Performance management systems in modern organizations

Arshad, Farah

Document version:
Publisher's PDF, also known as Version of record

Publication date:
2020

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Performance Management Systems in Modern Organizations

FARAH MAHAM ARSHAD
Performance Management Systems in Modern Organizations

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. K. Sijtsma, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie aan Tilburg University op dinsdag 23 juni 2020 om 13.30 uur door

Farah Maham Arshad

geboren te Karachi, Pakistan
Promotores:
prof. dr. B.C.G. Dierynck, Tilburg University
prof. dr. E. Cardinaels, Tilburg University

Promotiecommissie:
prof. dr. J.J.M. Potters, Tilburg University
prof. dr. V.S. Maas, University of Amsterdam
prof. dr. A. Van den Abbeele, KU Leuven
dr. A. Presslee, University of Waterloo
Acknowledgements

Now that my time as a PhD candidate at Tilburg University is coming to an end, I can look back and see it as one of the most challenging but rewarding journeys of my life. There have been several people who have been instrumental during this journey and who deserve my heartfelt gratitude. However, before doing so, I would like to go back to the start of my PhD journey. In 2014, I came to the Netherlands with plans to pursue a master’s degree and get a job afterwards. While I liked research and teaching, PhD was more of a distant dream that I did not plan to realize yet. However, I was encouraged to apply to the research master’s program by Bart Dierynck, who at the time was my instructor for the Management Control Systems course as well as the second reader for my master’s thesis. I am grateful to Bart for seeing the potential in me and for encouraging me to take that first step towards realizing my PhD dream.

I was also lucky to later end up with Bart as my supervisor during the PhD program. He has been a great mentor and is always available to share his valuable insights with me. One thing that I really appreciated was how Bart was always open for discussion. He provided me with valuable feedback and advice but also allowed me the room to think creatively and independently. Bart also encouraged me to think more broadly about the management accounting literature and practice and how we could use experiments to provide innovative solutions to different problems. I would also describe Bart’s supervision style as empathetic and supportive of work-life balance. As working can easily take over the life as a researcher, Bart was always a strong proponent of downtime and always encouraged me to take a break after all the hard work. I am really grateful for Bart’s mentorship and hope to always keep a strong connection as I progress further in my career.

My special thanks also go to my second supervisor, Eddy Cardinaels. It was during my second year of the research master’s program that I had a proper interaction with Eddy and started working with him on a joint project along with Bart. I consider Eddy as a valuable mentor and have gained tremendous insights from his experience. Given his experience as an editor as well as a well-published author with various projects in a variety of research areas, Eddy was always quick to share references of other studies that might be helpful in my own research. Despite being extremely busy, Eddy was also always open for a meeting to share his knowledge and discuss different ideas. One would think that research discussions are dull and boring; however, Eddy always managed to incorporate a touch of humor during my chats and discussions with him. Eddy’s passion for research reflected prominently in our discussions and
it was something I really appreciated. Eddy has played an important supportive role during my PhD and I am really thankful for his help and support.

I would also to express thanks to my dissertation committee members: Victor Maas, Alexandra van den Abbeele, Adam Presslee and Jan Potters. They have provided me with useful feedback and suggestions, which have greatly helped me in improving my papers. My initial interaction with Victor during my first conference as an accounting researcher in 2017. Since then, I have met and interacted with Victor in many conferences and PhD colloquia and always greatly appreciate his valuable insights and perspectives about experimental management accounting. Victor was also kind enough to offer valuable advice to improve my single-author paper before the start of the job market. His advice on my research projects and the job market has been really beneficial for me. My special thanks also go to Adam for going the extra mile to help me improve my papers. Adam not only provided me with detailed review reports with valuable suggestions for each paper but also sent valuable references along with marked up copies of my papers detailing his thoughts. During all of my interactions with Adam, he has always been really approachable, kind and helpful. Also, I would like to thank Alexandra for always being forthcoming with useful and critical advice. Alexandra was one of the faculty members who provided valuable feedback to management accounting PhDs during the EAA colloquia that I attended in 2019. During the intensive PhD colloquia encompassing several days, I was able to learn a lot and benefitted significantly from her thoughts and feedback. I am also grateful for the helpful comments and suggestions from Jan Potters. Jan was one of my instructors during the experimental economics course during the research master’s program. I was the only student from accounting following that course and really appreciated the fact that Jan was so open to ideas and perspectives from another field. I was able to gain many insights from Jan during the start of my research career and I am really grateful for the learning opportunity.

I would also like to thank my two remaining co-authors, Victor van Pelt and Joel Berge. Victor was a fellow PhD at Tilburg University till last year and has been a great colleague, co-author and friend. I always enjoyed all our long discussions in the PhD office and elsewhere. In fact, these discussions were often so lively that we ended up annoying other PhDs sharing our office and had to shift our discussion to the café or another area. Victor is an excellent researcher with a brilliant mind. I really respect Victor for always being straightforward with his thoughts and advice. His feedback has been helpful to me on numerous occasions. I also fondly remember the coffee walks around the campus with Victor and Bart to discuss and reflect on various research ideas. Even though Victor has recently moved to Germany as a
junior faculty member at WHU, I am happy that we have not lost touch with each other and manage to find time to talk via Skype and WhatsApp. I would also like to thank Joel for being an awesome co-author and friend. I first met Joel during the PhD course by Robert Bloomfield. Joel has an innovative mind and always has interesting ideas based on interesting insights from other domains, especially economics. His passion for research also shines through in all my discussions with him. It was during one of my discussions with Joel at the end of a conference presentation that we came up with an idea that we were fortunate to continue working on during Joel’s visit to Tilburg University in 2019. We shared an office during his visit to Tilburg and I was able to foster a great friendship with him along with his wife and daughter. I am especially thankful to Joel and his family for being great hosts during my job market visit to NHH. I greatly enjoy working with Joel and hope that we will continue to collaborate on various ideas in the future.

My thanks also go to several fellow PhDs over the years at Tilburg University. Even though I was the only one to start the PhD program in accounting in my year, my senior PhDs, Victor van Pelt, Yusiyu Wang and Ties de Kok, were really welcoming, helpful and always willing to share their experience. Even after graduating from the PhD program, all three of them are always willing to lend a sympathetic ear and help out. I am especially grateful to them for their advice regarding the job market. I would also like to say thanks to the PhDs and research master students at the accounting department in Tilburg University with whom I interacted with during the years: Jingwen, Martin, Tim, Ruishen, Christian, Qinnan, Jesse, Mustafa, Qinwei, Cardin, Fabien, and Eva. All of you have made my time at Tilburg more enjoyable and memorable. I would also like to thank PhDs from other departments: Georgi, Pranav, Vilma, Vijaya, and Clemens. It was always nice to hang out with you guys and I am grateful for all our chats and interactions over the years. I am also thankful to Grace, Irina, and Yanjia for being great buddies and friends during the job market.

My gratitude also goes to the accounting department at Tilburg University. There is a strong connection between all the members of the department and everyone from the faculty members to the secretaries is always open and eager to help in any way they can. The supportive and transparent environment at the department has been instrumental in my development. I would especially like to thank Stephan and Philip, for always being willing to impart their wisdom and advice no matter how busy they were. I would also like to thank Robin, Harm, Marti, Ruidi, Lucas, Christoph Sextroh, Christoph Hörner and all the other faculty members for always being ready to provide advice, feedback and suggestions, especially during the job market. I will also miss my pantry chats and lunch talks with the department members. Great
thanks to Sofie, Bob, Khurram, Dirk, and Wim for always being kind and supportive. I would
also like to thank Hetty and Agnes for providing great administrative support throughout my
time as a PhD. I consider myself to be very fortunate to have been a part of the accounting
department at Tilburg University and to have learned from some of the best researchers. I am
also really grateful to CentER for awarding me with the generous Koopmans scholarship for
my research master’s and for funding my research projects during the PhD.

I also wish to thank my family, especially my parents who always believed in me and
gave me their unconditional support. They have always encouraged me to pursue my dreams
and supported me when I wanted to come to Netherlands for higher education. Even though, I
was living far away from them, I could rely on them for support whenever things got stressful.
I have learned some of the best life lessons about hard work, persistence, and kindness from
my parents. Mama and papa, this PhD would not have been possible without your
encouragement and love. I would also like to thank my friends Eveline, Poppy, and Julienne
for their great friendship and for allowing me de-stress from the PhD whenever I needed it.

Life has its ups and downs. The coronavirus crisis in the last few months has changed
our lives tremendously. In these challenging times of lockdowns and isolation, my family,
supervisors, friends, and colleagues have been tremendously supportive and caring. I am really
thankful for that. I hope that we continue to rely on each other, and that humanity recovers
from this difficult situation soon!

Farah Maham Arshad
Tilburg, April 2020
Contents

Chapter 1 Introduction.............................................................................................................................................. 1

Chapter 2 Managers’ Self-Serving Incentives: Information Avoidance in Performance Evaluation................................................................. 9
  2.1 Introduction ....................................................................................................................................................... 11
  2.2 Literature review and hypothesis development ............................................................................................ 16
    2.2.1 Information avoidance under manager self-serving incentives ......................................................... 16
    2.2.2 Self-evaluations by employees .................................................................................................................. 19
  2.3 Experimental design ......................................................................................................................................... 21
    2.3.1 Overview .................................................................................................................................................... 21
    2.3.2 Participants ............................................................................................................................................... 24
    2.3.3 Experimental manipulations and procedure .......................................................................................... 25
  2.4 Results ............................................................................................................................................................ 27
    2.4.1 Descriptive statistics ................................................................................................................................. 27
    2.4.2 Main findings ........................................................................................................................................... 29
    2.4.3 Supplemental analysis .............................................................................................................................. 30
      2.4.3.1 Information avoidance under good luck versus bad luck ............................................................... 30
      2.4.3.2 Time spent looking at information ................................................................................................. 33
      2.4.3.3 Employee performance and intrinsic motivation .............................................................................. 34
  2.5 Summary and discussion .............................................................................................................................. 37
  2.6 References .................................................................................................................................................... 42

Chapter 3 Demand-driven Feedback Systems, Recordkeeping and Easy Task Prioritization .............................................................................. 55
  3.1 Introduction .................................................................................................................................................... 57
  3.2 Literature review and hypothesis development ........................................................................................... 62
    3.2.1 Easy task prioritization in demand-driven feedback systems ............................................................. 62
    3.2.2 Recordkeeping and easy task prioritization in demand-driven feedback systems ......................... 64
    3.2.3 Planning, dynamic sequencing and easy task prioritization in demand-driven feedback systems .... 65
  3.3 Main experiment ............................................................................................................................................ 67
    3.3.1 Overview ................................................................................................................................................. 67
    3.3.2 Participants .............................................................................................................................................. 69
    3.3.3 Experimental manipulations and procedures ....................................................................................... 69
  3.4 Results of main experiment ........................................................................................................................ 71
3.4.1 Descriptive statistics ........................................................................................................71
3.4.2 Main findings ......................................................................................................................71
3.4.3 Supplemental analysis ........................................................................................................72
  3.4.3.1 Performance effects .......................................................................................................72
  3.4.3.2 Mindfulness and Recordkeeping ......................................................................................72
  3.4.3.3 Demand for feedback ......................................................................................................73
3.5 Extension of main experiment ...............................................................................................74
  3.5.1 Overview ..........................................................................................................................74
  3.5.2 Participants .......................................................................................................................75
  3.5.3 Experimental manipulations and procedures .................................................................75
  3.5.4 Results .............................................................................................................................77
    3.5.4.3 Perceived choice and planning .....................................................................................78
3.6 Summary and discussion .........................................................................................................78
3.7 References ..............................................................................................................................81

Chapter 4 Facing A Calibration Committee: The Impact on Costly Information Collection and Subjective Performance Evaluation ......................................................... 99
4.1 Introduction ..............................................................................................................................101
4.2 Literature review and hypothesis development ......................................................................106
  4.2.1 Calibration committee and subjective performance evaluation .....................................106
  4.2.2 Information collection in the calibration committee: ....................................................109
  4.2.3 Presence of a HR-manager in the calibration committee: .............................................111
4.3 Experimental design .............................................................................................................113
  4.3.1 Overview .........................................................................................................................113
  4.3.2 Performance evaluation and costly information collection ............................................114
  4.3.3 Final performance evaluations and experimental manipulations ................................116
  4.3.4 Participants and experimental procedures .....................................................................118
  4.3.5 Differentiation between subjective performance evaluations of employees: ............119
4.4 Results ..................................................................................................................................121
  4.4.1 Descriptive statistics .......................................................................................................121
  4.4.2 Main findings ...................................................................................................................123
    4.4.2.1 Information transfer and impact of the HR-manager ................................................126
4.5 Summary and discussion .......................................................................................................129
4.6. References ..........................................................................................................................133
Chapter 1

Introduction
The tremendous progress in technology, digitization and globalization in the last few decades has transformed organizations. First, there is an increased unpredictability and uncertainty in the current business environment, which has influenced how organizations operate and how they are structured. Modern organizations are changing into flatter structures that have independent, flexible teams with fast decision-making and learning capabilities (McKinsey & Company, 2018; Capgemini Consulting, 2017). Moreover, the advancement in technology and digitization has facilitated the availability of information in organizations. In the past, decision makers in organizations, such as managers and employees, had to rely on the limited information supplied and available to them. However, decision makers in modern organizations can demand additional information due to the ample amount of information accessible nowadays. All these transformations have implications for performance management systems, such as performance evaluation and feedback systems in organizations. This dissertation presents three studies that use laboratory experiments to examine performance management systems in modern organizations.

Chapter 2 examines how managers’ self-serving incentives, which are becoming more prevalent in modern organizations, affect managers’ evaluation behavior. Self-serving incentives for managers can arise when it is costly for them to give high evaluations to employees, because it comes at the expense of their own payoff, for instance, the manager of a profit center or the manager of an independent profit-accountable team. In this chapter, I examine the impact of managers’ self-serving incentives on the collection and use of information for the purpose of subjective performance evaluation. I predict and find that managers with self-serving incentives collect less information for subjective performance evaluation compared to managers with no self-serving incentives. I also find that when managers with self-serving incentives collect all available information, they interpret it in a more self-interested manner. Next, I examine how employee self-evaluations, which is a
modern evaluation technique (Toegel and Conger, 2003; Vregelaar, 2017), can mitigate the negative effects of such self-serving incentives. My findings show that managers’ information avoidance under self-serving incentives is mitigated when employees evaluate their own performance and managers observe these employee self-evaluations afterwards. Overall, this chapter increases our knowledge about the role of subjective performance evaluations in modern organizational contexts where managers might have self-serving incentives, such as business units operating as profit centers and profit-accountable teams.

Chapter 3 examines whether demand-driven feedback systems, which are becoming a popular performance management tool in many organizations, induce employees to engage in easy task prioritization. Due to technological progress, the feedback process in organizations has changed from a discretionary process that happens once a year at the supervisor’s initiative to a demand-driven process often allowing employees to ask feedback whenever they want. Employees’ enhanced control over the feedback process under these demand-driven feedback systems have increased employee satisfaction and motivation (Impraise, 2017). We predict and find that under demand-driven feedback systems, employees are more likely to choose easier tasks over difficult tasks (which we label as easy task prioritization). Next, we examine whether the unintended consequence of easy task prioritization under demand-driven feedback systems can mitigated through recordkeeping whereby employee has a record of easy vs. difficult tasks he/she completed. We find that recordkeeping can reduce easy task prioritization induced by demand-driven feedback systems. In an extension of the experiment, we also examined whether easy task prioritization can be reduced by further modifications to demand-driven feedback systems, i.e. when employees can plan what tasks to do (planning) and when the system suggests the next task based on the plan (dynamic sequencing). However, these modifications do not have an incremental effect over recordkeeping in reducing easy task prioritization. Overall, this chapter uncovers how demand-driven feedback, a popular
performance management system in modern organizations, has altered the behavior of employees and how the unintended consequences of these demand-driven feedback systems can be mitigated.

Chapter 4 examines the role of calibration committees, a popular performance management system in modern organizations, in supervisors’ evaluation behavior. Calibration committees to review and correct the initial subjective performance evaluations made by the immediate supervisors of the employees working in flexible teams (Hastings, 2012). In this chapter, we use an experiment to examine how calibration committees affect supervisor collection of costly information about the employee for evaluation purposes. We predict and find that a calibration committee instigates supervisors to collect additional costly information that helps to better explain the performance of their employees. We also find that the presence of a calibration committee leads to better differentiated performance evaluations through supervisors’ collection of additional costly information. We also study two different types of calibration committees; those consisting of only supervisors compared to those with both supervisors and a third party in the form of a HR-manager. Our results reveal that the presence of a third party (i.e. HR-manager) in the calibration committee leads to better information transfer during discussion in the calibration committee. Specifically, supervisors are less likely to anchor on their initial ratings and more likely to consider other supervisors’ information about employees to reach a consensus about their evaluations when a third party is present. Overall, this chapter opens the black box of calibration committees by eliciting behavioral mechanisms that instigate supervisors to make more thorough evaluations.

I use experiments to examine my research questions as experiments allow me to better analyze information gathering processes. For instance, in chapter two and four, my experimental setting enables me to examine how managers and supervisors collect and use employee performance information to make subjective performance evaluation. These
processes are harder to study and often not observable with field data. Experiments also allow me to provide ex-ante evidence about potential solutions that can mitigate the unintended consequences of recent developments in technology and management practices. For instance, I use an experiment in chapter two to provide ex-ante evidence of how employee self-evaluations could mitigate strategic behavior of managers when they have self-serving incentives. Moreover, in chapter three, an experimental setting allows me to show how recordkeeping could mitigate task selection bias under real-time feedback system.

In summary, my dissertation is organized as follows. In chapter two, I present my single-author paper titled “Managers’ Self-Serving Incentives: Information Avoidance in Performance Evaluation”. In chapter three, I present a study co-authored with Bart Dierynck titled “Demand-driven Feedback Systems, Recordkeeping and Easy Task Prioritization”. Chapter four, presents a study co-authored with Bart Dierynck and Eddy Cardinaels titled “Facing a Calibration Committee: The Impact on Costly Information Collection and Subjective Performance Evaluation”.

References:


Chapter 2

Managers’ Self-Serving Incentives: Information Avoidance in Performance Evaluation
2.1 Introduction

In many organizational contexts, self-serving incentives can arise for managers whereby managers may refrain from giving high evaluations to employees because it hurts their own payoff. For instance, when managers have profit responsibility in business units or when managers have flexible and independent profit-accountable teams (McKinsey & Company, 2018; Parker et al., 2008), high employee bonuses would reduce the total profit thereby reducing managers’ own compensation\(^1\). Another example of self-serving incentives are bonus pools where managers are the residual claimant. While analytical models have examined these self-serving incentives by assuming that manager is the residual claimant (Bull, 1983; MacLeod and Malcomson, 1989; Baker, et al., 1994), there is a lack of empirical evidence on whether and how these self-serving incentives impact manager’s collection and use of information for subjective performance evaluation. My study first tries to address this gap. Next, I examine how any negative consequences of self-serving incentives on information collection and use for subjective performance evaluation can be mitigated.

Previous research has shown that managers collect additional information to better understand how employee actions map to noisy performance measures (Bol, 2008, 2011; Bol and Smith, 2011; Maas et al., 2012; Wang and Yin, 2017). Self-serving incentives, however, might distort this process of collecting and using additional information for effective subjective performance evaluation. Drawing on information avoidance theory, which predicts that people willingly avoid information when they have incentives to do so, I predict that managers with self-serving incentives will avoid collecting additional information when they observe the low realization of the performance measure (Golman, Hagman and Loewenstein, 2017). This is because collecting any additional information might reveal that the employee faced bad luck,

\(^1\) These kinds of self-serving incentives are becoming increasingly prevalent in modern organizations, as survey evidence suggests that 83 percent of the organizations want to implement more flexible profit-accountable units (Capgemini Consulting, 2017).
compelling the manager to adjust the employee’s compensation upwards at the expense of his/her own payoff, in order to maintain the manager’s self-image as a fair-minded individual. Willfully opting not to collect information allows self-serving managers the moral wiggle room to justify their self-interested evaluation behavior (Dana et al., 2007; Grossman, 2014; Grossman & van der Weele, 2017; Golman et al., 2017).

Instead of outright avoidance of information by simply not collecting it, managers with self-serving incentives could collect information but avoid drawing the most logical conclusions from it (Golman, Hagman and Loewenstein, 2017). Specifically, managers could interpret and weigh the additional information collected in a self-interested manner (Peyshakhovich and Karmarkar, 2015; Babcock et al. 1995). When the employee’s realization of the performance measure is low and the additional information shows that employee had some bad luck (for example, in the form of a difficult project), then self-serving incentives might prompt the manager to underestimate the bad luck faced by the employee. Therefore, I predict that conditional on full information collection, managers with self-serving incentives are less likely to adjust employee’s evaluations upward when information indicates bad luck, compared to managers with no self-serving incentives.

Next, I investigate whether the effect of self-serving incentives of managers on information collection and use can be mitigated through self-evaluations by employees. Instead of only managers evaluating employees, nowadays organizations also require employees to

---

2 In the paper, I only make formal hypothesis based on low performance realization and bad luck. Previous research has found that principals reciprocate when they see a good outcome regardless of effort. Therefore, it is ex ante difficult to predict what happens when manager observes high performance measure realization and good luck. Under high performance measure realization, reciprocity might confound effect of self-serving incentives on manager’s information collection and interpretation. Another reason that I focus on low performance measure realization and bad luck is that failure and bad luck is usually a major cause of concern in the business and academic circles. Adjustments for good luck are also rarely observed in practice. Kelly, Webb, and Vance (2015, p.2), for example, state that they focus on the use of ex post goal adjustments to filter out negative uncontrollable events and not positive uncontrollable events because adjusting for positive events is rarely observed in practice given the fairness concerns that would likely arise (Gibbs et al. 2004; Merchant 1989). Although I do not make formal hypothesis about it, in the supplemental analysis on p.20 I do provide insights on what happens in case of the high performance measure realization and good luck.
evaluate their own performance and allow managers to ex-post observe what the employees proposed as their discretionary adjustment. In fact, employees evaluating themselves is becoming increasingly popular in 360-degree assessments and in self-managed, independent teams (Cohen, Ledford and Spreitzer, 1996; Toegel and Conger, 2003; Vregelaar, 2017). These employee self-evaluations can improve manager’s information collection and use for subjective performance evaluation in two ways. First, presence of employee self-evaluations can enhance the moral context of the evaluation process and the importance of social or distributional preferences in this process such that the manager is more likely to consider the employee’s opinion (van der Weele, 2014). Second, when the manager ex-post observes a difference between his/her own discretionary adjustment and the employee’s self-evaluation, it signals disagreement about the employee’s performance. Given that employee knows more about his/her performance, this disagreement can trigger feelings of discomfort or cognitive dissonance in the manager (Festinger, 1957; Newcomb, 1953; Osgood and Tannenbaum, 1955). The manager, anticipating this discomfort, will look for ways to avoid such potential disagreement between the employee’s self-evaluation and his/her own discretionary adjustment by collecting and using more information, which would mitigate information avoidance. Therefore, I predict that the presence of employee self-evaluations makes managers with self-serving incentives (1) collect more information when they observe low performance measure realization (2) more likely to adjust employee evaluations upward when the information indicates bad luck.

I conducted a laboratory experiment using oTree software to test my hypothesis (Chen et al., 2016). Participants are randomly assigned to the roles of manager and employee and are randomly re-matched at the start of each period. The employee performs a real-effort task in which the employee’s performance is partially determined by the state of nature —in other words, by luck. The manager can make a discretionary adjustment to the employee’s
compensation and can collect additional information that allows him/her to better understand the employee’s performance. I use a between-subjects nested design: a control condition where managers do not have self-serving incentives (No SSI); a second condition where managers have self-serving incentives (SSI); and a third condition where managers have self-serving incentives and employees make a proposal about what their discretionary adjustment should be, and these self-evaluations are ex-post visible to the managers (SSIDA).

My results find information avoidance among managers with self-serving incentives. Compared to managers with no self-serving incentives, managers with self-serving incentives are less likely to collect information when they observe a low performance measure realization. I also find that managers with self-serving incentives interpret information in a more self-serving way. That is, conditional on full information collection and compared to managers with no self-serving incentives, managers with self-serving incentives are less likely to adjust employee evaluations upward when the information indicates bad luck. My findings also show that employee self-evaluations mitigate information avoidance by managers with self-serving incentives, as these managers are more likely to collect information when they observe low performance measure realization. Conditional on full information collection, managers with self-serving incentives and whose employees make self-evaluations are also more likely to use information to make upward adjustments when the information indicates bad luck, compared to managers in the self-serving incentives condition.

Manager’s discretionary adjustment to the employee’s compensation also has consequences for employee performance, because the way managers make these discretionary adjustments can impact the incentive effect of performance-based compensation (Hölmstrom, 1979; Berger et al., 2013; Bol, 2011). Ex-ante, it is not clear how managers’ self-serving incentives impact employee performance. My findings show that employee performance and intrinsic motivation are higher when managers have self-serving incentives. My rationale for
this result is that employees in the self-serving incentives condition know that managers have incentives to not give them high evaluations and, therefore, they must exert more effort to earn an acceptable level of compensation. I also find that employee performance is lower in the self-serving incentives with employee self-evaluations condition than in the self-serving incentives condition. However, employee performance in the self-serving incentives with employee self-evaluations condition is still higher than in the condition with no self-serving incentives. Moreover, the presence of employee self-evaluations under self-serving incentives does not negatively affect employees’ intrinsic motivation.

This study makes several contributions. First, I contribute to the management accounting literature on complementarities in organizational design choices (Milgrom and Roberts, 1995; Abernethy et al., 2004; Moers, 2006, Bouwens and van Lent, 2007; Indjejikian and Matejka, 2012). Managers with profit responsibility are often evaluated on aggregated financial and profit measures. I show how self-serving incentives that arise in such a situation can influence managers’ collection and use of information for the subjective performance evaluation of their employees. Besides the effect of self-serving incentives on information collection and use, I find that managers’ self-serving incentives strengthen the incentive effect of performance-based compensation for employees, yielding higher employee performance and motivation.

My study also contributes to the subjective performance evaluation literature in several ways. First, previous studies show that manager’s evaluations might be biased owing to high information gathering costs, such as time or effort (Bol, 2008; Bol, 2011). My study provides additional insights to this stream of research by showing that subjective performance evaluations might be biased downwards owing to the effects of manager’s self-serving incentives on information collection and use. Second, previous research finds that managers have social preferences that motivate them to collect additional costly information to evaluate
employees based on effort (Maas et al., 2012). My study adds to this research by demonstrating that managers might avoid information to have the moral wiggle room to justify their self-interested evaluation behavior.

Finally, my study contributes to the literature on employee self-evaluations, which are becoming increasingly popular in organizations. I show how employee self-evaluations can mitigate information avoidance in managers with self-serving incentives when they make subjective performance evaluations. The presence of employee self-evaluations under self-serving incentives also does not negatively affects employees’ intrinsic motivation and has a lower negative impact on employee performance than under no self-serving incentives. Thus, having employee self-evaluations under self-serving incentives might be the best alternative of the three conditions I study here, yielding greater collection and use of information by managers for making subjective performance evaluations while minimizing any negative effects on employee performance and intrinsic motivation.

2.2 Literature review and hypothesis development

2.2.1 Information avoidance under manager self-serving incentives

Subjective performance evaluations in organizations often take the form of discretionary bonus adjustments or subjective weights on different performance measures (Gibbs et al., 2004). To make effective subjective performance evaluations, managers need to learn how employee actions map to noisy performance measures as uncontrollable factors often impact this mapping (Baker, 1992, 2000; Hölmstrom, 1979). Previous research has shown that managers can collect additional information for this purpose (Bol, 2008, 2011; Bol and Smith, 2011; Maas et al., 2012; Wang and Yin, 2017). One important aspect that could impact information collection for subjective performance evaluation is the manager’s incentives. Specifically, managers might have self-serving incentives that make it costly for them to give employees high evaluations. These self-serving incentives of managers to maximize their own
payoff at the expense of employee evaluations might, therefore, affect how they collect and use additional information and consequently their subjective performance evaluations.

Models of rational decision making that depend exclusively on preferences over monetary outcomes suggest that people weakly prefer to have more (free) information because it allows them to make informed decisions. Unfortunately, an explanation based on these models provides little insights for cases where people avoid costless or free information. In fact, several behavioral economics papers have empirically shown that in social dilemmas people avoid obtaining information about the negative social effects of their self-interested decisions precisely to justify making these decisions (e.g., see Dana et al., 2007; Grossman, 2014). These results are explained by information avoidance theory (Golman, Hagman and Loewenstein, 2017) that suggests that people avoid information as a kind of commitment device, because they anticipate that it will affect their future actions. Suppose that an individual can take an action that is personally beneficial and can collect information to know whether it is unethical or socially harmful. Avoiding such information allows the individual to commit to the personally beneficial action because if he/she knows that it is unethical or socially harmful then his/her social preferences and self-image concerns would prevent him/her from taking that action. Thus, avoiding socially relevant information offers people the moral wiggle room to justify their opportunistic behavior while maintaining a positive self-image (Dana et al., 2007; Grossman, 2014; Grossman & van der Weele, 2017; Golman et al., 2017). People who avoid information benefit materially from their behavior without lowering their self-image, because they can credibly claim that if they had known the information then they would have done the right thing. This explains why people are often reluctant to know about climate change consequences of their actions (Grossman & van der Weele, 2017).

The literature on information avoidance, therefore, suggests that managers with self-serving incentives may avoid information that could justify upward discretion ary adjustments
to the employee payoff because doing so offers them the moral wiggle room to refrain from giving upward discretionary adjustments. Thus, when managers with self-serving incentives observe low realization of the performance measure, they will be less likely to collect additional information and will instead focus on the low realization to justify not giving any upward discretionary adjustment. Specifically, a manager will avoid collecting additional information when he/she observes the low performance measure realization because it is likely that any additional information will reveal bad luck and compel the manager, out of his/her social preferences, to make an upward discretionary adjustment to the employee’s payoff at the expense of his/her own payoff. Thus, if managers with self-serving incentives avoid information then there will be lower collection of information when the performance measure realization is low.

**H1a:** Managers with self-serving incentives are less likely to collect information when they observe a low performance measure realization compared to managers who do not have self-serving incentives.

Alternatively, it is possible that managers will not completely avoid collecting information because choosing to do so is considered an act of opportunism that reflects negatively on their self-image (Church et al., 2014; Grossman & van der Weele, 2017; Berge, 2018). In fact, van der Weele (2014) shows that in morally rich contexts, people feel that they should collect information. Specifically, instead of willfully opting not to collect information, managers with self-serving incentives could collect information but avoid drawing the most logical conclusions from it (Golman, Hagman and Loewenstein, 2017). In other words, managers could interpret and weigh the additional information collected in a self-interested manner (Babcock et al. 1995; Peysakhovich and Karmarkar, 2015). As a result, managers might collect information to maintain a positive self-image but then insufficiently process the information about bad luck that indicates an upward discretionary adjustment to the employee’s
payoff. Thus, if performance measure realization is low and the additional information shows that the project was difficult, then self-serving incentives might prompt the manager to underestimate the bad luck the employee faced in the form of a difficult project. This leads to the following prediction:

**H1b**: Managers with self-serving incentives are less likely to use information when it indicates upward discretionary adjustment compared to managers who do not have self-serving incentives.

### 2.2.2 Self-evaluations by employees

Next, I examine whether self-evaluations by employees can be one way to mitigate self-serving incentives of managers. Instead of just having managers evaluate their employees, organizations are increasingly requiring employees to evaluate their own performance as part of their evaluation systems. In fact, employees evaluating themselves is becoming increasingly popular in 360 assessments and in self-managed, independent teams (Cohen, Ledford and Spreitzer, 1996; Toegel and Conger, 2003; Vregelaar, 2017). Allowing employees to evaluate their own performance by deciding what they think should be their own discretionary adjustments and allowing managers to observe these self-evaluations after they make discretionary adjustments about the employees can mitigate information avoidance in two ways. First, employee self-evaluations enhance the moral context and the importance of social or distributional preferences in the evaluation process (van der Weele, 2014). Specifically, the manager with self-serving incentives might be more likely to think about the employee’s opinion if there is employee self-evaluation. Employee self-evaluation, thus, acts as an external factor that encourages the manager to, as they say, walk a mile in the employee’s shoes.\(^3\) Doing

---

\(^3\) These arguments draw parallels with the self-concept maintenance theory about dishonest behavior. Self-concept maintenance theory states that people can be dishonest without influencing their self-concept when it is easier for them to re-interpret their selfish behavior because of the context (categorization malleability), and when moral standards are less accessible to people (attention to moral standards). Based on self-concept...
so would make social preferences and fairness concerns more salient for the manager and reduce the manager’s information avoidance and downward bias in discretionary adjustments.

Second, when the manager ex-post observes a difference between his/her own discretionary adjustment and the self-evaluation of the employee, then it might signal disagreement about the employee’s performance. Previous research in psychology has shown that people want others to agree with their opinions and tend to avoid disagreements (Matz and Woods, 2005; Davis, 1963; Festinger, 1957; Newcomb, 1953). Theories in psychology suggest that disagreement with others produces a feeling of discomfort known as cognitive dissonance or imbalance (Festinger, 1957; Newcomb, 1953; Osgood and Tannenbaum, 1955). Previous studies have tried to examine the reasons for this cognitive dissonance caused by disagreement. These studies have shown that people have strong preferences for having other people agree with them because validation from others about our opinions make us feel better about our opinions. Specifically, people desire agreement with others to achieve a coherent, favorable self-image (Chaiken, Giner-Sorolla, and Chen, 1996; Cialdini and Trost, 1998); and to verify and validate their own attitudes and understanding (Newcomb, 1953; Tajfel, 1978; Tajfel and Turner, 1986). Regardless of whether cognitive dissonance is caused indirectly by disagreement with others or is an indirect product of the validity concerns posed by disagreement with others, it might generate tension and therefore might influence attempts to restore agreement (Matz and Woods, 2005). Given that employees know more about their own performance, managers in the presence of employee self-evaluations would want their own discretionary adjustment to be as close as possible to the employee’s self-evaluation. This allows them to reduce any cognitive dissonance caused by differences in the two adjustments.
ex post. As a result, managers might look for ways to avoid a potential disagreement by collecting and using more information in the presence of employee self-evaluations. Thus, when they observe low performance measure realization, managers with self-serving incentives whose employees make self-evaluations are more likely to collect information compared to managers with self-serving incentives whose employees do not make self-evaluations. Conditional on full information collection, presence of employee self-evaluation also make manager with self-serving incentives more likely to use information to give upward adjustments when information indicates upward adjustment i.e. bad luck information.

However, it is possible that employee self-evaluations will not mitigate self-serving incentives, when, for example, manager reduce the cognitive dissonance caused by ex-post difference in adjustments by dismissing as inaccurate any employee self-evaluations that differs from their own and by rationalizing and validating their own adjustments as objective and accurate. In fact, such beliefs could be reinforced over time such that a manager with self-serving incentives becomes increasingly confident about his/her own biased discretionary adjustments. Overall, I predict that:

**H2a:** Compared to the absence of employee self-evaluations, in the presence of employee self-evaluations, managers with self-serving incentives are more likely to collect information when they observe low performance measure realization.

**H2b:** Compared to the absence of employee self-evaluations, in the presence of employee self-evaluations, managers with self-serving incentives are more likely to use information when it indicates upward discretionary adjustment.

### 2.3 Experimental design

#### 2.3.1 Overview

I conducted a laboratory experiment using oTree software (Chen et al., 2016). Participants are randomly assigned the role of either manager or employee. Employees perform
a real-effort task where they get two matrices and have to find the highest number in each matrix and add these numbers. There are six periods. Each employee gets four tasks in each period and has to solve these four tasks in ninety seconds. To allow for variation in luck, employees might get either all four tasks with 2x2 matrices; two tasks with 2x2 matrices and other two tasks with 6x6 matrices; all four tasks with 6x6 matrices (see Appendix). At the start of each period, the state of nature (i.e., luck) determines which of the three combinations an employee gets in that period. After each period, the manager sees the performance measure realization for the employee. If the employee correctly solves all four tasks then his/her performance measure realization in the period is 1; otherwise, it is 0. The manager has the option to collect additional information about the tasks solved and the luck in the period. Specifically, after each period, the manager sees the performance measure realization for the employee and has the option to collect additional information by clicking a button to reveal the number of tasks correctly solved by the employee. Once the manager has seen the number of tasks correctly solved, he/she can then choose to click another button to reveal more detailed information, namely, the combination of 2x2 matrices tasks versus 6x6 matrices tasks the employee got during the period. The employee and the manager each earn a base salary of 1 experimental currency (EC) and 2 EC in each period, respectively. Both employee and manager can also earn a bonus of 2 EC based on the employee’s performance measure realization in the period. Managers can make discretionary adjustments ranging from –1 EC to +1 EC to the employee’s compensation at the end of each period.

4 The binary performance measure realization based on tasks correctly solved allows managers to have a simple benchmark of good or bad performance.

5 As the information is button-click away, it is almost costless to collect. This design choice allows us to provide a clean test of our information avoidance theory. Previous papers based on information avoidance theory usually make information costless to acquire (e.g. see Dana et al., 2007). If information is costly, then managers in self-serving incentives condition might ignore it either because (1) they do not want to pay for it, or (2) they do not want to know that employee had bad luck to avoid giving positive adjustment to him/her. Costless information collection is also an interesting feature in many modern organizations as they have reduced the cost to access information by making information in many cases click-of-a-button away.
To test my hypotheses, I use a between-subjects nested design. My control condition comprises managers who do not have self-serving incentives (No SSI); my second condition comprises managers who have self-serving incentives (SSI); and my third condition comprises managers who have self-serving incentives and employees who make a proposal about what their discretionary adjustment should be, such that employee self-evaluations are ex-post visible to the managers (SSIDA). I also ensure that the manager is re-matched with an employee after each period using random matching such that the probability of being matched with the same employee in the next period is low. This re-matching provides a clean test of my theory on self-serving incentives and allows me to avoid reputation-building concerns. In the no self-serving incentives (No SSI) condition, manager’s discretionary adjustments only impact employee compensation, whereas in the self-serving incentives (SSI) condition, manager’s discretionary adjustments not only impact employee compensation but also manager compensation. Specifically, in the SSI condition, any upward adjustment made by the manager to the employee’s compensation reduces the manager’s compensation while any downward adjustment by manager increases manager’s compensation. Thus in the No SSI condition the payoffs are: \[ \text{Manager’s payoff} = 2 \text{EC} + 2 \text{EC}^*X; \text{Employee payoff} = 1 \text{EC} + 2 \text{EC}^*X + DA; \]
where X is the performance measure realization which is 1 or 0 depending on whether the employee correctly solved all four tasks and DA is the discretionary adjustment made by the manager that ranges from −1 EC to +1 EC. For my SSI condition, the manager’s payoff is a function of the manager’s discretionary adjustment such that: \[ \text{Manager’s payoff} = 2 \text{EC} + 2 \text{EC}^*X − DA; \text{Employee payoff} = 1 \text{EC} + 2 \text{EC}^*X + DA; \]
where X is the performance measure realization which is 1 or 0 depending on whether the employee correctly solved all four tasks and DA is the discretionary adjustment made by the manager that can vary from −1 EC to +1.
EC⁶. Then for my SSIDA condition, managers still have self-serving incentives as in the second condition (SSI), but the employee also proposes what his/her discretionary adjustment should be at the end of every period while the manager decides the discretionary adjustment. The employee’s self-evaluation is visible to the manager after he/she makes his/her discretionary adjustment decision.

### 2.3.2 Participants

I recruited business and economics undergraduate students who responded to an email invitation.⁷ Participants’ ages range from nineteen to twenty-nine years old, and 64 percent of the participants are male. A large majority of the participants indicated that they have part-time or full-time work experience (97 percent), and all participants had completed at least one math course, economics course, accounting course, and finance course at the university level. As a show-up incentive, I introduced a modest amount of bonus course credit on top of their total grade. In addition to this show-up incentive, each participant earned monetary compensation based on their and other participants’ decisions/actions in the experiment.⁸ Before conducting the study, I obtained approval to run the study from the research institute.

In total, 13 sessions were conducted and there were 146 managers and 146 employees. In all conditions, I always have two-person groups of manager-employee: In No SSI, I had 43 managers and employees each, in SSI, I had 46 managers and employees each, and in SSIDA, I had 57 managers and employees each. The experiment took less than 60 minutes and participants on average earned € 7.2.

---

⁶ For example, suppose the performance measure realization is 0 and the manager makes a discretionary adjustment of +1 EC to employee’s payoff, then employee’s payoff will be 2 EC. If the manager is in No SSI condition, then the discretionary adjustment will not affect his/her payoff and his/her payoff will be 2 EC. However, if the manager is in the SSI condition, then the discretionary adjustment of +1 EC will reduce his/her payoff by -1 EC and his/her payoff will be 1 EC.

⁷ Participants were all recruited from the same management accounting course at bachelor’s level.

⁸ I paid €1.00 per 2 Experimental Currency (EC) earned in the experiment. Payouts were contingent on situational factors and on participants’ decisions.
2.3.3 Experimental manipulations and procedure

Upon their arrival in the research institute’s computer lab, participants had to wait briefly in a waiting room, where they publicly received some basic instructions (e.g., no talking, and no electronic devices) before moving into the lab. On entering the computer lab, participants randomly chose one cubicle containing a computer. Once I began the study, participants entered an instruction phase where they read information about their role, payoffs, and specific information related to their treatment condition. This phase included instructions and a few basic control questions. When participants gave wrong answers to the control questions, the software provided them feedback. I carefully designed a concise set of instructions and control questions that only checked whether participants understood the basics of the study. All participants also solved one 2x2 matrices task and one 6x6 matrices task as practice during the instruction phase.

After the instruction phase, participants received a warning that the first period was about to start (see, Appendix B for the example screens), and they would be randomly re-matched with another participant in each of the six periods. At the start of each period, the participants in the role of the employee entered the task phase, where they were given four tasks to complete in ninety seconds. The state of nature (i.e., luck) decided which of the three possible combinations of tasks (four tasks with 2x2 matrices; two tasks with 2x2 matrices and other two tasks with 6x6 matrices; or four tasks with 6x6 matrices) the employee would receive in any period. Thus, if the employee got all four tasks with 2x2 matrices then it is considered good luck; if employee got the two tasks with 2x2 matrices and other two tasks with 6x6 matrices then it is considered neutral luck; and if the employee all four tasks with 6x6 matrices then it is considered bad luck. For the sake of simplicity in my analysis, I kept the pattern of luck similar across all sessions and conditions: periods 1 and 4 had good luck, periods 2, 3 and 5 had bad luck, and period 6 had neutral luck. Employees were made aware of the time left
through a message that appeared when thirty seconds were remaining. After ninety seconds the task phase ended, and both the manager and the employee observed whether or not the employee had correctly solved all four tasks in the period. Participants in the role of the manager could then make a discretionary adjustment to the employee’s compensation and could collect additional information for this purpose. Specifically, by clicking a button the manager could find out the number of tasks the employee solved correctly, and then by clicking a second button, the manager could find out the combination of 2x2 and 6x6 matrices tasks that the employee got in the period. While managers were making their discretionary adjustments to employee’s compensation, employees in the self-serving incentives with employee self-evaluations (SSIDA) condition could propose what their discretionary adjustments should be. These employee self-evaluations did not impact the final payoffs but were visible to the manager at the end of the period. After the managers in first two conditions (No SSI and SSI) made their discretionary adjustments, the period ended and both the manager and the employee could observe a summary of the results in the period: whether the employee correctly solved all four tasks; the manager’s discretionary adjustment; and the employee’s and manager’s payoffs in the period. In the third condition (SSIDA), after the managers chose the discretionary adjustment to the employee’s compensation and the employees proposed their self-evaluation, the period ended and both the manager and the employee could observe a summary of the period’s results: whether the employee correctly solved all four tasks; the manager’s discretionary adjustment; the employee’s self-evaluation; and the employee’s and manager’s payoffs in the period. After every period, each manager was re-matched with another participant in the session who was in the role of employee.

After the six periods ended, participants completed an ex-post questionnaire containing several items intended to provide insight into the thoughts and feelings of the participants
during the study. At the end of the experiment, participants were informed about their total earnings in the experiment in euros.

2.4 Results

2.4.1 Descriptive statistics

Table 1 presents descriptive statistics wherein manager panel observations are grouped by experimental condition. Since managers could first collect information about number of tasks correctly solved by employee and then collect more information about the luck, table 1 provides statistics for both managers’ collection of task information and all information (i.e., luck and task). Managers in the no self-serving incentives (No SSI) condition collected on more task information than managers in the self-serving incentives (SSI) condition \( (Z = 3.143; \text{two-tailed } p\text{-value }= 0.002) \). Table 1 reports similar results for all information collection when comparing the SSI and the No SSI conditions \( (Z = 1.856; \text{two-tailed } p\text{-value }= 0.063) \). Under the self-serving incentives with employee self-evaluations (SSIDA) condition, managers collect more task information \( (Z = 2.153; \text{two-tailed } p\text{-value }= 0.031) \) as well as all information \( (Z = 3.764; \text{two-tailed } p\text{-value }= 0.000) \) than managers in the SSI condition. I also report the ordered categorical variable, Information Collection, with 2 representing both task and luck information collected, 1 representing only task information collected, and 0 representing no information collected. Results are similar to all information collection dummy variable. Managers in the No SSI condition collect more information than managers in the SSI condition \( (Z = 2.796; \text{two-tailed } p\text{-value }= 0.005) \), and managers in the it the SSIDA condition collect more information than managers in the SSI condition \( (Z = 3.691; \text{two-tailed } p\text{-value }= 0.000) \). Next, the discretionary adjustments by managers in the No SSI condition were more upwards than those by managers in the SSI condition \( (Z = 8.372; \text{two-tailed } p\text{-value }= 0.000) \). In fact, managers in the No SSI condition, gave on average, a positive discretionary adjustment of 0.275, whereas managers in the SSI condition gave, on average, a negative discretionary adjustment of \(-0.237\).
Managers in the SSIDA condition gave slightly negative discretionary adjustments of \(-0.085\) on average. The discretionary adjustments in the SSIDA condition were, thus, less upwards than those in the SSI condition \((Z = 2.926; \text{two-tailed } p\text{-value } = 0.003)\). The No SSI condition had lower employee performance realization \((Z = -2.564; \text{two-tailed } p\text{-value } = 0.010)\) and number of correct employee tasks \((Z = -2.511; \text{two-tailed } p\text{-value } = 0.012)\) than the SSI condition.\(^9\) The SSIDA condition also had lower employee performance realization \((Z = -2.492; \text{two-tailed } p\text{-value } = 0.013)\) and number of correct employee tasks \((Z = -2.169; \text{two-tailed } p\text{-value } = 0.030)\) than the SSI condition.

--- Table 1 about here ---

Figure 1 shows manager information collection for each of three conditions. Manager information collection is on the vertical axis as an ordered categorical variable from 0 to 2, with 0 representing no information collected, 1 representing only task information collected, and 2 representing both task and luck information collected. The figure displays that, on observing low performance measure realization, managers in the SSI condition collect less information than the managers in both the No SSI and SSIDA conditions. This finding provides some preliminary support for my hypotheses H1a and H2a about managers’ information collection.

--- Figure 1 about here ---

Figure 2 shows the managers’ discretionary adjustments when they observe bad luck information in the three experimental conditions. Conditional on full information collection, managers in the No SSI condition give higher adjustments than managers in the SSI condition. This finding provides preliminary support for my hypothesis H1b as managers with self-serving incentives (i.e., in the SSI condition) seem to use bad luck information that indicates

---

\(^9\) Employee performance realization is 1 when employee correctly solves all 4 tasks in a period, else 0, and number of correct employee tasks is the number of tasks that the employee correctly solves in a period.
upward adjustment less than managers in the No SSI condition. Moreover, managers in the SSIDA condition give adjustments that are higher than the adjustments under self-serving incentives (i.e., in the SSI condition) but lower than the adjustments under no self-serving incentives (i.e., in the No SSI condition). This finding provides preliminary support for my hypothesis H2b that the presence of employee self-evaluations under self-serving incentives mitigates manager’s low use of information that indicates an upward adjustment to the employee’s compensation.

--- Figure 2 about here ---

2.4.2 Main findings

Table 2 shows the formal tests for my hypotheses. I test H1a and H2a using Information collection as my dependent variable, where Information collection is an ordinal categorical variable that equals 0 when managers do not collect any additional information, 1 when managers collect only information about the correct number of tasks solved, and 2 when managers collect information about both the correct number of tasks solved and the luck during the period. I estimated a mixed-effects ordered logistic regression with random effects at the manager level, as the reported likelihood ratio test shows that there is enough variability between managers to favor this regression over a standard ordered logistic regression and the interclass correlation of 0.92 justifies the manager random effects. The regression also includes robust standard errors clustered by manager. Since my hypotheses H1a and H2a are conditional on managers observing low performance measure realization, I only included the observations where performance measure realization is 0. In support of H1a, column 1 of Table 2 shows that managers in the No SSI condition collect more information than the managers in

---

10 Mixed-effect models are essentially multi-level that can capture unobserved heterogeneity across units by allowing for random variation in intercepts or slopes (Luke 2004). Also, the errors of observations for same unit are likely correlated. The random effects in a mixed-effects model can, therefore, be viewed conceptually as a way of controlling for additional sources of variation (or error) that we are otherwise unable to account for (Luke 2004). Therefore, this study utilized a mixed-effect model with random effects for managers in the analysis.
the SSI condition when they observe a low performance measure realization (\(b = 1.889\), two-tailed \(p\)-value = 0.077). Column 1 of Table 2 also shows that managers in the SSIDA condition collect more information than managers in the SSI condition when they observe a low performance measure realization (\(b = 1.685\), two-tailed \(p\)-value = 0.086). This finding provides support for hypothesis H2a.

Next, I test hypotheses H1b and H2b using the dependent variable Information interpretation, which is the manager discretionary adjustment ranging from \(-1\) to \(+1\) conditional on the manager collecting full information. I estimated a mixed-effects regression with random effects at the manager level because and reported likelihood ratio test shows that there is enough variability between managers to favor this regression over a standard linear regression and the interclass correlation of 0.58 justifies the manager random effects. I also include robust standard errors clustered by manager. As hypotheses H1b and H2b are conditional on information that indicates upward discretionary adjustments, I included in my analysis only the bad luck periods. In support of H1b, column 2 of Table 2 shows that conditional on full information collection, managers in the No SSI condition are more likely to adjust upward than the managers in the SSI condition when they collect information that indicates an upward adjustment (\(b = 0.390\), two-tailed \(p\)-value = 0.001). Column 2 of Table 2 also supports hypothesis H2b. Managers in the SSIDA condition are more likely to adjust upward than managers in the SSI condition when they collect information that indicates an upward adjustment (\(b = 0.205\), two-tailed \(p\)-value = 0.066).

--- Table 2 about here ---

2.4.3 Supplemental analysis

2.4.3.1 Information avoidance under good luck versus bad luck

My main results show that when managers observe low performance measure realization, they collect less information if they have self-serving incentives. This is because
information about low performance measure realization probably reveals that the employee faced bad luck. In fact, the descriptive statistics for the bad luck and good luck periods in Table 3 show that the average performance measure realization is almost 1 (i.e., 0.914) under good luck, whereas the average performance measure realization under bad luck is close to 0 (i.e., 0.118). Since, the performance measure realization is highly correlated with the luck (correlation of 0.72, \textit{p-value} = 0.000), it is important to know what happens when managers observe high performance measure realization indicating that the employee probably had good luck during the period.

Table 3 shows that in all the three conditions, managers’ information collection was 1.11 in good luck periods vs. 1.32 in bad luck periods. Managers also make upward adjustments, 0.238, in the good luck periods and downward adjustments, –0.221, in the bad luck periods. This finding is contrary to the explanation that high performance measure realization under good luck should indicate downward adjustment to employee’s compensation. In fact, this result can be explained by trust reciprocity (Berg et al., 1995; Dufwenberg and Gneezy, 2000; Fehr and Schmidt, 2003). When the employee achieves high performance measure realization, it also benefits the manager by increasing his/her payoff and thus triggers trust reciprocity in managers. Managers observing high performance measure realization in the good luck period, therefore, would want to reciprocate the employee’s high performance by making an upward adjustment to the employee’s compensation. As a result, they focus on the high realization of the performance measure and reward it through upward adjustments without looking for information that explains the employee’s high performance, such as good luck. These results closely mirror the findings by Rubin and Sheremeta (2016), which show that managers base their adjustments on outcome (effort + random shock) rather than on effort. As a result, adjustments are increasing in the shock: with a positive shock, the outcome is high, which motivates greatest reciprocity from principals. Thus, once employees
have a high realization of the performance measure, trust reciprocity becomes important for managers and affects the way they collect and use information for subjective performance evaluation. This is not the case for the low performance measure realization, where trust reciprocity is not triggered and thus does not affect the managers’ collection and use of information.

--- Table 3 about here ---

Next, I compare conditions to analyze the information collection of managers with self-serving incentives (i.e., in the SSI condition) when they observe high performance measure realization. Table 3 shows that information collection is lower for managers in the SSI condition under both good and bad luck periods. This demonstrates that, compared to the No SSI condition, managers with self-serving incentives in the SSI condition collect less information when they observe high performance measure realization. However, the difference in information collection between the No SSI and SSI conditions is not that high under good luck (1.15 vs. 0.99) compared to bad luck (1.36 vs. 1.15). Regressions results in column 1 in Table 4 also show that, compared to managers with no self-serving incentives (i.e., in the No SSI condition), managers with self-serving incentives (i.e., in the SSI condition) do not significantly differ in information collection when they observe high performance measure realization (b = 0.644; two-tailed p-value=0.327). Thus, when trust reciprocity is a concern, managers with self-serving incentives do not make use of the information collection channel to excuse their self-interested evaluation behavior as doing so might be considered a salient act of opportunism.

However, it is possible that it is not that through the collection but the use of information that managers with self-serving incentives exhibit their opportunistic behavior under good luck. Results in column 2 in Table 4 support this by showing that when managers collect information that indicates good luck, managers in the SSI condition are more likely to adjust employee’s
compensation downward than managers in the No SSI condition \((b = 0.499; \text{ two-tailed } p\text{-value}=0.000)\). Thus, results indicate that manager with self-serving incentives interpret the good luck information such that they put more weight on good luck to justify not reciprocating high performance of the employee. As a result, upon collecting good luck information, managers in the SSI condition are less likely to reciprocate high performance measure realization of employee, compared to managers in the No SSI condition.

--- Table 4 about here ---

2.4.3.2 Time spent looking at information

It is possible that, managers in the SSI condition are less likely to use the information collected to make upward adjustment compared to managers in the No SSI condition because they spend less time looking at it. Column 1 in Table 5 shows that managers in the SSI condition spent the same amount of time, in seconds, looking at the information collected, as did the managers in the No SSI condition \((b=-0.0009; \text{ two-tailed } p\text{-value}=0.636)\). Therefore, it is not the time spent looking at the information collected that leads managers in the SSI condition to make lower adjustments. This provides support for my argument that managers with self-serving incentives interpret and weigh the additional information collected in a manner that fulfills their self-interests (Babcock et al. 1995; Peysakhovich and Karmarkar, 2015).

Next, I examine whether there is a difference in time spent looking at the information between managers in the self-serving incentives with employee self-evaluation (SSIDA) condition and managers in the self-serving incentives (SSI) condition. Column 1 in Table 5 shows that managers in the SSIDA condition spent more time, in seconds, looking at the information compared to managers in the SSI condition, but this difference is not significant \((b=0.0036; \text{ two-tailed } p\text{-value}=0.237)\). However, further analysis in column 2 in Table 5 based on type of information shows that managers in the SSIDA condition spent significantly more time, in seconds, looking at the information about luck compared to managers in the SSI
condition (b= 0.0051; two-tailed p-value=0.078). Thus, the presence of employee self-evaluations might prompt managers with self-serving incentives to be more considerate of their employee’s perspectives and ex-ante spend more time on information about the kind of luck that the employee faced during the period in order to give him the benefit of the doubt.

--- Table 5 about here ---

### 2.4.3.3 Employee performance and intrinsic motivation

Previous literature has documented that evaluation behavior of managers has implications for employee performance (Berger et al., 2013; Bol, 2011). According to agency theory, linking pay to performance will cause employees to exert effort in order to get higher pay for higher performance (Hölmstrom, 1979). Managers’ discretionary adjustments to employees’ bonus, therefore, have consequences for the performance of employees, as how managers make these discretionary adjustments can influence the incentive effect of bonuses or performance-based compensation. Simple economic logic suggests that when managers have self-serving incentives to bias employee bonuses downward, employees will have lower incentives to exert effort, as they anticipate that the manager might not collect and use information to make appropriate adjustments and may not reward high performance (Bol, 2008). Thus, manager’s self-serving incentives weaken the incentive effect of employee bonuses. Moreover, employees’ perceptions that managers’ adjustments under self-serving incentives are unfair could reduce employee motivation and incentives to perform (Akerlof and Yellen, 1988; Colquitt et al., 2001). These procedural fairness perceptions might arise even before experiencing outcomes and might be an important determinant of performance as employees might reciprocate unfair adjustments with lower effort (Kelly et al., 2015). On the other hand, manager’s self-serving incentives could strengthen the incentive effect, as the employee may anticipate that the manager is less likely to be lenient and will give the employee a systematically lower bonus. Thus, in the presence of managers’ self-serving incentives,
employees must exert more effort to reach the same level of compensation as in the absence of managers’ self-serving incentives.\textsuperscript{11} Moreover, based on arguments in Arnold and Artz (2015) and Burt et al. (2015), employees who know that their manager will be lenient and adjust positively for any bad luck, reduce effort when they get bad luck in anticipation of the positive adjustment. Thus, compared to employees working with managers with no self-serving incentives, employees of managers with self-serving incentives will work harder because they don’t anticipate positive adjustments for bad luck. In this way, the presence of managers’ self-serving incentives could be comparable to employees having a difficult target whereas the absence of managers’ self-serving incentives could be comparable to employees having an easy target (Blanchard et al., 1986, Bol, 2011). Therefore, ex-ante it is difficult to predict how manager’s self-serving incentives might impact employee performance and motivation.

I analyze employee performance using two variables: \textit{Employee performance realization} and \textit{Employee correct tasks}. \textit{Employee performance realization} equals 1 if the employee correctly solved all the four tasks in the period and 0 otherwise. \textit{Employee correct tasks} is the number of tasks out of the four tasks in the period that the employee solved correctly. Column 1 and 2 in Table 6 show that in the \textit{No SSI} condition, \textit{Employee performance realization} ($b=-1.049$; two-tailed p-value=0.000) and \textit{Employee correct tasks} ($b=-0.299$; two-tailed p-value=0.006) are significantly lower than in the \textit{SSI} condition\textsuperscript{12}. This finding shows

\textsuperscript{11} In other words, employees in the self-serving incentives need to work harder to earn more than their reservation utility. If the manager makes full negative adjustment of -1 EC, employees earn 0 under low performance measure realization and 2 EC under high performance measure realization. Employees whose managers have no self-serving incentives can expect a positive payoff under low performance measure realization as their manager might take any bad luck into account and is less likely to make full negative adjustments to their payoff. However, employees whose managers have self-serving incentives can expect a payoff of 0 under low performance measure realization as their manager benefits from making full negative adjustments to their payoff. Assuming employees do not want to earn zero but some positive payoff that is equal to or above their reservation utility, they will work harder to avoid getting a low performance measure realization in the self-serving incentives condition.

\textsuperscript{12} To ensure that the results are not impacted by certain employees, the models contain robust standard errors clustered by employee. One might argue that ability affects the results; however, results are unchanged when I control for employee ability by including employee’s number of mistakes on practice tasks and time spent on practice tasks. Employee’s number of mistakes on practice tasks and time spent on practice tasks do not have a significant effect on performance (i.e., \textit{Employee performance realization} or \textit{Employee correct tasks}). However, when we include interactions of employee’s number of mistakes on practice tasks with No SSI and with SSI in
that the employee performance is higher when managers have self-serving incentives indicating that the incentive effect might be strengthened under the SSI condition. Employees anticipate that they cannot rely on the manager with self-serving incentives to give them high bonuses. As a result, they exert more effort in the SSI condition leading to better performance\(^\text{13}\). I also measured employee’s intrinsic motivation for the tasks at the end of the experiment using the intrinsic motivation inventory scale by Deci et al. (1994). Column 3 in Table 6 shows that employees in the No SSI condition have significantly lower intrinsic motivation (b= \(-0.592\); \textit{two-tailed} p-value=0.001) than employees in the SSI condition.

I also compare the employee performance and intrinsic motivation in the SSI condition to that in the SSIDA condition. The results in column 1 and 2 of Table 6 show that in the presence of employee self-evaluations under self-serving incentives, \textit{Employee performance realization} (b= \(-1.138\); \textit{two-tailed} p-value=0.000) and \textit{Employee correct tasks} (b= \(-0.279\); \textit{two-tailed} p-value=0.016) are significantly lower than in compared to the self-serving incentives (SSI) condition. However, it is important to note that the negative effect on performance under the SSIDA condition is lower than under the No SSI condition\(^\text{14}\). Moreover, column 3 in Table 6 reveals that compared to the SSI condition, the intrinsic motivation of employees is not significantly lower (b= \(-0.204\); \textit{two-tailed} p-value=0.123) under the SSIDA condition.

--- Table 6 about here ---

\(^\text{13}\)The performance effects materialize quite immediately in the experiment indicating that employees anticipate that they cannot rely on a manager with self-serving incentives to give them high bonuses. In the first half of the experiment, employee performance realization in no self-serving incentives condition is lower than that in self-serving incentives condition (b= \(-1.915\), two-sided p-value<0.01). The same is true for employee correct tasks (b=\(-0.754\), two-sided p-value<0.01). The results are similar when I include only the first two periods.

\(^\text{14}\)For \textit{Employee performance realization}, under the No SSI condition the performance measure realization of employees was 0.35 times that of employees under the SSI condition. Under the SSIDA condition, the performance measure realization of employees was 0.32 times that of employees under the SSI condition. Both 0.35 and 0.32 are odd’s ratios from the logistic regression. As for \textit{Employee correct tasks}, under the No SSI condition, employees correctly solved 0.299 fewer tasks than under the SSI condition. Under the SSIDA condition, employees correctly solved 0.279 fewer tasks than under the SSI condition. Thus, compared to SSI, the negative impact on performance was lower in SSIDA than in No SSI.
2.5 Summary and discussion

In this study, I examine the impact of managers’ self-serving incentives on the collection and use of information for the purpose of subjective performance evaluation. Consistent with the information avoidance theory (Golman, Hagman and Loewenstein, 2017), I find that managers with self-serving incentives collect less information than managers with no self-serving incentives. Conditional on full information collection, I also find that managers with self-serving incentives interpret information in a more self-interested manner by making lower upward adjustments to employees’ compensation than managers with no self-serving incentives. However, I find that this information avoidance under self-serving incentives is mitigated when employees propose self-evaluations to their compensation and when managers with self-serving incentives observe these self-evaluations at the end of the period. Additional analysis also shows that manager’s self-serving incentives strengthen the incentive effect of performance-based compensation for employees leading to higher employee performance and motivation. However, I find that the difference in employee performance between the self-serving incentives under employee self-evaluation condition and the self-serving incentives condition is less negative than the performance difference between the no self-serving incentives condition and the self-serving incentives condition. Moreover, employees’ intrinsic motivation is not significantly different under the employee self-evaluations with self-serving incentives condition than under the self-serving incentives condition. Thus, the presence of employee self-evaluations under self-serving incentives yields the best overall outcome in terms of both employee performance and information collection and use for subjective performance evaluation.

This study makes several contributions to literature. I contribute to the literature on organizational design choices, particularly delegation and incentives (Milgrom and Roberts, 1995; Abernethy et al., 2004; Moers, 2006; Bouwens and van Lent, 2007; Indjejikian and
According to this literature, the delegation of profit responsibility of a business unit to managers goes hand in hand with using aggregated financial measures to incentivize managers. My study builds on the premise that managers’ self-serving incentives might be an unintended spillover effect of these complementary organizational design choices, such that it is costly for managers with business unit responsibility to give systematically high evaluations to their employees. I show that these self-serving incentives could have serious implications for the way managers collect and use information for subjective performance evaluation. Specifically, I contribute to the management accounting literature by showing that self-serving incentives, which arise because of certain organizational design choices, can encourage managers to avoid information when making subjective performance evaluations. However, I also find that managers’ self-serving incentives offer the benefit of strengthening the incentive effect of performance-based compensation for employees, leading to higher employee performance and motivation.

My study also contributes to the subjective performance evaluation literature in several ways. Contrary to analytical studies wherein the evaluator is the residual claimant, in most prior empirical studies, the manager as an evaluator is not the residual claimant and therefore does not feel disadvantaged by giving high evaluation or bonus (Moers, 2005; Bol, 2008; Bol and Smith, 2011; Bol et al., 2016). I, however, focus on other situations wherein it is costly for managers to give high evaluations to employees, for example, situations wherein the manager has profit responsibility in a business unit, the manager has a profit-independent team in an agile organizational structure, or the manager is the residual claimant in a bonus pool. The self-serving incentives in my study, therefore, draw parallels with incomplete contract settings where the principal promises to pay the agent well if a subjectively assessed performance standard is met but ex-post has a clear incentive to claim that the standard has not been met in order to save on wages (Bull, 1983; MacLeod and Malcomson, 1989; Baker, et al., 1994).
Previous research has shown that effective subjective performance evaluation requires managers to know how employee actions map to noisy performance measures as uncontrollable factors can impact this mapping (Baker, 1992, 2000; Hölmstrom, 1979). Previous research has shown that managers can collect additional information for this purpose (Bol, 2008, 2011; Bol and Smith, 2011; Maas et al., 2012; Wang and Yin, 2017). Thus, one source of bias in manager’s evaluations is the cost of collecting additional information. In fact, previous empirical literature documents an upward bias in evaluations due to the cost of information collection (Bol, 2008; Bol, 2011). I argue that subjective performance evaluations might be biased downwards because managers’ self-serving incentives might lead them to collect less information about the employee even when that information is almost costless to acquire. Moreover, I find that the downward bias may also result from managers strategically interpreting any additional information they do collect in a manner that allows them to avoid making upward adjustments to their employees’ compensation.

My study also provides additional insights to previous research on how managers’ social preferences make them willing to collect costly information for subjective performance evaluation (Maas et al., 2012). The findings of this research show that due to social preferences of fairness and trust reciprocity, managers might be willing to incur a cost to collect information to reward employees properly based on their effort. On the other hand, my study shows that managers might willfully choose not to collect costless information to avoid being confronted with the situation in the future where their social preferences compel them to make costly evaluation decision to maintain a positive self-image. Specifically, managers might avoid

\[\text{An important way that my study is different from Maas et al. (2012) is based on the research setting. In Maas et al. (2012), managers pay to collect information but giving high employee evaluation does not impact manager’s payoff. Previous research in economics shows that when information is costly such that a rational agent would never choose to pay to get information then people with social preferences would collect costly information to avoid being pooled with rational types. On the other hand, in cases where information is costless, information collection does not have signaling value for prosocial types anymore as a rational agent would get the costless information and not take the costly social action. Another aspect in Maas et al. (2012) is that it deals with relative performance – employees sharing a pie. Perhaps not shielding an employee from uncontrollable negative event is more acceptable if the uncontrollable is non-human.}\]
information collection to have the moral wiggle room to justify their self-interested behavior while maintaining a positive self-image.

My study also adds to the nascent stream of literature about strategic behaviors of managers (Hecht et al., 2018). These research findings show that managers make strategic promotion decisions to increase unit’s profit such that there is an increased probability that low performing employees are promoted out of their unit, and excellent employees stay in their unit. My results provide evidence of another kind of strategic behavior in performance evaluation by showing that when managers have self-serving incentives not to give an upward adjustment to employee compensation, they avoid collecting and using information that indicates an upward adjustment.

I also contribute to the literature on the increasingly popular practice of employee self-evaluations. My study shows how employee self-evaluations can mitigate information avoidance in managers with self-serving incentives when they make subjective performance evaluations. In fact, the presence of employee self-evaluations under managers with self-serving incentives induces the same intrinsic motivation in employees as under managers with self-serving incentives only, and no employee self-evaluation. The presence of employee self-evaluations under managers’ self-serving incentives also has a lower negative effect on employee performance, in contrast to the negative effect on performance observed under the condition with no self-serving incentives. Thus, employee self-evaluations in conjunction with managers’ self-serving incentives might be the best scenario for maximizing the collection and use of information for evaluations purposes while minimizing any negative effects on employee performance.

Like all research, my study is subject to limitations. One limitation that presents opportunities for future research is that information collection in my study is almost costless. Future research could make information costly, for example, through an effort task, and then
investigate whether making information costly affects information collection under the self-serving incentives condition. It is possible that managers under the self-serving incentives condition have greater effort aversion than managers under the no self-serving incentives condition, in which case we would find even lower information collection under self-serving incentives. Another limitation of my study is that managers observe the employee self-evaluations only ex-post and therefore cannot change their discretionary adjustments for the employee in that period. Future research can investigate whether providing the manager with the chance to change his/her evaluations after observing the employee’s self-evaluation would influence information collection and use by managers with self-serving incentives. Moreover, I consider three states which were equally likely: good luck, neutral luck, and bad luck. Previous research by Bol et al. (2015) indicates that managers are less likely to adjust for bad luck that has a high likelihood of occurring over time, in order to encourage employees to take actions to manage future bad luck events. Future research can examine whether impact of managers’ self-serving incentives on information avoidance changes when bad luck is more likely to occur overtime.
2.6 References


**TABLE 1**  
Descriptive statistics

<table>
<thead>
<tr>
<th>Conditions</th>
<th>N</th>
<th>Manager Collection of Task Information</th>
<th>Manager Collection of All Information</th>
<th>Manager Information Collection</th>
<th>Manager Discretionary Adjustments</th>
<th>Employee Performance Realization</th>
<th>Number of Correct Employee Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Self-serving Incentives (No SSI)</td>
<td>258</td>
<td>0.864 [1.00]</td>
<td>0.419 [0.00]</td>
<td>1.28 [1.00]</td>
<td>0.275 [0.50]</td>
<td>0.411 [0.00]</td>
<td>2.80 [3.00]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.343)</td>
<td>(0.494)</td>
<td>(0.689)</td>
<td>(0.650)</td>
<td>(0.493)</td>
<td>(1.23)</td>
</tr>
<tr>
<td>Self-serving Incentives (SSI)</td>
<td>276</td>
<td>0.757 [1.00]</td>
<td>0.341 [0.00]</td>
<td>1.10 [1.00]</td>
<td>-0.237 [1.00]</td>
<td>0.522 [1.00]</td>
<td>3.03 [4.00]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.430)</td>
<td>(0.475)</td>
<td>(0.759)</td>
<td>(0.676)</td>
<td>(0.500)</td>
<td>(1.24)</td>
</tr>
<tr>
<td>Self-serving Incentives with employee self-evaluations (SSIDA)</td>
<td>342</td>
<td>0.828 [1.00]</td>
<td>0.491 [0.00]</td>
<td>1.32 [1.00]</td>
<td>-0.085 [0.00]</td>
<td>0.421 [0.00]</td>
<td>2.85 [3.00]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.378)</td>
<td>(0.500)</td>
<td>(0.745)</td>
<td>(0.582)</td>
<td>(0.495)</td>
<td>(1.22)</td>
</tr>
</tbody>
</table>

|                                         |     | p=0.002 [0.002]                        | p=0.063 [0.063]                       | p=0.005 [0.005]                 | p=0.000 [0.000]                  | p=0.010 [0.010]                  | p=0.012 [0.012]                 |
|                                         |     | p=0.031 [0.031]                        | p=0.000 [0.000]                       | p=0.000 [0.000]                 | p=0.003 [0.003]                  | p=0.013 [0.013]                  | p=0.030 [0.030]                 |

p-levels are two-tailed. The numbers within square brackets are medians, and the numbers within the period parentheses are the standard deviations. The comparisons tests are non-parametric tests (Wilcoxin ranksum test).

*Manager Collection of Task Information*: 1 if manager collects task information, else 0.

*Manager’s Collection of All Information*: 1 if manager collects both task and luck information, else 0

*Manager Information collection*: 2 when manager collects both task and luck information, 1 when manager collects task information, 0 when manager collect no information.

*Employee performance realization*: 1 when employee correctly solves all 4 tasks in a period, else 0.

*Number of Correct Employee Tasks*: Number of tasks that the employee correctly solves in a period.
The figure displays manager information collection when managers observe low performance measure realization. The y-axis shows *Information collection* measured as a categorical variable with 0 as no information collection and 2 as full information collection, and the x-axis shows the experimental conditions, where No SSI = No self-serving incentives; SSI = Self-serving incentives; and SSIDA = Self-serving incentives with employee self-evaluations.
FIGURE 2

Manager information use

The figure displays manager information interpretation when manager observe bad luck information. The y-axis shows Information use measured as manager discretionary adjustment conditional on full information collection, and the x-axis shows the experimental conditions, where No SSI = No self-serving incentives; SSI = Self-serving incentives; and SSIDA = Self-serving incentives with employee self-evaluations.
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) Information Collection</th>
<th>(2) Information Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>No SSI</td>
<td>1.889*</td>
<td>0.390***</td>
</tr>
<tr>
<td></td>
<td>(1.068)</td>
<td>(0.117)</td>
</tr>
<tr>
<td>SSIDA</td>
<td>1.685*</td>
<td>0.205*</td>
</tr>
<tr>
<td></td>
<td>(0.981)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Employee performance realization</td>
<td>-</td>
<td>0.332**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.132)</td>
</tr>
<tr>
<td>Employee correct tasks</td>
<td>0.092*</td>
<td>0.280*</td>
</tr>
<tr>
<td></td>
<td>(0.160)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Constant</td>
<td>5.386</td>
<td>-0.924***</td>
</tr>
<tr>
<td></td>
<td>(1.652)</td>
<td>(0.102)</td>
</tr>
<tr>
<td>Wald χ²</td>
<td>4.11</td>
<td>189.9***</td>
</tr>
<tr>
<td>N</td>
<td>482</td>
<td>209</td>
</tr>
</tbody>
</table>

*p*-levels are two-tailed, * p < 0.10, ** p < 0.05, *** p < 0.01; The period parentheses contain robust standard errors clustered by manager; model 1 is mixed effects ordered logistic regression, whereas model 2 mixed effects regression. Model 1 is conditional on performance measure realization of 0, whereas model 2 is conditional on full information collection in bad luck periods. I control for performance measure realization and employee correct tasks in model 2. All models include random effects to control for unobserved heterogeneity over time. The baseline condition is SSI (self-serving incentives).

No SSI: 1 when no self-serving incentives condition, else 0.
SSIDA: 1 when self-serving incentives and employee self-evaluations condition, else 0.
Information collection: 2 when manager collects both task and luck information, 1 when manager collects task information, 0 when manager collect no information.
Information use: manager’s discretionary adjustment conditional on full information collection
Employee performance realization: 1 when employee correctly solves all 4 tasks in a period, else 0.
Employee correct tasks: Number of tasks that the employee correctly solves in a period.
### TABLE 3
Descriptives based on luck

**Panel A: Good luck**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Manager Information Collection</th>
<th>Manager Discretionary Adjustments</th>
<th>Employee Performance Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Conditions</strong></td>
<td>292</td>
<td>1.11 [1.00] (0.744)</td>
<td>0.238 [0.50] (0.647)</td>
<td>0.914 [1.00] (0.280)</td>
</tr>
<tr>
<td><strong>No SSI</strong></td>
<td>86</td>
<td>1.15 [1.00] (0.712)</td>
<td>0.643 [0.80] (0.491)</td>
<td>0.872 [1.00] (0.336)</td>
</tr>
<tr>
<td><strong>SSI</strong></td>
<td>92</td>
<td>0.99 [1.00] (0.748)</td>
<td>–0.083 [0.05] (0.705)</td>
<td>0.923 [1.00] (0.267)</td>
</tr>
<tr>
<td><strong>SSIDA</strong></td>
<td>114</td>
<td>1.18 [1.00] (0.759)</td>
<td>0.193 [0.40] (0.532)</td>
<td>0.938 [1.00] (0.241)</td>
</tr>
</tbody>
</table>

**Panel B: Bad luck**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Manager Information Collection</th>
<th>Manager Discretionary Adjustments</th>
<th>Employee Performance Realization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>All Conditions</strong></td>
<td>438</td>
<td>1.32 [1.00] (0.735)</td>
<td>–0.221 [-0.30] (0.598)</td>
<td>0.118 [0.00] (0.324)</td>
</tr>
<tr>
<td><strong>No SSI</strong></td>
<td>129</td>
<td>1.36 [1.00] (0.672)</td>
<td>–0.006 [0.00] (0.611)</td>
<td>0.078 [0.00] (0.268)</td>
</tr>
<tr>
<td><strong>SSI</strong></td>
<td>138</td>
<td>1.15 [1.00] (0.773)</td>
<td>–0.368 [-0.50] (0.636)</td>
<td>0.196 [0.00] (0.398)</td>
</tr>
<tr>
<td><strong>SSIDA</strong></td>
<td>171</td>
<td>1.41 [2.00] (0.733)</td>
<td>–0.264 [-0.30] (0.506)</td>
<td>0.088 [0.00] (0.283)</td>
</tr>
</tbody>
</table>

In the last three columns, the numbers without brackets are means, the numbers within square brackets are medians, and the numbers within the period parentheses are the standard deviations.

*No SSI*: 1 when no self-serving incentives condition, else 0.

*SSI*: 1 when no self-serving incentives condition, else 0.

*SSIDA*: 1 when self-serving incentives and employee self-evaluations condition, else 0.
**TABLE 4**  
*Regression results for high performance realization and good luck*

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) Information Collection</th>
<th>(2) Information Use</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>No SSI</em></td>
<td>0.644 (0.657)</td>
<td>0.499*** (0.142)</td>
</tr>
<tr>
<td><em>SSIDA</em></td>
<td>0.857 (0.641)</td>
<td>0.136 (0.140)</td>
</tr>
<tr>
<td><em>Employee performance realization</em></td>
<td>-</td>
<td>0.630** (0.263)</td>
</tr>
<tr>
<td><em>Employee correct tasks</em></td>
<td>-</td>
<td>0.197** (0.078)</td>
</tr>
<tr>
<td><em>Constant</em></td>
<td>−1.992 (0.476)</td>
<td>−1.190*** (0.141)</td>
</tr>
</tbody>
</table>

| Wald $\chi^2$                                | 1.92                        | 150.53***           |
| *N*                                          | 394                         | 99                  |

*p*-levels are two-tailed, *p* < 0.10, **p* < 0.05, ***p* < 0.01; The period parentheses contain robust standard errors clustered by manager; model 1 is mixed effects ordered logistic regression, whereas model 2 mixed effects regression. Model 1 is conditional on performance measure realization of 1, whereas model 2 is conditional on full information collection in good luck periods. I control for performance measure realization and employee correct tasks in model 2. All models include random effects to control for unobserved heterogeneity over time. The baseline condition is *SSI* (self-serving incentives).

*No SSI*: 1 when no self-serving incentives condition, else 0.

*SSIDA*: 1 when self-serving incentives and employee self-evaluations condition, else 0.

*Information collection*: 2 when manager collects both task and luck information, 1 when manager collects task information, 0 when manager collect no information.

*Information use*: manager’s discretionary adjustment conditional on full information collection.

*Employee performance realization*: 1 when employee correctly solves all 4 tasks in a period, else 0

*Employee correct tasks*: Number of tasks that the employee correctly solves in a period.
TABLE 5
Time spent looking at information

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) All Information</th>
<th>(2) Luck Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No SSI</strong></td>
<td>−0.0009 (0.0019)</td>
<td>−0.0004 (0.0019)</td>
</tr>
<tr>
<td><strong>SSIDA</strong></td>
<td>0.0036 (0.0030)</td>
<td>0.0051* (0.0030)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>0.0188 (0.0038)</td>
<td>0.0083*** (0.0020)</td>
</tr>
<tr>
<td><strong>F-statistic</strong></td>
<td>4.17**</td>
<td>3.93**</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>876</td>
<td>876</td>
</tr>
</tbody>
</table>

*p*-levels are two-tailed, *p* < 0.10, **p* < 0.05, ***p* < 0.01; The period parentheses contain robust standard errors clustered by manager; model 1 and model 2 are mixed effects regression. I control for performance measure realization and employee correct tasks in both models. All models include random effects to control for unobserved heterogeneity over time. The baseline condition is SSI (self-serving incentives).

**No SSI**: 1 when no self-serving incentives condition, else 0.
**SSIDA**: 1 when self-serving incentives and employee self-evaluations condition, else 0.
**All information**: Time (in seconds) that the manager spent looking at all the information, i.e., the information about the number of tasks that the employee correctly solved as well as the information about the luck in the period.
**Luck Information**: Time (in seconds) that the manager spent looking at the information about luck in the period, i.e., the combination of 2x2 matrices vs. 6x6 matrices tasks that the employee got during the period.
### TABLE 6

**Employee’s performance**

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) Employee Performance Realization</th>
<th>(2) Employee Correct Tasks</th>
<th>(3) Employee Intrinsic Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No SSI</strong></td>
<td>-1.067*** (0.295)</td>
<td>-0.302*** (0.107)</td>
<td>-0.592*** (0.169)</td>
</tr>
<tr>
<td><strong>SSIDA</strong></td>
<td>-1.155*** (0.305)</td>
<td>-0.283** (0.116)</td>
<td>-0.204 (0.131)</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>1.841 (0.329)</td>
<td>2.187*** (0.115)</td>
<td>4.971*** (0.106)</td>
</tr>
<tr>
<td><strong>Wald $\chi^2/F$-statistic</strong></td>
<td>148.9***</td>
<td>779.3***</td>
<td>4.65***</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>730</td>
<td>730</td>
<td>146</td>
</tr>
</tbody>
</table>

*p*-levels are two-tailed, *p < 0.10, **p < 0.05, ***p < 0.01; The period parentheses for model 1 and 2 contain robust standard errors clustered by employee; model 1 is mixed effects logit regression, whereas model 2 is mixed effects regression. Model 1 and 2 include random effects to control for unobserved heterogeneity over time. I control for lag manager discretionary adjustment, lag manager information collection, and luck during the period. Model 3 is ordinary least squares regression evaluating employee intrinsic motivation for the tasks at the end of the experiment. I control for average manager discretionary adjustment in all periods. The baseline condition in all models is SSI (self-serving incentives).

**No SSI**: 1 when no self-serving incentives condition, else 0  
**SSIDA**: 1 when self-serving incentives and employee self-evaluations condition, else 0  
**Employee performance realization**: 1 if employee correctly solves all 4 tasks in a period, else 0  
**Employee correct tasks**: Number of tasks the employee correctly solved in a period  
**Employee Intrinsic motivation**: Mean employee intrinsic motivation for the tasks
Appendix A

2x2 Matrices Task

![2x2 Matrices Task](image)

This figure depicts an example of real-effort task with 2x2 matrices, which were labelled as box 1 and box 2. To solve the task, participants had to find the highest number in each of the two boxes and add these two numbers.

6x6 Matrices Task

![6x6 Matrices Task](image)

This figure depicts an example of real-effort task with 6x6 matrices, which were labelled as box 1 and box 2. To solve the task, participants had to find the highest number in each of the two boxes and add these two numbers.
Appendix B

In this appendix, we provide an example of four screens presented to managers and employees in period 1.

**Employee**

At the start of the period, the employee gets a warning screen.

![Tasks - Warning. Round 1 of 6](image)

After clicking on the button to start working on tasks, employee gets the task screen with 4 tasks (partial view with one task shown below):

![Your tasks. Round 1 of 6](image)
Manager

For the sake of this example, suppose employee did not correctly solve all 4 tasks in the period. The example is from no self-serving incentives condition (No SSI) so manager’s payoff does depend on discretionary adjustment and would be 2.00 in this case. For SSI and SSIDA, the only thing different on the screen would the manager’s payoff calculation.

Results

For the sake of this example, suppose employee did not correctly solve all 4 tasks in the period and manager’s adjustment was -0.40. The following screen is from No SSI condition so manager’s payoff remains 2.00.
Chapter 3

Demand-driven Feedback Systems, Recordkeeping and Easy Task Prioritization

Co-authored with Bart Dierynck
3.1 Introduction

Previous research has shown that people need to have control and autonomy over their behavior and actions at the workplace in order to feel engaged and motivated (Ryan and Deci, 2000; Gagné and Deci, 2005; Deci and Ryan, 2014). One important aspect of job where employees could exercise control is feedback. Traditionally, employees have had little control over feedback and were subject to annual review process with a single instance of feedback annually. However, recent technological advancements have changed the feedback process in organizations from a discretionary process that happens once a year at the initiative of the supervisor to a demand-driven process often allowing employees to ask feedback whenever they want. Employees’ enhanced control over the feedback process under these demand-driven feedback systems have increased employee satisfaction and motivation (Impraise, 2017). While demand-driven feedback systems could increase employee motivation, this study examines whether there is an unintended consequence of demand-driven feedback systems in a setting where employees have discretion over managing, organizing, and completing their tasks. Specifically, we investigate whether demand-driven feedback systems influence employees’ susceptibility to prioritize easy tasks over difficult tasks (labelled as easy task prioritization). We next examine whether and how to mitigate this unintended consequence of demand-driven feedback systems.

To examine the effect of demand-driven feedback systems on easy task prioritization, we develop a laboratory setting using a counting task that allows us to vary whether the task is easy (takes shorter time to complete and requires less cognitive effort). Participants had to complete counting tasks in a common or never-ending queue for 10 minutes and could choose between easy vs. difficult counting tasks each time they proceed to the next task. Participants were paid for their time and earned a fixed compensation. To examine the influence of demand-driven feedback systems, in our main experiment, we consider a baseline case in our research
setting where individuals get no feedback during the period and only get feedback at the end of a period (*No demand-driven feedback system*) and a case where individuals have discretion over whether and how frequently they get feedback during the period (*Demand-driven feedback system*). Feedback provides information about individual’s performance on a task.\textsuperscript{16}

Information avoidance theory suggests that feedback might directly enter an individual’s utility function, leading to immediate gratification when this positive feedback can be obtained quickly (see, Golman, Hagmann, and Loewenstein, 2017; Andries and Haddad, 2017). As easy tasks can be successfully completed in a short time, they are more likely to generate positive feedback quickly. As a result, immediate feedback-related gratification is easier in case of easy versus difficult tasks. Compared with traditional feedback systems, which delay feedback-related gratification until the end of the period, demand-driven feedback systems allow to get immediate feedback-related gratification. Thus, if individuals have a tendency towards completing easy tasks over difficult tasks, then we expect this tendency to worsen in the presence of a demand-driven feedback systems. As a result, demand-driven feedback would increase people’s tendency for easy task prioritization. Results of our experiment confirm our predictions by showing that participants complete a higher percentage of easy tasks than difficult tasks in the demand-driven feedback condition compared to no demand-driven feedback condition. Supplemental analysis also shows that performance on difficult tasks is significantly lower under demand-driven feedback, potentially leading to an adverse impact on long-term productivity and learning.

\textsuperscript{16} We use outcome feedback in our experiment as it closely matches our setting where demand-driven feedback solves the lack of motivation problem. This is because outcome feedback has been shown to impact motivation and self-efficacy without much impact on leaning course of action or strategies (Briers et al., 1999; Earley et al., 1990; Kluger and DeNisi, 1996). Using outcome feedback, therefore, allows us to provide a clean test of our theory about immediate feedback-related gratification from demand-driven feedback without introducing confounds such as learning.
In the next part of our main experiment, we investigate how demand-driven feedback systems can be modified to mitigate the unintended consequence of easy task prioritization. We, therefore, consider whether recordkeeping mitigates easy task prioritization strengthened by demand-driven feedback systems by studying a third condition where the employees have a record of the percentage of easy vs. difficult tasks they already completed (*Demand-driven feedback with recordkeeping*). Since easy task prioritization is a self-control problem, any solution that increases self-control in individuals would mitigate easy task prioritization under demand-driven feedback system. One way to do so is by improving working memory capacity as working memory has been shown to affect people’s ability to exert self-control in their behavior (Cauffman, Steinberg, and Piquero, 2005; Hoffmann et al., 2008; Schmeichel, Volokhov, and Demaree, 2008). Previous research in psychology shows that external representations such as records or written down information serve as memory aids by freeing up working memory capacity (Zhang and Norman 1994; Zhang, 2000). Accounting research supports this by showing that recordkeeping improves memory of previous tasks solved and facilitates decision coordination (Basu and Waymire, 2006; Basu, Dickhaut, Hecht, Towry and Waymire, 2008). Once employees observe the record of easy tasks completed versus difficult tasks completed, it might influence them to do fewer easy tasks since improved memory capacity will make them less likely to get swept away by their preference for immediate gratification. Findings show that recordkeeping helps in mitigating easy task prioritization that is strengthened by the use of demand-driven feedback systems. Specifically, the demand-driven feedback with recordkeeping condition has a significantly lower percentage of easy tasks

---

17 While easy task prioritization emerges as an unintended consequence of demand-driven feedback system, estimating the true impact of easy task prioritization is outside the scope of our study. However, there could be cases where easy task prioritization might have a negative impact on an organization by reducing long-term productivity e.g. sacrificing R&D activities (difficult task) in favor of following established routines and strategies (easy task).

18 In our experiment, this record is kept by the demand-driven feedback system and displayed to the employee so he has the history of tasks he/she completed.
completed compared to the demand-driven feedback condition. To provide a strong test of our theory, we also analyze how variation in individual memory capacity impacts the usefulness of recordkeeping. Consistent with our theory that working memory capacity improves with mindfulness, further analysis shows that people who are less mindful exhibit the strongest decrease in easy task prioritization when there is recordkeeping of tasks completed in the past. Therefore, recordkeeping is most useful in reducing easy task prioritization for people who are less mindful.

Next, as an extension to our main experiment, we consider whether two additional modifications of demand-driven feedback systems have incremental effects over recordkeeping in reducing easy task prioritization strengthened by the presence of demand-driven feedback systems. The first modification is nested within the demand-driven feedback system with recordkeeping: in demand-driven feedback system with planning, individuals also make a plan beforehand about the percentage of easy vs. difficult tasks they want to do (Demand-driven feedback with planning). The second modification is nested within the demand-driven feedback system with planning and provides individuals with an advice which type of task they should complete to realize their plan before choosing the next type of task (Demand-driven feedback with dynamic sequencing). Based on previous research in economics and psychology, planning tasks prospectively instead of selecting tasks in the moment allows all consequences of task selection to be sufficiently far in the future such that immediate gratification is less of a problem (Read, Loewenstein and Kalyanaraman, 1999; O’Donoghue and Rabin, 2006). Thus, when individuals make a plan, they are more likely to choose a diverse mixture of easy and difficult tasks. On the other hand, individuals might anticipate the benefits they derive from immediate gratification and plan a higher percentage of easy than difficult tasks as it will allow them to get more positive feedback over time. Another reason planning might not work is because individuals face cognitive limitations in selecting the next task to
meet their plan, so they revert to selecting tasks based on immediate gratification. Dynamic sequencing, where the system suggests the next task based on the plan, might help in such a case. Findings show that both planning and dynamic sequencing conditions do not incrementally reduce easy task prioritization over recordkeeping in demand-driven feedback systems.

This study makes several contributions. Firstly, we contribute to accounting research by providing evidence of how modern feedback systems impact employee behavior. Previous research in accounting has looked at feedback where the feedback timing and frequency is determined ex-ante in the system (Casas-Arce et al, 2017; Thornock, 2017). With the advent of technology, however, employees have the choice whether, and how frequently to seek feedback through, for instance, mobile apps. Although recent survey evidence suggests that 85% of the millennials feel more confident with demand-driven feedback systems (TriNet, 2015), we show these demand-driven feedback systems might have unintended consequences as such systems strengthen the tendency to select easy tasks over difficult tasks or easy task prioritization.

Secondly, by studying possible modifications to demand-driven feedback systems, we provide theory-grounded guidance on how to mitigate easy task prioritization in the presence of demand-driven feedback systems. By doing so, we contribute to the scant accounting literature on recordkeeping (Clark, 1984; Basu and Waymire, 2006; Basu, Dickhaut, Hecht, Towry and Waymire, 2008) and also contribute to the literature on making choices prospectively rather than in the moment (Simonson, 1990; Read and Loewenstein, 1995; Read, Loewenstein and Kalyanaraman, 1999; Bernartzi and Thaler, 1995; Kahneman and Lovallo, 1993). Next to these contributions to the academic literature, our findings about the modifications to demand-driven feedback systems provide guidance to the developers of mobile apps on how the apps can be changed when easy task prioritization could be an issue.
Finally, we contribute to prior accounting research about how different types of individuals make different accounting decisions (e.g., Feichter 2016; Wang and Yin 2017; Arshad, Dierynck, and van Pelt, 2018) or respond to accounting systems differently (e.g. Upton 2009; Hales, Wang, and Williamson 2015; Wang 2017). Several archival accounting studies have also shown how heterogeneity in style across managers impacts corporate decision making (e.g., Bertrand and Schoar 2003; Fee, Hadlock, and Pierce 2013; Jia, Lent, and Zeng 2014). Apart from providing a strong test for our theory on recordkeeping, our findings on mindfulness trait add to this stream of literature by showing that recordkeeping especially works in reducing easy task prioritization in demand-driven feedback systems for employees with low level of mindfulness. These findings can help the developers of mobile apps to adapt the demand-driven feedback systems to the personal characteristics of employees.

3.2 Literature review and hypothesis development

3.2.1 Easy task prioritization in demand-driven feedback systems

Information that emerges from performance measurement and control systems is often communicated to employees in the form of feedback (Luckett and Eggleton, 1991; Pitkanen and Lukka, 2011; Cases et al., 2015; Loftus and Tanlu, 2017). Traditionally this feedback was provided to employees at the end of the period. With the advent of technology, organizations are using more demand-driven feedback systems to provide employees more autonomy over getting feedback so that they can get information whenever they want it. For instance, in firms such as PwC employees can ask for performance feedback from their supervisors at any given point in time instead of waiting for the annual performance review. Many firms, such as Deloitte and Adobe, have switched from an annual review to demand-driven feedback systems leading to an increase in employee satisfaction and motivation (Impraise, 2017).

In many organizations, employees have discretion over the selection and scheduling of tasks because there is a higher information asymmetry between supervisor and employee
leading to uncertainty about task management. In these cases, employees might have better information to implement their activities as it is too costly for the supervisors to structure tasks for each employee (van Donselaar et al. 2010, Campbell and Frei 2011, Kim et al. 2015, Phillips et al. 2015). Even though such employee discretion over tasks can be value-increasing, operations research also shows that employees may choose to prioritize tasks that they expect to complete faster as completing tasks quickly allows for positive emotion or gratification and sense of progress. Such behavior is referred to as easy task prioritization in our study. Such easy task prioritization could potentially come at the expense of long-term productivity (Ibanez et al., 2017; KC et al., 2018).

As it is easier to get immediate gratification by getting instant feedback, demand-driven feedback systems might lead to unintended consequence of exacerbating easy task prioritization. Recent theoretical developments in information avoidance suggest that if utility depends on not only material outcomes but also on beliefs and attention paid to them then feedback might directly enter the utility function of employees (Andries and Haddad, 2015; Golman et al., 2016). Feedback, therefore, directly satisfies the belief-based component of utility such that individuals would acquire feedback because of curiosity and motivated attention. This means that individuals get utility from fulfilling their curiosity as well as utility from getting information that is expected to be good (and conversely avoiding information that is expected to be bad). Consistent with Information avoidance theory, demand-driven feedback allows individuals to satisfy belief-based component of utility and derive immediate gratification as feedback can be obtained quickly. Immediate feedback-related gratification is easier in case of easy versus difficult tasks because easy tasks are more likely to generate positive feedback quickly. Higher utility and immediate gratification from receiving positive

---

19 Studying the full impact of easy task prioritization is outside the scope of this paper. However, in some cases easy task prioritization might lead to negative impact on long-term productivity (Ibanez et al., 2017; KC et al., 2018).
feedback might therefore motivate employees to seek the experience again through completing easy tasks over difficult tasks and then seeking feedback. Thus, we predict that the demand-driven feedback system will strengthen easy task prioritization in a setting where employees have discretion over tasks.

**H1**: *Compared to no feedback system, employees are more likely to select easy tasks over difficult tasks in the presence of demand-driven feedback system.*

### 3.2.2 Recordkeeping and easy task prioritization in demand-driven feedback systems

Since easy task prioritization under demand-driven feedback systems is a self-control problem, so any intervention that increases self-control in individual should mitigate easy task prioritization. One way to increase self-control is by increasing working memory capacity in an individual. Working memory is a cognitive system with a limited capacity that is responsible for temporarily holding information available for processing and enabling controlled attention to maintain information in an active, quickly retrievable state. Research has shown that working memory is a strong predictor of higher self-control in behavior (Cauffman, Steinberg, and Piquero, 2005; Hoffmann et al., 2008; Schmeichel, Volokhov, and Demaree, 2008). Maintaining self-control requires working memory resources as working memory may be needed to both inhibit strong impulsive responses as well as to maintain controlled dispositions in an active state (Hoffmann et al., 2008). People’s ability to exert higher self-control in their behavior, therefore, is determined by higher working memory.

Previous literature in psychology has shown that external representations (such as written or displayed information) can serve as a memory aid by freeing up working memory capacity (Zhang and Norman 1994; Zhang, 2000). In our study, we use one such external representation in the form of recordkeeping whereby employee has a record of easy vs. difficult tasks he/she completed. Role of recordkeeping in facilitating memory is also supported by accounting literature. Basu and Waymire (2006) state that basic recordkeeping enables
transformation of information into hard records and a “paper trail” that institutionalizes memory and verification of past events. Recordkeeping has been shown to sustain trust and reciprocity in complex economic interactions between strangers over time by supplementing imperfect human memory and thus facilitates increasingly coordinated actions between people in exchange relations (Basu and Waymire, 2006; Basu, Dickhaut, Hecht, Towry and Waymire, 2008). Since recordkeeping aids working memory, it can therefore improve self-control in behavior and curb the preference for immediate gratification.

Consistent with the idea that recordkeeping might enable higher self-control by supplementing working memory capacity, individuals will be less likely to get swept away by their preference for immediate gratification when they observe their record of the tasks (easy tasks versus difficult tasks) they completed in the past. Although it is possible that individuals ignore or disregard the records of type of tasks they completed in the past, such that recordkeeping has no effect on easy task prioritization in demand-driven feedback systems, we formulate the following hypothesis:

**H2: Compared to demand-driven feedback system without recordkeeping, employees are less likely to select easy tasks over difficult tasks in demand-driven feedback system with recordkeeping**

### 3.2.3 Planning, dynamic sequencing and easy task prioritization in demand-driven feedback systems

Apart from recordkeeping, it is important to see if more costly interventions to the self-control problem of easy task prioritization would work better in our setting. For this purpose, we extend the main study to examine solutions to self-control problem discussed in the fields of economics and consumer behavior such as planning. Previous research shows that when individuals make choices prospectively ahead of time as opposed to making choices sequentially in time, they choose a more diverse consumption bundle (Simonson, 1990; Read and Loewenstein, 1995; Read, Loewenstein and Kalyanaraman, 1999); show higher risk-
seeking behavior (Bernartzi and Thaler, 1995; Kahneman and Lovallo, 1993); are more attentive of delayed consequences (Ainslie, 1992; Heyman, 1996; Donoghue and Rabin, 2006). This is in part because while sequential in-the-moment choices might mean pandering to short-time inclinations of what to do, people pay greater attention and place more weight on delayed consequences when they make prospective choice (Read, Loewenstein and Kalyanaraman, 1999; Donoghue and Rabin, 2006). When individuals plan the percentage of easy and difficult tasks prospectively instead of selecting tasks in the moment, they might therefore perceive consequences of task selection to be sufficiently far in the future such that searching for immediate gratification has a lower impact on behavior. Thus, when employees are prompted to plan the percentage of easy and difficult tasks, they are more likely choose a diverse mixture of easy and difficult tasks.

On the other hand, planning might not work because of several reasons. Based on goal setting theory, people might set less difficult goals for themselves than if they were assigned goals, possibly because of low self-efficacy (Locke, Frederick, Lee, and Bobko, 1984; Schunk, 1990). Employees might therefore select an easier plan to accomplish by planning to complete a higher percentage of easy than difficult tasks. Moreover, setting such a plan where they select a higher percentage of easy tasks to complete would increase their utility because it not only will be easier to complete but also will allow them to get more positive feedback over time. Thus, planning of tasks in the presence of demand-driven feedback systems with recordkeeping might not work in the best possible way to mitigate easy task prioritization or might even worsen it.

Another reason planning might not lower easy task prioritization is because people face cognitive limitations hindering their ability to select the next task that best meets their plan inducing them to revert to selecting tasks based on immediate gratification. Planning to execute a particular percentage of easy and difficult tasks requires that, every time an individual starts
a new task, he/she determines which task to execute next to meet the planned percentage of easy and difficult tasks. Such a calculation is not straightforward. Therefore, a system that calculates which type of task people should select next to meet their plan should reduce the cognitive resources that people need to spend to figure out which task to select next. Such dynamic sequencing with planning might then help in reducing easy task prioritization in demand-driven feedback systems. However, such dynamic sequencing might not work to mitigate easy task prioritization if people choose an easier plan for the reasons mentioned above. Dynamic sequencing might also be worse if people perceive that they have less choice as the feeling of reduced autonomy can reduce intrinsic motivation. We therefore arrive at the following research questions:

RQ:

a) Does planning incrementally affect employees’ tendency to select easy tasks over difficult tasks, compared to demand-driven feedback system with recordkeeping?

b) Does dynamic sequencing incrementally affect employees’ tendency to select easy tasks over difficult tasks, compared to demand-driven feedback system with planning?

3.3 Main experiment

3.3.1 Overview

We conducted a laboratory experiment with o-Tree software. In order to pursue our research goals, we have a nested between-subjects design. Our main experiment tests hypothesis 1 and 2. In the baseline case, there is no demand-driven feedback during the period and only feedback at the end of the period- no demand-driven feedback system (NF). In the condition with a demand-driven feedback system, participants have choice whether, and when to get feedback during the period- demand-driven feedback system (DTF). We then vary whether the participants also see a record of the percentage of completed easy tasks vs. difficult tasks- demand-driven feedback system with recordkeeping (DTFR).
All the participants were randomly assigned to one of the three conditions. We used a modified version of the problem-solving task in Weber and Schram (2017). Each participant observes two matrices on the screen and must find the highest number in the left matrix and the highest number in the right matrix and then add the two numbers. For the summation, participants are not allowed to use calculators. In our experiment, each participant can choose between two types of the task varying in the time it takes to complete them: 5x5 matrices and 10x10 matrices (see Appendix I). If the participant chooses the easy (difficult) task, he/she sees two 5x5 matrices (10x10 matrices) on the screen that are filled with randomly generated three-digit numbers. Participants have 10 minutes to solve these tasks and have to make choices regarding the type of task they want to do next every time they want to start a new task. There are several reasons why this task provides a good test of our theory. First, by keeping the way of solving each task the same (i.e. addition of the largest number in each matrix), we keep constant the analytical ability to solve the two tasks. The two tasks mainly differ in how long it takes a participant to complete the task, with the 5x5 matrices tasks taking a shorter time to complete than 10x10 matrices task. This provides a clean test of our theory that easy tasks allow participants to get positive feedback quickly, thereby exacerbating the preference for immediate gratification. Second, as the nature of the task does not vary across the two types, our task also allows us to rule out the confounding effect of learning as the reason why people would choose easy tasks over difficult tasks, especially in the beginning of the period.

All the participants were paid a fixed compensation because we wanted to isolate how demand-driven feedback affects the way individuals choose easy tasks over difficult tasks without any spillover effects of monetary incentives on task selection. Incentives based on task

---

20 The original task was developed in o-Tree by Philipp Chapkovski.
21 In the experiment, the words, easy task or difficult task, are never used to avoid any demand effects. We only label the tasks as 5x5 matrices task and 10x10 matrices task.
22 Analysis shows that correctly completing a 5x5 matrices task took on average approximately half the time than correctly completing a 10x10 matrices task.
complexity might interact with our manipulations and thereby confound the effect of demand-driven feedback on easy task prioritization. Thus, for an internally valid test of our theory on demand-driven feedback systems, we use a fixed compensation instead of incentives based on task complexity.\textsuperscript{23}

\textbf{3.3.2 Participants}

We recruited business and economics undergraduate students who responded to an email invitation. Participants’ age ranges from 19 to 30 years old, and 61 percent of the participants are male. A large majority of the participants indicated to have part-time or full-time work experience (91.5 percent), and all participants had completed at least one math course, an economics course, an accounting course, and a finance course at university level. As a show-up incentive, we introduced a modest amount of bonus course credit on top of their total grade. In addition to this show-up incentive, each participant also earned fixed monetary compensation of EUR 6 for participation in the experiment. Before running the study, we obtained approval to run the study from the research institute.

\textbf{3.3.3 Experimental manipulations and procedures}

Upon their arrival in the lab of the research institute, participants had to wait briefly in a waiting room and publicly received some basic instructions before moving into the computer lab (e.g., no talking, and no electronic devices). When entering the computer lab, participants randomly chose one cubicle containing a computer. When we launched the study, participants first entered an instruction phase where they read information about their role, payoffs, and specific information related to their treatment. This phase included instructions and a few basic

\textsuperscript{23} This design choice is also reflective of practice as many organizations pay for employee’s time by giving a fixed salary, especially in lower level jobs where demand-driven feedback is more likely to be used. To use incentive contracting, the principal needs to have sufficient information about the state of the nature to develop a target. Often, the principal does not have that information. For instance, in the service industries, whether one can meet a certain target of easy jobs does not only depend on the agent but also on fluctuations in the demand and supply for tasks. Contracting on performance measures related to task management might, thus, be too difficult and costly.
control questions. When participants gave wrong answers to the control questions, our software provided feedback to participants. We carefully designed a concise set of instructions and control questions that only checked whether participants understood the basics of the study. The participants also solved one easy task (5x5 matrices) and one difficult task (10x10 matrices) during the instruction phase as practice before the actual task phase.

After the instruction phase, participants in the three conditions (NF, DTF, DTFR) entered the task phase, where they complete counting tasks for 10 minutes. During the task phase, participants in all conditions first choose the type of task they want to do next (5x5 matrices or 10x10 matrices) and then they solve the task. After solving each task, participants in the no demand-driven feedback condition (NF) proceed to selecting the next type of task without getting feedback on whether they were correct. Participants in the two other conditions with demand-driven feedback (DTF, DTFR) can choose after solving each task, whether they want to proceed directly to selecting the next type of task, or whether they want to get feedback on their answer and then proceed to selecting the next type of task (see, Appendix II). In the demand-driven feedback with recordkeeping condition (DTFR), when selecting the next type of task, participants also always see a history bar that shows a record of the percentage of 5x5 matrices and the percentage of 10x10 matrices they have completed so far.⁴ In all conditions, participants were made aware of the time left through a message that appeared when 5 minutes were remaining and when 1 minute was remaining.

After 10 minutes, participants in all conditions observe the number of tasks they completed and the number of tasks they correctly completed. Afterwards, they completed an

---

⁴ One argument could be that since the system keeps a record of the easy versus difficult tasks completed in the recordkeeping condition (DTFR), it is now possible for the principal to contract on it. However, it should be noted that the record in the system is based on the task labels i.e., number of 5x5 matrices tasks versus 10x10 matrices tasks completed. The difficulty level of each task completed is only known and experienced by the employee. Thus, the system still does not know the difficulty level of the different tasks completed for the principal to be able to contract on it.
ex-post questionnaire containing several items intended to provide some insight into the thoughts and feelings of the participants during the study.

### 3.4 Results of main experiment

#### 3.4.1 Descriptive statistics

Table 1 presents descriptive statistics by experimental condition. Participants in the no demand-driven feedback condition (NF) completed 13.12 tasks on average of 8.72 which were easy tasks and 4.40 were difficult tasks, whereas, participants in the demand-driven feedback condition (DTF) completed approximately the same number of tasks (13.02) but relatively more easy tasks (10.70) and less difficult tasks (2.32) on average. This provides some preliminary support for our first hypothesis. In the demand-driven feedback for recordkeeping condition (DTFR), the average total number of tasks completed by participants was 11.37 of which 7.61 were easy tasks and 3.76 were difficult tasks. Table 1 also shows the total number of tasks that were correctly completed in each experimental condition. Demand-driven feedback (DTF) had the highest average total number of tasks correct, 9.65, followed by 9.40 average total number of tasks correct under no demand-driven feedback (NF).

--- Insert Table 1 here ---

#### 3.4.2 Main findings

Table 2 shows the formal tests for first hypothesis and research questions. We measure easy task prioritization through the dependent variable *Percentage completed easy tasks*, which is number of easy tasks completed divided by the total number of tasks completed. In the presence of demand-driven feedback (DTF), participants select a significantly higher percentage of easy tasks than when there is no demand-driven feedback (NF) ($Z = 2.619$, two-tailed *p*-value $= 0.008$). This provides support for our first hypothesis that, compared to a setting without demand-driven feedback (NF), employees are more likely to select easy tasks over difficult tasks in the presence of a demand-driven feedback system (DTF). The results
presented in Table 2 also show that participants select a significantly lower percentage of easy tasks under a demand-driven feedback system with recordkeeping \( (DTFR) \), than under a demand-driven feedback system \( (DTF) \) \( (Z = 2.855, \text{ two-tailed } p\text{-value} = 0.004) \). This provides evidence for our second hypothesis that easy task prioritization induced by demand-driven feedback systems is mitigated by demand-driven feedback systems with recordkeeping.

--- Insert Table 2 here ---

3.4.3 Supplemental analysis

3.4.3.1 Performance effects

Previous research has documented that easy task prioritization is detrimental and comes at the expense of long-term productivity (Ibanez et al., 2017; KC et al., 2018). Therefore, it is important to see whether and how the increase in easy task prioritization in demand-driven feedback systems affects performance across tasks. Figure 2 shows the performance effects across the different experimental conditions. The percentage of difficult tasks correctly solved is significantly lower in demand-driven feedback \( (DTF) \) compared to no demand-driven feedback \( (NF) \) \( (Z = 2.782; \text{ two-tailed } p\text{-value} = 0.005) \) whereas it is the opposite case for percentage of easy tasks correctly solved\(^{25}\). These findings about significantly lower performance on difficult tasks indicate a potentially adverse impact on learning and long-term productivity under demand-driven feedback systems. Results also show that this adverse performance effect is mitigated by demand-driven feedback systems with recordkeeping \( (DTFR) \) \( (Z = 2.821; \text{ two-tailed } p\text{-value} = 0.005) \).

3.4.3.2 Mindfulness and Recordkeeping

While we find that recordkeeping mitigates easy task prioritization induced by demand-driven feedback systems, it is important to provide a strong test for our theory on recordkeeping. As working memory capacity is difficult to directly measure, we could test our

\(^{25}\) Results are consistent when using correct difficult (easy) tasks as percentage of total completed tasks.
theory by using a personality trait that has been shown to depend on working memory capacity and then analyze how variation in this personality trait impacts the usefulness of recordkeeping. We expect that mindfulness could be such a personality trait with interesting implications for recordkeeping. Mindfulness is commonly defined as the state of being attentive to and aware of what is taking place in the present. Previous research shows that mindfulness meditation improves working memory capacity (Jha et al., 2010; Mrazek et al., 2013) and reduces the susceptibility to behavioral biases (Kiken and Shook, 2011; Hafenbrack, Kinihas, and Barsade, 2014; Leuke and Gibson, 2015). Summarized, people who are less mindful are more likely to exhibit behavioral biases with lower working memory capacity being an important reason for this effect. In the ex-post questionnaire, we measure mindfulness using instrument developed by Brown and Ryan (2003), where high scores reflect more mindfulness. When splitting the recordkeeping condition based on the mindfulness score, we find that recordkeeping had the strongest effect ($Z= 2.862$, two-tailed $p\text{-value}=0.004$) in reducing easy task prioritization for participants with a low level of mindfulness (i.e. score of 3 or less on a Likert scale going from 1 to 6) (results not tabulated). Recordkeeping also worked for participants with a higher level of mindfulness, but to a smaller extent ($Z= 1.717$, two-tailed $p\text{-value}=0.086$). Thus, in line with the idea that recordkeeping supplements human working memory capacity, we can conclude that recordkeeping works especially well for people with low mindfulness.

3.4.3.3 Demand for feedback

We also analyzed the demand for feedback across different feedback conditions. Demand for feedback is measured as the percentage of times the participants asked for feedback after completing a task during the period. Figure 3 shows that the percentage demand for feedback across all feedback conditions is quite high. About 80% of the times people ask for feedback across all feedback conditions. However, there is no significant different in the demand for feedback across the different feedback conditions.
It is also interesting to examine the demand for feedback for easy and difficult tasks across the feedback conditions. It was not clear ex-ante what we should expect about the demand of feedback for easy vs difficult tasks and which of the two would dominant in any specific feedback condition. Based on the belief-based component of utility, curiosity or motivated attention to get (avoid) good (bad) feedback might drive the demand for feedback under demand-driven feedback (*DTF*) and consequently its effect on easy task prioritization. The analysis in Figure 4 shows that percentage demand for feedback is higher for difficult tasks than easy tasks under demand-driven feedback condition (*DTF*). This means that curiosity is driving the demand for feedback under demand-driven feedback systems, as one would expect to observe a higher demand for feedback when curiosity about the outcome is the higher. Since participants in the demand-driven feedback condition do fewer difficult tasks, they might end up being most curious about their performance in difficult tasks. As a result, demand for feedback for difficult tasks is higher than that for easy tasks under demand-driven feedback. On the other hand, we observe the opposite for recordkeeping (*DTFR*). Since participants perform fewer easy tasks, they would e more curious about the outcome of these easy tasks. Thus, demand for feedback for easy tasks is higher than that for difficult tasks under demand-driven feedback with recordkeeping (*DTFR*).

### 3.5 Extension of main experiment

#### 3.5.1 Overview

Our main experiment shows that the simple intervention of recordkeeping mitigates easy task prioritization in demand-driven feedback systems. Given that easy task prioritization is a self-control problem at its core, we next wanted to examine whether more sophisticated solutions for self-control problem proposed by the fields of economics and psychology would provide incremental benefit over recordkeeping. Based on previous research in economics and psychology, planning tasks prospectively is one such solution that makes immediate
gratification less of a problem (Read, Loewenstein and Kalyanaraman, 1999; O’Donoghue and Rabin, 2006) In the first condition, participants plan the percentage of easy tasks vs. difficult tasks they want to complete during the period and observe this plan in a demand-driven feedback system with recordkeeping- demand-driven feedback system with planning (DTFP). Lastly, in the condition with a demand-driven feedback system that consists of both recordkeeping and planning, we also allow participants to observe the next type of task they should pursue in order to complete their plan before they choose the next type of task- demand-driven feedback with dynamic sequencing (DTFD). Like the main experiment, we used the modified version of the problem-solving task in Weber and Schram (2017) and all the participants were paid a fixed compensation.

3.5.2 Participants

As in the main experiment, we recruited business and economics undergraduate students who responded to an email invitation. Participants’ age ranges from 18 to 31 years old, and 56 percent of the participants are male. A large majority of the participants indicated to have part-time or full-time work experience (97.8 percent), and all participants had completed at least one math course, an economics course, an accounting course, and a finance course at university level. As a show-up incentive, we introduced a modest amount of bonus course credit on top of their total grade. In addition to this show-up incentive, each participant also earned fixed monetary compensation of EUR 6 for participation in the experiment. Before running the study, we obtained approval to run the study from the research institute.

3.5.3 Experimental manipulations and procedures

Upon their arrival in the lab of the research institute, participants had to wait briefly in a waiting room and publicly received some basic instructions before moving into the computer lab (e.g., no talking, and no electronic devices). When entering the computer lab, participants randomly chose one cubicle containing a computer.
Like the main experiment, once the study is launched, participants first entered an instruction phase where they read information about their role, payoffs, and specific information related to their treatment. This phase included instructions and a few basic control questions. After the instructions phase, participants enter the task phase where they complete counting tasks for 10 minutes. Before starting the task phase, participants in both the conditions (DTFP, DTFD) first select on a scale the percentage of easy tasks (5x5 matrices) vs. percentage of difficult tasks (10x10 matrices) they plan to complete during the 10 minutes. They then proceed to the task phase. As in the demand-driven feedback conditions in the main experiment, participants in both conditions (DTFP, DTFD) can choose after solving each task, whether they want to proceed directly to selecting the next type of task, or whether they want to get feedback on their answer and then proceed to selecting the next type of task. In addition to seeing a record when selecting the next type of task, participants in the two planning conditions (DTFP, DTFD) also see a bar that shows the percentage of 5x5 matrices vs. 10x10 matrices they planned to complete. Participants in the last condition of demand-driven feedback with dynamic sequencing (DTFD) also observe the type of task they should do next to meet their plan along with the record and their plan, when selecting the next task. In DTFD, participants are not obliged to do the type of task that is suggested to them. In all conditions, participants were made aware of the time left through a message that appeared when 5 minutes were remaining and when 1 minute was remaining.

After 10 minutes, participants in all conditions observe the number of tasks they completed and the number of tasks they correctly completed. Afterwards, they completed an ex-post questionnaire containing several items intended to provide some insight into the thoughts and feelings of the participants during the study.
3.5.4 Results

3.5.4.1 Descriptive statistics

Table 3 reports the descriptive statistics for the two conditions (DTFP, DTFD) in the extension of the main experiment. Demand-driven feedback with planning (DTFP) and demand-driven feedback with dynamic sequencing (DTFD) had approximately a similar average total number of tasks completed (11.67 and 12.11). The easy tasks (difficult tasks) completed under DTFP and DTFD were 8.13 (8.98) and 3.54 (3.13), respectively. The total number of tasks that were correctly completed were 9.27 and 9.11 in DTFP and DTFD, respectively.

3.5.4.2 Main findings

Next, we investigate whether planning the percentage of easy tasks one wants to complete has an incremental effect over recordkeeping in reducing easy task prioritization induced by demand-driven feedback systems. Demand-driven feedback with planning (DTFP) leads to a significantly lower percentage of easy tasks completed than demand-driven feedback (DTF) ($Z = 2.037$, two-tailed $p\text{-value} = 0.042$) but the percentage of easy tasks is not significantly different from demand-driven feedback with recordkeeping (DTFR) ($Z = 0.909$, two-tailed $p\text{-value} = 0.363$). As a result, we can conclude that planning the percentage of easy tasks one wants to complete does not work better than recordkeeping alone in reducing easy task prioritization induced by demand-driven feedback systems. Moreover, adding dynamic sequencing (DTFD) has no incremental effect in reducing easy task prioritization compared to demand-driven feedback system with planning (DTFP) ($Z = 0.876$, two-tailed $p\text{-value} = 0.381$). Also, the percentage of easy tasks one wants to complete in the condition with demand-driven feedback with dynamic sequencing is not significantly different from demand-driven feedback condition (DTF) ($Z = 1.206$, two-tailed $p\text{-value} = 0.227$).
3.5.4.3 Perceived choice and planning

When comparing the planning condition with the dynamic sequencing condition, we find that dynamic sequencing has no incremental effect. One potential explanation is that people feel restricted and thereby less intrinsically motivated when there is a system telling them what to do next as is the case in dynamic sequencing.\textsuperscript{26} We find that this is indeed the case based on our untabulated analysis. In our ex-post questionnaire, we used a modified version of the perceived choice subscale from the intrinsic motivation inventory in Ryan (1982) and Deci et al. (1994). Despite the fact that participants had the same degree of freedom to choose the next type of task, we find that participants perceive significantly lower freedom of choice in their selection of activities under the demand-driven feedback with dynamic sequencing (\textit{DTFD}) compared to the demand-driven feedback with planning (\textit{DTFP}) ($t_{92} = 1.776$, two-tailed $p$-value = 0.079). The impact of the dynamic sequencing feature on perceived choice combined with the impact of the planning feature difficult might explain why planning with dynamic sequencing (\textit{DTFD}) has no significant effect on easy task prioritization compared to both the demand-driven feedback condition (\textit{DTF}) and the demand-driven feedback with planning condition (\textit{DTFP}).

3.6 Summary and discussion

In this study, we investigate easy task prioritization is strengthened by demand-driven feedback systems, where people have choice about whether, when, and how frequently they want to get feedback. Consistent with the idea that people are searching for immediate gratification and the fact that demand-driven feedback systems satisfy the preference for immediate gratification, we find that easy task prioritization is worsened in the presence of demand-driven feedback systems compared to a situation without demand-driven feedback.

\textsuperscript{26} Perceived choice is theorized to be a positive predictor of both self-reported and behavioral measures of intrinsic motivation (Deci et al., 1994; Cottright et al. 2013)
We also investigate possible ways to mitigate easy task prioritization in demand-driven feedback systems. We find that recordkeeping works best to mitigate easy task prioritization in demand-driven feedback systems, while planning and dynamic sequencing do not work better than recordkeeping to mitigate easy task prioritization. Additional analysis also shows that recordkeeping works especially well to reduce easy task prioritization among people with a low level of mindfulness.

The study contributes to accounting research by providing evidence of how accounting systems impact employee behavior. Previous research in accounting has looked at feedback (Casas et al., 2017; Thornock, 2017) where the feedback timing and frequency are determined ex ante in the system. As technological progress allows employees to choose whether, and how frequently to seek feedback through mobile apps, it is important to examine how these modern feedback systems impact employee behavior. Up to our knowledge, we are the first to analyze the behavioral implications of demand-driven feedback systems in a structured way. Interestingly, we show that demand-driven feedback systems can have negative behavioral consequences by amplifying the propensity of employees to choose easy tasks over difficult tasks.

In our main experiment, we also identify recordkeeping under demand-driven feedback systems that can help mitigate the problem of easy task prioritization induced by demand-driven feedback systems. Our findings about the impact of recordkeeping contribute to the scant literature in accounting on recordkeeping (Clark, 1984; Basu and Waymire, 2006; Basu, Dickhaut, Hecht, Towry and Waymire, 2008). In the extension of our main experiment, we also examine whether other modifications (planning and dynamic sequencing) based in the field of economics and psychology have an incremental benefit over recordkeeping in mitigating the problem of easy task prioritization induced by demand-driven feedback systems. Thus, we also contribute to the extensive literature on prospective choice over sequential choice.
in economics and psychology (Simonson, 1990; Read and Loewenstein, 1995; Read, Loewenstein and Kalyanaraman, 1998; Bernartzi and Thaler, 1995; Kahneman and Lovallo, 1993). Our findings about recordkeeping, planning and dynamic sequencing should also be of interest to the developers of mobile apps that support demand-driven feedback systems. Each of our modifications can be programmed into these mobile apps leading to better demand-driven feedback systems. On top of that, our findings about the differential impact of recordkeeping on easy task prioritization can help the developers of the mobile apps to better tailor the app to the individual needs of employees.

Like all research, we recognize that ours is subject to limitations. One limitation which yields opportunities for future research is that we focus on problem solving tasks that differ in duration to isolate easy task prioritization. While such tasks are part of employee’s work schedule, there might be other features that vary across tasks, such as importance and difficulty. Future research can investigate how demand-driven feedback systems impact task selection when tasks vary with respect to other features. Another limitation of our study is that the demand-driven feedback provided in our setting is not a subjective assessment coming from a supervisor. It is possible that reputation-building involved in the interaction between supervisor and employee impacts how demand-driven feedback systems affect easy task prioritization. Moreover, to provide an internally valid test of theory on demand-driven feedback systems we do not include incentives in our experiment. Future research could study how incentives interact with demand-driven feedback systems to affect easy task prioritization.
3.7 References


## TABLE 1

Descriptive statistics for main experiment

<table>
<thead>
<tr>
<th>Conditions</th>
<th>N</th>
<th>Tasks completed</th>
<th>Easy tasks completed</th>
<th>Difficult tasks completed</th>
<th>Tasks correct</th>
<th>Easy tasks correct</th>
<th>Difficult tasks correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>No demand-driven feedback</td>
<td>40</td>
<td>13.12</td>
<td>8.72</td>
<td>4.40</td>
<td>9.40</td>
<td>6.55</td>
<td>2.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.96)</td>
<td>(5.13)</td>
<td>(3.23)</td>
<td>(3.44)</td>
<td>(4.26)</td>
<td>(2.37)</td>
</tr>
<tr>
<td>Demand-driven feedback</td>
<td>44</td>
<td>13.02</td>
<td>10.70</td>
<td>2.32</td>
<td>9.65</td>
<td>8.27</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(4.05)</td>
<td>(5.03)</td>
<td>(2.02)</td>
<td>(3.67)</td>
<td>(4.15)</td>
<td>(1.30)</td>
</tr>
<tr>
<td>Demand-driven feedback with recordkeeping</td>
<td>46</td>
<td>11.37</td>
<td>7.61</td>
<td>3.76</td>
<td>8.39</td>
<td>5.84</td>
<td>2.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.98)</td>
<td>(4.92)</td>
<td>(2.71)</td>
<td>(3.50)</td>
<td>(4.02)</td>
<td>(1.94)</td>
</tr>
</tbody>
</table>

In the last six columns, the numbers without brackets are means and the numbers within the round parentheses are the standard deviations.
### TABLE 2
Results of main experiment

<table>
<thead>
<tr>
<th></th>
<th>Percentage completed easy tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NF</td>
</tr>
<tr>
<td></td>
<td>63.354</td>
</tr>
<tr>
<td></td>
<td>(29.245)</td>
</tr>
<tr>
<td>N=</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NF vs. DTF</th>
<th>DTF vs. DTFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>2.619</td>
<td>2.855</td>
</tr>
<tr>
<td>p</td>
<td>0.008</td>
<td>0.004</td>
</tr>
</tbody>
</table>

*p*-levels are two-tailed, the numbers within the round parentheses are the standard deviations. The comparisons tests are non-parametric tests (Wilcoxin ranksum test).

*Percentage completed easy tasks*: Percentage of easy tasks completed calculated by:

\[
\text{Percentage} = \frac{\text{Total no. of easy tasks completed}}{\text{Total no. of tasks completed}} \times 100
\]

*NF*: No demand-driven feedback

*DTF*: Demand-driven feedback

*DTFR*: Demand-driven feedback with recordkeeping
The figure displays easy task prioritization in different experimental conditions. The y-axis shows easy task prioritization calculated as *Percentage completed easy tasks*. The x-axis shows the different experimental conditions, where *NF*: No demand-driven feedback; *DTF*: Demand-driven feedback; *DTFR*: Demand-driven feedback with recordkeeping.
The figure displays the performance effects across the different conditions. The blue bars show the percentage of correct difficult tasks and the red bars show the percentage of correct easy tasks. The x-axis shows the different experimental conditions, where NF: No demand-driven feedback; DTF: Demand-driven feedback; DTFR: Demand-driven feedback with recordkeeping.
The figure displays the percentage demand of feedback across conditions. The x-axis shows the different experimental conditions, where NF: No demand-driven feedback; DTF: Demand-driven feedback; DTFR: Demand-driven feedback with recordkeeping.
The figure displays the percentage demand for feedback for easy and difficult tasks across feedback conditions. The blue bars show the percentage demand for easy tasks and the red bars show the percentage demand for difficult tasks. The x-axis shows the different experimental conditions, where NF: No demand-driven feedback; DTF: Demand-driven feedback; DTFR: Demand-driven feedback with recordkeeping.
TABLE 3  
*Descriptive statistics for extension of main experiment*

<table>
<thead>
<tr>
<th>Conditions</th>
<th>N</th>
<th>Tasks completed</th>
<th>Easy tasks completed</th>
<th>Difficult tasks completed</th>
<th>Tasks correct</th>
<th>Easy tasks correct</th>
<th>Difficult tasks correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demand-driven feedback with planning</td>
<td>48</td>
<td>11.67</td>
<td>8.13</td>
<td>3.54</td>
<td>9.27</td>
<td>6.58</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.89)</td>
<td>(3.67)</td>
<td>(2.97)</td>
<td>(2.47)</td>
<td>(2.93)</td>
<td>(2.40)</td>
</tr>
<tr>
<td>Demand-driven feedback with dynamic sequencing</td>
<td>46</td>
<td>12.11</td>
<td>8.98</td>
<td>3.13</td>
<td>9.11</td>
<td>7.02</td>
<td>2.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3.70)</td>
<td>(4.92)</td>
<td>(2.83)</td>
<td>(3.35)</td>
<td>(4.00)</td>
<td>(2.04)</td>
</tr>
</tbody>
</table>

In the last six columns, the numbers without brackets are means and the numbers within the round parentheses are the standard deviations.
Appendix I

Easy task

This figure depicts an example of counting task that takes easy to complete (easy task). There were two 5x5 matrices shown, which were labelled as box 1 and box 2. To solve the task, participants had to find the highest number in each of the two boxes and add these two numbers.

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>169</td>
</tr>
<tr>
<td>120</td>
<td>164</td>
</tr>
<tr>
<td>144</td>
<td>126</td>
</tr>
<tr>
<td>102</td>
<td>107</td>
</tr>
<tr>
<td>177</td>
<td>156</td>
</tr>
<tr>
<td>101</td>
<td>134</td>
</tr>
<tr>
<td>117</td>
<td>154</td>
</tr>
<tr>
<td>128</td>
<td>131</td>
</tr>
<tr>
<td>132</td>
<td>155</td>
</tr>
<tr>
<td>125</td>
<td>142</td>
</tr>
<tr>
<td>189</td>
<td>120</td>
</tr>
<tr>
<td>115</td>
<td>175</td>
</tr>
<tr>
<td>140</td>
<td>188</td>
</tr>
<tr>
<td>106</td>
<td>102</td>
</tr>
<tr>
<td>171</td>
<td>136</td>
</tr>
<tr>
<td>160</td>
<td>109</td>
</tr>
<tr>
<td>133</td>
<td>115</td>
</tr>
<tr>
<td>158</td>
<td>116</td>
</tr>
</tbody>
</table>

Difficult task

The above figure shows an example of counting task that takes longer time to complete (difficult task). There were two 10x10 matrices shown, labelled as box 1 and box 2. Similar to the easy task, participants had to find the highest number in each of the two boxes and add these two numbers.

<table>
<thead>
<tr>
<th>Box 1</th>
<th>Box 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>148</td>
<td>149</td>
</tr>
<tr>
<td>154</td>
<td>103</td>
</tr>
<tr>
<td>132</td>
<td>150</td>
</tr>
<tr>
<td>143</td>
<td>186</td>
</tr>
<tr>
<td>184</td>
<td>112</td>
</tr>
<tr>
<td>166</td>
<td>116</td>
</tr>
<tr>
<td>117</td>
<td>113</td>
</tr>
<tr>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>129</td>
<td>124</td>
</tr>
<tr>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>107</td>
<td>148</td>
</tr>
<tr>
<td>108</td>
<td>134</td>
</tr>
<tr>
<td>134</td>
<td>133</td>
</tr>
<tr>
<td>108</td>
<td>134</td>
</tr>
<tr>
<td>144</td>
<td>151</td>
</tr>
<tr>
<td>103</td>
<td>109</td>
</tr>
<tr>
<td>150</td>
<td>151</td>
</tr>
<tr>
<td>152</td>
<td>153</td>
</tr>
<tr>
<td>158</td>
<td>158</td>
</tr>
<tr>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>131</td>
<td>177</td>
</tr>
<tr>
<td>112</td>
<td>132</td>
</tr>
<tr>
<td>152</td>
<td>144</td>
</tr>
<tr>
<td>149</td>
<td>146</td>
</tr>
<tr>
<td>138</td>
<td>128</td>
</tr>
<tr>
<td>142</td>
<td>175</td>
</tr>
<tr>
<td>138</td>
<td>176</td>
</tr>
<tr>
<td>160</td>
<td>188</td>
</tr>
<tr>
<td>190</td>
<td>197</td>
</tr>
<tr>
<td>149</td>
<td>140</td>
</tr>
<tr>
<td>112</td>
<td>128</td>
</tr>
<tr>
<td>174</td>
<td>116</td>
</tr>
<tr>
<td>113</td>
<td>110</td>
</tr>
<tr>
<td>124</td>
<td>137</td>
</tr>
<tr>
<td>140</td>
<td>104</td>
</tr>
<tr>
<td>137</td>
<td>136</td>
</tr>
<tr>
<td>136</td>
<td>136</td>
</tr>
<tr>
<td>124</td>
<td>136</td>
</tr>
<tr>
<td>143</td>
<td>143</td>
</tr>
<tr>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td>129</td>
<td>128</td>
</tr>
<tr>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>143</td>
<td>143</td>
</tr>
<tr>
<td>138</td>
<td>138</td>
</tr>
<tr>
<td>151</td>
<td>133</td>
</tr>
<tr>
<td>126</td>
<td>133</td>
</tr>
<tr>
<td>150</td>
<td>133</td>
</tr>
<tr>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>129</td>
<td>129</td>
</tr>
<tr>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>154</td>
<td>154</td>
</tr>
<tr>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>114</td>
<td>114</td>
</tr>
<tr>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>157</td>
<td>157</td>
</tr>
<tr>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>132</td>
<td>132</td>
</tr>
<tr>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>103</td>
<td>103</td>
</tr>
<tr>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>134</td>
<td>134</td>
</tr>
<tr>
<td>151</td>
<td>151</td>
</tr>
<tr>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>119</td>
<td>119</td>
</tr>
<tr>
<td>107</td>
<td>107</td>
</tr>
<tr>
<td>135</td>
<td>135</td>
</tr>
<tr>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td>154</td>
<td>154</td>
</tr>
<tr>
<td>163</td>
<td>163</td>
</tr>
<tr>
<td>162</td>
<td>162</td>
</tr>
<tr>
<td>133</td>
<td>133</td>
</tr>
<tr>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td>158</td>
<td>158</td>
</tr>
<tr>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>131</td>
<td>131</td>
</tr>
<tr>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>121</td>
<td>121</td>
</tr>
<tr>
<td>152</td>
<td>152</td>
</tr>
<tr>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>128</td>
<td>128</td>
</tr>
<tr>
<td>190</td>
<td>190</td>
</tr>
</tbody>
</table>
Appendix II

Task Screen – Demand-driven feedback conditions (DTF and DTFR)

The figure above shows an example of the task screen in both the demand-driven feedback conditions (DTF and DTFR) when an easy task has been selected by the participant. Participant can enter their answer in the input box available and then either click the ‘Proceed directly’ button to proceed without getting feedback or click ‘Proceed with feedback’ button to get feedback.

Task selection after clicking ‘Proceed directly’ – Demand-driven feedback condition (DTF)

The figure above shows an example of the task selection window in the demand-driven feedback condition (DTF) after the participant has clicked the ‘Proceed directly’ button.
Task selection after clicking ‘Proceed with feedback’ – Demand-driven feedback condition (DTF)

The figure above shows an example of the task selection and feedback window in the demand-driven feedback condition (DTF) after the participant has clicked the ‘Proceed with feedback’ button.

Task selection after clicking ‘Proceed directly’ – Demand-driven feedback with recordkeeping condition (DTFR)

The figure above shows an example of the task selection window in the demand-driven feedback condition (DTF) after the participant has clicked the ‘Proceed directly’ button. The window displays a record of easy vs. difficult tasks completed.
Task selection after clicking ‘Proceed with feedback’ – Demand-driven feedback with recordkeeping condition (DTFR)

The figure above shows an example of the task selection and feedback window in the demand-driven feedback condition (DTF) after the participant has clicked the ‘Proceed with feedback’ button. The window displays a record of easy vs. difficult tasks completed.
Chapter 4

Facing A Calibration Committee: The Impact on Costly Information Collection and Subjective Performance Evaluation

Co-authored with Eddy Cardinaels and Bart Dierynck
4.1 Introduction

In many organizations, employees’ contribution to the organization is subjectively assessed by their immediate supervisor. These organizations have recently started installing calibration committees to review and correct the initial subjective performance evaluations made by the immediate supervisors of the employees (Hastings, 2012). The popular business press suggests that certain rater biases in supervisor’s performance evaluations can be addressed when their evaluations are reviewed and discussed in a calibration committee (Risher, 2011; Risher, 2014; Grote, 2005). However, research till date has not examined whether the mere presence of a calibration committee relative to its absence has an impact on the effort that supervisors spend before they make their evaluations. Crucial for performance evaluations in these organizations is that next to performance measures, which are often incomplete proxies of the employee’s performance, supervisors gather additional non-verifiable information to properly assess an employee’s performance potential27. Such information, however, is often costly to acquire and when not collected may lead to performance evaluations that do not properly reflect employee’s performance (Bol, 2008; Bol 2011). We feel that an important neglected benefit of calibration committees is that its presence may stimulate supervisors to work harder on making their evaluations. The first aim of our study is therefore to examine if the mere presence of a calibration committee can bring additional non-verifiable information about an employee to the table, which can ex-ante affect initial evaluations. We further examine how this affects the second stage i.e., discussion in the calibration committee where information is being shared and where final evaluations are determined.

The second aim of this study is to examine whether presence of an impartial third party in the calibration committee may further enhance information collection by supervisors and/or

---

27 For example, information on employee’s citizenship, about how he or she functions in the team, about the difficulty of the tasks the employee is often information that can complement incomplete performance measures.
further facilitate information transfer among supervisors during the discussion stage in the calibration committee. While calibration committees at many organizations consist of the immediate supervisors of the employees that are being evaluated, some organizations add a third party in the form of a higher-level manager or a HR-manager to the calibration committee. While such a third party can serve a vital role in consensus building and information transfer, research till date has not examined the impact of such a third party in the calibration committee.

We use an experimental setting to pursue our research goals. When using field data, it is difficult to draw causal inferences about the impact of calibration committees on supervisor behavior because of the absence of data about the counterfactual (i.e. the situation without a calibration committee)28. Our experimental setting also allows us to elicit and explicitly measure the collection of additional costly information by the supervisor. We vary the presence or absence of a calibration committee. Nested within the calibration committee present condition, we vary whether or not a third party (i.e. HR-manager) is actively present in the calibration committee discussions. Participants act as supervisors and need to subjectively assess the performance of two hypothetical employees who have worked for them in different projects. Each supervisor is the lead supervisor for one of the two employees. Each supervisor receives the performance measure, client satisfaction score, for each employee, but can choose to get additional non-verifiable information at a cost to put the client satisfaction score into context. In the condition without calibration committee (No CC), supervisors see the initial evaluations made by each supervisor but don’t get together to discuss them before each lead supervisor makes the final performance evaluation of her employee. When a calibration committee is present, supervisors discuss their initial performance evaluations in a committee meeting either with two of them only (CCI) or under the presence of a third participant labeled

---

28 Data about the counterfactual is important as supervisors likely behave differently when they anticipate the presence of a calibration committee.
as HR-manager (CC2) before each lead supervisor makes the final performance evaluations of her employee.

We predict that the mere presence of a calibration committee relative to its absence increases supervisor’s willingness to collect costly additional information about the employee performance. Previous research from psychology shows that the prospects of a discussion and justification to other members may prompt a more thorough critical reflection (Schlenker, 1986; Mero and Motowidlo, 1995; Mero, Motowidlo, and Anna, 2003; Lerner and Tetlock, 1999) and more information acquisition (Levine and Russo, 1995; Nemeth and Rogers, 1996). Justification increases impression management concerns (Mero et al., 1995; Mero et al., 2003; Lerner and Tetlock, 1999). Due to these two reasons, anticipated justification in a calibration committee would make the supervisor more willing to collect additional costly information. However, there is sufficient tension with regard to this prediction as the absence of any monetary incentives for the supervisors means that traditional economics would not predict a difference in supervisor’s willingness to collect additional costly information regardless of whether a calibration committee is present or not. In addition, based on research that verifiable quantitative information is more persuasive (Kadous, Koonce, and Towry, 2005), the supervisor might feel that the performance measure, client satisfaction score, available to her ex-ante is enough to convince other committee members about her evaluations. The results, however, confirm our prediction. Compared to the absence of a calibration committee, supervisors in the two calibration committee conditions have a higher willingness to collect additional costly information about the employees’ performance. Importantly, we find that in the presence of a calibration committee, additional information that is acquired by supervisors allows supervisors to distinguish between employees in their initial performance evaluations based on the variation in underlying performance between employees. These initial performance evaluations are highly correlated with the final performance evaluations of
employees. These findings thus suggest that the effect of calibration committees on the extent to which final performance evaluations properly distinguish between employees based on performance differences, is strongly driven by higher collection of additional costly information by supervisors before any discussion in calibration committee takes place.

Next, we examine the impact of a third party in the calibration committee by comparing the calibration committee in which supervisors discuss their initial performance evaluations in a meeting with supervisors only (CC1), compared to the calibration committee where the meeting takes place with both supervisors and a HR-manager (CC2). We formulate a research question regarding the presence of the third party, as its presence can impact either information collection stage, the discussion stage, or both stages. The presence of a HR-manager in the calibration committee might make supervisors feel actively monitored so that they collect more additional costly information and thereby put more effort in making initial performance evaluations. On the other hand, HR-manager might only have an impact when he/she is involved in the discussion to arrive at the final performance evaluations.

Our results indicate that compared to the calibration committee with only supervisors, a calibration committee with a HR-manager does not impact in the information collection stage. However, our results suggest a strong impact of a HR-manager in the discussion stage. Specifically, the differentiation between final evaluations of employees in the presence of a HR-manager is less likely to depend on the differentiation between initial evaluations of employees, especially when both supervisors collect additional costly information. This finding is consistent with the explanation that the presence of a HR-manager prompts information transfer between the supervisors during the discussion, such that they are less likely to anchor on their initial evaluations and more likely to consider the information from the other party to reach a consensus. We also find that the presence of an HR-manager increases the number of words used by each supervisor during discussions when both supervisors collect additional
costly information. Thus, the HR-manager as an impartial third party can properly moderate the discussions, facilitate information transfer, and ensure better involvement of all parties in the discussion by actively asking supervisors to explain their initial performance evaluations. This is consistent with decision making and negotiation literature about the role of a third party as mediator (Fisher, 1972; Fisher and Keashly, 1991; Valley, White, and Iacobucci, 1992; Lewicki, Weiss, and Lewin, 1992; Carnevale and Pruitt, 1992; Goltsman et al., 2009).

We make several important contributions to the literature. Our study adds to the literature on subjective performance evaluation. Often, one key concern is that evaluators largely differ to the extent to which they engage in costly search of information that may put performance into a different context (Maas, Rinsum, and Towry, 2012). Performance measures are often incomplete proxies of the employee’s contribution to the firm. Moreover, employees may have distorted the performance measures, or external circumstances may lead to lower scores even though the employee provided sufficient effort (Bloomfield, 2018; Bentley 2018 Bul, 2008). For supervisors, it is important to gauge the extent to which realized performance is an accurate reflection of the employees’ potential. We offer a mechanism for organizations to instigate supervisors to engage in this costly search of information and to address potential performance measures imperfections by showing that calibration committees augment the justification pressure on the supervisor. As a result of this pressure, more information goes into the subjective performance evaluations, leading to evaluations that are a better picture of employee’s overall performance.

This study further contributes to the limited stream of research investigating the effects of calibration committees on performance evaluations. To our knowledge, not much research work has been done in this area. Exceptions are empirical studies based on field data by Bol et al. (2018), Demere, Sedatole, and Woods (2018), and Grabner, Kunneke, and Moers (2016) which investigate the adjustments made by the calibration committee to correct for rater biases.
in evaluations such as favoritism, leniency or centrality bias. Yet, these studies have difficulties to investigate differences in data gathering as a result of having a calibration committee and zooming into the process of how revisions are being made is also often not possible in the field. Specifically, we offer additional insights in two important ways: First, by using an experiment, we show that calibration committees could serve as a type of action accountability control that helps firms deal with the lack of motivation problem of supervisors failing to collect additional costly information during the evaluation process. We show a strong “anticipation effect” of calibration committees suggesting that supervisors ex-ante put more effort in collecting additional non-verifiable information and use that information into their initial performance evaluations when a calibration committee is present. Thus even prior to the calibration meeting, supervisors might already distinguish between employees in their initial evaluations based on the variation in their underlying performance. This, in turn, would also lead to better final performance evaluations. Secondly, we further contribute to literature on calibration committees by being further able to investigate how the different types of calibration committees (calibration committees with discussion between supervisors only vs. calibration committees with discussion between both supervisors and HR-manager) impact supervisor behavior. Our findings show that compared to the condition without a HR-manager, the presence of a HR-manager makes a difference during the discussion by increasing information transfer and consensus building.

4.2 Literature review and hypothesis development

4.2.1 Calibration committee and subjective performance evaluation

While managerial discretion in performance evaluation can improve contracting by allowing managers to incorporate non-contractible information (Hopwood, 1972; Baiman and Rajan, 1995; Gibbs et al., 2004; Fisher et al., 2005; Rajan and Reichelstein, 2009), it also comes at a cost. Specifically, managers can engage in favoritism, rate according to their own personal
perceptions of performance, or give lenient or compressed performance evaluations. These biases in performance evaluations can be quite costly for organizations (Moers, 2005; Bol, 2008). Organizations are becoming aware of the various rater biases that are associated with subjective performance evaluations. Recently business press and practitioners have suggested a role for calibration committee arguing that one of the primary purposes of a calibration committees is to reduce different rater biases in supervisor evaluations by adjusting these evaluations and to ensure that employees are evaluated in a consistent manner (Grote, 2005; Risher, 2011; Risher, 2014). Any adjustments made in the calibration committee based on distributional properties should promote organizational justice and restore fairness in a subjective evaluation system (Schappe, 1998; McFarlin and Sweeney, 1992; Colquitt, 2001; Colquitt et al., 2006; Colquitt et al. 2013).29

To date, few studies have investigated calibration committees (Demere et al., 2018; Grabner et al., 2016; Bol et al., 2018). For instance, Demere et al. (2018), Grabner et al. (2016), and Bol et al. (2018) study calibration committees in a specific organizational context. All three papers provide useful insights on the adjustments to the initial performance evaluations of supervisors during the discussion in calibration committees. While most studies focus on the benefits of calibration committee on revision of initial evaluations during discussion, we investigate an important neglected aspect. We examine how the prospect of discussing their initial performance evaluations impacts the effort that supervisors spend when making their initial performance evaluations. Specifically, there are two stages in calibration committee. In the first stage individual supervisors make initial performance evaluations on the employees and for this purpose, they can incorporate additional information that is costly to acquire

29 The process of a calibration committee is quite consistent across organizations. After the immediate supervisor of an employee makes the initial performance evaluations of the employee, the calibration committee convenes and reviews the performance evaluations through collective discussions between committee members. If required, the calibration committee revises the performance evaluations up or down before disseminating them to the employees and/or using them for bonus determination (Risher, 2011; Risher, 2014).
(information collection). In the second stage calibration committee members get together to discuss initial performance evaluations to arrive at final performance evaluation for each employee (information transfer). An important aspect of the process towards the initial performance evaluations is the effort that the supervisor invests in collecting additional information that can help to explain the performance measure realization of the employee (information collection). Thus, if manager might reduce effort in collecting information about the employee’s performance, then incomplete information enters the performance evaluation (Bol, 2008; Bol, 2011). These imperfect performance assessments are costly for companies since they reduce the pay to performance link, lead to unfairness concerns among employees, and stimulate suboptimal employee behavior (Moers, 2005; Bol, 2008; Maas et al., 2012).

In this paper, we create an experimental framework. This allows us to investigate and zoom in on whether supervisors change how they evaluate their employees when being faced with a calibration committee. We investigate (1) whether the presence of calibration committees relative to its absence changes the way supervisors collect additional costly information to form initial performance evaluations and (2) how information collection under calibration committees impact final performance evaluations compared to absence of a calibration committee. Next, we also investigate the role of a third party in the calibration committee and its impact on information collection and information transfer. In practice various types of calibration committees exist and sometimes (more independent) third parties enter into a calibration committee. Yet, the importance of having this third party has not been scrutinized either. These questions are quite difficult to study by means of field data as it does not allow comparison of the situation with and without a calibration committee and it is harder to directly observe or isolate the two stages of information collection and information transfer with field data. Furthermore, field data is less useful to address our research questions as organizations often change other aspects of the performance evaluation when introducing calibration
committees. Our experimental setting allows us to causally attribute changes in the outcome of performance evaluation to the unique features of calibration committees.

4.2.2 Information collection in the calibration committee:

Because there are no monetary incentives for a supervisor to collect additional costly information, traditional economics would predict little or no difference in the willingness to collect additional information regardless of whether a calibration committee is absent or present. Moreover, if the supervisor already has access to verifiable quantitative information in the form of client satisfaction scores (without exerting costly effort), he or she may be satisfied with that information to form evaluations. Since verifiable quantitative information is more persuasive (Kadous et al., 2005), supervisors may feel that they can convince other supervisors with the verifiable quantitative information of client satisfaction scores such that there is no need for collecting additional non-verifiable information at a cost.

However, research in psychology suggests that such calibration committees may have an impact because of justification. After supervisors make initial performance evaluations about employees, supervisors face the prospect of justifying their evaluation decisions to other members of the committee through elaborate descriptions. Based on research in psychology, justification pressure induced by the presence of a calibration committee can impact additional information collection due to two reasons. First, previous studies show that justification to an audience can further trigger critical reflection (Schlenker, 1986; Mero et al., 2003; Lerner and Tetlock, 1999) and that any anticipated interaction in a group can lead to more information acquisition by the individual (Levine and Russo, 1995; Nemeth and Rogers, 1996). Thus, using these arguments from psychology we argue that supervisors will think more carefully about evaluation criteria and engage in more thorough search of relevant information when anticipating a discussion in a calibration committee. Secondly, justification increases impression management concerns. When the supervisor presents his/her performance
evaluations to the committee, he or she would want to project a positive image, appear competent and avoid being awkward or appearing foolish when asked to justify his evaluations (Mero et al., 1995; Mero et al., 2003; Lerner and Tetlock, 1999).

In sum, there are two ways in which the prospect of justification in the committee discussion could affect the initial performance evaluation of the supervisor and prompt him/her to collect more additional costly information about employee performance: (1) by increasing supervisor’s concerns to portray a better image of himself in front of others and (2) by facilitating the supervisor in the process of thinking more carefully about the criteria he/she uses to make decisions. Thus, calibration committee could serve as a type of action accountability control that helps firms deal with the lack of motivation problem of not collecting additional costly information that supervisors experience during the evaluation process. We, therefore, expect that anticipating the discussion under the calibration committee, supervisors will be more willing to collect additional costly information when a calibration committee is present relative to when it is absent

**H1:** Supervisors will be more willing to obtain additional costly information on employee performance when the calibration committee is present compared to when it is absent.

Next, it is interesting to examine how the presence of a calibration committee affects the extent to which subjective performance evaluations reflect the variation in underlying performance between employees. Previous research by Demere et al., 2018 investigates whether after discussion in a calibration committee, the employee evaluations are properly differentiated, but find no evidence of such an effect. Apart from the discussion, the second channel through which a calibration committee might affect final performance evaluations is through the information collection by supervisors. We argue that additional non-verifiable information about employees that is costly to acquire is more likely to see the surface when calibration committees are being present. Specifically, when immediate supervisors anticipate
discussion in calibration committee, it will change how they develop their initial performance evaluations in the information collection stage. The mere presence of a calibration committee may already ex-ante improve supervisor evaluation behavior, which can be important step towards improved performance evaluations (see, Bol, 2008; Bol, 2011). According to Bol (2011). Relying on more complete information leads supervisors to make performance evaluations that properly distinguish between employees based on the variation in underlying performance between employees. Thus, before any committee discussion, higher information collection by supervisors may already translate into initial performance evaluations which properly reflect variation between employees’ performance and consequently lead to better differentiated final performance evaluations.

Therefore, we predict that the presence of calibration committee would allow supervisors to properly distinguish between employees through the information collection effect. As a result, we derive the following hypothesis:

**H2: Supervisors’ collection of additional information mediates the effect of the presence of a calibration committee on the extent to which subjective performance evaluations distinguish between employees based on the variation in underlying performance between employees.**

4.2.3 Presence of a HR-manager in the calibration committee:

For calibration committees that also contain the HR-manager, there is an additional aspect of active monitoring by a third party compared to calibration committees where only immediate supervisors meet. According to agency theory, monitoring ensures that the agent does not act in a self-interested way and exerts effort in an appropriate way. Thus, the active presence of a HR-manager in the calibration committee can further increase the willingness of supervisors to put effort in developing the performance evaluations and collect additional costly information about employee performance.
On the other hand, it is possible that the presence of a third party in the form a HR-manager does not have an impact on additional information collection by supervisors. Instead the HR-manager could have more impact in the information transfer once he/she is involved in the discussion in calibration committee. Previous research in decision making and negotiations has looked at the role of a neutral third party in the negotiation process as a mediator (Fisher, 1972; Fisher and Keashly, 1991; Valley, White, and Iacobucci, 1992; Lewicki, Weiss, and Lewin, 1992; Carnevale and Pruitt, 1992; Goltsman et al., 2009). Under mediation, the impartial third party gathers information from the negotiating parties by hearing their arguments and facilitates the settlement between parties. Thus, as an impartial third party, the HR-manager may gather information by actively asking supervisors to explain their initial performance evaluations. In this way, the HR manager could properly mediate the discussions, by facilitating information transfer and ensuring better involvement of all parties in the discussion. As a result, supervisors are less likely to anchor on their initial evaluations based on their own information and more likely to consider and accept the information from the other party in order to reach a consensus on performance evaluations. This could result into final performance evaluations that properly reflect the variation in the underlying performance between employees based on all the information.

**RQ 1a)** Compared to the absence of HR-manager, does the presence of HR-manager in the calibration committee affect the extent to which final performance evaluations distinguish between employees based on the variation in employees’ performance, through costly information collection?

---

30 It should be notes that previous research in decision making and negotiations implicitly assumes that there is an information asymmetry between the third party and the two negotiating parties before the discussion. Thus, that the third party only gathers information from the parties during the negotiation discussion and is uninformed before it.
RQ 1b) Compared to the absence of HR-manager, does the presence of HR-manager affect the extent to which final performance evaluations distinguish between employees based on the variation in employees’ performance, through information transfer in calibration committee?

4.3 Experimental design

4.3.1 Overview

We conducted a laboratory experiment with z-Tree software. We use a nested between-subjects design to test our predictions. We first look at the presence of the calibration committee by comparing the No Calibration committee condition (No CC), against two conditions with a Calibration committee present (CC1 and CC2). Nested within the calibration committee present conditions, we varied the type of the calibration committee: a calibration committee in which only supervisors are present (CC1) and a calibration committee in which both supervisors and a HR-manager are present condition (CC2).

All the participants were randomly assigned to one of the three roles at the start of the experiment labeled as Supervisor X, Supervisor Y, and HR-manager. Supervisors had to make performance evaluations of two hypothetical employees. The HR-manager’s role was to oversee the performance evaluations made by the supervisors. There were two periods. In period one, supervisors made performance evaluations of Employee A and Employee B. Supervisor X was assigned as the lead supervisor of Employee A whereas Supervisor Y was the lead supervisor of Employee B. In period two, participants stayed in their respective roles, but we use random re-matching to ensure that each participant is matched with a new other participant. In period two the performance of two different employees labeled as Employee C and Employee D was assessed. In the second period, Supervisor X was the lead supervisor of Employee C and Supervisor Y was the lead supervisor of Employee D. Below, we describe the evaluation process for the hypothetical employees A and B. This process is repeated for employees C and D in the second period.
4.3.2 Performance evaluation and costly information collection

Both employees worked for the two supervisors on different projects. This setting is illustrated by Figure 1, where the employee and his/her lead supervisor are connected by a solid blue line. The two supervisors each had to subjectively evaluate the performance of Employee A and B by giving a performance score on a 7-point scale with the similar labels as the scale used in Banker et al. (2004). The labels on the 7-point scale are: 0 = Should be fired: sufficient improvement unlikely; 1 = Very Poor: considerably below expectation, needs considerable improvement; 2 = Poor: somewhat below expectations; 3 = Average: meets expectations; 4 = Good: somewhat above expectations; 5 = Very Good: considerably above expectations; and 6 = Excellent: far beyond expectations. The performance score should reflect a supervisor’s assessment of the employee’s contribution to the project in which he/she has worked under his/her supervision.

Performance assessment by the supervisors was done in two stages: the initial stage and the final stage. In the initial stage, each supervisor made a decision about the initial performance evaluation for Employee A and Employee B. Each supervisor was told that they would receive a client satisfaction score of each employee for the project where the employee had worked under his/her supervision. The client satisfaction score indicated how satisfied the client was with the employee on a range from 0 (=not satisfied at all) and 100 (=very satisfied), but could be a noisy reflection of performance. Next to the client satisfaction score, each supervisor could also bid for diary logs at a cost. These logs contained additional non-verifiable information that put client satisfaction score into context (see, Appendix 1). These logs offered extra information about the project and the employee’s functioning during the project such as the type of task the employee performed during the project, any environmental factor that

---

31 We decided to use this flexible team setting because the use of calibration committees fits very well with such a flexible team setting where different lead supervisors meet the same employees on different projects.
32 We used client satisfaction score because this is the criterion which is usually used to evaluate employees according to our interviews with different consultants.
affected the employee’s performance, and how he functioned during the project. As the employees worked in different projects under the supervision of each supervisor, the client satisfaction scores and the diary logs that a supervisor observed differed from those that the other supervisor observed. However, they all indicate similar types of information such as information on task difficulty, and whether the employee was cooperative towards other team members or was a team player.

The diary logs were private information. Participants were made aware that the HR-manager and the other supervisor can never observe their diary logs nor whether they had consulted these diary logs. We use the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964; Bohm et al., 1997) to assess the supervisor’s willingness to obtain additional costly information in the diary logs for subjective performance evaluation. This mechanism, which has also been used in Maas et al. (2012), makes information available at a cost. The cost is randomly determined by a computer between 1 EC to 5 EC. At the start of each period and before observing the client satisfaction scores, the supervisor had to decide whether he/she wanted to consult the diary logs by making an offer that varies from 0 EC to 5 EC. This price offer of the supervisor in our study represents our proxy for the manager’s willingness to acquire costly information. If the supervisor made an offer that was higher than or equal to the randomly determined number, then the cost for the supervisor to consult the diary logs is equal to the randomly determined number. If the supervisor’s offer was below the randomly determined number, then the supervisor cannot consult the diary logs. Participants were informed that the chance of obtaining the diary logs increases linearly with the offer of the supervisor: for an offer of 0 a supervisor is certain not to receive the diary logs, while for an offer of 5 EC the supervisor is 100 percent certain to receive the diary logs. If a supervisor obtained the diary logs, then he/she obtained the diary logs for both employees. Thus, the cost

---

33 These logs were developed based on our interviews with different consultants.
a supervisor had to pay for the diary logs is for both employees. The procedure to determine whether a supervisor obtained the diary logs was repeated at the start of the second period.

The design choice of making an offer before getting the client satisfaction scores allows us to ensure that willingness to get additional information is not contingent on the specific satisfaction scores of employees observed by the supervisor.\(^{34}\) Important to us is that there is no monetary incentive for supervisors to obtain additional information in the setting. If supervisors only have utility for wealth, then we would expect them to never obtain any additional costly information. We made this design choice because we wanted to investigate how the psychological mechanisms of justification and monitoring can influence a supervisor’s decision to obtain additional costly information solely for the purpose of improving the subjective performance evaluation.

Each supervisor’s payoff for a period was based on a fixed endowment of 20 EC minus any cost of consulting the diary logs. Thus, if supervisor’s offer $\geq$ randomly determined number: Supervisor’s period payoff = 20 EC– randomly determined number. If supervisor’s offer $<randomly determined number$: Supervisor’s period payoff = 20 EC. The HR-manager always received a fixed payoff of 20 EC per period.

4.3.3 Final performance evaluations and experimental manipulations

The final stage of the performance evaluation was different between the three different conditions. In all three conditions, at the start of the final stage each supervisor observed the initial performance evaluation and client satisfaction scores of the other supervisor for both employees. Afterwards, in the no calibration committee condition (No CC) each supervisor determined the final performance score for the employee he/she was the lead supervisor of. Supervisor X determined the final score for Employee A and Supervisor Y the final score of

\(^{34}\) In reality, supervisors collect diary logs during the course of the project before the client satisfaction scores are realized.
Employee B. The HR-manager only stored the final performance scores, implying that he/she cannot change the final performance scores. The HR-manager observed the client satisfaction scores since this was freely available public information, but he/she could not observe the initial performance scores or the diary logs of the two supervisors. This signifies that the HR-manager did not have the final authority in the performance evaluation.

Participants in the two calibration committee conditions (CC1 and CC2) knew that they have to discuss their initial performance evaluation in a committee. They were told that they could write notes before starting the discussion explaining why they assign a specific performance score to an employee, to prepare for their discussion in a committee in the final stage of performance assessment. To serve as guidance, these notes were visible during the discussion to the supervisor who wrote them but were not visible to other participants. Participants in the no calibration committee condition (No CC) did not have to write such preparation notes since they did not face a calibration committee. Note taking is part of the calibration committee manipulation to trigger justification pressure among supervisors. We also made the notes not visible to others so that the discussion between participants in the calibration committee conditions (CC1 and CC2) would be richer and more interactive.

In the calibration committee with supervisors only condition (CC1), the committee composed of the two supervisors discussed the initial performance scores of each supervisor. This discussion was conducted through messages in the chat boxes of z-Tree software. Finally, each supervisor determined the final performance score for the employee he/she is the lead supervisor of. We only varied the discussion between supervisors in the calibration committee with supervisors only condition (CC1) and kept the passive role of the HR-manager the same as in the no calibration committee condition (No CC). That is the HR-manager only stored the final performance scores and could only view the public information on client satisfaction scores. In the calibration committee with supervisors and HR-manager (CC2), the HR-manager
also participated in the discussion. Specifically, the committee composed of the HR-manager and the two supervisors discussed the initial performance scores of each supervisor. To make sure that the amount of information to the HR-manager was constant across conditions, the HR-manager can again only observe the client satisfaction scores, but he/she could not observe supervisors’ initial performance scores, their diary logs and the accompanying notes of the supervisors. However, as the HR-manager was actively present in the committee, he/she could ask questions during the discussion about the information he/she did not know. Finally, as in other conditions, each lead supervisor determined the final performance score for their respective employee. The HR-manager then stored the final performance scores.  

4.3.4 Participants and experimental procedures

The participants are undergraduate students from a university in The Netherlands. In total, 175 students participated in the study. The mean age of the participants was 21.2 years, with the youngest participant being 18 and the oldest 37. About 37 percent of the participants were female. In total 12 sessions were conducted and there were 146 different supervisors and 29 HR-managers (24 managers for CC2 and 5 managers for No CC and CC1). In CC2, we have always three-person groups. For the 48 supervisors, we recruited 24 HR-managers. For No CC (also 48 supervisors) and CC1 (50 supervisors), the role of HR was passive. To economize on participants, we did not assign every third participant the role of an HR manager and had two-person groups of supervisors instead of three-person groups of supervisors and HR manager for the sessions with No CC and CC1 conditions. We then conducted a separate session in the

---

It is important to mention that our manipulations are much weaker than what is observed in practice. For reasons of experimental control, we perform the discussion in the form of chat messages between anonymous participants. In reality, committee discussions are face to face between known individuals, creating a stronger justification pressure. Active presence of the HR-manager in practice also puts a higher pressure on supervisors due to career related concerns. Thus, our design choices are conservative choices, which may bias against finding predicted effects. To isolate the effect of justification pressure and active monitoring due to the presence of calibration committee, it is crucial to minimize the influence of other factors such as career concerns and experience.
end with HR managers only where we recruited 5 participants to be HR-managers who stored the evaluations made by supervisors in the No CC and CC1 sessions earlier.

The participants were students who had taken business courses and have some part time/full time working experience. Each student self-registered as participants in response to an invitation on the university’s laboratory website. Course credits were used as a show-up fee. The actual pay-out in Euros for the participants was determined by converting the experimental unit payoff from both periods to Euros using an exchange rate of 0.2. The average amount earned over the two periods was €7.59 in total.

Upon arrival in the laboratory, the students entered a waiting room. An instructor explained the basic procedures and told the participants that they would find detailed sets of instructions in their cubicles. The participants were afterwards randomly assigned to a cubicle. The instructions explained the roles as supervisors or HR-manager, the task, and the procedures for determination of the participants’ pay-off. It was emphasized that they would be interacting with each other and that there was no deception of any kind. Then the participants had to answer a set of questions about the experimental task to enhance task understanding. Participants could only continue after answering all questions correctly. After all participants had successfully completed this check, the first period round began, and the computer task started automatically.

After the last period, the participants completed an exit questionnaire. This questionnaire was used to gain a better understanding of the motives behind participants’ decisions. On average, the two periods of the experiment took the participants 60 minutes. Finally, after completing the questionnaire, participants collected their money and left.

4.3.5 Differentiation between subjective performance evaluations of employees:

We use the proxy of differentiation between subjective performance evaluations between employees to measure the extent to which subjective performance evaluations distinguish between employees based on the variation in underlying performance between
employees. In period 1, the variation in customer satisfaction scores between employees is small but the supervisors can distinguish between employees through additional information, which reflects a larger variation in underlying performance between employees (see, Appendix 1). Thus, if supervisors make evaluations which account for all information then the differentiation proxy measuring the extent to which subjective performance evaluations of the two employees reflect the variation in their underlying performance should be higher in period 1. On the other hand, in period 2, the variation in customer satisfaction scores between employees is larger but the additional information reflects a smaller variation in underlying performance between employees (see, Appendix 1). Thus, the differentiation proxy measuring the extent to which subjective performance evaluations of the two employees reflect the variation in their underlying performance should be smaller in period 2.36 Specifically, the additional information in period 2 was designed to reduce the difference in the performance of the employees depicted by the client satisfaction score in the second period.

Based on the above discussion, to calculate differentiation between initial performance evaluations of employees (SD_IS), we used the standard deviation of a supervisor’s initial performance evaluations for the two employees in period 1. Since the additional costly information for the evaluation of the performance of the employees in period 2 suggested lower variation between performance evaluations of employees, we reverse code the average standard deviation in period 2 when calculating the differentiation proxy. In conclusion, to measure the extent to which subjective performance evaluations reflect the variation in underlying performance between employees, the overall proxy of differentiation between initial evaluations of employees (SD_IS) is the standard deviation of a supervisor’s initial performance evaluations for the two employees in period 1 and negative of the standard deviation.

---

36 Using two evaluation periods allows us to provide robust results such that the results to be driven by our choice of the performance measure variation between employees as well as the additional information.
deviation of a supervisor’s initial performance evaluations for the two employees in period 2. This shows that subjective performance evaluation of employees should be more distinguished in period 1 and less distinguished in period 2. Similarly, $SD_{FS}$, differentiation between final performance evaluations of employees, was the standard deviation of the final performance evaluations for the two employees in period 1 and the negative of the standard deviation in period 2.

4.4 Results

4.4.1 Descriptive statistics

We have 292 supervisor-period observations from the pool of 146 participants (48 participants for No CC, 50 for CC1, and 48 for CC2). Table 1 shows the descriptive statistics. From the range of 0 to 5 EC, supervisors offered to pay an average of 2.65 EC for additional information at the beginning of each period. Each supervisor then gave an average initial performance evaluation of 3.675 from the 7-point scale of 0 (=should be fired) to 6 (=excellent), while the final performance evaluation given to an employee by his/her lead supervisor was 3.547 on average. Both initial and final performance evaluations are around the middle of the 7-point scale. Descriptive statistics show that on average, there was some differentiation between initial and final performance evaluations of employees. The means of $SD_{IS}$ and $SD_{FS}$ are negative which might indicate that standard deviation in period 1 was dominated by the reverse coded standard deviation in the second period. This shows that even though additional information in period 2 was designed to lower the large difference in client satisfaction scores, supervisors still anchor on the objective quantitative measure in their evaluations leading to evaluations between employees that are still quite differentiated in the second period.

--- Insert Table 1 here ---
Figure 2 shows the mean price offer for each period in the three different conditions. While the mean price offered by the supervisors in No CC condition remains approximately constant during the two periods (2.375 EC in period 1 and 2.395 EC in period 2), the mean price offered for both conditions where a calibration committee is present (CC1 and CC2) increases in the second period to around 2.9 EC. Since the two periods were independent with different employees (employees C & D as opposed to employees A & B) and each participant’s role was randomly re-matched, it can be concluded that the increase in period 2 for the calibration committee conditions might be attributed to each participants’ experience with a calibration committee in period 1. Thus, on average each supervisor who was in the calibration committee conditions increased his/her price offer after being exposed to calibration committee earlier.

--- Insert Figure 2 here ---

Descriptive results in Table 2 show support for our hypothesis 1. We pool the observations from the two calibration committee conditions to get the calibration committee condition (CC). Comparing to the condition where calibration committee is absent (No CC), the average price offer over the two periods for additional costly information by a supervisor is 14 percent higher when a calibration committee is present (CC). The difference ($p<0.013$) is significant at 5 percent level. This supports our first hypothesis which states that supervisors are more willing to obtain additional costly information when a calibration committee is present compared to when it is absent. Next, we compare each type of calibration committees to the condition without calibration committee. Compared to no calibration committee condition (No CC), the mean overall price offer is 12 percent higher when a calibration committee consisting of only supervisors is present (CC1). Similarly, mean overall price offer is 16 percent higher when a calibration committee consisting of supervisors and HR-manager is present (CC2) compared to no calibration committee (No CC). Table 2 shows that that the differences in the overall results are mainly driven by period 2. When we compare the two types of calibration
committees, we can see that the mean overall price offer is almost the same for the two calibration committees (2.71 for CC1 and 2.85 for CC2). This suggests that the presence of a HR-manager does not increase additional information collection.

--- Insert Table 2 here ---

Table 3 shows descriptive results for the additional information collection by a supervisor (Acq_t), additional information collection by both supervisors (Both_Acq), differentiation between initial evaluations of employees (SD_IS) and differentiation between final evaluations of employees (SD_FS). We pool the observations from the two calibration committee conditions to get the calibration committee condition (CC). Comparing to the condition where calibration committee is absent (No CC), Acq_t is significantly higher and Both_Acq is marginally higher in the presence of CC. Acq_t and Both_Acq are not significantly different when comparing the two types of committees, CC1 and CC2. As for SD_IS and SD_FS, they are not significantly different in the presence of CC. We find similar results for SD_IS and when comparing the two types of committees, CC1 and CC2. However, since this univariate analysis ignores the collection of additional information, we cannot make conclusions without doing further analysis that also considers additional information collection.

--- Insert Table 3 here ---

4.4.2 Main findings

Next, we provide regression results for hypothesis 1 in Table 4. First, we test hypothesis 1 in column 1-2. We pool the observations from the two calibration committee conditions to get the calibration committee condition (CC). The dependent variable is the price offer made by the supervisors to acquire additional costly information about employee performance. In column 1, our main independent variable is the dummy for the calibration committee (CC) with no calibration committee (No CC) as the baseline case. We estimate a random effects regression
on the two period panel data. Consistent with hypothesis 1, we find that the price offer of a supervisor is significantly higher in the calibration committee compared to no calibration committee ($b=0.395$, two-tailed $p<0.10$). Since Figure 2 and the descriptive results in Table 2 shows that the effects are more pronounced in period 2, we also run a regression with robust standard errors where the price offer in period 2 is the dependent variable. Similar to results in column 1, the price offers of supervisors in period 2 are significantly higher in calibration committee compared to no calibration committee ($b=0.336$, two-tailed $p<0.10$).

For column 3 and 4, we have the data for the two types of calibration committee conditions $CC1$ and $CC2$. We take calibration committee with supervisors only $CC1$ as the baseline case. In column 3, we estimate panel random effects regression with price offer as the dependent variable.\textsuperscript{37} We do not find any significant difference in the price offer between the two types of calibration committees. Column 4 estimates regression with robust standard errors when price offer in period 2 is the dependent variable. Results are similar to column 3 and there is no significant difference in the supervisors’ price offer in period 2. Thus, the analysis shows that there is no difference between $CC1$ and $CC2$ in terms of acquiring costly information\textsuperscript{38}.

--- Insert Table 4 here ---

As additional tests, we used the ex-post questionnaire about whether the participants felt more pressure to collect additional information because of the presence of the other supervisor. Results (untabulated) show that compared to the calibration committee absent condition ($No CC$), participants felt significantly more pressure ($Z= 2.595$, $p<0.01$) because of the presence

\textsuperscript{37} The random effect is at the supervisor level. The random effects account for unobserved heterogeneity across supervisors as the likelihood ratio test shows that there is enough variability between supervisors to favor this random-effects regression over standard regression.

\textsuperscript{38} As robustness check, we performed a logit regression using High Offer (1 if Offer is higher than median and 0 otherwise) as alternative dependent variable in the models of Table 4. The results (untabulated) were qualitatively similar to those in Table 4, suggesting that under the presence of a calibration committee ($CC1$ and $CC2$), people are more likely to make a high price offer compared to $No CC$ (model 1, model 2). Comparison between $CC1$ and $CC2$, suggest again no difference (smallest $p=0.18$).
of other supervisor when a calibration committee was present (CC1 and CC2). This offers additional support for our theory on justification pressure. Specifically, the impact of presence of the other supervisor to collect additional costly information is more prominent when supervisors face a calibration committee. Results of the mediation analysis with Price Offer as dependent variable and pressure due to the presence of the other supervisor (Pressure Supervisor) as mediator, suggest that the relationship between CC (CC1 and CC2 compared to No CC as baseline) and Price Offer is fully mediated by the pressure due to the presence of the other supervisor in the process. The mediator Pressure Supervisor is significant (Z=3.27, p<0.01), while the main effect of CC on Price Offer becomes non-significant (Z=1.01, p=0.31) consistent with the criteria of mediation by Baron and Kenny (1986).

Next, we test hypothesis 2 related to whether the presence of calibration committee leads supervisors to distinguish between employees in their initial performance evaluations, through collection of additional costly information (mediation). It is important to study initial performance evaluations of employees by supervisors as initial performance evaluations for an employee presented to committee are highly correlated with the final adjusted performance evaluations of that employee (correlation coefficient of 0.731, p<0.01). Moreover, the differentiation between initial performance evaluations of employees is also highly correlated with the differentiation between final performance evaluations of employees (correlation coefficient of 0.802, p<0.01). Figure 3 shows the results of the mediation analysis, where the dependent variable is the differentiation between the initial performance evaluations of employees, the mediator is the acquisition of additional costly information and independent variable is the dummy variable for one of the conditions. We first take the combination of the two calibration committee conditions (CC) as independent variable with No CC as the baseline case. Results in round brackets show that the indirect effect of CC on differentiation between initial performance evaluations of employees through collection of additional costly, is positive
and significant. This is in consistent with hypothesis 2. Moreover, when comparing CC2 with CC1, we do not find a significant indirect effect of CC2 on differentiation between initial performance evaluations of employees through collection of additional costly information. This is consistent with earlier results the presence of HR-manager does not lead to more collection.39

--- Insert Figure 3 here ---

4.4.2.1 Information transfer and impact of the HR-manager

The results above show that presence of a HR-manager with supervisors in the calibration committee (CC2) compared to the committee in which he or she is absent (CC1) does not impact collection of additional costly information and subsequently the differentiation between initial evaluations of employees. Therefore, we next examine whether the presence of a HR-manager with supervisors in the calibration committee (CC2) affects the discussion stage in a committee and in turn the differentiation between final performance evaluations of employees, compared to calibration committee with supervisors only (CC1).

Table 5 shows the results of the ANOVA analysis, where the dependent variable is $SD_{FS}$, differentiation between the final performance evaluations of employees. We consider two models. In model 1, we consider independent variables, CC2 with CC1 as baseline, and $SDIS$, differentiation between initial evaluations of employees at the committee level.40 To examine

---

39 We also checked whether the presence of a calibration committee reduces leniency (mean initial performance evaluations) through the collection of additional costly information. We do not find any such evidence. This is not surprising given that the experiment was designed such that the additional information acquired by the supervisor enables him/her to add context to the freely available client satisfaction score and to distinguish between the two employees. Thus, the direct result of the additional information acquired would be expected to be stronger on the variation between performance evaluations of employees rather than on the leniency in the performance evaluations. As the mean initial performance evaluations are between 3.6 on the scale of 0 to 6, leniency does not seem to be a big problem in our setting.

40 Our dependent variable $SD_{FS}$, differentiation between final performance evaluations of the two employees is at the committee level as the lead supervisor only makes the final evaluation of his/her employee. Therefore, the independent variable needs to be at the committee level as well. However, $SD_{IS}$, differentiation between initial performance evaluations of the two employees, is at the supervisor level, as each supervisor makes initial evaluations for both employees. Therefore, to calculate differentiation between initial evaluation of the two employees at the committee level, SDIS, we calculate the average of $SD_{IS}$ from the two supervisors.
how the presence of HR-manager affects performance evaluations during the discussion, we
include an interaction term between SDIS and CC2. Results show that SDIS is the main
predictor of SD_FS, again highlighting that differentiation between final evaluations of
employees is significantly determined by differentiation between initial evaluations of
employees. We find that, CC2 significantly affects the relation between SDIS and SD_FS. In
fact, the top margin plot for model 1 in Figure 4 shows that CC2 significantly reduces the
relation between SDIS and SD_FS, compared to CC1. Thus, presence of a HR-manager
prompts information transfer between the supervisors during the discussion, such that they are
less likely to stick to the initial evaluations based on their own information and more likely to
consider the information from the other party to reach a consensus. As a result, the
differentiation between final evaluations in the presence of a HR-manager is less likely to
depend on the differentiation between initial evaluations of employees before the discussion.

--- Insert Table 5 and Figure 4 here ---

If information transfer is the reason why the presence of HR-manager weakens the
relationship between SDIS and SD_FS, then it should be most prominent when both supervisors
acquire additional costly information. For this purpose, we next consider model 2 in Table 5,
where we include CC2, SDIS, and Acq_Both (equal to 1 when both supervisors collect
additional costly information, else 0), along with all the interaction terms in the ANOVA
model. Table 5 show that the three-way interaction between CC2, SDIS and Acq_Both is
significant (two-tailed p<0.10). The bottom margin plot in Figure 4 for the ANOVA results in
model 2 shows us that indeed CC2 significantly weakens the relationship between SDIS and
SD_FS, only when both supervisors acquire information (i.e. Acq_Both = 1). This confirms the
explanation that compared to no HR-manager in a committee, the presence of a HR-manager
facilitates information transfer between supervisors so that they more likely to share costly
information during discussion and consider other supervisor’s information in making the final
evaluations and thus anchor less on their own initial rating resulting from their own information they have available or collected. Our findings therefore show that presence of HR-manager impacts final performance evaluations through information transfer and better transfer of costly non-verifiable (soft) information in discussion stage rather than information collection.

We also analyzed supervisors’ word count during the committee discussions to provide additional evidence for the explanation that compared to no HR-manager in a committee, the presence of a HR-manager facilitates information transfer between supervisors. Table 6 shows the results of the analysis, where the dependent variable is Supervisor’s Word Count. We estimate a random effects regression on the two period panel data and control for Acq_t, i.e., whether the supervisor acquired costly additional information during the period as it would impact the supervisor’s word count. In column 1, we consider independent variables, CC2 with CC1 as baseline, Acq_Both (equal to 1 when both supervisors collect additional costly information, else 0), and include the interaction term of CC2 with Acq_Both. We find that the interaction term is significant (b=23.31, two-tailed p=0.069) showing that number of words used by supervisor when both supervisors acquire additional costly information is significantly higher when an HR-manager is present in the committee. Thus, consistent with the information transfer explanation, supervisors are more likely to talk and share the costly additional information they collected when an HR-manager is present in the committee.

Next, it is also interesting to analyze the variation within the committees in which an HR-manager is present (CC2) and show whether supervisor’s word count in CC2 varies with the HR-manager’s word count and questions. In column 2 and 3, we consider independent variables, HR_Questions and HR_WordCount respectively. We find that supervisor’s word count significantly increases when the HR-manager asks more questions (b=6.14, two-tailed p<0.01) and uses more words (b=3.79, two-tailed p<0.01) during the committee discussions.
This shows that by talking and asking more questions, an HR-manager plays an important role in facilitating information transfer during calibration committee discussions.

4.5 Summary and discussion

The paper investigates whether and how the presence of calibration committee affects the supervisor’s initial performance evaluation. We find that supervisors are more willing to obtain additional costly information on employee performance when a calibration committee is present as compared to when it is absent. We, thus, find support for our theory suggesting that the prospect of the discussion about supervisors’ initial performance evaluations in calibration committees already influences supervisors’ costly information acquisition. We also compare the different types of calibration committees: a calibration committee with supervisors only and calibration committees with both supervisors and a HR-manager. We find that the supervisor’s willingness to obtain additional information does not differ between the two types of calibration committees. Our results, however, suggest that active presence of the HR-manager impacts the evaluation process through information transfer in the discussion stage.

With regards to the use of the additional information to distinguish between employees, we find that presence of calibration committee increases the extent to which supervisors’ performance evaluations properly distinguish between employees based on their performance differences, through the acquisition of additional information. This provides support for the prediction that the initial performance evaluations by the supervisors properly incorporate the variation in underlying performance between employees when a calibration committee is present. This is an important finding since Demere et al. (2018) show that the adjustments made by a calibration committee are not effective in reducing the compression of performance evaluations. Through our experiment, we provide more context to these findings as our results suggest that the final performance evaluations are only significantly differentiated when supervisors collect additional information, and that they do not differ when no information is
being collected. Another explanation is that the initial performance evaluations which are the input of the committee discussions are already significantly differentiated, so that the calibration committee then might not adjust them to increase differentiation.

The study contributes to limited but nascent stream in accounting research regarding the use of calibration committees for performance evaluation. Prior studies like Bol et al., 2018, Grabner et al., (2016), and Demere et al. (2018) have examined the adjustment process in calibration committees where committees adjust initial performance evaluations to final performance evaluations. We contribute to this new stream of literature by using an experimental design to study how the presence of a calibration committee changes the information gathering of the supervisor prior to the determination of the initial performance evaluation. As such, we provide new insights about calibration committee by studying how different psychological mechanisms are activated by the presence of a calibration committee. To our knowledge, this is the first study that provides causal evidence regarding the role of calibration committees in subjective evaluation.

We offer evidence that bringing together supervisors in a calibration committee can help organizations ensure that supervisors anchor less on objective performance measures that are often far from complete measures of performance when evaluating their performance. The presence of a calibration committee can instigate managers to work harder and collect more additional information that can make ex-ante initial evaluations already better prior to any discussion taking place. Anchoring on imperfect or incomplete performance measures can have serious costs for organizations (Bloomfield, 2018; Bol, 2008). Thus, one big benefit of a calibration committee that the literature has largely ignored is the potential to address performance measure imperfections by instigating managers to collect additional costly non-verifiable performance information that puts the objective performance measure into context. Additionally, the presence of a third party in the calibration committee discussion could ensure
that supervisors stick less to their initial evaluation based on their own information when revising initial to final evaluations. A third party can potentially further result in performance evaluations that better reflect employees’ overall performance. Lastly, as calibration committees improve performance measures imperfections by incorporating additional non-verifiable information, they may also be considered fairer by employees.

We feel that our research can extend to other decision contexts such labor market decisions, employee selection, lending decisions, investment decisions etc. These decision contexts often involve situations where people also need to acquire costly soft information (e.g. soft information about clients in a banking context) next to hard and more verifiable information (e.g. collateral) that is available to the decision maker. We argue that justification pressure in these environments can vary too. One path for future research is to investigate for example to whom a decision maker needs to ratify or justify his or her decision and whether that has an effect on the amount soft information that is available for decision making and the extent to which soft information improves decision outcomes.

Future research can also study other aspects of calibration committees. We heavily focused on performance evaluations and the process of acquiring additional information and abstract away from leniency biases and favoritism. Future research can for example examine how expressions of favoritism towards certain employees (Prendergast and Topel, 1996) might be mitigated as a result of calibration committees. A supervisor who has a well performing employee might be reluctant to share soft information he or she has collected if that information depicts a less favorable picture of the employee. For reasons of experimental control, we also have people negotiate over a computer network. Impression management and humanization effects may play a bigger role in face-to-face settings or hierarchical settings where people justify to a person higher in the hierarchy (Kachelmeier and Towry, 2002). These effects may potentially increase the willingness to collect costly additional performance information.
Finally, next to performance evaluation decisions, promotion decisions are also part of the discussion and could be investigated in future studies on the topic of calibration committees and its use.
4.6. References


The figure shows the three experimental conditions. The dotted box represents the calibration committee. In calibration committee 1 (CC1), supervisors meet with one another in the calibration committee. In Calibration committee 2 (CC2), also the HR-manager takes part in the calibration committee. To test H1 and H2, we compare the no calibration committee against the two calibration present conditions. In RQ 1a and RQ 1b, we focus on the type of calibration committee we compare CC1 against CC2. The roles of supervisors and HR-manager are played by real participants. In period 1, the performance of hypothetical employees A and B had to be assessed. In period 2, re-matching takes place, and performance of hypothetical employees C and D had to be assessed by the supervisors.
FIGURE 2

Mean price offer

The figure shows the mean price offered at the start of each period by the supervisors for the additional costly information about employee performance in the three experimental conditions.
### Table 1

**Descriptive statistics**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Obs.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Price Offer</em></td>
<td>292</td>
<td>2.650</td>
<td>1.474</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><em>Initial performance evaluations</em></td>
<td>292</td>
<td>3.675</td>
<td>0.626</td>
<td>1.5</td>
<td>5.5</td>
</tr>
<tr>
<td><em>Final performance evaluations</em></td>
<td>292</td>
<td>3.647</td>
<td>1.053</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td><em>SD_IS</em></td>
<td>292</td>
<td>-0.068</td>
<td>1.316</td>
<td>-4.242</td>
<td>3.536</td>
</tr>
<tr>
<td><em>SD_FS</em></td>
<td>146</td>
<td>-0.111</td>
<td>1.236</td>
<td>-3.536</td>
<td>2.828</td>
</tr>
</tbody>
</table>

*Price Offer* is the offer made at the start of each period by a supervisor for additional information on employees.

*Initial performance evaluations* is the average of the initial performance evaluations given by a supervisor to both employees from the scale of 0 (=should be fired) to 6 (=excellent).

*Final performance evaluations* is the final performance evaluations of an employee given by his/her lead supervisor from the scale of 0 (=should be fired) to 6 (=excellent).

*SD_IS* is Differentiation between initial performance evaluations of employees equal to the standard deviation of a supervisor’s initial performance evaluations for the two employees in period 1 and negative of the standard deviation in period 2. We reverse code the standard deviation in period 2 because the additional information if acquired was designed to reduce the large difference in the performance of the employees depicted by the client satisfaction score in that period.

*SD_FS* is Differentiation between final performance evaluations of employees equal the standard deviation of the final performance evaluations for the two employees in period 1 and negative of the standard deviation in period 2. We again reverse code the standard deviation in period 2.
### TABLE 2
**Descriptive results – price offer**

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Price Offer</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No CC</td>
<td>2.375 (1.524)</td>
<td>2.395 (1.553)</td>
<td>2.385 (1.531)</td>
</tr>
<tr>
<td></td>
<td>[n=48]</td>
<td>[n=48]</td>
<td>[n=96]</td>
</tr>
<tr>
<td>CC1</td>
<td>2.500 (1.199)</td>
<td>2.920 (1.367)</td>
<td>2.710 (1.297)</td>
</tr>
<tr>
<td></td>
<td>[n=50]</td>
<td>[n=50]</td>
<td>[n=100]</td>
</tr>
<tr>
<td>CC2</td>
<td>2.770 (1.490)</td>
<td>2.937 (1.642)</td>
<td>2.854 (1.562)</td>
</tr>
<tr>
<td></td>
<td>[n=48]</td>
<td>[n=48]</td>
<td>[n=96]</td>
</tr>
<tr>
<td>CC vs. No CC</td>
<td>Z= -1.264</td>
<td>Z= -2.229</td>
<td>Z= -2.484</td>
</tr>
<tr>
<td></td>
<td>p= 0.206</td>
<td>p= 0.026</td>
<td>p= 0.013</td>
</tr>
<tr>
<td>CC1 vs. No CC</td>
<td>Z= -0.717</td>
<td>Z= -1.934</td>
<td>Z= -1.893</td>
</tr>
<tr>
<td></td>
<td>p= 0.473</td>
<td>p= 0.053</td>
<td>p= 0.058</td>
</tr>
<tr>
<td>CC2 vs. No CC</td>
<td>Z= -1.476</td>
<td>Z= -1.921</td>
<td>Z= -2.414</td>
</tr>
<tr>
<td></td>
<td>p= 0.140</td>
<td>p= 0.055</td>
<td>p= 0.016</td>
</tr>
<tr>
<td>CC1 vs. CC2</td>
<td>Z= -1.265</td>
<td>Z= -0.490</td>
<td>Z= 1.236</td>
</tr>
<tr>
<td></td>
<td>p= 0.206</td>
<td>p= 0.624</td>
<td>p= 0.217</td>
</tr>
</tbody>
</table>

The numbers without brackets are the mean values. The numbers in round brackets are standard deviation. The square brackets show the number of observations. We use non-parametric test (Wilcoxon rank-sum test) to compare conditions.

No CC equals 1 for the absence of calibration committee, else 0.
CC1 equals 1 for the presence of calibration committee with supervisors only, else 0.
CC2 equals 1 for the presence of calibration committee with supervisors and HR-manager, else 0.
Price Offer is the offer made at the start of each period by a supervisor for additional information on employees.
TABLE 3
Descriptive results- other variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>No CC</th>
<th>CC1</th>
<th>CC2</th>
<th>No CC vs. CC</th>
<th>CC2 vs. CC1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acq_t</td>
<td>0.437</td>
<td>0.580</td>
<td>0.552</td>
<td>Z= 2.067</td>
<td>Z= 0.393</td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.496)</td>
<td>(0.499)</td>
<td>p=0.039</td>
<td>p= 0.694</td>
</tr>
<tr>
<td></td>
<td>[n=96]</td>
<td>[n=100]</td>
<td>[n=96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Both_Acq</td>
<td>0.271</td>
<td>0.420</td>
<td>0.354</td>
<td>Z= 1.387</td>
<td>Z= 0.665</td>
</tr>
<tr>
<td></td>
<td>(0.446)</td>
<td>(0.496)</td>
<td>(0.480)</td>
<td>p=0.165</td>
<td>p= 0.506</td>
</tr>
<tr>
<td></td>
<td>[n=48]</td>
<td>[n=50]</td>
<td>[n=48]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD_IS</td>
<td>-0.058</td>
<td>-0.035</td>
<td>-0.110</td>
<td>Z= 0.093</td>
<td>Z= 0.422</td>
</tr>
<tr>
<td></td>
<td>(1.234)</td>
<td>(1.454)</td>
<td>(1.258)</td>
<td>p=0.926</td>
<td>p= 0.673</td>
</tr>
<tr>
<td></td>
<td>[n=96]</td>
<td>[n=100]</td>
<td>[n=96]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD_FS</td>
<td>-0.206</td>
<td>-0.099</td>
<td>-0.029</td>
<td>Z= 0.635</td>
<td>Z= 0.228</td>
</tr>
<tr>
<td></td>
<td>(1.148)</td>
<td>(1.456)</td>
<td>(1.081)</td>
<td>p= 0.525</td>
<td>p= 0.820</td>
</tr>
<tr>
<td></td>
<td>[n=48]</td>
<td>[n=50]</td>
<td>[n=48]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The numbers without brackets are the mean values. The numbers in round brackets are standard deviations. The square brackets show the number of observations. We use non-parametric test (Wilcoxon rank-sum test) to compare conditions.

Acq_t equals 1 when the supervisor himself/herself acquired additional costly information, else 0.
Both_Acq equals 1 when both supervisors acquire additional costly information, else 0.
SD_IS is Difference between initial performance evaluations of employees, equal to the sum of the standard deviation of a supervisor’s initial performance evaluations for the two employees in period 1 and negative of the standard deviation in period 2. We reverse code the standard deviation in period 2 because the additional information if acquired was designed to reduce the large difference in the performance of the employees depicted by the client satisfaction score in that period.
SD_FS is Difference between final performance evaluations of employees, equal to the sum of the standard deviation of the final performance evaluations for the two employees in period 1 and negative of the standard deviation in period 2. We again reverse code the standard deviation in period 2.
No CC equals 1 for the absence of calibration committee, else 0.
CC1 equals 1 for the presence of calibration committee with supervisors only, else 0.
CC2 equals 1 for the presence of calibration committee with supervisors and HR-manager, else 0.
CC equals 1 for the presence of calibration committee (both CC1 and CC2), else 0.
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>(1) Price Offer</th>
<th>(2) Price Offer in Period 2</th>
<th>(3) Price Offer</th>
<th>(4) Price Offer in Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>0.395*</td>
<td>0.336*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.238)</td>
<td>(0.180)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC2</td>
<td>-</td>
<td>-</td>
<td>0.144</td>
<td>-0.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.262)</td>
<td>(0.237)</td>
</tr>
<tr>
<td>Period</td>
<td>0.205**</td>
<td>-</td>
<td>0.295**</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.093)</td>
<td></td>
<td>(0.122)</td>
<td></td>
</tr>
<tr>
<td>Offer in Period 1</td>
<td>-</td>
<td>0.759***</td>
<td>-</td>
<td>0.721***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.082)</td>
<td></td>
<td>(0.121)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.077***</td>
<td>0.591***</td>
<td>2.710***</td>
<td>1.115***</td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td>(0.266)</td>
<td>(0.191)</td>
<td>(0.363)</td>
</tr>
<tr>
<td>N</td>
<td>292</td>
<td>146</td>
<td>196</td>
<td>98</td>
</tr>
<tr>
<td>$\chi^2$/F statistic</td>
<td>7.56**</td>
<td>54.15***</td>
<td>6.15**</td>
<td>17.91***</td>
</tr>
</tbody>
</table>

$p$-levels are two-tailed, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. For the regression of Price Offers in both periods, we have panel data so we use random effects model to account for the individual heterogeneity (Column 1 and 3). The numbers within the round parentheses are the robust standard errors (Column 2 and 4). The baseline case in column 1 and 2 is No CC, while the baseline case in column 3 and 4 is CC1.

Price Offer is the offer made at the start of each period by a supervisor for additional information on employees.

Price Offer in period 2 equals offer made by the supervisor for additional information in second period.

Offer in period 1 equals offer made by the supervisor for additional information in first period.

CC equals 1 for the presence of either type of calibration committee, and equals 0 for the absence of calibration committee.

CC1 equals 1 for the presence of calibration committee with supervisors only, else 0.

CC2 equals 1 for the presence of calibration committee with supervisors and HR-manager, else 0.
The figure shows the mediation analysis for testing hypothesis 3 that states that presence of calibration committee affects differentiation between initial performance evaluations of employees through acquisition of additional costly information. We use S.E.M with maximum likelihood estimation method. We test two models:

Round brackets show the results for model 1 with Condition = CC with No CC as baseline case. Square brackets show the results for model 2 with Condition = CC1 with CC2 as baseline case.

For model 1 (model 2), N=292 (196) and global fit statistics are RMSEA= 0.000(0.000), CFI=1.000(1.000), SRMR= 0.000(0.000). The indirect effects of Condition on Differentiation in initial performance evaluation (a1*b1) for both models are shown in the box above.
### TABLE 5
ANOVA – Effect of presence of HR-manager

<table>
<thead>
<tr>
<th>Source</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SS</td>
<td>df</td>
<td>F-stats</td>
<td>p-value</td>
</tr>
<tr>
<td>SDIS</td>
<td>116.9</td>
<td>1</td>
<td>331.04</td>
<td>0.000</td>
</tr>
<tr>
<td>CC2</td>
<td>0.341</td>
<td>1</td>
<td>0.970</td>
<td>0.328</td>
</tr>
<tr>
<td>CC2*SDIS</td>
<td>1.381</td>
<td>1</td>
<td>3.910</td>
<td>0.051</td>
</tr>
<tr>
<td>Both_Acq</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDIS*Both_Acq</td>
<td>0.025</td>
<td>1</td>
<td>0.070</td>
<td>0.787</td>
</tr>
<tr>
<td>CC2*Both_Acq</td>
<td>0.007</td>
<td>1</td>
<td>0.020</td>
<td>0.889</td>
</tr>
<tr>
<td>CC2<em>SDIS</em>Both_Acq</td>
<td>1.256</td>
<td>1</td>
<td>3.650</td>
<td>0.059</td>
</tr>
<tr>
<td>Residual</td>
<td>33.20</td>
<td>94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*SDIS* is Differentiation between initial performance evaluations of employees at committee level, equal to the sum of the average standard deviation of the two supervisor’s initial performance evaluations for the two employees in period 1 and negative of the average standard deviation in period 2. We reverse code the standard deviation in period 2 because the additional information if acquired was designed to reduce the large difference in the performance of the employees depicted by the client satisfaction score in that period. To provide a sensible test for *CC2* in ANOVA model, we subtract the mean from the SDIS variable.

*SD_FS* is Differentiation between final performance evaluations of employees, equal to the sum of the standard deviation of the final performance evaluations for the two employees in period 1 and negative of the standard deviation in period 2. We again reverse code the standard deviation in period 2.

*Both_Acq* equals 1 when both supervisors acquire additional costly information, else 0.

*CC2* equals 1 for the presence of calibration committee with supervisors and HR-manager, 0 for calibration committee with supervisors only.
The figure above shows the predictive margins plot to interpret the interactions in the ANOVA model 1 in Table 5. CC2=0 shows the results for CC1, calibration committees with supervisors only, while CC2=1 is for CC2, calibration committee with supervisors and HR-manager.

The figure above shows the predictive margins plot to interpret the interactions in the ANOVA model 2 in Table 5. Both_Acq equals 1 when both supervisors collect additional costly information about employees, else 0.
### TABLE 6
Regressions results of supervisor’s word count during committee discussions

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>CC1 vs. CC2</th>
<th>CC2 only</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC2</td>
<td>-3.25</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(9.30)</td>
<td></td>
</tr>
<tr>
<td>CC2* Both_Acq</td>
<td>23.31 asterisk</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(12.81)</td>
<td></td>
</tr>
<tr>
<td>Both_Acq</td>
<td>-35.71 asterisk</td>
<td>-17.22</td>
</tr>
<tr>
<td></td>
<td>(11.36)</td>
<td>(12.22)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.19)</td>
</tr>
<tr>
<td>HR_Questions</td>
<td></td>
<td>6.14 asterisk</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.89)</td>
</tr>
<tr>
<td>HR_WordCount</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acq_t</td>
<td>38.65 asterisk</td>
<td>36.04 asterisk</td>
</tr>
<tr>
<td></td>
<td>(9.33)</td>
<td>(12.64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(12.75)</td>
</tr>
<tr>
<td>Constant</td>
<td>79.98 asterisk</td>
<td>64.13 asterisk</td>
</tr>
<tr>
<td></td>
<td>(7.13)</td>
<td>(8.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.38)</td>
</tr>
<tr>
<td>N</td>
<td>196</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>96</td>
</tr>
<tr>
<td>$\chi^2$ statistic</td>
<td>20.81 asterisk</td>
<td>22.40 asterisk</td>
</tr>
<tr>
<td></td>
<td>(7.13)</td>
<td>(8.96)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9.38)</td>
</tr>
</tbody>
</table>

$p$-levels are two-tailed, "p < 0.10, "p < 0.05, ""p < 0.01. For the regression of a supervisor’s word count during chats in both periods, we have panel data, so we use random effects model to account for the individual heterogeneity. The baseline case in column 1 is CC1. In columns 2 and 3 only CC2 is considered. We also control for whether the supervisor himself/herself acquired additional costly information during the period.

Supervisor’s Word Count is the number of words that the supervisor used during the committee discussions. CC2 equals 1 for the presence of calibration committee with supervisors and HR-manager, 0 for calibration committee with supervisors only. Both_Acq equals 1 when both supervisors acquire additional costly information, else 0. HR_Questions is the number of questions that the HR-manager asked during the committee discussions. HR_WordCount is the number of questions that the HR-manager asked during the committee discussions. Acq_t equals 1 when the supervisor himself/herself acquired additional costly information, else 0.
## Appendix 1

### Period 1:

<table>
<thead>
<tr>
<th>Supervisor X</th>
<th>Employee A</th>
<th>Supervisor Y</th>
<th>Employee B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Client Satisfaction Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>During the period, Employee A has worked under you in a consulting team on <em>project alpha</em>. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was <strong>69% satisfied</strong> with the performance of Employee A.</td>
<td></td>
<td><strong>Client Satisfaction Score</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Diary logs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>From your diary that you maintained during the course of project alpha, you know that:</td>
<td></td>
<td>From your daily diary during the course of the project beta, you know that</td>
<td></td>
</tr>
<tr>
<td>- Compared to other team members, Employee A</td>
<td></td>
<td>- The major player in the client team who had access to information about</td>
<td></td>
</tr>
<tr>
<td>Supervisor Y</td>
<td>Employee B</td>
<td>Supervisor X</td>
<td>Employee B</td>
</tr>
<tr>
<td><strong>Client Satisfaction Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee B has also worked under you in a consulting team on <em>project beta</em>. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was <strong>72% satisfied</strong> with the performance of Employee B.</td>
<td></td>
<td>Employee B has also worked under you in a consulting team on <em>project gamma</em>, for which you were the supervisor. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was <strong>73% satisfied</strong> with the performance of Employee B.</td>
<td></td>
</tr>
<tr>
<td><strong>Diary logs</strong></td>
<td></td>
<td><strong>Diary logs</strong></td>
<td></td>
</tr>
<tr>
<td>Based on daily logs during the course of the project gamma, you know that</td>
<td></td>
<td>Based on diary logs during the course of the project delta, you know that</td>
<td></td>
</tr>
<tr>
<td>- Employee B had a mediocre task that was not really challenging compared to other team members</td>
<td></td>
<td>- Compared to other team members, Employee B worked on a really easy task</td>
<td></td>
</tr>
</tbody>
</table>

**Supervisor X**

Employee B has also worked in a consulting team on *project gamma*, for which you were the supervisor. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was **73% satisfied** with the performance of Employee B.

**Supervisor Y**

Employee B has also worked under you in a consulting team on *project delta*. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was **68% satisfied** with the performance of Employee B.

**Diary logs**

From your diary during the course of the project delta, you know that:

- Employee B had a mediocre task that was not really challenging compared to other team members.
worked on a difficult task during project alpha.
- In addition, there were an unusually large number of junior consultants on the team during project alpha. Employee A had to train these junior consultants apart from his regular duties related to the project. He always spent time helping other team members especially these junior consultants even though it was harder to complete his own task as a result. His cooperation and help allowed the team to complete the project on time.

- different aspects of project X dropped out of the client team in the middle of the project.
- Employee A also had a task that required special input from the client team member who dropped out. Thus, Employee A had a particularly difficult time completing the task at hand. However, he was still able to complete his task.

- during the course of project B.
- Employee B was also not a team player during the project. He was often involved in disruptive behavior such as not completing some of his regular duties on time. Due to this behavior, the other members faced difficulty in finishing related tasks and there was some delay in the progress of activities during the project.

- to other members on project Y.
- Employee B was also from the same university and undergraduate program as the client manager, and therefore had really good personal relations with the client manager.
**Period 2:**

<table>
<thead>
<tr>
<th>Supervisor X</th>
<th>Supervisor Y</th>
<th>Supervisor X</th>
<th>Supervisor Y</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Client Satisfaction Score</strong></td>
<td><strong>Client Satisfaction Score</strong></td>
<td><strong>Client Satisfaction Score</strong></td>
<td><strong>Client Satisfaction Score</strong></td>
</tr>
<tr>
<td>During the period, Employee C has worked under you in a consulting team on <em>project theta</em>. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance on a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was 62% satisfied with the performance of employee.</td>
<td>During the period, Employee C has worked under you in a consulting team on <em>project sigma</em>. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance on a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was 60% satisfied with the performance of employee.</td>
<td>Employee D has also worked in a consulting team on <em>project lambda</em>, for which you were the supervisor. After the project was completed, the client gave feedback on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance on a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was 81% satisfied with the performance of employee.</td>
<td>Employee D has also worked under you in a consulting team on <em>project phi</em>. After the project was completed, you received feedback from the client on each individual member of the team. The client manager provided how satisfied he/she was with each member’s performance on a 0% (=not satisfied at all) to 100% (=very satisfied) scale. According to feedback you received from the client manager, the client was 79% satisfied with the performance of employee.</td>
</tr>
<tr>
<td><strong>Diary logs</strong></td>
<td><strong>Diary logs</strong></td>
<td><strong>Diary logs</strong></td>
<td><strong>Diary logs</strong></td>
</tr>
</tbody>
</table>
| From your diary that you maintained during the course of project theta, you know that:  
- Compared to other team members, Employee C was responsible for a more | From your daily diary during the course of the project sigma, you know that  
- During the course of project sigma, Employee C was responsible for more tasks than the other team | Based on diary logs during the course of the project lambda, you know that  
- During the course of project lambda, Employee D was responsible for a task that was less challenging | From your diary during the course of the project phi, you know that  
- Compared to other team members of project phi, Employee D was responsible for a task that |
- A demanding and challenging task during project theta.
  - In addition, one of the team members, who was assisting Employee C with the difficult task faced serious health concerns during the project and had to drop out before he could complete his part of the task. Therefore, on top of fulfilling his own task obligations, Employee C also had to complete the other member’s unfinished duties. As a result, he faced difficulty in completing his own task. Despite this, Employee C was a really productive member of the team during the project.

  - Members. This is because at the start of the project he volunteered to take care of overall maintenance of records for the team in addition to his regular project duties.
  - During the project, the client wanted Employee C to complete some additional tasks which were not entirely related to the current project. Since he was already really busy with several duties and had a higher than average workload, he was not able to effectively complete these additional requests of the client.

  - Compared to the tasks of the other team members.
  - Employee D delegated part of his task to the junior members of the team and received a lot of assistance from them in completing the task. Thus, he didn’t put in a lot of effort and was easily able to finish his obligations on time.

  - Didn’t require a lot of effort on his part.
  - Employee D was responsible for training some junior team members but he was generally uncooperative and unhelpful towards them. As a result, the team had to face some issues in completing the project efficiently.
## CENTER DISSERTATION SERIES
CentER for Economic Research, Tilburg University, the Netherlands

<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>ISBN</th>
<th>Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>579</td>
<td>Julius Rüschenpöhler</td>
<td>Behavioural Perspectives on Subsistence Entrepreneurship in Emerging Markets</td>
<td>978 90 5668 580 5</td>
<td>January 2019</td>
</tr>
<tr>
<td>580</td>
<td>Khulan Altangerel</td>
<td>Essays on Immigration Policy</td>
<td>978 90 5668 581 2</td>
<td>January 2019</td>
</tr>
<tr>
<td>581</td>
<td>Kun Zheng</td>
<td>Essays on Duration Analysis and Labour Economics</td>
<td>978 90 5668 582 9</td>
<td>January 2019</td>
</tr>
<tr>
<td>582</td>
<td>Tatiana Zabara</td>
<td>Evolution of Entrepreneurial Teams in Technology-Based New Ventures</td>
<td>978 90 5668 583 6</td>
<td>February 2019</td>
</tr>
<tr>
<td>583</td>
<td>Yifan Yu</td>
<td>Essays on Mixed Hitting-Time Models</td>
<td>978 90 5668 584 3</td>
<td>April 2019</td>
</tr>
<tr>
<td>584</td>
<td>Daniel Martinez Martin</td>
<td>Unpacking Product Modularity, Innovation in Distributed Innovation Teams</td>
<td>978 90 5668 585 0</td>
<td>April 2019</td>
</tr>
<tr>
<td>585</td>
<td>Katalin Katona</td>
<td>Managed Competition in Practice Lessons for Healthcare Policy</td>
<td>978 90 5668 586 7</td>
<td>April 2019</td>
</tr>
<tr>
<td>586</td>
<td>Serhan Sadikoglu</td>
<td>Essays in Econometric Theory</td>
<td>978 90 5668 587 4</td>
<td>May 2019</td>
</tr>
<tr>
<td>587</td>
<td>Hoang Yen Nguyen</td>
<td>Emotions and Strategic Interactions</td>
<td>978 90 5668 588 1</td>
<td>May 2019</td>
</tr>
<tr>
<td>588</td>
<td>Ties de Kok</td>
<td>Essays on reporting and information processing</td>
<td>978 90 5668 589 8</td>
<td>May 2019</td>
</tr>
<tr>
<td>589</td>
<td>Yusiyu Wang</td>
<td>Regulation, Protest, and Spatial Economics</td>
<td>978 90 5668 590 4</td>
<td>June 2019</td>
</tr>
<tr>
<td>590</td>
<td>Ekaterina Neretina</td>
<td>Essays in Corporate Finance, Political Economy, and Competition</td>
<td>978 90 5668 591 1</td>
<td>June 2019</td>
</tr>
<tr>
<td>591</td>
<td>Ruth Wandhöfer</td>
<td>Technology innovation in Financial Markets: Implications for Money, Payments and Settlement Finality</td>
<td>978 90 5668 592 8</td>
<td>June 2019</td>
</tr>
<tr>
<td>592</td>
<td>Andinet Worku Gebreselassie</td>
<td>On communicating about taboo social issues in least developed countries: The case of Ethiopia</td>
<td>978 90 5668 593 5</td>
<td>June 2019</td>
</tr>
<tr>
<td>593</td>
<td>Filip Bekjarovski</td>
<td>Active Investing</td>
<td>978 90 5668 594 2</td>
<td>June 2019</td>
</tr>
<tr>
<td>594</td>
<td>Miguel Sarmiento</td>
<td>Essays on Banking, Financial Intermediation and Financial Markets</td>
<td>978 90 5668 595 9</td>
<td>June 2019</td>
</tr>
<tr>
<td>No.</td>
<td>Author</td>
<td>Title</td>
<td>ISBN</td>
<td>Published</td>
</tr>
<tr>
<td>-----</td>
<td>-------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>--------------</td>
<td>-----------------</td>
</tr>
</tbody>
</table>
| 595 | Xiaoyin Ma              | Essays on Alternative Investments                                    | 978 90 5668  
596 6 | June 2019                |
| 596 | Victor van Pelt         | A Dynamic View of Management Accounting Systems                      | 978 90 5668  
597 3 | June 2019                |
| 597 | Shuai Chen              | Marriage, Minorities, and Mass Movements                             | 978 90 5668  
598 0 | July 2019                |
| 598 | Ben Gans                | Stabilisation operations as complex systems: order and chaos in the interoperability continuum | 978 90 5668  
599 7 | July 2019                |
| 599 | Mulu Hundera            | Role Conflict, Coping Strategies and Female Entrepreneurial Success in Sub-Saharan Africa | 978 90 5668  
600 0 | August 2019              |
| 600 | Hao Hu                  | The Quadratic Shortest Path Problem – Theory and Computations        | 978 90 5668  
601 7 | September 2019           |
| 601 | Emerson Erik Schmitz    | Essays on Banking and International Trade                            | 978 90 5668  
602 4 | September 2019           |
| 602 | Olga Kuryatnikova       | The many faces of positivity to approximate structured optimization problems | 978 90 5668  
603 1 | September 2019           |
| 603 | Sander Gribling         | Applications of optimization to factorization ranks and quantum information theory | 978 90 5668  
604 8 | September 2019           |
| 604 | Camille Hebert          | Essays on Corporate Ownership and Human Capital                      | 978 90 5668  
605 5 | October 2019             |
| 605 | Gabor Neszveda          | Essays on Behavioral Finance                                         | 978 90 5668  
606 2 | October 2019             |
| 606 | Ad van Geesbergen       | Duurzame schaarste - Een kritische analyse van twee economische duurzaamheids-paradigma’s geïnspireerd door de filosofie van Dooyeweerd | 978 90 5668  
607 9 | October 2019             |
| 607 | Richard T. Mason        | Digital Enrollment Architecture and Retirement Savings Decisions: Evidence from the Field | 978 90 5668  
608 6 | November 2019            |
| 608 | Ron Triepels            | Anomaly Detection in the Shipping and Banking Industry               | 978 90 5668  
609 3 | November 2019            |
| 609 | Feng Fang               | When performance shortfall arises, contract or trust? A multi-method study of the impact of contractual and relation governances on performance in Public-Private Partnerships | 978 90 5668  
610 9 | November 2019            |
| 610 | Yasir Dewan             | Corporate Crime and Punishment: The Role of Status and Ideology      | 978 90 5668  
611 6 | November 2019            |
<table>
<thead>
<tr>
<th>No.</th>
<th>Author</th>
<th>Title</th>
<th>ISBN</th>
<th>Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>611</td>
<td>Mart van Hulten</td>
<td>Aiming for Well-Being through Taxation: A Framework of Caution and Restraint for States</td>
<td>978 90 5668 612 3</td>
<td>December 2019</td>
</tr>
<tr>
<td>612</td>
<td>Carlos Sandoval Moreno</td>
<td>Three essays on poverty measurement and risk protection</td>
<td>978 90 5668 613 0</td>
<td>December 2019</td>
</tr>
<tr>
<td>613</td>
<td>Harmke de Groot</td>
<td>Core strength or Achilles’ heel: Organizational competencies and the performance of R&amp;D collaborations</td>
<td>978 90 5668 614 7</td>
<td>December 2019</td>
</tr>
<tr>
<td>614</td>
<td>Peter Brok</td>
<td>Essays in Corporate Finance and Corporate Taxation</td>
<td>978 90 5668 615 4</td>
<td>December 2019</td>
</tr>
<tr>
<td>615</td>
<td>Pascal Böni</td>
<td>On the Pricing, Wealth Effects and Return of Private Market Debt</td>
<td>978 90 5668 616 1</td>
<td>December 2019</td>
</tr>
<tr>
<td>616</td>
<td>Ana Martinovici</td>
<td>Revealing Attention: How Eye Movements Predict Brand Choice and Moment of Choice</td>
<td>978 90 5668 617 8</td>
<td>December 2019</td>
</tr>
<tr>
<td>617</td>
<td>Matjaz Maletic</td>
<td>Essays on international finance and empirical asset pricing</td>
<td>978 90 5668 618 5</td>
<td>January 2020</td>
</tr>
<tr>
<td>618</td>
<td>Zilong Niu</td>
<td>Essays on Asset Pricing and International Finance</td>
<td>978 90 5668 619 2</td>
<td>January 2020</td>
</tr>
<tr>
<td>619</td>
<td>Bjorn Lous</td>
<td>On free markets, income inequality, happiness and trust</td>
<td>978 90 5668 620 8</td>
<td>January 2020</td>
</tr>
<tr>
<td>620</td>
<td>Clemens Fiedler</td>
<td>Innovation in the Digital Age: Competition, Cooperation, and Standardization</td>
<td>978 90 5668 621 5</td>
<td>June 2020</td>
</tr>
<tr>
<td>621</td>
<td>Andreea Popescu</td>
<td>Essays in Asset Pricing and Auctions</td>
<td>978 90 5668 622 2</td>
<td>June 2020</td>
</tr>
<tr>
<td>622</td>
<td>Miranda Stienstra</td>
<td>The Determinants and Performance Implications of Alliance Partner Acquisition</td>
<td>978 90 5668 623 9</td>
<td>June 2020</td>
</tr>
<tr>
<td>623</td>
<td>Lei Lei</td>
<td>Essays on Labor and Family Economics in China</td>
<td>978 90 5668 624 6</td>
<td>May 2020</td>
</tr>
<tr>
<td>624</td>
<td>Farah Arshad</td>
<td>Performance Management Systems in Modern Organizations</td>
<td>978 90 5668 625 3</td>
<td>June 2020</td>
</tr>
</tbody>
</table>
The tremendous progress in technology, digitization and globalization in the last few decades has not only facilitated the availability of information in organizations but also changed their operations and structure. All these transformations have implications for performance management systems, such as performance evaluation and feedback systems, in organizations. This dissertation presents three studies that use laboratory experiments to examine performance management systems in modern organizations. The first study (chapter 2) examines how managers’ self-serving incentives, which are becoming more prevalent in modern organizations, affect managers’ evaluation behavior. The second study (chapter 3) investigates whether demand-driven feedback systems, which are becoming a popular performance management tool in many organizations, induce employees to engage in easy task prioritization. The third study (chapter 4) examines the role of calibration committees, another popular performance management system in modern organizations, in supervisors’ evaluation behavior.

**Farah Maham Arshad** (Karachi, Pakistan, 1991) obtained her bachelor’s degree in Accounting and Finance from Lahore University of Management Sciences (Pakistan) in 2012. She later graduated with a master’s degree in international management from Tilburg University in 2015 and obtained her research master’s degree in accounting from the same school in 2017. Since then, Farah has been a PhD candidate at the Accountancy department at Tilburg University.

ISBN: 978 90 5668 625 3
DOI: 10.26116/center-lis-2003