ARTICLE IN PRESS

Management Accounting Research xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Management Accounting Research

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

The moderating role of decision mode in subjective performance evaluation

Dennis D. Fehrenbacher[a,*], Axel K.-D. Schulz[b], Kristian Rotaru[a]

[a] Department of Accounting, Monash Business School, Monash University, 900 Dandenong Road, Caulfield East, VIC 3145, Australia
[b] Department of Accounting and Data Analytics, La Trobe Business School, La Trobe University, Plenty Road & Kingsbury Drive, Melbourne, VIC 3086, Australia

ARTICLE INFO

Keywords:
Subjective performance evaluation
Performance measurement
Spill-over effect
Eye tracking

ABSTRACT

We use eye tracking technology to provide a better understanding of cognitive processes behind biases in subjective performance evaluation. In our experiment, subjective performance evaluation involves a supervisor evaluating the office administration performance of a subordinate. Consistent with previous literature, we find that the subjective evaluation of subordinate performance is influenced by performance on an unrelated objective measure used to evaluate the subordinate (spill-over). We predict and provide evidence that the supervisor's decision modes (intuition *versus* deliberation) interact with the level of performance on the objective performance measure to determine the subjective performance evaluation and the magnitude of the spill-over. Specifically, we find that individuals who use more effortful, deliberate decision modes show lower levels of spill-over. As such, we contribute to the accounting literature by examining how biases in performance evaluation can be reduced and showing ways to capture cognitive processes accompanying those biases more precisely.

1. Introduction

The use of subjectivity in performance evaluation is an important part of performance measurement systems (Gibbs et al., 2004; Bol, 2011). Its importance is supported in the analytical and empirical literature as it allows performance evaluation to be supplemented with non-contractible information which can benefit performance (Baker et al., 1994; Baiman and Rajan, 1995; Van der Stede et al., 2006; Bol, 2008; Cheng and Coyte, 2014).[1]

Prior research, however, has documented a number of biases that arise with the use of subjectivity in performance evaluation (Moers, 2005; Bol and Smith, 2011; Woods, 2012; Franco-Santos et al., 2012). Studying these biases is important as they can reduce the benefits of including subjectivity in performance evaluation. Our guiding question is whether biases in subjective performance evaluation arising through subjectivity can be reduced. It has been suggested that biases can be mitigated if individuals exhibit more effortful cognitive processing (Kennedy, 1993; Kahneman and Frederick 2002, 2005; Kahneman, 2011). However, it has also been noted that not all biases can be mitigated by more effortful processing (Kennedy, 1993; Kahneman, 2011).

We study the spill-over bias which has been found in subjective performance evaluation. Spill-over occurs when supervisors bias their evaluation of a subjective measure consistent with the level of performance of an unrelated objective measure. Spill-over has been described to be associated with an 'evaluative disposition' formed by supervisors who rate an objective measure (Bol and Smith, 2011). The evaluative disposition is formed by information that is evaluated early (i.e. the objective measure). It then influences the processing of subsequent information (i.e. the subjective measure) in a coherent way with the disposition (Bond et al., 2007). Whether the strength of the disposition and hence the spill-over bias is also associated with the cognitive effort exerted by the supervisor has not been the focus of prior research. Thus, the literature lacks clarity as to whether this bias can be mitigated by more effortful cognitive processing.

There is evidence in the accounting literature that accounting related decision-making is influenced by different decision modes (e.g. Shields, 1980; Chenhall and Morris, 1991; Bailey et al., 2011). In accounting, and in the wider literature, a variety of typologies have been suggested to differentiate between multiple decision modes (Kahneman and Frederick, 2002). We use a typology by Horstmann et al. (2009) which distinguishes between two widely used decision modes: a rather slow and effortful deliberate decision mode (deliberation) *versus* a rather fast and effortless intuitive decision mode (intuition) (see also Kahneman and Frederick 2002; Evans, 2008). Horstmann et al. (2009) showed that these two decision modes can be approximated by fixation count, a measure obtained using eye tracking technology. A fixation

* Corresponding author.
E-mail addresses: Dennis.Fehrenbacher@monash.edu (D.D. Fehrenbacher), A.Schulz@latrobe.edu.au (A.K.-D. Schulz), Kristian.Rotaru@monash.edu (K. Rotaru).

[1] We use the terms 'non-contractible information' and 'subjective information' interchangeably to denote pieces of information which need to be considered subjectively in order to evaluate performance. In line with analytical literature (Rajan and Reichelstein, 2009), we define subjective information as being not verifiable for contracting purposes.

https://doi.org/10.1016/j.mar.2018.03.001
Received 31 March 2016; Received in revised form 12 March 2018; Accepted 14 March 2018

occurs when the eyes stay focused on a single location longer than a particular time threshold. The advantage of this measurement, as compared to survey-based measurements, is that it can be captured while individuals process the relevant information (Rotaru et al., 2018; Birnberg and Ganguly, 2012).

We argue that the evaluative disposition leading to spill-over has a stronger effect in an intuitive mode due to the fast and effortless cognitive processing of that mode leading to a rapid contextualization relying more on prior beliefs and knowledge (Evans, 2008). In contrast, spill-over is reduced if an individual's decision mode is more consistent with a deliberate mode. In a deliberate mode, supervisors take the time to consider what dimensions of subordinates' jobs performance measures capture (i.e., supervisors increase cognitive effort), and recognize that the measures capture distinct dimensions (i.e., the job dimensions and thus their measures are not significantly associated). As such, the evaluative disposition is mitigated. Consequently, we examine whether and how objective performance on an unrelated task interacts with the decision mode to influence the spill-over and thus the subjective performance evaluation.

We examine this in an employment setting, in which a subordinate's performance is measured on two unrelated dimensions (tasks): performance on one dimension is measured objectively, while the supervisor evaluates performance on the other dimension subjectively. The objective information is available before the supervisor subjectively evaluates performance of the subordinate on the other dimension. In such a setting, Bol and Smith (2011) showed that the level of performance on an objective measure can affect the subjective evaluation of performance on an unrelated dimension. Specifically, when performance on the objective measure is lower, supervisor's evaluation of information for the subjective measure is lower compared to the case when performance on the objective measure is higher.

We use an experiment in which participants assumed the role of a regional director with supervisory authority over district managers who had both sales and office administration related duties. Objective sales information in relation to the district manager's sales performance was given, after which participants were asked to evaluate the performance of the district manager's office administration duties. In half the cases the subordinate performance was low as reported by the objective measure; in the other half the subordinate performance was high. The experiment is based on a 2×1 between-subjects design.

Our study makes a number of contributions. First, prior accounting research has shown that subjective performance evaluation can be affected by cognitive biases (Bol and Smith, 2011; Woods, 2012; Ruggeri, 2012). We extend this line of research by reporting evidence that biases which are primarily associated with evaluative pre-dispositions and not effort, may still be mitigated by a decision mode that involves more cognitive effort. We provide empirical support for this in association with the spill-over bias. Hence, the spill-over bias may be stronger when evaluators use more intuitive and less effortful decision modes, and the size of the spill-over bias may not be universal across different decision modes. The result of this study provides opportunities for mitigating this and similar biases (e.g. memory bias, as discussed below) in subjective performance evaluation.

Second, Bol and Smith (2011) report that they do not find evidence that would suggest that the overall amount of time participants take to complete the experiment, used as a proxy for effort, interacts with the level of the objective performance measure. We argue that this is due to the overall time measure being too noisy and not precise enough to reflect the processing of the subjective information. The imprecision is due to the time measure capturing both time spent attending the subjective information as well as time spent attending information other than the subjective information. We extend their research by comparing our primary measure, fixation count, with the overall time measure and additional measures (fixation duration and time spent until page submit on the relevant screen). Consistent with Bol and Smith (2011) we find that the overall time measure does not interact with the level of the objective performance measure, however, we find that fixation duration and time spent on the screen that depicted the subjective information lead to similar results as fixation count. Thus, in our study the less noisy measures including the eye tracking measures provide additional insights. In accounting, the use of eye tracking devices is a recent method innovation. Our results provide support that it has the potential to further open the 'black box' and to get insights into cognitive processing and its effects on observed behavior (Birnberg and Ganguly, 2012).

In the next section, we present the background and hypothesis. Subsequently, we describe the experimental method, present the results, discuss the findings and outline avenues for future research.

2. Background and hypothesis development

2.1. Subjectivity in performance evaluation and spill-over

Incorporating subjectivity into the performance measurement process can help to overcome disadvantages of objective measures. While there are conditions under which the sole reliance on a single objective item of accounting information can be preferable (Luft et al., 2016), several analytical and empirical studies indicate that subjectivity can alleviate dysfunctional subordinate behavior resulting from incomplete performance measures (Baker et al., 1994; Baiman and Rajan, 1995; Cheng and Coyte, 2014).

Difficulties and biases, however, are also associated with subjectivity (Ittner et al., 2003; Ittner and Larcker, 1998; Lipe and Salterio, 2000). Supervisors' evaluations of current performance may be influenced by past performance evaluation (i.e., a memory effect, Woods, 2012), may understate performance differences between employees (i.e., a centrality effect, Moers, 2005; Golman and Bhatia, 2012), and may be influenced by the level of objective performance measures (i.e., a spill-over effect, Bol and Smith, 2011). We investigate the latter factor in this study.

Bol and Smith (2011) show that the level of performance on an objective measure can affect the subjective evaluation of performance on an unrelated measure. Literature on cognitive distortion explains spill-over such that the information evaluated earlier leads to an evaluative disposition. This disposition influences the processing of subsequent information (Bond et al., 2007). Specifically, when performance on an objective measure is low, individuals' subjective evaluation of information of an unrelated subjective measure is lower than when performance on an objective measure is high. Bol and Smith (2011) further show that this influence depends on the objective measure's controllability, i.e. the covariation of the measure with the subordinate's action (Banker and Datar, 1989). If a subordinate's controllability is low, the level of the objective measure does not influence the supervisor's evaluation of the subjective measure as much than when controllability is high. We extend this research by exploring whether the spill-over bias also depends on supervisors' decision modes. The potential reduction of the spill-over bias based on the supervisors' decision mode provides important guidance to organizations seeking to develop interventions to reduce this type of bias independent from environmental factors such as controllability.

2.2. Two modes of cognitive information processing

There is a broad body of literature which proposes two modes of cognitive information processing for decision-making (Kahneman and Frederick, 2005; Evans, 2008; Evans et al., 2013). Despite inconsistent conceptualization of dual process concepts, Kahneman and Frederick (2002 p. 51) acknowledge their similarities: "dual-process models come in many flavors, but all distinguish cognitive operations that are quick and associative from others that are slow and rule-governed."

The many flavors of different processing modes are also apparent in accounting research in which two different processing modes are

Table 1
Intuitive Decision Mode and Deliberate Decision Mode Comparison based on Horstmann et al. (2009).

Intuitive Mode	Deliberate Mode
Fast	Slow
Effortless	Effortful
High Capacity	Limited Capacity
Unconscious	Accessible to conscious awareness
Parallel	Sequential
Lower Fixation Count	**Higher Fixation Count**

Note: This table summarizes the qualifiers used by Horstmann et al. (2009) to describe cognitive processing that adopts two underlying decision modes: intuitive decision mode and deliberate decision mode. Horstmann et al. (2009) show that these two modes can be associated with the eye tracking measure *fixation count*. They evidence that a higher number of fixations indicate a higher degree of deliberation, while a lower number of fixations indicate a lower degree of deliberation. Thus, the extent of deliberation increases with a higher number of fixations. Other dual process models used to describe cognitive processing can be found in Evans (2006), Evans (2008), Kahneman (2011), Stanovich and Toplak (2012), Evans and Stanovich (2013).

commonly distinguished (e.g. Shields, 1980: non-compensatory *versus* compensatory; Chenhall and Morris, 1991: intuitive *versus* sensational; Bailey et al., 2011: anchoring approach *versus* integrative approach; Farrell et al., 2014: System 1 *versus* System 2). For instance, Shields (1980) investigates the effect of information load on managers' search behavior in performance evaluation. He finds increased use of search strategies consistent with non-compensatory decision models (e.g. combination, lexicographic or elimination-by-aspect models) when information load, in terms of alternatives (number of managers evaluated), is high. This rise in non-compensatory decision strategies is not observed when the number of attributes (number of performance measures per manager) is increased.

We extend this stream of literature by introducing the typology of intuition *versus* deliberation used by Horstmann et al. (2009). In this typology, intuition refers to less effortful cognitive information processing that an individual may use in judgment and decision-making while deliberation refers to a decision mode that involves relatively more effortful cognitive processing. As shown in Table 1, characteristics of the intuitive *versus* deliberation mode include fast, effortless, high capacity, unconscious and parallel for the former and slow, effortful, limited capacity, accessible to conscious awareness and sequential for the latter (Horstmann et al., 2009). An advantage of this typology is that it is associated with a more granular analysis of cognitive processing proxied by the eye tracking measure of fixation count, a measurement that can be captured while individuals process relevant information.

2.3. Spill-over effect and intuition versus deliberation

A more effortful decision mode, such as a deliberate decision mode, can improve judgment (Kennedy, 1993; Kahneman and Frederick 2002, 2005; Kahneman, 2011). A rather effortless decision mode, such as an intuitive decision mode, can be subject to bias. This is because by applying rules of thumb, heuristics and by accessing mental representations in fast ways, in the formation of judgments, some context-specific information cues might not be taken into consideration or might be influenced by the decision maker's prior knowledge about the decision-making context. The spill-over effect (Bol and Smith, 2011) can be seen as a bias because there should *not* be an influence of the performance level of an objective measure on a subjective measure when measures are meant to be independent. This leads us to the question of whether the spill-over bias can be mitigated by a more effortful, deliberate decision mode.

Based on a cross-disciplinary literature review, Kahneman (2011) suggests that not all biases can be reduced by cognitive processing associated with more effort. In auditing, Kennedy (1993) suggests that some judgment biases are more effort-related than others and that these biases can be mitigated by more effortful decision modes. Spill-over has not been associated with effort as a driving factor, but rather with an evaluative disposition. This may be partly a result of prior research using a noisy proxy for effort (e.g. total time, Bol and Smith, 2011). It raises the question of whether the influence of effort on spill-over can be shown using a more precise measure of effort.

Further, Evans (2008, p. 261) contends that faster and less effortful processes "rapidly contextualize problems with prior knowledge and belief". Thus, cognitive processing characterized by a larger extent of intuition (less effort) increases the likelihood that the evaluation of subjective information is rapidly contextualized with prior knowledge about the level of the objective measure. As such, the evaluative disposition resulting from the level of the objective measure is more likely to have a stronger influence on judgment when cognitive processing tends to be of a more intuitive decision mode than a deliberate decision mode that includes slower and more effortful processing.

Taking the perspective of a supervisor, in a deliberate mode a supervisor is more likely to spend the time to think about what dimensions of a subordinate's job are captured by the performance measures. In such a mode, a supervisor is more likely to take the time to assess a subordinate's performance on a second job dimension independent from a first job dimension thus using the second measure more independently. Hence, a supervisor is more likely to recognize that objective and subjective performance measures capture distinct job dimensions and thus their measures are less likely to be associated. Therefore, a less distorted evaluation is made when more deliberate rather than more intuitive processing is involved during the judgment formation process.

In other words, our expectation is that the spill-over is larger when information processing is more consistent with an intuitive decision mode, whereas it is smaller when information processing is more consistent with a deliberate decision mode. Thus, subjective performance evaluation may be considered as an interactive function of performance on an objective measure and the decision mode. Stated more formally:

Hypothesis 1. The difference in subjective performance evaluation between high performance on an objective measure and low performance on an objective measure is larger in an intuitive decision mode than in a deliberate decision mode.

3. Method

Participants were asked to assume the role of a regional director and complete a performance evaluation task for a hypothetical industrial pipe and fitting company. The experimental task was adapted from Bol and Smith (2011). We used students as proxies for managers in the context of our decision-making task consistent with prior literature (Ashton and Kramer, 1980; Guala, 2005; Arnold and Triki, 2018). One hundred and twenty-three second year undergraduate students with accounting majors took part in the study on a voluntary basis. Students were recruited from a department research subject pool and received two credit points for taking part in the study. For those students who chose not to take part in the study, an alternative task was made available for equivalent credit points. As we describe in our control analysis, the relative differences in our results are consistent with the results of Bol and Smith (2011) who use managers with an average of 13 years of supervisory experience. Our study received ethics clearance from the ethics committee of the university where the study was conducted.

3.1. Experimental design and eye tracking device

The analysis of the experiment is based on a between-subjects design involving two conditions, i.e., high and low performance on an objective measure. Participants were randomly assigned to one of the two conditions. The extent of deliberation (intuition versus deliberation) was measured using eye trackers. As explained later in more detail, consistent with Horstmann et al. (2009), the eye tracking measure fixation count was used to measure extent of deliberation.

Eye tracking is an established method in visual processing and psychology research (Rayner, 1998), in website usability analysis (Lorigo et al., 2008) and marketing (Wedel and Pieters, 2006). In accounting, the use of eye tracking devices is a recent methodological innovation (Rotaru et al., 2018). Thus, the literature suggests that neurophysiological devices, including eye trackers, may allow us to open the 'black box' and to get insights into cognitive processing and its effects (Birnberg and Ganguly, 2012). There is emerging evidence of the use of eye tracking technology in accounting (Grigg and Griffin, 2013; Chen et al., 2016). In the following section and in Appendix A, we provide information about the parameters used in our study in order to allow comparability with other studies.

We used a Tobii T120 eye tracking system to record eye movements while participants were processing the subjective information. The Tobii T120 eye tracking system uses infrared corneal and pupil reflection to follow the eyes on screen (Tobii Technology AB, 2010). The camera is built into the rim of a 17-in. TFT monitor in order not to distract subjects. Recordings are taken with a frequency of 120 Hz. Chin rests preventing participants head movements are not necessary with this technology as slow head motions do not significantly affect the quality of the measurements taken. Thus, the experimental set up resembles a situation where the respondents use their own computer without feeling constrained in their movements.

At the beginning of the experiment the eye tracker is calibrated using a nine-point fixation technique in order to adjust for participants' individual differences in seating position or their eye characteristics. Thus, despite the non-intrusiveness of the technology, participants are aware that their eyes are tracked.

3.2. Procedure

Participants completed the experiment in a behavioural laboratory of a large university. They were seated individually in carrels in front of monitors of eye tracking devices. After the calibration of the eye tracker the case was presented.

In the case, each participant was asked to assume a hypothetical situation in which she worked as a regional director in an industrial pipe and fitting company. The regional director had supervisory authority over sales managers. The sales managers in turn supervised a number of sales associates and office staff. An organizational chart depicted the organizational hierarchy and sales managers' work responsibilities were described.

Subsequently, the instructions noted that one of the duties of the regional director was to perform semi-annual performance evaluations for each supervised sales manager. The performance evaluation contained two measures: an objective measure of individual sales and a subjective measure of the sales manager's office administration performance. Both measures were scored on a 0–10 scale and the average yielded the overall performance score. Participants were told that the two measures were independent because they captured different job requirements. Participants were asked to conduct the performance evaluation of the office administration of one of the sales managers. Information on the objective measure was provided based on dollar sales figures associated with the sales score (0–10). Sales of the sales manager to be evaluated were manipulated to be either 2 (low) or 9 (high).

Regarding the information on the subjective measure, participants were given statements by the sales manager's staff. The statements were about the sales manager's office performance (in the results referred to as 'subjective information' or 'interviews'). The interviews were designed to depict an average level of office administration performance and were the same as in Bol and Smith (2011).[2]

After having read the case information, participants were prompted to evaluate the office administration performance on a scale from 0 to10.

3.3. Independent variables

3.3.1. Level of objective measure

The first independent variable is the level of the objective measure. The objective measure is the subordinate's individual sales for the most recent six months. The score relates to a 0–10 scale with each score linked to a particular range of sales. The level of the subordinate's objective measure is manipulated on this scale by either being low (2 out of 10) or high (9 out of 10).

3.3.2. Intuitive and deliberate decision mode and eye tracking measures

The second independent variable is the participant's *extent of deliberation* (intuitive *versus* deliberate processing mode) as a measured independent variable proxied by *fixation count* (obtained using an eye tracker device). Literature supports that people differ in the extent to which they process information and to which they are inclined to analytically reflect on a task (Frederick, 2005; Evans, 2006). As noted above, Horstmann et al. (2009) describe cognitive processing using two underlying decision modes: intuition *versus* deliberation. They empirically validate fixation count as a measure of processing mode with higher number of fixations indicating a higher degree of deliberation and a lower number of fixations indicating a lower degree of deliberation in individuals' cognitive processing.

Consistent with prior research, Horstmann et al. (2009) expect that gaze data, being the product of neurophysiological activity of the brain and dynamic eye-brain communication, can be used as a proxy for people's effort and information processing (Kahneman 1974, 2011; Van Gompel et al., 2007). Thereby, Horstmann et al. (2009) found evidence that cognitive processing is rather consistent with integrated models (Kahneman and Frederick, 2002; Evans, 2006) as opposed to distinct-process models (Sloman, 1996). This implies that the decision modes can be proxied using continuous measures. As shown in Table 1, Horstmann et al. (2009) describe the intuitive decision and the deliberate decision mode using a variety of qualifiers. At the measurement level, a higher number of fixations indicates a rather deliberate decision mode and a lower number of fixations a rather intuitive decision mode.

A fixation occurs when the eyes stay focused on a single location longer than a particular time threshold. We use Tobii Studio 3.2.1 software to analyze the eye tracking data. The IV-T fixation filter is adopted as part of Tobii Studio's global settings. We set key parameters of the IV-T Tobii filter to 'define fixations'. In our experiment, a fixation is recorded when participants fixate at least 60 ms on a particular spot on the screen. We set this threshold relatively low as short fixations are commonly present during reading and our key area of interest contains

[2] Before presenting the statements of the interviews, participants were asked whether they wanted to conduct interviews and receive the information in form of the statements. Participants could choose whether to conduct interviews or not. We included this choice as a potential measured independent variable for our study. After this, participants were told that their preference had been understood, but due to a newly introduced policy, the supervisor was still required to conduct interviews. We chose not to include this variable as only seven participants out of 123 chose not to conduct interviews. Including or excluding them did not statistically impact on our analysis. Along with the question whether or not to conduct interviews, participants were presented with personal notes about the sales manager they had taken during the last six months. In the personal notes another manipulation was performed by giving participants information on whether the sales manager agreed or disagreed with the supervisor's budgeting approach. This manipulation turned out to be too weak and only 73 out of 123 participants passed the manipulation check at the end of the instrument. As the manipulation was randomized, we dropped the analysis of this manipulation and analyzed the data only with respect to the manipulation of the objective measure. Including this manipulation as an independent variable did not change the significance levels of the existing factors in the model of Table 3 and was insignificant by itself (p = .694).

textual information (Rayner, 1998; Radach et al., 2008; Holmqvist et al., 2011). Sixty milliseconds is often set as a minimum threshold for classifying a single fixation (see technical appendix in Appendix A).[3]

Our *fixation count* measure is the sum of all fixations recorded while participants process the subjective information. It is used as a measure for *extent of deliberation* that is hypothesized to influence spill-over. We use the interview statements by staff members of the sales manager as a vehicle to display the information relating to the subjective measure. The statements are displayed on one screen within the case material. Participants could review this screen only once, but for as long as they wanted to. After clicking on the 'continue' button they could not go back to the statements. We define our area of interest as a rectangle covering the statements' text plus side/bottom/top margins of 30–70 pixels from the text. The side margins vary as the text is aligned left and the lines are not equally long. The size of the area of interest is 783 × 725 pixel.

Single fixations vary in length (Velichkovsky, 1999). Thus, another commonly used eye tracking measure is the sum of the length of all single fixations in an area of interest, commonly referred to as *(total) fixation duration* (in sec). In additional analysis we also use this measure. Further, since both fixation count and total fixation duration are related to the time individuals spend processing the area of interest, we may readily see an association with the overall time individuals spent on the screen containing the area of interest. This simple time measure has the appeal of being readily available from the computerized survey tool we use. Thus, in supplementary analysis we also use *time spent until page submit* as an additional measure.

3.4. Dependent variable

Sales managers' office administration performance (subjective measure) is our primary dependent variable. Participants were asked to evaluate the office administration performance on a scale from 0 to 10. The dependent variable is called *Subjective Performance Evaluation*.

4. Results

4.1. Manipulation checks and control analysis

The total number of participants in our experiment was 123. One participant had to be eliminated as we did not receive eye tracking data for this participant. This resulted in 122 usable responses. Thirteen subjects of the 122 participants failed the manipulation check. We conducted our analysis both including and excluding subjects who failed the manipulation check and our results did not change. We thus report all subsequent analysis based on the full sample.

The mean subjective performance evaluation is 4.85 for the low objective measure condition, whereas it is 6.32 for the high objective measure condition (refer to Table 2). This difference is significant ($df = 120$, $t = 4.45$, $p < .01$). This difference is consistent with the spill-over effect reported by Bol and Smith (2011). Bol and Smith (2011) reported subjective performance levels of 5.42 for the low objective measure condition and 6.90 for the high objective measure condition.

Further, we test whether the manipulated independent variable (level of objective measure) is associated with our observed independent variable (extent of deliberation). As shown in Table 2, the means reflecting extent of deliberation (fixation count) for the high (low) objective performance measure condition are 270.47 (246.89). The means are not statistically different ($df = 120$, $t = 0.947$, $p > .34$). Levene's test does not indicate differences in variances between the groups ($df1 = 1$, $df2 = 120$, $F = 0.894$, $p > .34$). Thus, the objective measure is not associated with information processing modes and can be used to examine the moderating effect of decision modes (intuitive vs. deliberate).

Table 2
Descriptive Statistics.

Objective Performance Measure		Subjective Performance Evaluation	Extent of Deliberation (Fixation Count)
Low	Mean	4.85	246.89
	N	62	62
	Std. Dev.	1.86	144.48
High	Mean	6.32	270.47
	N	60	60
	Std. Dev.	1.76	129.97
Total	Mean	5.57	258.48
	N	122	122
	Std. Dev.	1.95	137.48

Note: The objective performance measure is the manipulated variable with two conditions: low and high. The subjective performance evaluation is based on participants' evaluations on a scale from 0 to 10. Extent of deliberation is measured by the total fixation count observed on the area of interest.

4.2. Hypothesis testing

To test our hypothesis, we perform a linear regression analysis using subjective performance evaluation as the dependent variable and the level of the objective measure, degree of deliberation and their product term as independent variables. The product term is used to test the significance of the interaction. Significance relating to our hypothesis is evaluated one-tailed due to the directional nature of our prediction in our hypothesis indicating convergence and not divergence (see e.g. Darlington and Hayes, 2017; Elliott et al., 2015, 2017). We find evidence of a significant interaction as proposed in our hypothesis (Table 3, $t = -1.671$, $p < .05$). As such, the level of the objective measure and the extent of deliberation interactively influence subjective performance evaluation. The more cognitive processes are deliberate, the more the subjective performance evaluation between the treatment conditions converge, i.e. the smaller the spill-over effect gets.

Thus, consistent with our prediction, in a rather intuitive decision mode the subjective evaluation is more influenced by prior knowledge about the level of performance of the objective measure. As such, in the evaluation process the subjective information is more rapidly contextualized with the level of the objective measure and thus more distorted in a rather intuitive decision mode compared to a decision mode with a higher degree of deliberation, supporting our hypothesis.

We conduct further analysis to explore the nature of the interaction. Fig. 1 graphically depicts regression lines per level of the objective measure condition. The graph shows subjective performance evaluation as a function of extent of deliberation per treatment condition. To chart

Table 3
Summary Linear Regression on Subjective Performance Evaluation.

	B	t	Sig.
(Constant)	4.752	10.447	0.000
O – Objective performance measure	2.539	3.598	0.000
D – Extent of Deliberation (Fixation Count)	0.000	0.261	0.794
Interaction O x D	−0.004	−1.671	0.049

Note: Dependent Variable: Subjective performance evaluation; Independent Variables: Objective performance measure (0 = Low, 1 = High), extent of deliberation (measured by the total number of fixations on the subjective information), Interaction O x D; R square = 0.170; Significance relating to our hypothesis is evaluated one-tailed because of the directed prediction in our hypothesis indicating convergence and not divergence.

[3] In the analysis of the data, we had to make several choices associated with the optimal use of the eye tracking technology in the context of the given research design. Please refer to the technical appendix in Appendix A for a more in-depth discussion regarding the fixation filter used.

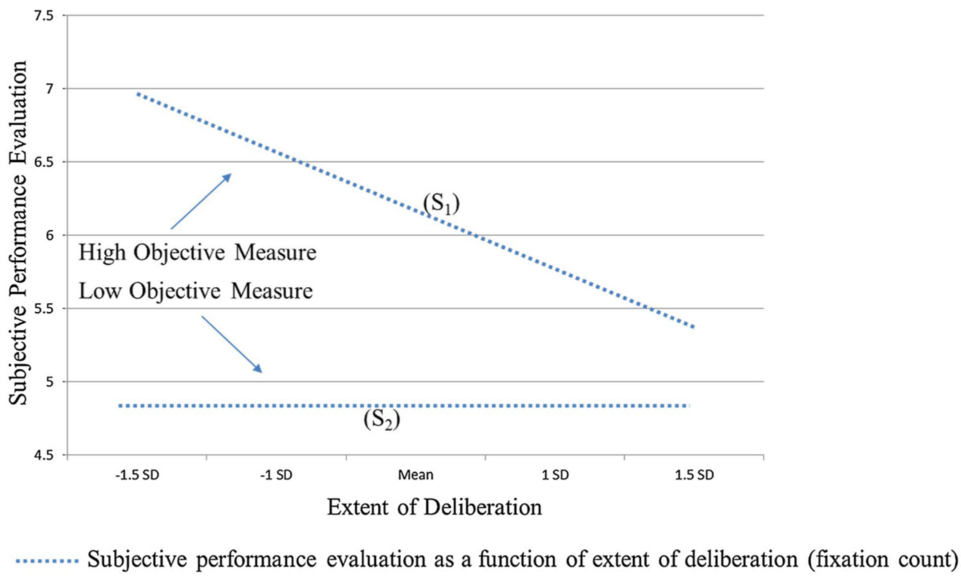

Fig. 1. Influence of Extent of Deliberation (Fixation Count) on Subjective Performance Evaluation.
Note: Dependent variable (y-axis): Subjective Performance Evaluation (based on participants' evaluations on a scale from 0 to 10). Independent variable (x-axis): Extent of deliberation measured by number of fixations on the subjective information. For (S_1) the sample in the objective measure treatment 'high' is used to estimate a linear regression function. For (S_2) the sample in the objective measure treatment 'low' is used to estimate a linear regression function.

the graph, two separate regression models were analyzed using subjective performance evaluation as the dependent variable and extent of deliberation as the independent variable. The first regression model (1) was run with the sample in the objective measure treatment 'high'. The second regression model (2) was run with the sample in the objective measure treatment 'low'.

The resulting regression terms are:

$$S_1 = 7.29 + (-.004) * D_1 \quad (1)$$

$$S_2 = 4.75 + (.000) * D_2 \quad (2)$$

where S = Subjective performance evaluation; D = Extent of Deliberation proxied by fixation count.

S as a function of D for both samples (high and low) are illustrated in Fig. 1. Only for the treatment involving 'high' level of objective measure, the B coefficient (−0.004) is significant (t = −2.101, p = .04), while it is not significant for the treatment involving 'low' level of objective measure (B = 0.000, t = 025, p = .80).[4] This indicates that the interaction is driven by participants in the 'high' condition, such that information processing consistent with a deliberate decision mode reduces participants' subjective performance evaluation. Our hypothesis predicts a convergence. This convergence is observable in Fig. 1. Yet, our theory does not lead us to predict that this convergence is asymmetric. In subsequent analysis, we explore this interaction further using our additional measures.

4.3. Supplementary analysis

As described above, Horstmann et al. (2009) show a link between the number of fixations (fixation count) and extent of deliberation. For this reason we adopted fixation count as our primary measure. Horstmann et al. (2009) do not report analysis regarding a link between (total) fixation duration and extent of deliberation.[5] Yet, total fixation duration is another widely used eye tracking measure. It is derived by summing up the duration of all individual fixations. Durations of fixations can vary, with short fixations typically classified < 150 ms, medium fixations ≥ 150 ms and < 500 ms, and long fixations ≥ 500 ms (Velichkovsky, 1999).

Since the total fixation duration measure contains the number of fixations, these two measures are significantly correlated (r = 0.956, p < .01) as shown in the correlation matrix in Table 4. Further, as both measures, total fixation count and total fixation duration, are associated with time, the time individuals spend on the page containing the area of interest should be associated with the two measures. We captured the time participants spent on the page that contained the interview statements using the survey tool. As shown in Table 4, the measure 'time until page submit' is also significantly correlated with total fixation count (r = 0.253, p < .01) and total fixation duration (r = 0.199, p < .05).

Consequently, we test whether these additional two measures (total fixation duration and time until page submit), used as proxies for the extent of deliberation, provide similar results for our hypothesis. Linear regression analysis using the same model as in Table 3 shows a significant interaction between the objective measure and total fixation duration (t = −1.798, p < .038, one-tailed) as well as between the objective measure and time until page submit (t = −2.043, p < .023, one-tailed). Thus, all three measures – fixation count, total fixation duration, and time spent until page submit – support our moderation hypothesis.

In order to analyze the nature of the interaction, we again ran two regression models per measure. Regression models (3) and (5) were calculated using the sample in the objective measure treatment 'high'. Regression models (4) and (6) were calculated using the sample in the objective measure treatment 'low'.

$$S_3 = 7.33 + (-.021) * FD_1 \quad (3)$$

$$S_4 = 4.91 + (-.001) * FD_2 \quad (4)$$

where S = Subjective performance evaluation; FD = Total Fixation Duration.

$$S_5 = 7.03 + (-.014) * TP_1 \quad (5)$$

$$S_6 = 3.88 + (.018) * TP_2 \quad (6)$$

where S = Subjective performance evaluation; TP = Time until page submit.

Models (3) and (4) using fixation duration are similar to models (1) and (2) using fixation count. As in model (1), in model (3) using the

[4] Note, we report B coefficients and not Betas because we use the B coefficients for visualization purposes in Figs. 1 and 2.

[5] While Horstmann et al. (2009) do not report the analysis of total fixation duration, they report the analysis of mean fixation duration that relates to the average duration of single fixations. They do not find an association between the instruction to adopt a particular decision mode and the average duration of single fixations. Since they find a difference in the number of fixations and no difference in average duration of single fixations, the total fixation duration (which is measured as the product of the number of fixations and the average duration of single fixations) should logically be similar to the results based on the number of fixations in their study (as it is for our study).

Table 4
Correlation Matrix (Pearson).

		Subjective Performance Evaluation	Extent of Deliberation (Fixation Count)	Fixation Duration	Time until Page Submit
Subjective Performance Evaluation	R	1			
	Sig.				
Extent of Deliberation (Fixation Count)	R	−0.062	1		
	Sig.	0.498			
Extent of Deliberation (Fixation Duration)	R	−0.128	0.956**	1	
	Sig.	0.162	0.000		
Extent of Deliberation (Time until page submit)	R	−0.004	0.253**	0.199*	1
	Sig	0.965	0.005	0.028	

Note: N = 122 **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed). The subjective performance evaluation is based on participants' evaluations on a score from 0 to 10. Extent of deliberation (fixation count) is measured by the total fixation count observed on the area of interest. Fixation duration is the sum of all fixations observed on the area of interest. Time until page submit is the time participants spend on the page that contained the area of interest.

Table 5
Summary of Results (Regression Models, Measures and T-Tests).

Main Effect Regression Models	Subsample (High or Low Objective Measure Treatment)	Proxy	T-tests of B Coefficients	T-tests of Interaction Terms of Regression Models (Full Sample that includes both Treatments)
(1) $S_1 = 7.29 + (-0.004) \times D_1$	High	Fixation count	t = −2.101, p = .04*	t = −1.671, 0.049*
(2) $S_2 = 4.75 + (.000) \times D_2$	Low	Fixation count	t = 0.25, p = .80	
(3) $S_3 = 7.33 + (-0.021) \times FD_1$	High	Total Fixation Duration	t = −2.724, p = .008**	t = −1.798, p < 0.038*
(4) $S_4 = 4.91 + (-0.001) \times FD_2$	Low	Total Fixation Duration	t = −0.165, p = .87	
(5) $S_5 = 7.03 + (-0.014) \times TP_1$	High	Time until page submit	t = −1.255, p = .22	t = −2.043, p < 0.023*
(6) $S_6 = 3.88 + (0.018) \times TP_2$	Low	Time until page submit	t = 1.670, p = .10	

where *S* = Subjective performance evaluation; *D* = Extent of Deliberation proxied by fixation count; *FD* = Total Fixation Duration; *TP* = Time until page submit; The *t*-tests of the interaction terms of the regression models follow the model as shown in Table 3; Note, we report B coefficients and not Betas because we use the B coefficients for visualization purposes in Fig. 1 and Fig. 2. * p < 0.05, ** p < .01.

'high' treatment sample, the B coefficient (−0.021) is significant (t = −2.724, p = .008). As in model (2), in model (4) using the 'low' treatment sample, the B coefficient (−0.001) is not significant (B = −0.001, t = −0.165, p = .87). This is consistent with the results using fixation count in that the interaction is driven by participants in the 'high' condition and that information processing consistent with a deliberate decision mode reduces participants' subjective performance evaluation in the 'high' condition.

This notion is not supported by the time until page submit measure, in that the B coefficients for both model (5) (B = −0.014, t = −1.255, p = .22) and model (6) (B = 0.018, t = 1.670, p = .10) are not significant. Table 5 summarizes the regression models and *t*-tests using the different measures.

Fig. 2 depicts all six (1–6) regression lines. Generally, all regression lines support our prediction because they are generally converging with higher values of the independent measures reflecting extent of deliberation. Interestingly, the only regression line with an upward trend in the 'low' treatment (despite not being significant) was derived using the time measure. Since our theory did not lead us to predict an asymmetric interaction, we may conclude that all three measures provide support for our interaction. On the other hand, we may also argue that the simple time measure may be less sensitive when detecting a particular nature of interaction because of the non-significance of the regression coefficients per treatment condition. Yet, the latter interpretation is very tentative and further research is necessary.

Bol and Smith (2011) also capture a time measure in their study based on the overall time participants took to complete the experiment. This time measure is different to the one above as it reflects the time taken to complete the entire task and not, as in our study, the time participants spent on the subjective information. In non-tabulated control analysis, Bol and Smith (2011) report that they do not find evidence that would suggest that the amount of time participants take to complete the experiment interacts with the level of the objective performance measure. We argue that this is likely due to the overall time measure being too noisy and not reflecting precisely enough the processing of the subjective information. Running the model in Table 3 using a time measure that captures the time participants took to complete our experiment (similar to Bol and Smith, 2011) does not result in a significant interaction term in our study either (t = 1.441, p = .152).

5. Discussion

Our study contributes to prior literature in subjective performance evaluation by investigating the influence of decision modes in subjective performance evaluation using the spill-over effect as a setting (Bol and Smith, 2011; Ruggeri, 2012). There are many settings in which supervisors have discretion to apply subjective judgment when evaluating performance that is difficult to be captured with objective measures. However, if supervisors are influenced by cognitive biases such as spill-over, their subjective evaluation is likely to be distorted. This can have important negative consequences for employee motivation, compensation or promotion (Bol, 2008; Bol and Smith, 2011). Thus, it is important to examine factors that mitigate such biases.

Our results confirm the presence of the spill-over effect in subjective performance evaluation, such that supervisors' subjective performance evaluation is influenced in the direction of their subordinates' objective performance levels. More importantly, we find that the spill-over bias is reduced when individuals process subjective information using a more deliberate decision mode as compared to a more intuitive decision mode. Thus, the size of the spill-over bias may not be universal across different decision modes. We further provide evidence as to which measures have the potential to capture cognitive processes adequately to show the theorized moderation.

In prior literature the spill-over bias has been attributed to an

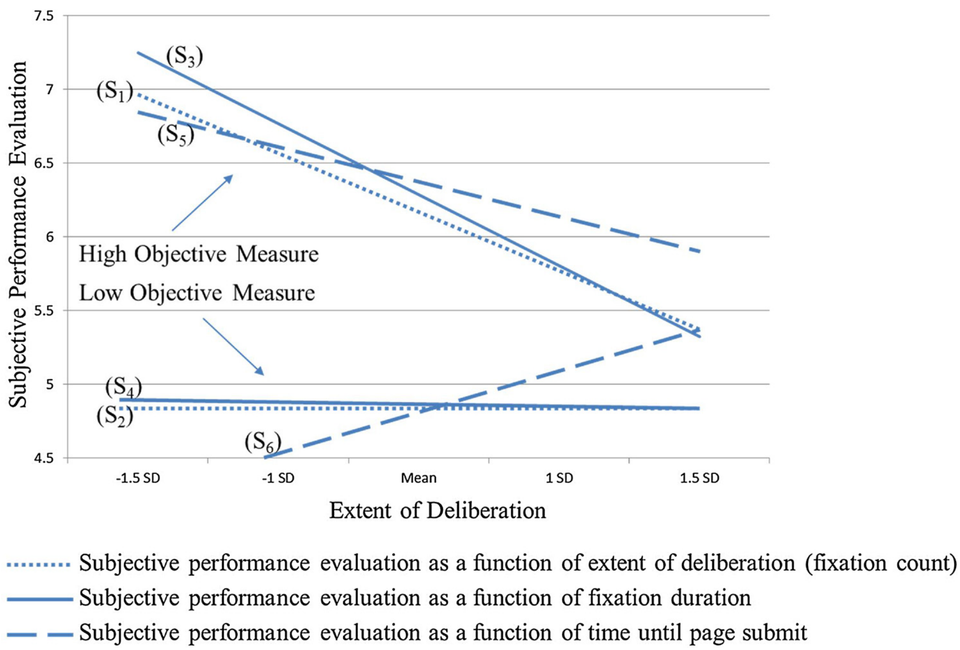

Fig. 2. Influence of Extent of Deliberation on Subjective Performance Evaluation (Comparison with Additional Measures).
Note: Dependent variable (y-axis): Subjective Performance Evaluation (based on participants' evaluations on a scale from 0 to 10). Independent variable (x-axis): Extent of deliberation measured by the number of fixations on the subjective information for (S_1) and (S_2), total fixation duration for (S_3) and (S_4) and time until page submit for (S_5) and (S_6). For (S_1), (S_3) and (S_5) the sample in the objective measure treatment 'high' is used to estimate linear regression functions. For (S_2), (S_4) and (S_6) the sample in the objective measure treatment 'low' is used to estimate linear regression functions.

evaluative disposition formed by managers when receiving information about a subordinate's performance on an objective measure early in an evaluation. Thus, knowledge available to an individual prior to attending to relevant information seems to drive this effect. In a more deliberate decision mode, prior knowledge is more effectively updated. Thus, this finding may also generalize to other biases such as the memory bias (Woods, 2012) or the 'halo effect' (O'Donnell and Schultz, 2005).

As described in Woods (2012), it has been repeatedly found that past performance evaluation often influences current performance evaluation; however, it is often unclear as to why. Woods argues that this effect is triggered by supervisors correcting measure deficiencies and making current performance evaluation consistent with prior performance, which signals that the current measure is less deficient and more certain. Our findings may also provide a potential way of reducing this bias. In a memory bias, prior performance of a subordinate may also form an evaluative disposition because it pertains to past signals. Thus, in a rather deliberate decision mode such an evaluative disposition stemming from signals in memory may be more effectively updated than in a rather intuitive decision mode. This could be an avenue for future research.

Further, the nature of the spill-over bias discussed and empirically demonstrated in this study shares certain properties with the halo effect reported in the accounting literature (e.g. O'Donnell and Schultz, 2005). The halo effect predicts a spill-over of the overall impression about a company or person on specific attributes of the company or person under investigation. The nature of the spill-over in our study is between two attributes associated with individual performance: subjective and objective performance measures. As such, our results may also generalize to the halo effect. This needs to be examined more closely in future research.

The comparison of fixation count with our additional measures of fixation duration and time participants spent on the relevant page leads us to conclude: first, that fixation count and fixation duration lead to very similar results, and second, that the time measure also shows similar results. In addition, the measure of time participants spent on the overall experiment (as used in Bol and Smith, 2011) appears to be too noisy to confirm results. Generally, this supports the notion that more precise measures including eye tracking measures can yield additional insights into the questions relevant to accounting research and practice, such as subjective performance evaluation.

Considering evidence on constructs that determine judgment performance in the accounting settings (e.g. Libby and Luft, 1993), the notion of decision mode is associated with cognitive effort. The latter, according to Libby and Luft, reflects the degree to which people engage their knowledge and abilities to achieve a certain level of judgment performance. Another determinant suggested by Libby and Luft is the environment. Future research may explore the characteristics of the environment that can be purposefully designed to reduce managers' propensity to cognitive biases when they conduct performance evaluation.

Another potential avenue for future research into subjective performance evaluation would be to investigate the enabling factors which allow an individual to channel or change decision modes and processes. As reported above, our results show that fixation count is independent from the level of the objective measure. Yet this may provide initial insights into how underlying cognitive processes associated with evaluative dispositions may work or may not work. High or low levels in quantified objective performance measures do not seem to trigger different cognitive processing modes. If it is possible for performance evaluation systems to influence supervisors' cognitive processing modes, then evaluations across supervisors could be made more consistent as biases are reduced or eliminated altogether. In order to improve decision-making, policy instructions may not only tell managers what information to consider, but also how information should be considered. Biases in subjective performance evaluation may be reduced by such instructions. However, the question remains whether accounting systems should advise managers on the decision processes to be used and whether such advice would be effective. While it may be argued that for certain types of decisions, including performance evaluation, made by a limited number of people in an organization a consistent approach across decision makers is preferable, additional policy instructions could constrain the flexibility of supervisors.

Our analysis is limited to the extent that our research instrument does not allow for a finer textual analysis disentangling different information cues of the subjective information. Analysis of the information cues of different valence presented to supervisors in the context of subjective performance evaluation could be another avenue for future research. Further, while we provide evidence that our student subjects behave in a similar way to experienced managers, it is clear that our subjects possess far less knowledge, experience, and expertise in performance evaluation. At the same time this may indicate that in real life situations biases are even more likely because managers come with more prior knowledge to evaluating measures that can influence evaluations in unwanted ways. Thus, in real life settings, issues around the distortion of subjective performance evaluation are likely to be further exacerbated through factors like incentives, personal relationships or other confrontations.

Acknowledgements

We thank Jake Birnberg, Mike Shields, Gary Hecht, Annie Farrell, Jasmijn Bol, Steve Smith, Steve Kaplan, Vicky Arnold, Steve Sutton, Alexander Brueggen, Kristy Towry, Kathryn Kadous, Heba Yousef Abdel-Rahim, the participants of the Global Management Accounting Research Symposium (GMARS), Sydney 2014, of the Annual Meeting of the American Accounting Association (AAA), Atlanta 2014, of the Conference of the Accounting and Finance Association of Australia and New Zealand (AFAANZ), Auckland 2014 and of the Monash Department of Accounting Brown Bag Series, for helpful comments and suggestions. We particularly thank Jasmijn Bol and Steve Smith for sharing their instrument. Further, we thank the Accounting and Finance Association of Australia and New Zealand for funding this research and the Monash Business School and Australian Research Council (ARC LIEF grant LE130100112) for funds associated with the equipment used in the Behavioural Laboratory.

Appendix A

Technical Appendix

Neuro-technology, such as eye tracking, allows us to open the 'black box' of human cognition and to get additional insights into how the information used as an input into decision-making is being processed and what its effects are on decision-making outcomes (Dickhaut et al., 2010; Birnberg and Ganguly, 2012; Farrell et al., 2014). By proxying cognitive processing through eye tracking measures, we extend the use of eye tracking as a new method in accounting research (Rotaru et al., 2018) and supplement behavioural survey data with eye tracking data. The use of eye tracking allows us to investigate processing of subjective information in a relatively directed and precise manner. In the following section we describe details of the eye tracking device and parameters used in the study.

To analyze the eye tracking data, we use the Tobii Studio 3.2.1 software package, developed by the producers of Tobii T120 eye tracking devices (Tobii Technology AB, Sweden). The process of fixation identification is supported by mathematical algorithms and statistical analysis approaches available in Tobii Studio 3.2.1 by the means of selecting one of the inbuilt fixation filters. The filters determine how the raw data reflecting eye fixations (e.g. fixation count, duration or location) is going to be analyzed by the software. Therefore, this choice is critical for appropriately capturing the eye movements (Tobii Technology AB, 2012). The filters define the start and the end points of the eye fixations and saccades. We use the IV-T fixation filter (Komogortsev et al., 2010; Tobii Technology AB, 2012), which is one of the most commonly used filters reported in the recent eye tracking literature that uses Tobii technology (e.g. Lappi et al., 2013). The I-VT Filter is based on the eyes' angular velocity and operates on eye movement data. As a result, the data is independent of screen size, screen resolution and the distance between the stimulus and eyes (Tobii Technology AB, 2015).

We set key parameters of the IV-T Tobii filter to define fixations. In our case, a fixation is recorded when participants spend at least 60 ms fixating their eye gaze on a particular spot. In line with Rayner (1998), Radach et al. (2008), Holmqvist et al. (2011), we set this threshold relatively low as short fixations are commonly present during reading, and our key area of interest contains textual information. The average length of a single fixation in our relevant area of interest is 184 ms across all participants. The velocity threshold is set at 30 visual degrees per second. As the eye is never completely still and experiences microsaccades or tremors (Rayner, 1998, Holmqvist et al., 2011) and environmental noise may occur in addition, the threshold is set at a level that is not expected to interfere with the classification of eye fixations and is appropriate given the varying levels of noise (Olsen, 2012; Olsen and Matos, 2012). Above the velocity threshold, an eye movement is classified as a saccade, which indicates the end of a fixation (Salvucci and Goldberg, 2000).

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.mar.2018.03.001.

References

Arnold, V., Triki, A., 2018. Use of student and online participants in behavioural accounting research. In: Libby, T., Thorne, L. (Eds.), The Routledge Companion to Behavioural Accounting Research. Routledge, pp. 253–272.
Ashton, R.H., Kramer, S.S., 1980. Students as surrogates in behavioral accounting research: some evidence? J. Acc. Res. 18 (1), 1–15.
Bailey, W.J., Hecht, G., Towry, K.L., 2011. Dividing the pie: the influence of managerial discretion extent on bonus pool allocation? Contemp. Acc. Res. 28 (5), 1562–1584.
Baiman, S., Rajan, M., 1995. On the usefulness of discretionary bonus schemes? Acc. Rev. 70 (4), 557–579.
Baker, G., Gibbons, R., Murphy, K.J., 1994. Subjective performance measures in optimal incentive contracts? Q. J. Econ. 109 (4), 1125–1156.
Banker, R.D., Datar, S.M., 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation? J. Acc. Res. 27 (1), 21–39.
Birnberg, J.G., Ganguly, A.R., 2012. Is neuroaccounting waiting in the wings? An essay. Acc. Organiz. Soc. 37 (1), 1–13.
Bol, J.C., Smith, S.D., 2011. Spillover effects in subjective performance evaluation: bias and the asymmetric influence of controllability? Acc. Rev. 86 (4), 1213–1230.
Bol, J.C., 2008. Subjectivity in compensation contracting. J. Acc. Lit. 27, 1–24.
Bol, J.C., 2011. The determinants and performance effects of managers' performance evaluation biases? Acc. Rev. 86 (5), 1549–1575.
Bond, S.D., Carlson, K.A., Meloy, M.G., Russo, J.E., Tanner, R.J., 2007. Information distortion in the evaluation of a single option? Organ. Behav. Hum. Decis. Process. 102 (2), 240–254.
Chen, Y., Jermias, J., Panggabean, T., 2016. The role of visual attention in the managerial judgment of balanced-Scorecard performance evaluation: insights from using an eye-Tracking device. J. Acc. Res. 54 (1), 113–146.
Cheng, M.M., Coyte, R., 2014. The effects of incentive subjectivity and strategy communication on knowledge-sharing and extra-role behaviours? Manage. Acc. Res. 25 (2), 119–130.
Chenhall, R., Morris, D., 1991. The effect of cognitive style and sponsorship bias on the treatment of opportunity costs in resource allocation decisions? Acc. Organiz. Soc. 16 (1), 27–46.
Darlington, R.B., Hayes, A.F., 2017. Regression Analysis and Linear Models: Concepts, Applications, and Implementation. Gilford Press, New York.
Dickhaut, J., Basu, S., McCabe, K., Waymire, G., 2010. Neuroaccounting: consilience between the biologically evolved brain and culturally evolved accounting principles? Acc. Horiz. 24 (2), 221–255.
Elliott, W.B., Rennekamp, K.M., White, B.J., 2015. Does concrete language in disclosures increase willingness to invest? Rev. Acc. Stud. 20 (2), 839–865.
Elliott, W.B., Grant, S.M., Rennekamp, K.M., 2017. How disclosure features of corporate social responsibility reports interact with investor numeracy to influence investor judgments? Contemp. Acc. Res. 34 (3), 1596–1621.
Evans, J.S.B.T., Stanovich, K.E., 2013. Dual-process theories of higher cognition: advancing the debate? Persp. Psychol. Sci. 8 (3), 223–241.
Evans, J.S.B.T., 2008. Dual-Processing accounts of reasoning, judgment, and social cognition? Annu. Rev. Psychol. 59 (1), 255–278.
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? Acc. Rev. 89 (6), 1979–2010.
Franco-Santos, M., Lucianetti, L., Bourne, M., 2012. Contemporary performance measurement systems: a review of their consequences and a framework for research? Manage. Acc. Res. 23 (2), 79–119.
Frederick, S., 2005. Cognitive reflection and decision making? J. Econ. Perspect. 19 (4), 25–42.
Gibbs, M., Merchant, K.A., Stede, W.A.V.D., Vargus, M.E., 2004. Determinants and effects of subjectivity in incentives. Acc. Rev. 79 (2), 409–436.
Golman, R., Bhatia, S., 2012. Performance evaluation inflation and compression Accounting. Organiz. Soc. 37 (8), 534–543.
Grigg, L., Griffin, A.L., 2013. A role for eye-tracking research in accounting and financial reporting? In: Horsley, M., Toon, N., Knight, B., Reilly, R. (Eds.), Current Trends in Eye Tracking Research. Springer Science + Business Media, Berlin/Heidelberg, pp. 225–230.
Guala, F., 2005. The Methodology of Experimental Economics. Cambridge University Press, Cambridge, UK.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., van de Weijer, J.,

2011. Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford University Press, Oxford, UK.
Horstmann, N., Ahlgrimm, A., Glöckner, A., 2009. How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes? Judgment Decis. Making 4 (4), 335–354.
Ittner, C.D., Larcker, D.F., 1998. Innovations in performance measurement: trends and research implications. J. Manage. Acc. Res. 10, 203–238.
Ittner, C.D., Larcker, D.F., Meyer, M.W., 2003. Subjectivity and the weighting of performance measures: evidence from a balanced scorecard. Acc. Rev. 78 (3), 725–758.
Kahneman, D., Frederick, S., 2002. Representativeness revisited: attribute substitution in intuitive judgment. In: Gilovich, D., Griffin, T., Kahneman, D. (Eds.), The Psychology of Intuitive Judgment. Cambridge University Press, New York, NW, pp. 49–81.
Kahneman, D., Frederick, S., 2005. A model of heuristic judgment. In: Holyoak, K., Morrison, R.G. (Eds.), The Cambridge Handbook of Thinking and Reasoning. Cambridge Univ. Press, Cambridge, UK, pp. 267–294.
Kahneman, D., 1974. Attention and Effort. Prentice-Hall, Inc., Prentice-Hall, Inc.
Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus, Giroux, New York, NY.
Kennedy, J., 1993. Debiasing audit judgment with accountability: a framework and experimental results. J. Acc. Res. 31 (2), 231.
Komogortsev, O.V., Gobert, D.V., Jayarathna, S., Koh, D.H., Gowda, S.M., 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. IEEE Trans. Biomed. Eng. 57 (11), 2635–2645.
Lappi, O., Pekkanen, J., Itkonen, T.H., 2013. Pursuit eye-movements in curve driving differentiate between future path and tangent point models. PLoS One 8 (7), e68326.
Libby, R., Luft, J., 1993. Determinants of judgment performance in accounting settings: ability, knowledge, motivation, and environment? Acc. Organiz. Soc. 18 (5), 425–450.
Lipe, M.G., Salterio, S.E., 2000. The balanced scorecard: judgmental effects of common and unique performance measures? Acc. Rev. 75 (3), 283–298.
Lorigo, L., Haridasan, M., Brynjarsdóttir, H., Xia, L., Joachims, T., Gay, G., Pan, B., 2008. Eye tracking and online search: lessons learned and challenges ahead? J. Am. Soc. Inf. Sci. Technol. 59 (7), 1041–1052.
Luft, J.L., Shields, M.D., Thomas, T.F., 2016. Additional information in accounting reports: effects on management decisions and subjective performance evaluations under causal ambiguity? Contemp. Acc. Res. 33 (2), 526–550.
Moers, F., 2005. Discretion and bias in performance evaluation: the impact of diversity and subjectivity? Acc. Organiz. Soc. 30 (1), 67–80.
O'Donnell, E., Schultz Jr, J.J., 2005. The halo effect in business risk audits: can strategic risk assessment bias auditor judgment about accounting details? Acc. Rev. 80 (3), 921–939.
Olsen, A., Matos, R., 2012. Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In: Santa Barbara, CA, USA. Proceedings of the Symposium on Eye Tracking Research and Applications – ETRA Vol. 12.
Olsen, A., 2012. Tobii I-VT Fixation Filter – Algorithm Description. tobii.com.
Radach, R., Huestegge, L., Reilly, R., 2008. The role of global top-down factors in local eye-movement control in reading? Psychol. Res. 72 (6), 675–688.
Rajan, M.V., Reichelstein, S., 2009. Objective versus subjective indicators of managerial performance? Acc. Rev. 84 (1), 209–237.
Rayner, K., 1998. Eye movements in reading and information processing: 20 years of research? Psychol. Bull. 124 (3), 372–422.
Rotaru, K., Schulz, A.K.-D., Fehrenbacher, D.D., 2018. New technologies for behavioral accounting experiments. In: Libby, T., Thorne, L. (Eds.), Routledge Companion to Behavioral Accounting Research. Routledge, pp. 253–272.
Ruggeri, D., 2012. Subjective weighting under organisational interdependence conditions: an experimental investigation. In: In: Davila, A., Epstein, M.J., Manzoni, J.F. (Eds.), Studies in Managerial and Financial Accounting 25. Emerald Group Publishing Limited, pp. 301–337.
Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols. Paper Presented at the Proceedings of the Symposium on Eye Tracking Research and Applications – ETRA '00.
Shields, M.D., 1980. Some effects of information load on search patterns used to analyze performance reports? Acc. Rev. 5 (4), 429–442.
Sloman, S.A., 1996. The empirical case for two systems of reasoning? Psychol. Bull. 119 (1), 3–22.
Stanovich, K.E., Toplak, M.E., 2012. Defining features versus incidental correlates of Type 1 and Type 2 processing? Mind Soc. 11 (1), 3–13.
Tobii Technology AB, 2010. Tobii Eye Tracking An Introduction to Eye Tracking and Tobii Eye Trackers. tobii.com.
Tobii Technology AB, 2012. TUser Manual – Tobii Studio Version 3.2. tobii.com.
Tobii Technology AB, 2015. Tobii Statistics. (Accessed from: http://www.tobii.com/de/eye-tracking-research/global/products/software/tobii-studio-analysis-software/statistics/).
Van Gompel, R.P.G., Fischer, M.H., Murray, W.S., Hill, R.L., 2007. Eye-Movement Research: Insights into Mind and Brain. Elsevier BV, The Netherlands (Chapter 1).
Van der Stede, W.A., Chow, C.W., Lin, T.W., 2006. Strategy, choice of performance measures, and performance? Behav. Res. Acc. 18 (1), 185–205.
Velichkovsky, B.M., 1999. From levels of processing to stratification of cognition: converging evidence from three domains of research. In: Challis, B.H., Velichkovsky, B.M. (Eds.), Stratification in Cognition and Consciousness. John Benjamins Publishing Company, Amsterdam, Netherlands, pp. 203–235.
Wedel, M., Pieters, R., 2006. Eye tracking for visual marketing? Found. Trends Market. 1 (4), 231–320.
Woods, A., 2012. Subjective adjustments to objective performance measures: the influence of prior performance? Acc. Organiz. Soc. 37 (6), 403–425.