.696	0.488
	SD	7.39	10.55	9.20		
	Min	0	0	0		
	Max	31	58	58		
	n	44	51	95		
Proportion favorable fixations (0.00-1.00)	Mean	0.511	0.536	0.525	0.674	0.502
	SD	0.206	0.164	0.184		
	Min	0.00	0.20	0.00		
	Max	1.00	1.00	1.00		
	n	44	51	95		

This table presents descriptive statistics for the dependent variable and for a number of eye-tracking measures. The statistics are presented for the overall sample and for each of the two experimental conditions separately. In the positive recommendation condition participants were asked to assume that they had advised in favor of the ratee's promotion to his current position. In the negative recommendation condition participants were asked to assume that they had advised against the ratee's promotion to his current position.

Continued on next page

Table 1 continued

The dependent variable (performance rating) refers to the overall rating of the ratee's performance provided by the participants. Performance was rated on a scale from 0 (*exceptionally bad performance*) to 100 (*exceptionally good performance*).

Absolute time spent on favorable measures refers to the number of seconds that participants spent fixated on the target values and the actual values of the four performance measures for which the actual level of performance exceeded the target level of performance.

Absolute time spent on unfavorable measures refers to the number of seconds that participants spent fixated on the target values and the actual values of the four performance measures for which the actual level of performance fell short of the target level of performance.

Relative time spent on favorable measures is the proportion of time that participants were fixated on favorable measures and is calculated as Absolute time spent on favorable measures divided by the sum of Absolute time spent on favorable measures and Absolute time spent on unfavorable measures.

Number of fixations on favorable measures is the number of times participants spent a minimum of 200 milliseconds looking at the target or actual value of a performance measure for which the actual level of performance exceeded the target level of performance.

Number of fixations on unfavorable measures is the number of times participants spent a minimum of 200 milliseconds looking at the target or actual value of a performance measure for which the actual level of performance fell short of the target level of performance.

Proportion favorable fixations is the proportion of fixations on measures that are favorable and is calculated as Number of fixations on favorable measures divided by the sum of Number of fixations on favorable measures and Number of fixations on unfavorable measures.

The *t*-values are from independent sample *t*-tests for the difference in means between the two conditions. The *p*-values are two-tailed.

TABLE 2
Bivariate Correlations

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
[1]: Recommendation										
[2]: Performance rating	0.337**									
[3]: Absolute time spent on favorable measures	0.058	-0.084								
[4]: Absolute time spent on unfavorable measures	0.062	-0.077	0.867**							
[5]: Relative time spent on favorable measures	0.075	-0.064	0.171	-0.144						
[6]: Number of fixations on favorable measures	0.053	-0.063	0.990**	0.863**	0.176					
[7]: Number of fixations on unfavorable measures	0.072	-0.072	0.846**	0.987**	-0.157	0.849**				
[8]: Proportion favorable fixations	0.070	-0.053	0.169	-0.135	0.979**	0.188	-0.162			
[9]: Age	0.160	0.114	0.140	0.148	-0.131	0.144	0.177	-0.133		
[10]: Gender	0.036	-0.012	0.071	0.021	0.133	0.043	0.005	0.119	0.129	
[11]: Work experience	-0.020	-0.130	0.124	0.042	0.177	0.119	0.049	0.170	0.251*	0.005

This table contains Pearson correlation coefficients. Recommendation is a dummy variable that indicates which experimental condition participants received. The dummy has value 1 for participants in the positive recommendation condition and 0 for participants in the negative recommendation condition. See the notes to Table 1 for a description of the experimental conditions and the non-demographic variables. *Age*, *Gender* and *Work experience* are participants' self-reported age (in years), gender and years of work experience.

** indicates the correlation is significant at the 0.01 level (2-tailed).

* indicates the correlation is significant at the 0.05 level (2-tailed).

TABLE 3
Hypotheses Tests

Unmediated model	β	<i>t</i> -statistic	<i>p</i> -value
Regression 1:			
Recommendation → Performance rating	0.337	3.451	<0.001
Mediation model	β	<i>t</i> -value	<i>p</i> -value
Regression 2:			
Recommendation → Relative time spent on favorable measures	0.075	0.723	0.471
Regression 3:			
Relative time spent on favorable measures → Performance rating	-0.090	0.915	0.363
Recommendation → Performance rating	0.344	3.507	<0.001

This table reports the results of the OLS regressions that were used to test the hypotheses. The table contains standardized β 's.
The *p*-values are two-tailed.

FIGURE 1
Balanced Scorecard Used in the Experiment

Balanced Scorecard containing the Key Performance Indicators for the year 2012 of WAI's Rotterdam store managed by Richard Berg		
	2012 Target	2012 Actual
Financial perspective		
Return on sales (+)	15.00 %	24.15 %
Sales revenue growth (+)	10.00 %	3.90 %
Customer perspective		
Repeat sales (+)	30.00 %	22.60 %
Average customer satisfaction rating, measured on a 1 – 100 scale (+)	75.00	93.50
Internal Business Processes Perspective		
Returns to suppliers (-)	5.00 %	2.21 %
Average markdowns (average % markdown from original retail price) (-)	10.00 %	15.58 %
Learning & Growth Perspective		
Hours of employee training per employee (+)	20.00	8.45
Employee suggestions per employee (+)	8.00	12.62

Please take your time to study the information presented above and then rate Richard Berg's overall performance as a store manager in 2012 on a scale from 0 (exceptionally bad performance) to 100 (exceptionally good performance). Use the numbers on the keyboard to pick any whole number between 0 and 100 and then press ENTER to confirm. Note that once you have provided an answer you cannot go back. So please be careful in entering your performance rating. My rating of Richard Berg's overall performance as store manager in 2012 is

This figure shows the screen with the balanced scorecard as it was presented to the participants during the experiment. The rectangles represent the AOIs used to measure the time spent on favorable and unfavorable performance indicators as captured by the eye-tracker. Participants did not see these AOIs on their screens. The plus and minus signs after the names of the measures indicate whether a higher score on that measures indicates better or worse performance. This was explained to the participants on a previous screen.