Subjectivity in Performance Evaluation and Group Identity as Antecedents of Employee Overwork

by

Ala Mokhtar

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Accounting

Waterloo, Ontario, Canada, 2023

EXAMINING COMMITTEE MEMBERSHIP

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Flora Zhou Associate Professor of Accounting Department of Accounting Bentley University
Supervisor	Tim Bauer Associate Professor of Accounting School of Accounting & Finance University of Waterloo
Supervisor	Adam Presslee Associate Professor of Accounting School of Accounting & Finance University of Waterloo
Internal Member	Alan Webb Professor Emeritus of Accounting School of Accounting & Finance University of Waterloo
Internal Member	Krista Fiolleau Associate Professor of Accounting School of Accounting & Finance University of Waterloo
Internal-External Member	Janet Boekhorst Associate Professor at the Conrad School of Entrepreneurship and Business University of Waterloo

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Employees often overwork by working longer than contractual or statutory standard working time for no immediate additional monetary gain. Despite the prevalence of overwork in firms, little is understood about why employees choose to work such long hours. Firms often have high overwork levels despite management encouraging employees to make use of work-life balance policies, and employees at such firms often believe that their long work hours are self-imposed. Employees likely would not feel that way if they are given explicit management directives to work long hours, indicating that other factors in the organization lead employees to overwork. I use an experiment to investigate how two key features of a firm's management control system – the subjectivity in performance evaluation and the strength of employees' identity with their colleagues (hereafter, group identity) – affect employees' level of overwork. I find that the effect of subjectivity in performance evaluation on the level of overwork is increasing in group identity strength, such that a positive effect is present only when group identity is stronger and not when it is weaker. I also find that group identity has a significant positive effect on the level of overwork at higher but not lower levels of subjectivity in performance evaluation. These results largely support my hypotheses. Finally, I employ a secondary experiment and provide evidence that subjectivity in performance evaluation impacts the level of overwork primarily through the effort heuristic mechanism. My study is important because understanding factors within firms that propagate overwork is consequential for firms that want to discourage such overwork due to its negative consequences. Understanding these factors also allows firms to have a more complete understanding of what motivates their employees.

ACKNOWLEDGMENTS

I am infinitely thankful to my supervisors Tim Bauer and Adam Presslee for their constant encouragement, support, and guidance. Their patience is what made this dissertation possible. I would also like to thank the members of my examining committee, Alan Webb, Krista Fiolleau, Flora Zhou, and Janet Boekhorst, for taking the time to provide me with valuable feedback. I am also grateful to Kate Patterson for her never-ending support, Amar Mahmoud for his valuable research assistance, as well as Bradley Pomeroy, Joyce Tian, Pei Wang, Jessie Ge, Yifei Kuang and Jillian Adams for their help throughout. I am thankful for comments and suggestions by participants in the School of Accounting and Finance's brown bag seminar, as well as helpful comments from participants at the AAA/Deloitte Foundation/J Michael Cook 2022 Doctoral Consortium. I would also like to thank participants and my reviewers and discussant at the 2023 Management Accounting Section Midyear Meeting. Finally, I would like to thank the CPA Ontario Centre for Sustainability Reporting and Performance Management (CSPM) for their generous funding.

I would like to express my deepest gratitude and appreciation to my friends and family in Waterloo and across the world, who have been an invaluable source of comfort and warmth throughout my time as a graduate student.

Finally, my PhD journey would have been impossible were it not for my parents, Mona and Ahmed, and my brothers, Ali and Omar. Their encouragement, endless support and unconditional love are my biggest blessings in life.

DEDICATION

This thesis is wholeheartedly dedicated to Mona, Ahmed, Ali and Omar, who are my unwavering support system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Employees working for excessively long hours is a notable feature in many firms (Surowiecki, 2014; Pozen, 2012; Cha, 2010). For example, a 2008 Harvard Business School survey found that 94% of surveyed professionals work more than 50 hours a week (Perlow & Porter, 2009). A more recent 2019 survey found that 91% of United Kingdom professionals at least sometimes work more than their contracted hours and that 90% of those professionals receive no form of additional compensation for working longer than their contracted hours (Morgan McKinley, 2019). Long-working hours were also exacerbated during the COVID-19 pandemic (Beheshti, 2021; Vershbow, 2021). The literature terms the phenomenon of working long hours as "overwork" (e.g., Cha & Weeden, 2014). Specifically, I define overwork as employees working beyond contractual or statutory standard working time for no immediate additional monetary gain ("Work-life balance", n.d.). Despite the prevalence of overwork in firms, little is understood about why employees choose to work such long hours (Lupu & Empson, 2015). Firms often have high overwork levels despite management encouraging employees to make use of work-life balance policies, and employees at such firms often believe that their long work hours are self-imposed (Blagoev et al., 2018; Empson, 2018; Ladva & Andrew 2014; Michel, 2011; Kellogg, 2009; Lewis, 2003). Employees likely would not feel that way if they are given explicit management directives to work long hours (Michel, 2011), indicating that other factors in the organization lead employees to overwork.

I experimentally investigate how two key features of a firm's management control system - the subjectivity in performance evaluation and the strength of employees' identity with their colleagues (hereafter, group identity) - affect employees' level of overwork. First, subjectivity in performance evaluation is the extent to which "judgment based on personal impressions, feelings,

and opinions" is used by management to evaluate employee performance (Bol, 2008, p. 2).[1] Incorporating subjective performance evaluation in contracts – as opposed to objective performance evaluation – is a common method by which employee performance is evaluated and by which employees are incentivized; in fact, it is more prevalent in incentive contracts than objective performance evaluation. Thus, it is a key component of employee incentive contracts (Bol, 2008; Gibbs et al., 2004; Prendergast 1999; Murphy & Cleveland, 1995). Second, group identity is the extent to which a person defines herself as a member of a group, such that she derives her sense of identity from being a member of that group (Chen & Li, 2009; Towry, 2003; Hogg & Turner, 1987). Group identity is an important cultural control studied in the accounting literature (e.g., Alvesson & Kärreman, 2004; Towry, 2003), and it is important to study in the overwork context since the literature has generally theorized how overwork is a social phenomenon impacted by group dynamics (e.g., Brett & Stroh, 2003).

Understanding the antecedents to employee overwork within a firm is critical so that management can guide employee overwork to an optimal level. When employees overwork, it can suggest great alignment between employee effort and firm objectives to the extent overworking leads to higher firm performance. Indeed, firms often desire overwork because it is cheaper to hire fewer employees who each overwork than to hire more employees to work the same number of hours (Dembe, 2009; Brett & Stroh, 2003). Conversely, management may want to minimize employee overwork due to its potential negative consequences and the impact of these consequences on firm value. Some of these consequences include lack of efficiency and employee

[1] Employee performance evaluations can be based on both objective and subjective performance evaluation together (Grabner, 2014; Gibbs et al., 2004). Performance evaluation that is higher in subjectivity relies more on subjective performance evaluation and less on objective performance measures.

burnout, which increase employee turnover, increase absenteeism, and lower firm performance (Carmichael, 2015; Golden, 2012; Dembe, 2009; Brett & Stroh, 2003; Kodz et al., 1998).

My first prediction examines subjectivity in employees' performance evaluation and its impact on employees' level of overwork. I argue that higher subjectivity in performance evaluation creates uncertainty in employees about how they can perform and be evaluated well (Bol, 2008), which results in employees using heuristics to anticipate how they will be evaluated. When individuals face uncertainty in a decision situation, they are more likely to rely on heuristics when making their judgments (Kruger et al., 2004; Tversky & Kahneman, 1974). An important heuristic that is relevant in work settings is the effort heuristic (Kruger et al., 2004): employees believe that the more effort they are seen to exert, the better their output will be evaluated. Specifically, employees will anticipate that if they increase their level of overwork, a visible form of effort, they will receive a higher performance evaluation. When employees perceive overwork as a metric for evaluation success, they will be driven to increase their level of overwork. Thus, I predict greater subjectivity in performance evaluation leads to greater levels of overwork.

My second prediction is that the positive effect of subjectivity in performance evaluation on the level of overwork is amplified when group identity is stronger versus weaker. Employee group identity varies both between and within organizations due to both deliberate management decisions to manage it (e.g., organizing team-building events) and organically from other decisions made by management (e.g., recruitment efforts). When group identity is stronger (weaker), an employee is more (less) prone to socially comparing against her colleagues because they are (are not) seen as relevant comparison targets (Brewer & Weber, 1994; Tesser, 1988). Importantly, as subjectivity increases, overwork becomes an increasingly important evaluation metric and one on which employees socially compare. Because an employee with a stronger group identity (versus a

weaker group identity) is motivated to derive a positive self-image from the social comparisons she makes, she will want to be at least as good as her colleagues on the overwork metric and to avoid being the first to stop working; thus, I argue that this motivates her to increase her level of overwork more in response to an increase in subjectivity in her performance evaluation.

My third prediction posits a main effect of group identity on overwork. As group identity with colleagues becomes stronger, this also increases the likelihood that employees will work longer and increase their level of overwork to try to benefit their group. This is based on economic theory that posits that employees gain utility simply from working with, and for the benefit of, a group with which they identify (Akerlof, 2016; Akerlof & Kranton, 2005; 2000).

Finally, although not my primary interest in this study, I examine the interactive effects of subjectivity in performance evaluation and group identity on performance, as mediated by overwork. *Ex-ante*, it is unclear how an increased level of overwork would be related to performance. On the one hand, the increased effort duration that comes with overwork may translate into better performance (Bonner & Sprinkle, 2002). On the other hand, employees may prioritize overworking as an end goal, rather than focusing on improving their actual performance, such that overwork may not translate to better performance. Thus, I pose a research question examining how subjectivity in performance evaluation and group identity interact to affect performance. However, I acknowledge that effects related to this research question may depend on my experimental parameters.

I test my hypotheses using a 2 x 2 between-subjects online experiment, with 290 Amazon Mechanical Turk workers. I manipulate subjectivity in performance evaluation at two levels, *Higher* vs *Lower*, by describing the evaluation of participants' task output as subjective versus objective, respectively. I also manipulate group identity at two levels, *Stronger* vs *Weaker*, by

describing participants as identifying more versus less with their hypothetical teammates. I measure overwork as the amount of time participants spend working on an assigned writing task minus the time that I *contract* them to spend working on the task for a fixed payment.

Experimental results partially support my first hypothesis that subjectivity has a positive main effect on the level of overwork (H1). Specifically, I find that subjectivity has a positive effect on the level of overwork when group identity is stronger but not when it is weaker. Further, results support my second hypothesis that the H1 effect is greater when group identity is stronger (H2). Results partially support my third hypothesis, that group identity has a positive main effect on overwork (H3). Specifically, I find that group identity has a positive effect on the level of overwork when subjectivity is higher but not when it is lower. Supplemental analysis provides support that the effort heuristic and social comparison act as mechanisms underlying my findings: I find that participants who are more inclined to use the effort heuristic show stronger support for H1 and H2; similarly, those who are more inclined to socially compare show somewhat stronger support for H2. Finally, in examining RQ1, I find no evidence that subjectivity in performance evaluation and group identity interact to impact participants' performance on the writing task.

I also employ a secondary experiment to provide further evidence that subjectivity in performance evaluation impacts the level of overwork primarily through the effort heuristic mechanism and through feelings of uncertainty. Asay et al. (2022) stress the benefits of using separate experiments to test theoretical process. Thus, I test the direct impact of *Higher* vs *Lower Subjectivity* on participants' use of the effort heuristic and on their perceptions of whether they feel certain about how they can get a good evaluation. I find, as expected, that subjectivity has a positive main effect on measures for the effort heuristic and uncertainty.

My research contributes to theory and practice in at least three ways. First, my study contributes to the overwork literature by showing, using causal evidence, how two key management control system features lead employees to work longer and contribute to the propagation of firm overwork cultures (e.g., Ladva & Andrew, 2014; Michel, 2011; Feldman, 2002). Importantly, there is little empirical evidence that examines management controls that lead to overwork and that examine a direct causal relationship between management controls and overwork. I contribute to filling this gap. My findings indicate that how employees are evaluated and how they identify with their colleagues constitute important reasons for their overwork. Understanding factors within firms that propagate overwork is consequential for firms that want to discourage employees from overworking due to its negative consequences, such as a lack of efficiency and employee burnout (e.g., Golden, 2012; Sweeney & Summers, 2002). It is also informative to firms designing their control systems, as it allows them to have a more complete understanding of what motivates their employees, thus addressing a key function of firms' management control systems (Van der Stede, 2015; Sprinkle, 2003).

Second, my study contributes to the subjective performance evaluation literature (e.g., Cai et al., 2022; Kelly et al., 2015). The current literature discusses the uncertainty that comes with subjective performance evaluation and its resulting contracting issues: rational employees may reduce effort because they do not see how their effort translates to compensation due to a lack of clear, objective measurement criteria (Grabner et al., 2020; Luft et al., 2016; Bol, 2008). In contrast, my study develops behavioural theory that suggests this uncertainty can result in judgment heuristics that lead employees to believe working longer will result in better evaluation. Further, a gap in the literature, as noted by Wick (2021), is that there is little research that looks at

the effects of subjectivity in performance evaluation on outcomes other than employee bonuses and performance. In studying employee overwork, my study contributes to filling this gap.

Finally, my study extends the identity literature that examines the impact of identity on employee behaviour (e.g., Brown et al., 2022; Estep, 2021). Current research in management accounting shows how stronger group identity can lead to group aligned behavior such as increased coordination, cooperation, and other-regarding preferences. This in turn can lead to either positive or negative effects on group performance (Shang et al., 2020; Kelly & Presslee, 2017; Towry, 2003). In contrast, I find that stronger group identity can result in behaviour that is not group-aligned, whereby employees being evaluated more subjectively wish to avoid having the lowest level of overwork in their group. Paradoxically, this creates a negative externality in the group because each group member then feels obliged to keep increasing their overwork levels to match that of the rest of the group. This behavior in turn generates exceedingly longer effort duration in the group, which may result in either positive or negative group performance effects.

The remainder of my dissertation is organized as follows. In Chapter 2, I review the overwork, subjective performance evaluation, and group identity literatures. In Chapter 3, I develop my predictions regarding the impact of subjectivity in performance evaluation and group identity on overwork, as well as my performance research question. In Chapter 4, I present the design of my main experiment. In Chapter 5, I present the results of my main experiment. In Chapter 6, I report the design and results of my secondary experiment. Finally, I conclude in Chapter 7.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this chapter, I review the literature examining my main constructs of interest: overwork, subjectivity in performance evaluation, and group identity. In section 2.2, I examine the antecedents and outcomes of overwork and how it has been studied in previous literature, both within accounting and outside. In section 2.3, I examine the literature on subjective performance evaluation and its impact on employee and firm outcomes. Finally, in section 2.4, I provide an overview of social identity theory and describe the antecedents and outcomes of group identity strength.

2.2 Overwork

In this section, I provide an overview of the literature on employee overwork. I expand on the broader management, economics, and psychology literatures that study antecedents to employee overwork. I then discuss consequences of overwork documented in the literature to expand on the importance of understanding overwork in firms. I also discuss overwork as it has been studied in the accounting literature, to highlight its importance to accounting academics.

2.2.1 Overview of Employee Overwork

In 1930, John Maynard Keynes, one of the most important economists of the 20th century, predicted that production would be so efficient by 2030 that people would not need to work more than 15 hours a week. In an ironic twist unanticipated by Keynes, 70-hour work weeks are commonplace just a few years from his predicted cut-off year (Karaian & Sorkin, 2021; Hewlett & Luce, 2006). This is exemplified in a 2008 Harvard Business School survey that found that 94% of surveyed professionals work more than 50 hours a week (Perlow & Porter, 2009). A more recent

2019 survey conducted in the UK found that 91% of professionals at least sometimes work more than their contracted hours (with thirty one percent of respondents *always* working more than contracted hours) and that 90% of those professionals receive no form of additional compensation for working longer than their contracted hours (Morgan McKinley, 2019). Although long working hours have been a long-standing issue, they were also exacerbated during the COVID-19 pandemic (Beheshti, 2021; Vershbow, 2021). For example, a year into the pandemic, analysts at Goldman Sachs, who were working an average of 95 hours a week, threatened to quit their jobs unless their conditions improved (Reuters Staff, 2021).

I define overwork as employees working beyond contractual or statutory standard working time, for no immediate additional monetary gain ("Work-life balance", n.d.; Feldman, 2002).[2] From a statutory perspective, some employees, notably professional employees, may be exempted from regulations that define standard working hours and the requirement to be paid for overtime hours (e.g., Canada Labour Code of 1985; U.S. Fair Labor Standards Act of 1938). However, if these employees' contracts do not make explicit their working hours, then statutory standard hours can nevertheless be used as a benchmark for their standard working time such that hours worked beyond that are considered overwork (Campbell & Charlesworth, 2012). Additionally, whether a contract makes explicit standard working hours, firms often have an internal policy of a certain number of standard hours (e.g., 37.5), acknowledging that hours worked beyond that are considered unpaid overtime (Ladva & Andrew, 2014). Thus, regardless of the exact number of

[2] Some papers define overwork as working more than 50 hours a week (e.g., Cortes & Pan, 2017; Cha, 2010). This is an effective operational definition for survey papers that need to use a clear definition understandable to survey takers, and for papers that rely on archival data that captures employees' working time. However, I use a more conceptual definition to be able to generalize my theory about overwork as supported by experimental findings.

hours beyond which an employee is considered to be overworking, the threshold of contractual or statutory standard working time provides a benchmark for my overwork construct.[3]

Note that overwork is related to the construct of overtime. Overtime is generally defined as "all hours worked in excess of the normal hours, unless they are considered in fixing renumeration in accordance with custom", with "normal hours" defined as "contractual working time, usual working time or statutory working time" (ILO, 2004, p. 1). Overtime can be paid or unpaid. My definition of overwork can thus be considered as employees working *unpaid* overtime.

Finally, as a note, the literature uses a variety of terms to capture the construct of overwork. These include 'long working hours', 'long hours', and 'lack of work-life balance'. I also review the audit literature which captures overwork through various constructs such as 'high workloads' and 'busy season'.

2.2.2 Settings with Overwork

Overwork is often found in the broad context of professional employees (e.g., Ladva & Andrew, 2014; Mazmanian et al., 2013; Michel, 2011; Golden, 2009; Perlow & Porter, 2009; Lewis, 2007; Perlow, 1999; Kunda, 1995; Coffey, 1994). This category of workers consists of employees who are well-educated and qualified, and generally apply their knowledge to specific customer problems (Michel, 2011). Their work tends to involve "high levels of customer engagement, extensive customization, knowledge intensity, and low levels of capital intensity" (Brandon-Jones et al., 2016, p. 9). These employees work in a wide range of jobs, such as

[3] Although the literature sometimes uses the word overwork to mean working beyond one's capacity (making it a subjective notion) (e.g., Golden, 2009), my definition of overwork can be objectively determined based on the threshold of contractual working time or statutory standard working time. Further, overwork captures how long an employee works and is separate from how intensely the employee works. This follows the same distinction made in the effort literature between effort duration and effort intensity (Bonner & Sprinkle, 2002). Thus, I make no assumptions about how intensely employees work when they overwork and, regardless of intensity or whether employees feel overworked, they would still be overworking per my definition if they work past the thresholds of contractual or statutory working time.

accounting, legal, architecture, engineering, technology, banking, consulting, and research and development. (Brandon-Jones et al., 2016). For example, Kunda (1995) examines tech engineers for whom "it is fairly well-established that long hours are the norm" (p. 352). Ladva and Andrew (2014) study the "web of control" produced by junior accountants that overwork, as a means of securing their own identities (p. 634). Michel (2011) examines the controls in banks that cause investment banking employees to experience their long work hours as self-chosen. Lupu et al. (2020) interview accounting and law professionals that "usually work over 50 hours per week and up to 100-hour weeks in busy periods, e.g., the audit season" (p. 10).

Although professional employees tend to exhibit high levels of overwork, I do not limit my theory to settings with professional workers. Rather, I predict that my theory generalizes to settings that exhibit high levels of my antecedents, namely subjectivity in performance evaluation and group identity, irrespective of whether they are settings with professional employees.

2.2.3 Proposed Reasons for Overwork

Extant research outside the accounting literature, in the fields of management, economics, and psychology, has pondered the important question of why employees overwork (for a brief review, see Burke, 2009). Feldman (2002) proposes a theoretical model that can help explain why managers work long hours. He proposes a four-pronged model consisting of economic factors, individual factors, job factors and organizational factors. Below I organize my review into these four factors and literature that falls into these categories.[4]

Economic factors: Feldman (2002) proposes three economic factors that firms may face and that in turn would lead employees in the firm to work longer. He proposes that firms that face

[4] Some factors (e.g., job factors and organizational factors) are particularly relevant from a management accounting perspective and can be considered at various stages in control system decisions.

competitive pressures from other firms, those that have a decline in corporate profits, and those that have threats of layoffs would be more likely to have managers that work longer hours. He proposes that under these organization-threatening conditions, managers may be more willing to work long hours to help ensure the success of the firm, or they may be pressured to do so by shareholders/stakeholders. Further, Golden (2009) and Golden and Altman (2008) also propose several economic-level factors that may determine workers' desired hours: (1) employees' current real wage rates: employees can afford to work fewer hours if their wages are higher, although he argues that this is a basic economic premise that does not do a good job of explaining current long hours; (2) employees' anticipated rewards for working long hours: employees will work longer hours if they expect this to lead to better career progression; (3) the prestige that comes with working long hours: employees will work longer hours when this leads to higher status in the long term, by allowing them to consume status-conferring goods and services; (4) employers' demand for long hours: employers dictate long work hours and workers may not have other options but to accept these hours.

Although economic factors are theorized to play an important role in predicting employees' long hours, studies stress that there are other critical determinants of overwork. I elaborate on these potential determinants in what follows.

Individual factors: Feldman (2002) also proposes that individual-level factors play a role in predicting managers' working time. These factors include gender (male managers are more likely to overwork), marital and family status (single managers and those without children are more likely to overwork), and the salience of the breadwinner role (the more salient this role the greater the overwork). Other factors are one's ability to adapt to behavior in a group (such that this positively predicts overwork), one's conscientiousness as a personality trait and one's achievement

motivation (such that both positively predict overwork), and one's investment in outside-work activities (such that this negatively predicts overwork). Complementing Feldman's (2002) individual-level factors, Golden (2009) theorizes that longer hours can be predicted by employees' intrinsic motivations or benefits acquired from work, such that employees will work longer if their work is more intrinsically rewarding.

Empirical evidence helps support that individual-level factors impact working hours. Ng and Feldman (2008) use a metanalysis to support an identity framework that predicts that variables impacting an individual's occupational identity are predictors of the number of hours worked. These variables include individuals' current salary, career satisfaction, number of promotions, work centrality to the self, career interruptions, education level, work experience, and networking. This is because these variables are associated with individuals' career success, focus and investment. They find that these variables are indeed the strongest predictors of hours worked. Similarly, Major et al. (2002) survey employees at a Fortune 500 company and find that stronger career identities and fewer responsibilities away from work are associated with more time at work. Further, Brett and Stroh (2003) use survey evidence and find that male managers who worked the longest hours received extrinsic rewards for doing so and were the most psychologically involved in their work, even holding their extrinsic monetary compensation constant. Female managers, in contrast, seemed to have multiple reasons for working long hours, including making a trade-off between leisure time and earning money, and extrinsic monetary rewards for the long hours they work. Kuroda and Yamamoto (2019) use panel survey data about Japanese workers and find that although longer work hours have a negative effect on workers' mental health, working longer than 55 hours a week is associated with increasing job satisfaction (proxied by workers' satisfaction with job promotion). The study implies that workers may choose to work extreme hours because

they overvalue the satisfaction they obtain from work and underestimate the negative effects on their mental health.

Job factors: Feldman (2002) further proposes job-level factors that predict managers' work hours. These are particularly relevant to the management accounting literature and to my study. He proposes that when managers' work is less tangible, they work longer hours to demonstrate their work. He also proposes that when managers' work is appraised less specifically and the appraisal criteria are less measurable, employees will be evaluated on their facetime and working hours. This is supported by Wick (2020), a dissertation study that uses an experiment to support that when employees' output cannot be objectively assessed, their work-day duration is used to evaluate the quality of their work, when the purpose of their evaluation is a bonus reward (rather than a promotion). Feldman (2002) also proposes that when managers are evaluated on their performance on discretionary and interpersonal behaviors, rather than just their performance on their tasks, then they are more likely to work longer hours because this helps them build relationships with their peers.

A 1998 report by the Institute for Employment Studies, an independent center for research in the UK, provides some empirical evidence that corresponds to Feldman's proposed job-level factors. The institute conducted a field study with 12 employers in the UK to understand the reasons behind the long hours employees worked (Kodz et al., 1998). Although this is not an academic study, it provides an idea of what employees feel are the reasons they overwork, and it corresponds with much of the extant academic literature. They find that employees overwork due to their high workload and work pressures, tight deadlines, the pressure to perform well at work, and high customer expectations. Employees also feel that working long hours is necessary to be promoted in their jobs.

Organizational factors: Finally, Feldman (2002) proposes organizational factors as predictors of long working hours. These are also especially relevant to the management accounting literature and to my study. Feldman (2002) proposes that firm culture predicts long working hours. He also proposes that employees whose personalities fit the long hours culture select into firms with such cultures and are less likely to quit, while the opposite holds true for employees who do not fit these cultures. In this way, the long hours culture is perpetuated. Relatedly, Afota et al. (2019) develop a theoretical model to explain how employees are influenced by their supervisors to work longer hours, through social contagion. They propose that employees will imitate their supervisors' long working hours when their supervisors' perceived status is higher, when work is more central to the employee, and when the subordinate identifies more strongly with their supervisor. Sullivan (2014) suggests that cultures at "greedy institutions" can be used to subtly prevent employees from taking advantage of family-friendly policies that the institutions themselves have adopted, such that overwork continues despite these policies.

Empirical evidence sheds more light on this. For example, Brett and Stroh's (2003) survey evidence indicates that female managers exhibit evidence of social contagion in their work hours, such that managers who work extreme hours are more likely to be in the financial services industry. They reason that the financial services industry is one that is "the bellwether of overwork" (p. 68), such that social interaction within the industry perpetuates the overwork culture. Similarly, Maume and Bellas (2001) survey a sample of families in Ohio in 1998 and 1999 and find that the best determinant of individuals' working hours is how demanding their supervisors are. Another case in point is a study by Eastman (1998), which examines the extent to which individuals' chosen work hours depends on how long their colleagues work rather than on their own desired number of hours. He conducts a survey with MBA students and asks respondents how many hours they

would work, with commensurate pay, if their colleagues worked different numbers of hours. He finds that respondents' chosen number of hours increases with an increase in the number of hours their colleagues worked, such that individuals' chosen hours is longer than they themselves desire.

Recent qualitative evidence also supports this. Lupu and Empson (2015), an interview study at an accounting firm in France, suggests that employees overwork because they are inadvertently taken in by the game of trying to enhance their status and obtain recognition in the field. Similarly, Peticca-Harris et al. (2015) study blogs by the spouses of game developers about the extreme working conditions in the industry. They suggest that video game developers work long hours due to a combination of a love for their work and "neo-normative control mechanisms" in the form of project-based work and the importance of meeting project deadlines.

Within the extensive literature on why employees work long hours, how management controls of organizations drive employees to overwork is seldom studied. This is despite the central use of management controls by organizations to ensure employees' objectives are aligned with those of the firm (Ladva & Andrew, 2014). Importantly, from a firm's perspective, understanding how management controls impact overwork is important to understand so management can direct employee behaviour to a desired level of overwork. Although employee overwork suggests increased alignment between employee and firm objectives (if overworking has positive returns to firm performance), management may also wish to decrease employee overwork due to its potential negative consequences to firm value.

One of the few papers that consider how management controls within organizations can lead to overwork is Ladva and Andrew's (2014) qualitative study. They propose that in the auditing context a combination of formal and informal controls combine to create the overwork culture in accounting firms. These include pushing auditors to work within budget, auditors' concern for

looking efficient in their work, and auditors' career aspirations. Still, a better understanding of the key features in management control systems that lead employees to overwork is essential to understanding how overwork develops in firms and how these features continuously act as viable control mechanisms in motivating employees.[5]

2.2.4 Consequences to Overwork

In this section, I briefly discuss some of the negative and positive consequences of overwork that have been documented in the literature to illustrate the importance of better understanding the antecedents to employee overwork.

Negative consequences

Negative consequences to overwork have been studied in the literature. Studies find that overwork can lead to a lack of efficiency and employee burnout, which have direct effects on employee turnover, absenteeism, and job performance (Lupu & Ruiz-Castro, 2021; Carmichael, 2015; Steinmetz et al., 2014; James, 2013; Golden, 2012; Burke, 2009; Brett & Stroh, 2003; Sweeney & Summers, 2002; Babbar & Aspelin, 1998; Kodz et al., 1998). For example, Golden (2012), writing for the International Labour Office, notes evidence that worker performance in a white-collar workforce decreases when they work 60 or more hours per week. He also notes that high overtime levels can impact employee morale, productivity and absenteeism, and that overworked employees are likely to make more mistakes at work. Conversely, he notes, shorter hours are associated with higher rates of output in many industries in the United States. This is consistent with Pencavel (2015), who examines a sample of munition workers during World War I and finds that after a threshold of 48 hours a week, there is a decline in the marginal product of

[5] In my study, I assume that there are no signals from management that overworking is expected. This is because I want to isolate other important factors that lead employees to work longer, besides simply responding to management instructions.

hours. Further, “those weeks without a day of rest from work had about 10% lower output than weeks when there was no work on Sunday holding weekly hours constant” (p. 2072). Similarly, Steinmetz et al. (2014), using data from a web survey, find that working overtime is associated with a lower intention to stay among healthcare employees in Belgium, Germany and the Netherlands. Overwork also has consequences to family life (Burke, 2009). For example, Major et al. (2002) conduct a survey on employees in a Fortune 500 firm and find that working longer hours is associated with more work-family conflict and is indirectly associated with psychological distress in employees. The literature, in turn, indicates the impact of work-family conflict on employee job satisfaction (Lapierre et al., 2008), an important Corporate Social Responsibility (CSR) consideration for firms and an indicator of future financial performance (Dhaliwal et al., 2012; Banker & Mashruwala, 2007).

On a macro-level, Cortes and Pan (2017) use cross-country evidence to show that when more skilled men work longer than 50 hours a week, the labor force participation rate of skilled, ever-married women goes down. They also find that when an occupation has more skilled men working longer than 50 hours a week, this reduces the percentage of skilled, ever-married women in that occupation. In the same vein, Cha (2010) examines the impact of overwork on women; specifically, she uses archival panel data to examine the effect on wives’ careers when their husbands work long hours and finds that husband overwork increases the likelihood that their wives will quit their jobs. The reverse effect does not seem to hold true, such that wives working long hours does not impact whether their husbands will quit. The effect of husbands working long hours on their wives quitting is even stronger when they have children, compared to if they are childless. She argues that long hours thus perpetuate arrangements in households where men are the breadwinners and women are the homemakers, when these households were once dual income.

Relatedly, Cha and Weeden (2014) use archival data to support that overwork contributes to the wage gap between men and women; this is because men are more likely to overwork than women, and because overwork comes with a wage premium.[6]

Positive consequences

Despite the negative consequences associated with overwork, it is evident that firms value and reward it, as supported by Gicheva (2013), who finds a positive association between long hours and wage growth in data from a panel survey. Among other things, this is likely because firms see overworking employees as committed to their jobs (Brett & Stroh, 2003). Further, as noted earlier, Kuroda and Yamamoto (2019) find that Japanese workers who work more than 55 hours a week have increasing job satisfaction with hours worked, at least as it relates to satisfaction related to job promotion. This suggests that at the individual level there may be positive consequences to overworking. At a more fundamental level, firms benefit from overwork because it is cheaper to hire fewer employees who each overwork than to hire more employees to work the same number of hours (Dembe, 2009). More broadly, from a management accounting perspective, overwork may be considered an indicator of increased effort since it is a form of effort duration. This in turn may be seen by firms as an indication of better alignment between employee and firm objectives. However, importantly, overwork is at a specific end of the effort duration construct, such that it can lead to either better or worse firm performance.

[6] This wage premium does not refer to an immediate monetary gain received by employees for the time they work above contractual or statutory standard working time. Rather, it refers to a wage premium trend that occurs due to several possibilities, including overwork being concentrated "among highly educated, professional, and managerial workers" who experienced a large wage growth over the years (p. 460), as well as tournament-style compensation systems that may use overwork as a proxy for workers' productivity and cause overworkers to "win" the competition (p. 460), or even due to macrostructural changes that mean that core employees who tend to overwork are paid more than employees "who work part-time, under subcontracts, or in temporary positions for lower pay" (p. 461).

2.2.5 Overwork in the Accounting Literature

Overwork is important to examine in an accounting context, particularly from a management accounting perspective. Management accounting research is broadly interested in how managerial controls can be used to mitigate employee control problems, such as a lack of employee motivation (Van der Stede, 2015; Sprinkle, 2003). Interestingly, in settings where overwork occurs, employees are seemingly very motivated, working beyond their contractual or statutory working time for no immediate additional monetary gain; yet, we have little understanding of what management controls lead to this overwork and why they do so. Below I discuss what has been examined about overwork in the broad accounting literature.

The audit literature examines constructs related to overwork, such as workload pressure and audit busy season (e.g., López & Peters, 2012; Agoglia et al., 2010). The auditing context exhibits overwork due to the "the tension between limited audit resources and the need to complete a high number of audit engagements within a limited time window" (López & Peters, 2012, p. 139). Generally, the literature indicates that overwork has negative consequences on auditors and their work quality. For example, Christensen et al. (2021), using archival data, find that higher audit team workloads have a negative effect on audit quality, particularly for workloads above 60 hours a week.

Studies also indicate the adverse effects of overwork on auditors and their perceptions of their work. Persellin et al. (2019) indicate using survey evidence that auditors perceive their high workloads to lead to worse audit quality and worse job satisfaction. Sweeney and Summers (2002) use a longitudinal survey and find that by the end of busy seasons, when auditors worked an average of 63 hours a week, there was a direct relationship between auditors' workload and job burnout, exhibited by emotional exhaustion and depersonalization in their job approach.

Interestingly, the relationship between workload and burnout did not hold for the pre-busy season, even though auditors were working an average of 49 hours per week. They propose that this could be because public accountants develop higher resistance for workload pressure.

More broadly, overwork is a component of effort duration, and the management accounting literature has examined effort duration as a dependent variable. For example, Awasthi and Pratt (1990) experimentally show that participants who are offered monetary incentives, compared to those under fixed pay, spend more time on their tasks. Sprinkle (2000), in a study examining employee learning and performance, finds a similar effect: participants receiving incentive-based contracts spend more time on the task than those receiving flat-wage contracts. Cloyd (1997) finds that accountability – being required to justify decisions to others – increases tax professionals' effort duration on a tax-research task. Chan et al. (2021) show that when employees are rewarded for the time they spend working, they increase their effort duration but decrease their effort intensity, due to their fairness concerns about their incentive system. Interestingly, Yatsenko (2022) examines the impact on workers' productivity of peers' effort duration being observable; he finds that social comparison incentivizes workers to *reduce* their effort duration (in the absence of information about peers' relative output) because workers want to compare favorably against each other on their *ability*, and ability is inversely related to effort duration.

Note that although overwork is a component of effort duration, as noted above, overwork is at a specific end of the effort duration construct, such that it can lead to either better or worse firm performance. Thus, it needs to be better understood as a distinct dependent construct, and my study examines it as such.

In sum, several factors have been predicted to lead employees to overwork, namely economic, individual, job and organizational level factors. Further, despite firms valuing

overwork, the literature documents undesirable consequences to overwork. Importantly, there is little empirical evidence that examines management controls that lead to overwork and that examine a direct causal relationship between management controls and overwork.

2.3 Subjectivity in Performance Evaluations

As my dissertation examines how subjectivity in performance evaluation impacts employee overwork, I provide an overview of the use of subjectivity in performance evaluation in incentive contracts and what the literature has documented on how subjectivity impacts employee and firm outcomes. In section 2.3.1, I provide an overview of subjectivity in performance evaluation. In section 2.3.2, I discuss the benefits of subjectivity, and in section 2.3.3, I discuss the costs of subjectivity, and how the downside of subjectivity contributes to the uncertainty it creates for employees.

2.3.1 Overview of Subjectivity in Performance Evaluation

Subjectivity in performance evaluation is the extent to which "judgment based on personal impressions, feelings, and opinions" is used in evaluating an employee's performance (Bol, 2008, p. 2). Incorporating subjective performance evaluation in contracts – as opposed to objective performance evaluation - is a common method by which employee performance is evaluated and by which employees are incentivized; in fact, it is more prevalent in incentive contracts than objective performance evaluation. Thus, it is a key component of employee incentive contracts (Bol, 2008; Gibbs et al., 2004; Prendergast 1999; Murphy & Cleveland, 1995).

There are generally three types of subjectivity in performance evaluation. The first is the use of subjective performance measures, which are "superiors' subjective judgments about qualitative performance indicators" (Moers, 2005, p. 68). The second is allowing discretionary ex-post adjustments in the weighting of objective performance measures, and the third is allowing ex-

post adjustments based on factors other than the ones specified before the contract period (Bol, 2008). My study is generally more focused on the first type of subjectivity; however, my review of the literature encompasses all three types, following previous literature reviews that do not differentiate between them in their discussions of the implications of subjectivity (Bol, 2008; Wick, 2021).

To begin, subjectivity brings benefits to compensation contracting (Baker et al., 1994). When compensation contracts rely solely on incomplete objective performance measures, this incentivizes employees to focus exclusively on the job dimensions that these performance measures capture and neglect the job dimensions that are more difficult, and costly, to objectively evaluate (e.g., teamwork, time management; Grabner et al., 2020; Gibbs et al., 2004; Prendergast, 1999). Subjectivity in performance evaluation thus allows supervisors to consider those important job dimensions not captured by objective measures when evaluating their employees and helps align the employee's objectives with that of the firm. I expand on these benefits in the next section.

Despite the benefits that come with subjectivity in performance evaluation, the literature has also identified costs to it (Wick, 2021; Bol, 2008). The main cost is that subjective performance evaluation can be inaccurate (Gibbs et al., 2004). This inaccuracy is either due to evaluators' cognitive limitations that limit their use of all the available information about an employee in their evaluation, or because they intentionally choose to be inaccurate, due to perceived costs of being more accurate (Bol, 2008). For example, supervisors may be more lenient to avoid unpleasant interactions with employees (Bol et al., 2010). They may also provide better ratings to employees with whom they have stronger relationships (Bol, 2008; 2011). I expand on these costs in section 2.3.3.

Below I present relevant literature on the impact of subjective performance evaluation on employee and firm outcomes.

2.3.2 Benefits of Subjectivity in Performance Evaluation

I start with the benefits of subjectivity in performance evaluation in employee compensation contracts, to justify their common use in firms. Subjectivity can be beneficial to employee outcomes, such as employee performance, through several mechanisms. One such mechanism is that it can be used to reduce the risk on employees from explicit incentive contracts that can be too noisy; that is, explicit contracts using only objective performance measures capture both employees' effort as well as uncontrollable events outside employees' control, increasing the risk on employees (Bol, 2008). Thus, subjectivity reduces the risk on employees and strengthens the incentive contract because it allows the employer to consider uncontrollable events *ex-post*. This function of subjectivity has been documented in archival studies. For example, Anderson et al. (2020) use survey and archival data from a large retail firm and find that non-executive managers are evaluated more highly on the *subjective* part of their performance evaluations when their *objective* measures of sales performance are more likely to capture noise due to uncertainty in the environment and when they have greater risk from more difficult performance targets. In this way, the subjective performance evaluations adjust for managers facing higher risk in their compensation from only relying on objective measures. Similarly, Gibbs et al. (2004) use data from 250 car dealerships and find that the use of subjectivity in bonus allocations is positively associated with the difficulty of meeting a consequential performance target, reducing the risk on employees from failing to meet the target.

Further, partly in reducing noise from uncontrollable events, subjectivity can increase employees' trust and fairness perceptions (Voußem et al., 2016). For example, Aranda et al. (2019)

find, using an experiment, that if employees set more demanding performance targets, managers may reduce the risk on these employees by subjectively rewarding them with a bonus; in turn, this subjectively rewarded bonus elicits trust from employees and increases employees' performance in subsequent periods. Similarly, Kelly et al. (2015) find that when there is subjectivity in ex-post goal adjustments, and goals are moderately difficult, employees perform better because they feel that having subjectivity entails better procedural justice. Along similar lines, Bol et al. (2010) use archival data from Korean business units and find that companies use their discretion to subjectively set more achievable targets for stores that face more risk, and they also use their discretion to increase fairness in the compensation contract by setting easier targets for stores with more perceived unfairness. Cai et al. (2022) find that when subjectivity is used to override objective rankings to reward employees, this is perceived by employees as favorable treatment compared to a formula-based reward, and they find that this elicits improved employee performance.[7]

Team settings are also important settings in which subjectivity can be used to enhance perceived fairness (Wick, 2021). Working in teams makes it harder to objectively assess each team member's individual performance and so subjectivity serves as an important mechanism for better tying rewards to individual contribution (Arnold et al., 2020; Arnold et al., 2018; Bol & Leiby, 2018; Maas et al., 2012; Fisher et al., 2005). For example, Arnold and Tafkov (2019) experimentally support that when a team task is not interdependent, subjectivity in allocating bonuses has a positive effect on team performance compared to an equal bonus allocation. This is because rewards are tied to each team member's contribution to the team's output and this makes

[7] However, they also find the reverse: when subjectivity is used to override objective rankings and *punish* employees, this is perceived by employees as unfavorable treatment compared to a formula-based reward, and this results in decreased performance.

the reward allocation fairer.[8] In a similar vein, Arnold et al. (2018) find that using team members' subjective communication to their manager enables the manager to better allocate a fair bonus to team members based on their contributions to the team output. However, they find that this benefit is reduced (although still positive) when team members' abilities are heterogeneous, because their views on fairness diverge and become less usable by the manager; this in turn weakens the link between contribution and reward and reduces the positive impact of team member communication on team performance.

Although subjectivity is used to enhance fairness perceptions, research finds that there is a limit to its usefulness in achieving this. Specifically, Voußem et al. (2016) use a time-ordered cross-sectional survey study to look at the impact of subjectivity on employees' perceptions of justice; they find that it follows an inverted u-shape, such that lower levels of subjectivity improve justice perceptions, but higher levels decrease it. This is mainly because the high levels of subjectivity are likely to come with the possibility of inaccurate ratings, as mentioned previously, as well as a lack of clarity about what is needed to be evaluated well. I expand on these downsides in the next section.

Finally, another important benefit to subjectivity in performance evaluation is that it elicits adaptive behavior from employees. For instance, Kelly et al. (2020) examine lawyers in a field study and find that when law firm partners' pay is not transparent, such that the impact of each partner's actions on the pay of other partners is less evident and social comparison concerns are less at play, subjectivity in performance evaluation reduces partners' accedence to client wishes

[8] Arnold and Tafkov (2019) also find that when the task *is* interdependent, which is often the case when work is organized by teams, team performance is negatively impacted by subjectivity in allocating bonuses. This is due to reduced cohesion in the team due to competitive tendencies that come with the subjectivity in allocating the bonus. As they note, this is ironic in that subjectivity becomes least useful in a context that most requires teamwork.

when they potentially deviate from best professional judgment. This is because subjectivity induces more focus on the long-term reputational impacts versus the short-term economic impacts of the partner's actions. Similarly, Cheng and Coyte (2014) find that subjectivity in the weighting of performance measures (versus using a formula-based incentive scheme) is more likely to induce employees to share their knowledge and perform desirable behaviors that are not explicitly rewarded.

2.3.3 Costs of Subjectivity in Performance Evaluation

Although much research has found benefits to subjectivity, there are also costs to using subjectivity in performance evaluation. An important cost to subjectivity in evaluating employees is that it creates uncertainty for the employee being evaluated (Bol, 2008). As opposed to objective performance measures, which set clear measurement criteria, subjective performance measures provide a murkier path for an employee's performance improvement (Bol, 2008). These measures provide general qualitative evaluation metrics with no clear parameters on how to be evaluated well (e.g., teamwork, leadership; Luft et al., 2016; Moers, 2005). The greater uncertainty that comes with higher subjectivity has been theorized to cause employees to reduce their effort because they see a weaker link between their actions and compensation (Bol, 2008).

The literature provides empirical evidence on the uncertainty inherent in subjectivity and some of its effects. For example, Luft et al. (2016) find that the uncertainty in subjective performance evaluation leads to failures in subordinates making decisions that their superiors will reward, when subordinates receive additional non-financial and external information. Moreover, the uncertainty also leads employees to expect their evaluations to be higher than their superiors actually evaluate them. This is because employees are more motivated than their superiors to see themselves as good performers, and they are consequently negatively surprised by their actual

evaluations. Likewise, Van Rinsum and Verbeeten (2012) use survey data from public sector managers in the Netherlands and find that subjectivity decreases managers' perceived mission clarity, which in turn reduces their motivation. Further, Kunz (2015) finds that subjectivity has a negative motivational effect on employee's effort when employees are low in autonomous motivation, because these employees benefit from having objective and precise performance assessments, rather than the uncertainty and lack of precision in subjective performance measures.

Another contributing factor to the uncertainty inherent in subjective performance evaluation is that it relies on the evaluator's impressions, by definition. These impressions are subject to the evaluator's cognitive limitations, or biases in evaluators' judgments, due to their use of judgment heuristics.[9] Furthermore, evaluators' impressions are constrained by the complexity of the evaluation process, leading the employee to be less certain about how they will be evaluated and what they need to do to be evaluated well (e.g., Grabner et al., 2020; Demeré et al., 2019; Jawaher & Williams, 1997).

Much literature has studied inaccurate performance assessments as an important cost of subjectivity. For example, Bol and Smith (2011) use an experiment and find that managers' subjective evaluations of their employees are biased in the direction of employees' performance on an objective measure. This objective measure captures a performance dimension that is unrelated to their subjectively evaluated performance, and so this spillover effect is not justified by its informativeness. In the same vein, Bloomfield et al. (2021) find that female analysts, but not male analysts, are subjectively evaluated as less promotable when they perform the same action,

[9] Heuristics are cognitive shortcuts, and biases are the resulting systematic errors in judgments from using heuristics (Kahneman, 2011).

an action that is seen as non-conforming to typical analyst behavior. This is due to a bias that results from individuals' tendency to use categorization as a shortcut in their judgments.

More importantly, studies also document performance costs of the inaccuracy in subjective performance evaluations. For example, Bol (2011) uses performance data from a financial services company to study the impact of this inaccuracy on employee performance. She studies the effects of leniency bias, the tendency to inflate employees' evaluations, and centrality bias, the tendency to compress employees' evaluations towards the center of the evaluation scale. Although she finds that leniency bias has a positive effect on performance improvement, likely due to better employee fairness perceptions, she also finds that centrality bias hurts employee performance improvement because it creates a disproportionate pay-to-performance ratio. Similarly, Ahn et al. (2010) use archival data about Korean public enterprises and find that subjectivity discriminates less between ratees (the public enterprises) due to centrality bias. In turn, ratees are less incentivized by subjectivity to improve their performance, compared to objective measures which show more discriminability. Further, Bachmann et al. (2020) find that when CEOs' incentives are linked to non-financial performance targets, which have a high degree of subjectivity, this has a negative relationship with subsequent firm performance. This is likely because these subjective targets do not accurately link CEO incentives to firm performance.

In sum, the literature finds both costs and benefits to subjectivity in performance evaluation in employee incentive contracts. One such cost is that there is uncertainty inherent in higher subjectivity in performance evaluation. A gap in the literature noted by Wick (2021) is that there is little research that looks at the effects of subjectivity in performance evaluation on outcomes other than employee bonuses and performance. In studying employee overwork, my study contributes to filling this gap.

2.4 Group Identity

In this section, I provide an overview of research on group identity. In section 2.4.1, I introduce the theories from which the construct of group identity derives, namely Social Identity Theory and Organizational Identification. In section 2.4.2, I briefly discuss the antecedents of employee group identity strength to present the sources for group identity heterogeneity. In section 2.4.3, I examine the consequences of stronger group identity documented in the accounting, organizational behavior, economics, and psychology literatures to elucidate its importance as an informal control, and I discuss how stronger group identity impacts social comparison tendencies.

2.4.1 Social Identity Theory and Organizational Identification

Group identity is the extent to which a person defines herself as a member of a group, such that she derives her sense of identity from being a member of that group (Chen & Li, 2009; Hogg & Terry, 2000; Hogg & Turner, 1987; Tajfel & Turner, 1979). Group identity is based on Social Identity Theory (SIT), which expands on the concept of an individual's social identity (Tajfel & Turner, 1979): an individual knowing she belongs to a social group, such that belonging to this group contributes to the individual's identity and such that this group has value and emotional significance to her (Hogg & Terry, 2000; Tajfel & Turner, 1979). Much SIT research examines the phenomenon of in-group bias, which is the tendency to favor one's in-group at the expense of the out-group, even with very little to trigger an awareness of one's group (Tajfel & Turner, 1979). For example, individuals tend to reward more money to anonymous, arbitrarily assigned in-group members, and they feel more positively about in-group versus out-group members (Chen & Li, 2009; Hogg & Turner, 1985; Tajfel & Turner, 1979).

Organizational identification research in turn is derived from SIT, and it specifically examines the role of employees' identification within organizations (Ashforth & Mael, 1989). It

argues that one's social identity can also be derived from their organization or from groups within the organization such as teams or departments. Organizational identity thus has implications on employees' utility functions and their behavior (Heinle et al., 2012; Ashforth et al., 2008; Ashforth & Mael, 1989). In the words of Akerlof & Kranton (2005), identifying with the organization means "workers are willing to put in high effort rather than low effort with little wage variation" (p. 10), and workers gain utility from acting towards the organization's objectives. This is exemplified in Abernethy et al. (2017), who use third-party survey data about financial controllers. They find that although performance-based compensation is associated with more earnings management, when financial controllers identify more strongly with their organization, they are less likely to manipulate earnings in response to performance-based compensation. They argue that this is because managers with higher organizational identity "will experience disutility when making reporting choices that may increase their own welfare, but may come at the expense of the firm [i.e., manipulating earnings]" (p. 3).

Overall, organizational identification outcomes include positive ones such as increased intrinsic motivation, effort, cooperation, and knowledge-sharing (Ashforth et al., 2008). As such, organizational identity can be used as an important form of managerial control, including in knowledge-intensive firms where more bureaucratic forms of control are less feasible on their own (Kärreman & Alvesson, 2004). Notably, overidentification with the organization may come with some detriments, including acting unethically on behalf of the organization (Ashforth et al., 2008).

Organizational identification within a firm includes organizational identity, group identity, team identity, and identity with specific persons (such as superior-subordinate identity; Ashforth et al., 2008; Rousseau, 1998). I focus the rest of my review on group identity and team identity;

that is, I focus on individuals' identity with their colleagues, rather than with the organization as a whole entity, as that is more directly related to my research question.

Notably, a person belonging to a group can vary in how much her sense of self derives from belonging to that group, such that she can have stronger or weaker group identity with her group (Shang et al., 2020; Mael & Ashforth, 1992). At a conceptual level, I am interested in how individuals who belong to a group vary in how strongly they feel an affiliation with the group. For example, an employee assigned to work with a team in her company is a member of that team (i.e., it is her in-group); however, she may vary in how much her sense of self derives from being a team member. In contrast, an employee who works on a temporary basis with another team is not a member of that team (i.e., it is her outgroup). I am interested in a setting where there is variation in the extent of belonging *within* an ingroup. This is because I am interested in settings where employees are working with colleagues in their own group, rather than with a group they do not belong to.[10]

2.4.2 Antecedents of Employee Group Identity Strength

Firms can deliberately manage group identity among their employees and as such can use group identity as an informal control. Some of the ways firms strengthen group identity include allowing employees to interact more frequently and organizing team-building events. For example, companies like Facebook (now Meta) and SalesForce use team-building experiences like adventure games, scavenger hunts, karaoke, and music competitions (Caprino, 2016). Further, group identity can fluctuate in a firm because of unrelated firm choices, such as hiring employees

[10] An individual can identify with multiple groups. For example, a person can identity as an employee of company X, female, and African, as well as identify as a member of a work group. A person tends to follow the most salient identity (or the identity activated in a certain situation) when making a decision (Forehand et al., 2002). In my context, the employees' other identities, besides their identity with their colleagues or work group, are not relevant to my theory, because they are less salient in the context of my study compared to their identity with their colleagues.

that have similar or dissimilar characteristics (e.g., age and educational background), or organizing work outcomes such that outcomes are team-driven versus individually driven (Kelly & Presslee, 2017; Konnikova, 2014; Akerlof & Kranton, 2008; Lembke & Wilson, 1998; Ashforth & Mael, 1989). Group identity also increases in groups that feel distinct or differentiated from other comparable groups, groups that feel prestigious, and in the presence of a salient out-group (Kärreman & Alvesson, 2004; Ashforth & Mael, 1989). Further, leadership style can impact group identity. For example, Mitchell et al., (2015) find that inclusive leadership (defined as "the inclusion of all team members in discussions and decisions and in which … divergent perspectives are explicitly valued and encouraged" (p. 220)) is positively associated with team identity in interprofessional teams because it causes all team members to feel like valuable contributors to the team.

2.4.3 Outcomes of Employee Group Identity Strength and Group Identity in the Accounting Literature

Group identity has been studied in the organizational behavior, economics, and psychology literatures, as well as in the accounting literature, due to the implications it has on employee behavior in group settings. These implications can be beneficial for firms and are the reason group identity is considered an informal control (Kärreman & Alvesson, 2004). When an individual identifies strongly with her group, she feels stronger ties with her group members, likes her group members more, believes her group members are similar and share similar attitudes to her, and feels a sense of belonging with her group members and that the group is more central to her identity (Estep, 2021; Kelly & Presslee, 2017; Cameron, 2004; Ashforth & Mael, 1989; Hogg & Turner, 1985; Hensley & Duval, 1976). Cameron (2004) empirically supports that group identity is comprised of three factors: centrality (how frequently the group comes to one's mind and the

importance of the group to one's self-definition), ingroup affect (the extent that a person feels good knowing they are part of a group), and ingroup ties (how bound an individual feels to the group).

Broadly, stronger group identity comes with benefits such as increased coordination, knowledge sharing, mutual monitoring, and cooperation (Shang et al., 2020; Akerlof & Kranton, 2005; Eckel & Grossman, 2005; Van der Vegt & Bunderson, 2005; Cabrera and Cabrera, 2002; Lembke & Wilson, 1998). It may also enhance commitment to the team, unity with and altruism towards group members, and appraisals of group members, as well as lead to better adherence to group values and a greater willingness to enforce group norms (Chen & Li, 2009; Goette et al., 2006; Ashforth & Mael, 1989). Interestingly, group identity can also contrarily lead to increased social comparison against group members because group members are seen as closer and more similar to each other (see, for example, Liu (2017), which I expand on below). This latter outcome of group identity forms a basis for my hypothesis development in Section 3.5.

In the management accounting literature, extant research has examined the outcome of group identity on employee performance and effort. For example, Towry (2003) examines team identity in the context of two incentive systems that rely on mutual monitoring – a vertical incentive system where peer observations are reported to management, and a horizontal system where peer observations are used to directly control each other's actions. Towry (2003) shows that when team identity is stronger, a horizontal system is more effective than a vertical system because team identity increases team member coordination. A vertical system relies on the assumption that team members will each be working for their own objectives (i.e., will not be coordinating) and it does not do well when identity is stronger. In contrast, a horizontal system relies on team members agreeing and enforcing high effort levels amongst each other, and so does better when team identity is stronger.

Relatedly, Shang et al. (2020) use a field-based dataset of archival and survey evidence from a Chinese state-owned enterprise to examine the interaction of group identity and performance transparency on group members' performance. They find that group identity can have two different performance effects on group members: they may want to increase their performance in order to work in the best interest of the group (the interest alignment effect), or they may conform to what other group members are doing to look more similar to group members (the conformity effect). In particular, when performance of other group members is transparent, this activates the conformity effect, which may decrease group performance as group members who were performing better than the rest of the group may decrease their performance to conform. However, when performance transparency is low, then this activates the interest alignment effect of stronger group identity, which increases performance of the group.

Furthermore, Kelly and Presslee (2017) use an experiment to study the impact of tournament group identity on tournament performance. They find that stronger group identity increases tournament participants' other-regarding preferences for their fellow participants, which in turn decreases their competitiveness in the tournament and their performance. This effect is even stronger when winner proportion is larger, when losing a reward in the tournament comes at a lower cost. Along similar lines, Brown et al. (2022) find that identifying with one's immediate sub-group (e.g., work team) and identifying with a superordinate group (e.g., one's company) both increase helping behavior between subgroups and interact so that combined they have a stronger effect.

The audit literature has also examined group performance on audit outcomes. Bauer and Estep (2019) show that when auditors and IT specialists see each other as part of separate teams, rather than one team, there is limited communication and coordination in completing the audit.

Conversely, Estep (2021) finds benefits to weaker team identity between auditors and IT specialists; specifically, she finds that a weaker team identity highlights the differences between auditors and IT specialists and emphasizes IT specialists' knowledge, "resulting in heavier weight of the specialist's input on IT-related issues" (p. 264). It also results in auditors differentially weighting higher and lower quality input for non-IT issues because they are less likely to use a trust heuristic with IT specialists with whom they weakly identify.

Finally, as evidenced by the literature, the strength of group identity means individuals with stronger group identity are prone to viewing group members as like them in attributes and opinions, and they are prone to feeling close to them (Estep, 2021; Hogg & Turner, 1985; Wilder, 1984; Hensley & Duval 1976). This in turn means that group members are also seen as relevant social comparison targets, and it further means that stronger group identity comes with more social comparison against group members (Garcia et al., 2013; Tesser & Campbell, 1982; Gastorf & Suls, 1978; Hoffman et al., 1954; Festinger, 1954). For example, Liu (2017), an accounting dissertation examining the effect of group identity on sabotage behavior, finds that group identity increases employees' tendency to sabotage their group members when employees are evaluated using relative performance evaluation. Theoretically, the study argues that this is because group members tend to compare themselves more against each other when they identify more strongly with the group. As I discuss in Section 3.5, the different impact of group identity on employee behavior (e.g., helping behavior versus sabotaging) can be largely attributed to the difference in the importance of the domain of comparison to the employee's self-esteem, such that a domain that is important to an employee's self-esteem is likely to induce social comparison concerns.

In sum, group identity can be deliberately managed in a firm and used as a management control to achieve firm objectives. While much research finds benefits to firm outcomes when

group identity is strengthened between employees, research has also documented some benefits to weaker group identity (see Figure 1 for a summary of the antecedents and outcomes of group identity strength). In Chapter 3, I outline how group identity acts as an antecedent to employee overwork.

CHAPTER 3: HYPOTHESES DEVELOPMENT

3.1 Introduction

In this chapter, I use theory on the effort heuristic and social comparison to develop hypotheses about the impact of subjectivity in performance evaluation and group identity on employees' level of overwork. First, I specify the setting that I study and in which I test my theory. My objective is to clarify the conditions under which I develop my theory and to which my hypotheses should generalize. Second, I discuss the relevance of my two antecedents of overwork – subjectivity in performance evaluation and group identity - in settings that often exhibit high levels of overwork. Third, I discuss the uncertainty that comes with subjective performance evaluation, and I examine the impact of uncertainty in the use of judgment heuristics. This in turn creates the link between subjectivity in performance evaluation and employees' level of overwork in Hypothesis 1. Fourth, I study the impact of employee group identity on employees' tendency to socially compare, and I examine the impact of social comparison on the relationship between subjectivity in performance evaluation and overwork. This provides the basis for group identity as a moderator of the relationship between subjectivity in performance evaluation and employees' level of overwork in Hypothesis 2. Fifth, I discuss the main effect of group identity on employees' level of overwork, forming Hypothesis 3. Finally, I examine the performance effects of subjectivity in performance evaluation interacting with group identity, as mediated by overwork, forming Research Question 1.

3.2 Setting Features

Similar to some prior accounting research (e.g., Bauch et al., 2021; Maas et al., 2012; Bol & Smith, 2011), I examine a single-period setting. This allows me to understand how overwork first develops in a work setting, which is the main interest of this study. I examine a setting with one task to provide a stronger test of theory. Further, my setting assumes employees care about their work and about how they are evaluated, due to inherent career and reputational concerns (Holmström, 2017). My setting is also one in which employees work in groups, where there is variation in the extent of belonging *within* an ingroup (and not, as noted in Chapter 2, where an employee works with an outgroup). Relatedly, although employees work in groups, each employee's work is evaluated on a standalone basis; this mirrors work settings where each employee's individual contribution to the group's work is known to a large extent by the evaluator(s) in the group (Arnold et al., 2020; Arnold & Tafkov, 2019). There is also no incentive pay tied to group performance. This is an assumption that simplifies the setting and removes other aspects of group dynamics aside from group identity, as I intend to show that group identity is consequential for employees even in the absence of explicit incentives for the whole group to perform well.

Further, to understand how high overwork levels develop, I assume that there are no norms in the work setting that signal to the employee that their level of overwork matters. This is an assumption that biases against finding support for my theory; yet I intend to show that, even in the absence of such signals, employees anticipate that overworking matters and they increase the extent they overwork. This is also an important setting to consider given what prior research has found about employees feeling that their overwork is self-imposed, and that employees choose to overwork despite firms implementing and encouraging work-life balance policies (e.g., Ladva &

Andrew, 2014; Michel, 2011). Finally, my setting is one in which employees have a neutral level of trust in their evaluator, because the evaluator has no explicit incentive to be untrustworthy. In a similar vein, I make no assumptions about how much employees identify with their evaluator.

3.3 Relevance of Subjectivity and Group Identity in Settings with High levels of Overwork

As noted in Chapter 2, my study is not limited to a particular category of workers. Rather, I expect that any setting that exhibits high levels of my antecedents, namely subjectivity in performance evaluation and group identity, would also exhibit high levels of overwork. However, because professional employees tend to exhibit high levels of overwork (e.g., Ladva & Andrew, 2014; Mazmanian et al., 2013; Michel, 2011), my two antecedents should have the potential to exist at a high level in professional settings. Notably, subjectivity in performance evaluation is common in professional roles because output tends to be difficult to evaluate objectively (Grabner et al., 2020; Ladva & Andrew, 2014; Kärreman & Alvesson, 2004). The work of professionals involves applying expertise to propose solutions to different problems and professionals tend to work in teams. The result is that their output is intangible (unobservable) and difficult to evaluate objectively at the individual level (Arnold et al., 2020; 2018; Bol & Leiby, 2018; Maas et al., 2012; Fisher et al., 2005). Similarly, their roles involve different aspects that are important to achieving firm objectives, but that cannot be objectively assessed. These include client service and technical ability (Buchheit et al., 2003; Anderson-Gough et al., 2000; Baker et al., 1994). An employee's group identity with her colleagues also has the potential to be strong amongst professionals, because they tend to work in close proximity with each other and often work in teams, and they also tend to work with colleagues who are like them in education level and knowledge (Hewlett & Luce, 2006; Kärreman & Alvesson, 2004).

3.4 Subjectivity in Performance Evaluation and Overwork

In this section, I develop theory supporting the positive main effect of subjectivity in performance evaluation on employee overwork. As noted in previous literature discussed in Chapter 2, subjectivity in evaluating employees creates uncertainty for the employee being evaluated (Bol, 2008). As opposed to objective performance measures, which set clear measurement criteria, subjective performance measures provide a murkier path for an employee's performance improvement (Bol, 2008). These measures provide general qualitative evaluation metrics with fewer clear parameters on how to be evaluated well (qualitative metrics include, for example, measures of employees' teamwork or leadership) (Luft et al., 2016; Bol, 2008; Moers, 2005). Subjective performance evaluation, by definition, relies on the evaluator's impressions; these impressions are subject to the evaluator's cognitive limitations and are constrained by the complexity of the evaluation process. This leads the employee to be less certain about how she will be evaluated and what she needs to do to be evaluated well, compared to if an objective evaluation were used (e.g., Grabner et al., 2020; Demeré et al., 2019; Jawaher & Williams, 1997). I argue that one way that an employee will respond to the uncertainty that comes with the perceived subjectivity in her evaluation is that she will use a heuristic to try to anticipate how she will be evaluated.

3.4.1 Uncertainty and the Use of Heuristics

Heuristics help simplify complex judgments, and when individuals face uncertainty in a decision situation, they are more likely to rely on heuristics when making judgments (Kruger et al., 2004; Chaiken and Maheswaran, 1994; Tversky and Kahneman, 1974). In fact, Neth and Gigerenzer (2015) argue that heuristics, rather than being irrational, *should be* used under conditions of uncertainty. Mousavi and Gigerenzer (2017) similarly note the following:

> [U]nder uncertainty, optimization is by definition infeasible, either because the full set of options and consequences or their probabilities cannot be known in advance or because the problem is computationally intractable… people's use of heuristics under uncertainty cannot be universally attributed to cognitive limitations. Rather, heuristics can be ecologically rational. (p. 365, 375)

The literature provides evidence of the use of heuristics in response to uncertainty. Kruger et al. (2004) find that participants are more likely to use an effort heuristic when there is ambiguity associated with the object they are assessing. In particular, they are more likely to judge a higher effort object as better than a lower effort object when there is more ambiguity associated with it. Likewise, Van den Berg et al. (2021) find that experimental participants are more likely to use heuristics in a cooperative social situation when there is uncertainty about the payoff of different cooperation strategies. Further, Chang (2004) finds that consumers rely on a country-of-origin heuristic when evaluating product quality only when advertising messages convey ambiguous information about the product. Thus, overall, evidence supports that when there is uncertainty in a context, individuals use heuristics to make judgments and decisions.

3.4.2 The Effort Heuristic

In my setting of interest, when an employee is uncertain about how she will be evaluated, she must make a judgment on what she expects will be important metrics in her evaluation. I posit that in doing so, she will default to using a heuristic to simplify this judgment. Specifically, she will use what is known as the effort heuristic, which is the belief that "[t]he more effort [is] invested in an object… the better it is deemed to be" (Kruger et al., 2004, p. 92).[11] The object in this context

[11] There are individual differences in the use of heuristics in decision-making (Jackson et al., 2016) such that the effort heuristic may differ across individuals. I make a general assumption that on average individuals will use the effort heuristic; although I measure its use in my experiment and account for differences in its use across individuals.

is the employee's output; thus, an employee would assume that if she is seen to exhibit more effort in producing her output, by increasing her level of overwork, then her output will be deemed to be of better value and it will be evaluated more highly. Overwork, a form of effort duration, is a visible form of effort, such that it is a good choice as a means for employees to exhibit their effort (as opposed to working more intensely, for example).

There is much research showing that people use effort exerted in creating an object as a heuristic for the value of the object (e.g., Inzlicht et al., 2018; Schrift et al., 2016; Buell & Norton, 2011; Yeung & Soman, 2007; Morales, 2005; Kruger et al., 2004). Much of this research studies the perspective of the person making a judgment on a specific object rather than the perspective of the person creating the object and anticipating how it will be evaluated. For example, Wick (2020) shows that managers evaluate employees' output as worse if employees work for less time compared to their colleagues, if the purpose of the evaluation is a bonus. Likewise, Kruger et al. (2004) show that objects are judged to be of higher quality when participants are told that they took more effort and time to produce.

Further evidence is found in the consumer and marketing literature. Yeung and Soman (2007) find that consumers evaluate an exercise program more highly if it lasts longer, when they also know the price of the program. Buell and Norton (2011) find that in some online self-service settings, customers prefer to wait longer for a service if the work that the website is doing in delivering the service is made apparent. This is because customers perceive more effort involved in delivering the service, and this increases the perceived value of the service. Different from the previous papers, Cheng et al. (2017) examine the perspective of the individual exerting effort when creating an object rather than the perspective of the person judging a specific object in which effort was exerted. The authors argue and find that some people are more inclined to believe that greater

cost, including effort, implies greater outcomes, based on an individual difference variable known as the Protestant Work Ethic.

Similarly, I posit that employees will anticipate that their output will be evaluated more highly by their superiors if they exhibit higher effort. This is because when employees face uncertainty due to subjectivity in their evaluation, they use the same heuristic that more effort means more value. Employees are also likely to anticipate that their managers will use the same heuristic. For example, accounting research provides evidence that individuals can anticipate how others will act and behave accordingly (e.g., Hecht et al., 2019). Specifically, employees will anticipate that if they increase their level of overwork, a visible form of effort, they will be evaluated more highly. Thus, employees will come to believe that the extent of their overwork is an important metric in their evaluation, and they will increase their level of overwork.

In contrast to when employees are evaluated subjectively, when employees are evaluated based on objective performance metrics, such as the quantity of output produced, they will focus on actions that these objective metrics measure (Baker et al., 1994; Holmström & Milgrom, 1991). Indeed, this is one of the main criticisms of using objective performance metrics, especially in multi-task environments: employees focus on actions that are measured and rewarded at the exclusion of other important but unmeasured tasks (Bol, 2008; Holmström & Milgrom, 1991). Therefore, under a more objective performance evaluation, it is not as necessary for employees to rely on judgment heuristics because they have a clearer path for how their performance will be evaluated well. Consequently, they do not need to rely on the effort heuristic, and they will find it less necessary to overwork to increase the perceived value of their output.[12]

[12] In the special case where employees' objective metric is some measure of their working time, then employees are likely to work longer (Chan et al., 2021). However, I do not consider this case in my study for two reasons. First,

Overall, I expect that when performance evaluation is more subjective, employees will use the effort heuristic. In turn, they will increase the extent of their overwork, since they believe that this will lead to better performance evaluation. Thus, I predict the following (see Figure 2 for the conceptual model):

> **Hypothesis 1:** Greater subjectivity in performance evaluation will lead to a greater level of overwork.

There is tension in the above hypothesis. Rational employees may reduce effort when there is greater subjectivity in their evaluation because they do not see how their effort translates to compensation, due to a lack of clear, objective measurement criteria (Grabner et al., 2020; Luft et al., 2016; Bol, 2008). However, few papers in the literature empirically test the direct relationship between subjectivity and effort (Kunz, 2015; Gibbs et al., 2004). Those that test this relationship support arguments by MacLeod (2003) that there will be less agency costs associated with subjectivity if the employee and supervisor both "have shared values regarding what constitutes good performance" (p. 218). For example, Gibbs et al. (2004) show that subjectivity increases (has no effect on) managers' productivity if managers have (not) had a long tenure with the firm, with tenure acting as the proxy for trust between the manager and the firm. Thus, if there is sufficient trust between the employee and their evaluator, this reduces the cost of uncertainty in subjective

---

many professional employees overwork without being evaluated on chargeable hours (e.g., investment bankers; Baker et al., 1994). This is also supported by public accountants underreporting their time, but still working long hours (Agoglia et al., 2015; Gonzalez, 2014; Nelson & Tan, 2005; Anderson-Gough et al., 2001; Ponemon, 1992). Second, even when chargeable hours are used as an indicator of performance, it is typically one small metric within a much wider reliance on subjective performance metrics, with no direct path between compensation and working hours (Grabner et al., 2020; Buchheit et al., 2003). For example, in auditing firms, where employees may be partly evaluated on their utilization, their evaluation places much more weight on subjective metrics, such as client service and technical ability; even their utilization metric does not reward them for working longer, but rather for being more efficient (Ladva & Andrew, 2014; Buchheit et al., 2003). Law firms also have only a subtle link between hours worked and compensation (Campbell & Charlesworth, 2012). Thus, it remains an important question why employees overwork when they are not explicitly compensated or incentivized to work longer.

performance evaluation. As noted above, my setting of interest is one in which employees have a neutral level of trust in their evaluator, because the evaluator has no explicit incentive to be untrustworthy.[13]

Another source of tension is an argument based on goal theory (Locke & Latham, 2002). This argument may predict that employees who are evaluated based on objective metrics have more specific goals and so should exert more effort than employees evaluated based on "do your best goals"; the latter may be more congruent with the uncertainty inherent in subjective performance evaluation. However, with objective metrics, the goal is specifically the metric measured (e.g., sales figures), such that the effort duration of employees on the task is only a secondary or tertiary factor in achieving that specific goal.[14] Thus, the level of overwork is not the goal when subjectivity in performance evaluation is low. In contrast, with greater subjectivity in performance evaluation, overwork becomes the primary and most salient goal (due to the effort heuristic). Thus, with an increase in subjectivity in performance evaluation, employees' goals change to overwork goals. Thus, this source of tension does not change my prediction in Hypothesis 1.

3.5 Interactive Effect of Group Identity and Subjectivity on Overwork

My second antecedent to employee overwork is an employee's group identity with her colleagues, which I predict will moderate the relationship between subjectivity and overwork (see Figure 2 for the conceptual model). As noted in Chapter 2, group identity is the extent to which a

[13] In my operationalized setting, I am silent on the extent of trust the participant should have in the evaluator. However, I expect my participants will not view the evaluator as untrustworthy because the evaluator has no explicit incentive to be so.

[14] In achieving these goals, employees may also trade-off between effort duration and a more invisible form of effort, which is effort intensity (Locke & Latham, 2002); the latter would not translate to overwork. For example, employees may work more intensely to achieve a sales goal without overworking.

person defines herself as a member of a group, such that she derives her sense of identity from being a member of that group (Kelly & Presslee, 2017; Towry, 2003; Hogg & Terry, 2000; Hogg & Turner, 1987; Tajfel & Turner, 1979).

I posit that group identity moderates the relationship between subjectivity and overwork, such that the effect of subjectivity on an employee's level of overwork will be greater under stronger, versus weaker, group identity. Briefly, this is based on the premise that an employee who identifies more strongly with her colleagues sees her colleagues as relevant social comparison targets and is more prone to comparing herself against them. She is thus motivated to compare favorably in that comparison (Brewer & Weber, 1994; Tesser, 1988; Tesser & Campbell, 1982; Festinger, 1954; Festinger et al., 1954). In my context, she becomes concerned with her group members' work rather than just her own and she is increasingly motivated to compare favorably on her level of overwork as subjectivity in her performance evaluation increases. I expand on the details of the theory supporting this in the subsections that follow.

3.5.1 Group Identity and Social Comparison

I start by outlining why stronger group identity leads to more social comparison, which is the mechanism by which I posit group identity moderates the relationship between subjectivity in performance evaluation and the level of overwork. In essence, group identity impacts social comparison because identifying more strongly with group members leads group members to feel closer to each other. In turn, feeling closer to group members makes them feel like more relevant social comparison targets (Buunk et al., 2012; Smith, 2000; Kruglanski & Mayseless, 1990; Festinger, 1954).

In particular, closeness increases with anything that links one individual with another, such as group membership, physical proximity, demographic similarities, or common characteristics

(Tesser & Campbell, 1982, p. 262). Because one of the drivers of social comparison is a basic desire to evaluate one's abilities and opinions, it is more useful to evaluate oneself by socially comparing against close others: those one has a relationship with or is psychologically closer to (as opposed to strangers) and similar (as opposed to dissimilar) individuals (Garcia et al., 2013; Suls et al., 2002; Tesser & Campbell, 1982; Gastorf & Suls, 1978; Hoffman et al., 1954; Festinger, 1954). I expand on the closeness construct and how it relates to social comparison in the following paragraphs.

Individuals who are psychologically closer (e.g., friends) as opposed to farther (e.g., strangers) are more relevant comparison targets, and when one compares against someone who is psychologically closer, this impacts their self-evaluations to a larger extent. This is exemplified in a few studies. For instance, Tesser and Smith (1980) hypothesize that socially comparing against a friend, versus a stranger, hurts one's self-evaluation if the friend is perceived to be outperforming them on a task. In support of this, they find that lab participants are more likely to help strangers versus close friends on a task that they are told measures important characteristics (such as verbal intelligence). Similar support is found in other studies (e.g., Tesser et al., 1984; Tesser & Campbell, 1982). Overall, evidence indicates that there is a greater incentive to maintain a positive self-image in a comparison with psychologically closer individuals (Tesser, 1988).

Likewise, individuals are more likely to socially compare and be impacted by their social comparison when they compare against similar others (Garcia et al., 2013; e.g., Tesser & Campbell, 1980). For example, Brown et al. (1992) find that female college students appraise their own attractiveness differently depending on whether they are shown a picture of an attractive or unattractive person, but only if the picture they are shown is that of a female. They posit that this is because females compare their attractiveness to those that are similar to them, i.e., other

females. Complementing this notion, Kilduff et al. (2010) study a more intense form of social comparison – rivalry – and find that among national college basketball teams, the most intense rivalry occurs between teams that are similar in terms of geography (closer to each other or in the same state), current and historic basketball status, and university academic quality.

Importantly, both psychological closeness and perceived similarity are directly impacted by group identity. When an individual identifies strongly with her group, she feels stronger ties with her group members (i.e., feels psychologically closer), likes them more, and believes they are similar and share similar attitudes to her (Estep, 2021; Kelly & Presslee, 2017; Cameron, 2004; Ashforth & Mael, 1989; Hogg & Turner, 1985; Hensley & Duval, 1976). Thus, overall, her group members are seen as relevant social comparison targets, and they socially compare against each other more. In my broader theoretical model, this increased social comparison forms the basis for the moderating effect of group identity in the relationship between subjectivity in performance evaluation and the level of overwork.

Finally, individuals are motivated to socially compare on a domain of interest (Brewer & Weber, 1994; Tesser, 1988; Tesser & Campbell, 1982; Festinger, 1954). The domain of interest provides the basis for whether individuals compare against other individuals to begin with. In fact, if the domain is not relevant to the individual, then individuals may derive increased self-evaluation from a close other's good performance on the domain, rather than feel worse about their own inferior performance. This is supported by several studies (e.g., Tesser & Campbell, 1982; Pleban & Tesser, 1981). For instance, Tesser and Campbell (1982) find that individuals will predict their friends' performance on a task to be lower when the task is one that is relevant versus one that is irrelevant to the individual.

Because individuals socially compare on domains that they find important or relevant, in the context of my study, a requirement for social comparison to occur is that employees care about their work or at least find their evaluation on their work to be relevant to their self-evaluation. I believe this assumption is met in my setting, such that employees care about how they are evaluated due to inherent career and reputational concerns (Holmström, 2017).[15,16]

It is important to note that individuals are motivated to maintain a positive self-image when comparing themselves against others (Tafkov, 2013; Beach & Tesser, 1995; Tesser, 1988). This forms the basis of the self-evaluation maintenance model by Tesser (1988), which assumes that people behave to maintain or increase their self-evaluation, and that peoples' relationships impact their self-evaluation. Thus, individuals are motivated to work to ensure they maintain their positive self-image and avoid being the worst in their social comparison (Festinger et al., 1954).

3.5.2 Group Identity as a Moderator

Overall, I expect the effect of subjectivity on an employee's level of overwork will be greater under stronger versus weaker group identity due to increased social comparison processes. Specifically, I expect the employee to be increasingly motivated to increase her level of overwork in response to greater subjectivity in performance evaluation when she socially compares. In comparing, she becomes concerned with not just her own work but also how her work compares to that of her group members, and social comparison happens more when she has *stronger* (versus weaker) group identity. To maintain a positive self-image in that comparison, the stronger (versus

[15] I operationalize the importance of the domain of interest in my experiment by explicating in the instructions that the evaluations participants receive on the experimental task are an indication of analytical and critical thinking skills, meaning that the comparison domain should be important to the person because their performance is indicative of an important ability (Tafkov, 2013).

[16] Although I assume this requirement to be met even at weaker levels of group identity, it is probable that with stronger group identity the comparison domain - employees' work and their work evaluation - becomes more important to them because employees gain utility from working for the benefit of a group they identify with (Akerlof & Kranton, 2005).

weaker) group identity employee will want to be at least as good as her colleagues on the overwork metric (Festinger et al., 1954). This means that the employee cannot be the first to stop working and she needs to continue increasing the extent of her overwork until she surpasses a minimum standard set by another group member (i.e., someone else must stop working first, before she can stop).[17] Note that at lower levels of subjectivity, when employees are assessed on more objective performance metrics, stronger group identity will lead employees to socially compare and increase their work on whatever they are objectively assessed [e.g., sales figures], rather than on their level of overwork.

On the other hand, when group identity is *weaker*, higher subjectivity in performance evaluation will have a smaller impact on the employee's level of overwork because she will be less concerned about socially comparing the extent of her overwork against her colleagues' level of overwork. She will instead set her own arbitrary limit on how much she should overwork and stop when she believes she has overworked enough.

Overall, my model describes subjectivity in performance evaluation as the factor that pushes employees to increase their level of overwork, while stronger group identity dials up this relationship by increasing the motivation of employees to increase their level of overwork. Thus, I predict the following:

> **Hypothesis 2**: The positive effect of subjectivity in performance evaluation on the level of overwork will be higher when group identity is stronger compared to weaker.

See Figure 3 for a graph of the hypothesized results.

[17] Meanwhile, other group members are also trying to avoid being the first to stop working, which continues to increase the minimum standard of overwork in the group, up to some limit.

Tension exists in the above hypothesis. Because stronger group identity can increase other-regarding preferences between individuals (Kelly & Presslee, 2017), it is possible that the level of overwork will not increase more in response to subjectivity in performance evaluation when group identity is stronger compared to weaker. Instead, other-regarding preferences may lead group members to be reluctant to increase their level of overwork in response to subjectivity due to an awareness that doing so would compel other group members to increase their own level of overwork. Thus, this awareness of a possible negative externality from increasing one's own level of overwork could prevent the positive interaction effect hypothesized in Hypothesis 2. However, I do not expect this source of tension to be strong. This is because, as noted above, I examine a setting that employees find important to their self-evaluation, such that social comparison concerns should be primary to them (rather than other-regarding preferences). Thus, any sacrifices in their evaluations for their group members are likely to be perceived as significantly psychologically costly.

3.6 Main Effect of Group Identity on Overwork

My final prediction posits a main effect of group identity on overwork. In particular, as group identity with colleagues becomes stronger, this also increases the likelihood that employees will work longer, and increase their level of overwork, to try to benefit their group. This is based on economic theory that posits that employees gain utility simply from working with, and for the benefit of, a group with which they identify (Akerlof, 2016; Akerlof & Kranton, 2005; 2000). Thus, my prediction is as follows:

Hypothesis 3: Stronger group identity will lead to a greater level of overwork.

There is some tension in this prediction. In particular, employees may prefer to spend time outside of work with a work group with whom they strongly identify, rather than overworking to spend time with them at work. However, I still predict that on average, within the work context, employees would overwork to a larger extent when they identify more strongly (versus more weakly) with their group.

3.7 Performance Effects

Although not my primary interest in this study, my setting of interest allows for me to examine the interactive effects of subjectivity in performance evaluation and group identity on performance, as mediated by overwork. Ex-ante, it is unclear how an increased level of overwork would be related to performance. On the one hand, the increased effort duration that comes with overwork may translate into better performance (Bonner and Sprinkle, 2002). On the other hand, employees may prioritize overworking as an end goal, rather than focusing on improving their actual performance, such that overwork may not translate to better performance. In fact, research reveals that managers cannot differentiate between an employee who works long hours and one that only pretends to work those hours, implying that overwork is not necessarily correlated with performance (Reid, 2015). Thus, I pose the following research question:

> **Research Question 1**: What is the interactive effect of subjectivity in performance evaluation and group identity on task performance?

I acknowledge that my setting is a one task, single period one and so generalizability of effects related to Research Question 1 may be limited outside this type of setting. I also acknowledge that there may be other mechanisms by which subjectivity and identity may impact performance, such as through changes in effort intensity, work strategy, etc.; however, these mechanisms are outside the scope of my study.

3.8 Conclusion

To conclude, this chapter develops three hypotheses that predict that subjectivity in performance evaluation and employees' group identity with their colleagues are antecedents of employee overwork. The first hypothesis predicts that greater subjectivity in performance evaluation leads to greater levels of overwork; this prediction is based on psychology theory, specifically on the effort heuristic. The second hypothesis predicts that employees' group identity with their colleagues acts as a moderator between subjectivity and the level of overwork, such that the positive effect of subjectivity in performance evaluation on the level of overwork will be higher when group identity is stronger versus weaker. This is also based on psychology theory, specifically on social comparison. My third hypothesis predicts a main effect of group identity on overwork, based on economic theory linking group identity and utility. Finally, I pose a research question on the performance effects of subjectivity in performance evaluation interacting with group identity, as mediated by overwork.

CHAPTER 4: MAIN EXPERIMENT RESEARCH METHOD

4.1 Introduction[18]

In this chapter, I describe my main experiment which I use to test Hypotheses 1, 2, and 3, and to explore Research Question 1 (RQ1). Section 4.2 provides an overview of the experiment. Section 4.3 provides details about participant recruitment. Section 4.4 describes the experimental procedures and task details. Section 4.5 describes the independent variables and manipulation check measures. Section 4.6 describes the dependent variables, followed by Section 4.7., which describes the post-experimental measures. Section 4.8 details the pilot studies testing the efficacy of my independent variable manipulations. I conclude in Section 4.9.

4.2 Overview

I test my hypotheses using a 2 x 2 between-subjects online experiment.[19] I manipulate the strength of group identity as either *Stronger* or *Weaker*. I manipulate subjectivity in performance evaluation as either *Higher* or *Lower*. Both manipulations are done using vignettes, in which participants are told to imagine that they are part of a scenario. Participants then complete a writing task. My primary dependent variable, level of overwork, is how long participants spend on the writing task beyond how long I contract them to spend on the task.

---

[18] All experiments (including all pilot tests conducted) received ethics clearance through a University of Waterloo Research Ethics Board (REB #43366 and #44103).

[19] The strength of using an experiment to test my hypotheses is that it allows me to test my causal model by isolating my independent variables of interest and randomly assigning participants to my manipulated conditions (Sprinkle & Williamson, 2006). A survey, on the other hand, would not be conducive to testing the impact of my two management controls on overwork, as it would be difficult to isolate the effects of subjectivity in performance evaluation and group identity from other variables (such as task type, employees' views on the management control system, etc.). Although there may be challenges associated with generalizing findings about time from the lab to the real world, prior accounting research has examined time variables in the lab without compromising construct validity (e.g., Waddoups, 2022; Hecht et al., 2020). In the same way, my construct of overwork is amenable to operationalizing in the lab as it only depends on the time for which participants are contracted to work as a benchmark.

4.3 Participant Recruitment

I recruit my participants from Amazon Mechanical Turk (MTurk, via CloudResearch). My experimental task does not require subject matter expertise, and so MTurk workers are an appropriate participant group. Online labour market participants have been shown to exert similar effort levels to students and are good proxies for non-expert workers (Farrell et al., 2017). They are also more diverse and representative of the general population than student participants (Buchheit et al., 2018; Paolacci & Chandler, 2014). However, as will be noted later in this chapter, online participants may be less susceptible to my group identity manipulation, compared to other participant groups. As such, I pilot test my manipulations with a group of MTurk workers (see Section 4.8).

In recruiting participants, I screen for CloudResearch approved MTurk workers (i.e., MTurkers who have been vetted by CloudResearch for their evidence of attention), who have an approval rating higher than 95% and who have completed more than 100 Human Intelligence Tasks (HITs) on the platform (Bentley, 2021; Eyal et al., 2021).[20] I use these screens to reduce inattentive participants in my study. Only participants in the United States can participate. I restrict participation to workers in the United States to reduce noise that could come from having workers with heterogeneous backgrounds (Bentley, 2021).[21] I do not allow participants who participated in the pilot studies I conduct before the main experiment to participate in this study.

[20] HITs are the tasks that workers complete on Amazon Mechanical Turk and for which they get paid.

[21] Nevertheless, I expect my theory to generalize beyond the United States unless there are cultures with beliefs that would not lead to the use of the effort heuristic. This has not been explicitly examined in prior literature, although Cheng et al. (2017; discussed in Chapter 2) replicate their Protestant Work Ethic results across participants in the United States and Asia.

4.4 Experimental Procedures and Task Details

To begin, participants consent to participate in the study and answer a reCAPTCHA question to ensure they are not bots.[22] Then, participants are told that they will be asked to complete a writing task as part of the study, and that their writing output evaluation will be sent to them after they complete the study. They are told their evaluation will either be Outstanding, Acceptable or Poor. Participants' anticipation that they will be evaluated is intended to make them invested in *how* they will be evaluated (i.e., at a higher or lower level of subjectivity). This in turn has the implication of allowing participants to feel the uncertainty (lack of uncertainty) that comes with higher (lower) subjectivity in evaluation and provides a strong test of my theory. Evaluating participants also allows for a social comparison requirement to be met, that is, the importance of the comparison domain, because it underlines to participants that the task is of importance to their self-evaluation. This also allows for a stronger test of theory. Finally, evaluating participants increases the external validity of my experiment since in a real work setting employees are invested in how they are evaluated.

Participants are then told to imagine that they are in the scenario outlined on the screens that follow while they are completing the writing task. Here, participants are randomly assigned to the *Stronger* or the *Weaker Identity* condition. Group identity is manipulated using vignettes describing a hypothetical team in a company that participants are told to assume they are working with (see Independent Variables section). In both conditions, participants are told to assume that they are a member of a work team and that their team members are named Kai, Aly, and Fin. Following that, participants are asked to briefly describe what they think it would be like to work

[22] Bots are "computer programs that fill out web-based surveys with random responses" (Xu et al., 2022, p. 343). In other words, they are not real human participants completing the study.

with the team in the vignette. The purpose of this open response is to increase the saliency of the identity manipulation (Estep, 2021). My group identity manipulation checks follow the open-response question, on a separate screen (measures are described in the Independent Variables section).

As part of the experimental setting, participants are then told that they and their team members work for a consulting company and that they will be working on a report for a client. They are told that each team member will be working on a section of the report, and each section will be evaluated separately. This ties back to my setting of interest, where employees work in a team but not necessarily on the same task (i.e., it is not a group task where they are all evaluated together). Participants are told that they will receive fixed payment for their work.[23]

At this point, participants receive either the *Higher* or *Lower Subjectivity* manipulation. Each manipulation is captured in a different vignette (see Independent Variables section below). Participants are told either that their evaluation will be evaluated subjectively (*Higher Subjectivity* condition) or objectively (*Lower Subjectivity* condition). On the same screen, they are given the subjectivity manipulation check (measures are described below). Next, participants are told that they are contracted to work on their section of the report for 5 minutes.[24] Participants are informed that they can choose to work longer than 5 minutes on their task. Then, they see the screen with

[23] I do not offer participants performance contingent pay as I expect that doing this would prompt all participants to stay longer in anticipation of a reward, such that it would obscure any real test of my theory within a relevant range that I can test in an experiment. In other words, if I did this, I would need a much longer experiment to be able to test my theory.

[24] I pilot test the time I contracted participants to work (5 minutes), to ensure that it allows for variation in the time worked on the writing task. Specifically, I test whether the *Higher Subjectivity/ Stronger Identity* condition differs significantly in the level of overwork compared to the *Lower Subjectivity/ Weaker Identity* condition. These are the two conditions that should theoretically differ the most from each other. I find that under a contracted time of 5 minutes, these conditions differ significantly in their level of overwork. This provides confidence that a contracted time of 5 minutes allows me to have a strong test of my theory.

the comprehension check questions to test their understanding of the instructions. These comprehension checks comprise the following (True/False responses): "In this hypothetical scenario, I am asked to assume that I am a member of a team in a consulting company"; "My actual writing evaluation grade (Outstanding, Acceptable, or Poor) will be sent to me in a few days through Amazon Mechanical Turk"; "I am contracted to work on my section of the report for 5 minutes, for which I will be paid a fixed payment. However, I may work on my section for longer than 5 minutes, if I choose to". If the participant responds incorrectly to these comprehension checks (i.e., if they answer 'False' for any of them), they are told that their response is incorrect and can answer the question again (I use these comprehension check questions to drop inattentive participants. This is explained more fully in Chapter 5).

Participants then receive the writing task, where they are given a writing prompt (*Please provide your opinion on the following statement: "People communicate with each other less effectively now than in the past because of social media"*).[25] On the same screen, they are given information about their hypothetical team members, specifically the vital piece of information about whether their hypothetical team members are still working on the task (see Appendix C for the task screen participants see). Each hypothetical team member has a mark next to their name indicating to participants whether their team members are still working on the task, to allow for social comparison of time worked. Team members still working on the task are indicated using a green mark next to their names.[26] I keep these green marks constantly present for all hypothetical

[25] Although it is unlikely that firms would ask their employees to write a response to this particular prompt, the prompt I choose is one on which most people should have an opinion and to which they should be able to write. Similarly, in a real work setting, employees would have tasks for which they are qualified and for which they are able to produce output (even if they are not writing tasks).

[26] Although I expect this operationalization to capture any setting where employees can observe their team members and socially compare their level of overwork against them, a real work setting that directly compares to my experiment is one in which employees work from home and can observe whether their team members are still logged in and "online" or if they have logged off.

team members throughout the experiment, in all conditions. This means that participants can see that their hypothetical team members are working on the task until participants submit the writing task and move to the next screen. By keeping this element constant throughout the time of the writing task and across all conditions, I reduce unnecessary noise and strengthen the test of my theory. Note that participants would only notice and be concerned that their other team members are still working on the task if they are already engaging in social comparison of overwork; otherwise, they would complete the writing task and submit it without considering whether their hypothetical team members are still working. Finally, on this same screen, participants are also given a timer showing them their own time working on the task. Their timer stops and they stop being able to see other team members' timers once they submit their writing task. They can only submit their task after five minutes have concluded.

Importantly, to understand how high overwork levels develop, I assume that there are no norms in the work setting that signal to the employee that their level of overwork matters. This is an assumption that biases against finding support for my theory; yet I intend to show that even in the absence of these signals employees themselves anticipate that overworking matters, and they increase the extent they overwork.

The experiment ends with a post-experiment questionnaire and demographic questions. Participants submit the experiment to receive their compensation. All participants are paid within two days of completing the study. Participants also receive their evaluation grade through Amazon Mechanical Turk (via CloudResearch) within a few days of completing the task. Note that for participants who are in the *Lower Subjectivity* condition, I evaluate their writing output using a

writing software, Grammarly, to grade the writing output.[27] In contrast, in the *Higher Subjectivity* condition, I depend on my own personal view on their writing output for their evaluation grade.

Importantly, two features in my experiment should prompt social comparison concerns to affect behaviour. First, participants are told that all sections of the report worked on by them and their team members are of the same length and difficulty. This fulfills the condition that the task performed should be similar across individuals to allow for a meaningful comparison (Tafkov, 2013). Second, participants are told at the beginning of the study that they may find the evaluation they receive on their writing task useful, as writing task evaluations are an indication of analytical and critical thinking skills. This fulfills the social comparison condition that the comparison domain should be important to the person (Tafkov, 2013). This means that participants will want to compare favourably on their performance in the task because their performance is indicative of an important ability. These two features are present across all conditions. These two experimental features are also found in my setting of interest: employees that work in groups tend to work on similar tasks (especially if they are at a similar level in the group hierarchy), and their evaluation on their work is meaningful and indicative of their ability at work. Finally, the impact of social comparison on a person's behaviour also increases if the comparison target is similar to the person on attributes that predict performance (Tafkov, 2013). This should differ by group identity condition (*Stronger* vs *Weaker*), according to my theory.

[27] https://www.grammarly.com/

4.5 Independent Variables and Manipulation Check Measures

4.5.1 Group Identity

Group identity is manipulated at two levels: *Stronger* vs *Weaker*. This manipulation follows the team identity manipulation used in Estep (2021). For each condition, participants are provided with a vignette that describes a hypothetical team the participant is told to assume they work with. The vignettes touch on three factors of group identity derived from the group identity literature: (1) ingroup ties or attitude similarity, (2) liking of ingroup members or ingroup affect, (3) belonging to the group or the group's centrality to the self (Estep, 2021; Kelly & Presslee, 2017; Cameron, 2004; Ashforth & Mael, 1989; Hogg & Turner, 1985; Hensley & Duval, 1976). The *Stronger* (*Weaker*) *Identity* vignette describes the participant as having high (low) levels of these factors in relation to the hypothetical team. My vignette manipulation of group identity thus changes different pieces of information at two levels: how much participants have in common with their team (a lot / not much), how much they like and trust their team members (very much / not particularly) and how close they feel to them (very close / not close). I also change how important the team is to them (very important / not very important) and how happy they are to be a part of the team (very happy / not very happy). Please see Appendix C for the exact wording of the vignettes in both conditions.

Immediately after reading the vignettes, participants answer three group identity manipulation check questions adopted from Kelly and Presslee (2017): the extent they would be happy to be part of the team (*Happy*), the extent they would feel like a member of the team (*Belong*), and the extent they would like their team members (*Like*), all measured on a 7-point scale from (1) 'not at all' to (7) 'a great extent' with an unlabeled midpoint. I also use the Aron et al. (1992) *Inclusion of Other in the Self* Scale (*IOS Overlap* measure), a validated measure of

identity (Estep, 2021; Bauer, 2015; Tropp & Wright, 2001). I show participants seven images of two interconnecting circles, one representing the participant and one the team described in the vignette: the first image is coded 1 (no overlap between circles / weakest group identity) and the last coded 7 (near-complete overlap / strongest group identity).

4.5.2 Subjectivity

Subjectivity in performance evaluation is manipulated at two levels: *Higher* vs *Lower*. My manipulation broadly follows Gorenflo and Crano (1989). Specifically, I manipulate participants' perceptions of whether their work will be evaluated more subjectively or objectively. In the *Lower Subjectivity* condition, participants are told that the evaluator will be able to objectively evaluate their output from the writing task because they have objective writing criteria that they will use (i.e., the number of spelling mistakes, grammar and punctuation mistakes in the report section, and the conciseness and formality of the writing). They are told that their evaluation will not be based on the evaluator's personal views or opinions on each section of the report. In contrast, in the *Higher Subjectivity* condition, participants are told that the evaluator will *not* be able to objectively evaluate their output from the writing task, even though they have some objective writing criteria that they might refer to (the same criteria as the *Lower Subjectivity* condition). They are told that their evaluation will be based on the evaluator's personal views and opinions on each section of the report.[28] Please see Appendix C for the exact wording of the vignettes in both conditions.

[28] I evaluate participants' writing task and send them their evaluation grade after they complete the task (see Procedures section). As noted in the previous section, my method of evaluation does in fact vary depending on whether participants are in the *Higher* or *Lower Subjectivity* condition. In the *Higher Subjectivity* condition, I depend on my own personal views on participants' writing output. In contrast, in the *Lower Subjectivity* condition, I use a writing software, Grammarly, to grade the writing output. Grammarly outputs a numerical score on a scale from 0 to 100. Thus, within the *Lower Subjectivity* condition, I translate any score above 70 to be 'Outstanding' and any other score to be 'Acceptable', unless the participant did not write a response that fits with the prompt, in which case a 'Poor' is awarded. I thus send participants their score as 'Outstanding', 'Acceptable' or 'Poor'. Within the *Higher Subjectivity* condition, the grades I award participants based on my subjective evaluation is highly correlated with the evaluation

Immediately after the manipulations, participants answer a subjectivity manipulation check measure: the extent they agree with the statement "I feel that my section of the report will be evaluated subjectively, based on the evaluator's personal views and opinions on each section of the report", measured on a 7-point scale from (1) 'strongly disagree' to (7) 'strongly agree' with all points labelled.

4.6 Dependent Variable

My dependent variable of overwork is operationalized as the time participants spend on the writing task minus the 5-minute time I contract them to work on the task (*DV_Overwork*, in seconds). I use two dependent variables for performance for RQ1: the first is the score rewarded by the Grammarly writing software for all participants in all conditions (*Grammarly_Performance*), and the second is the score rewarded by an independent rater for all participants in all conditions, on a scale from 1 (weak) to 7 (excellent) (*Rater_Performance*). The independent rater is a PhD student (not the author of this dissertation) who was blind to condition when rating participants' writing output. *Grammarly_Performance* and *Rater_Performance* are highly positively correlated ($r(284) = 0.47$, $p < 0.001$, Table 9, Panel C).

4.7 Post-experimental Measures

At the end of the experiment, I ask participants to answer measures in the post-experimental questionnaire. All items are measured on a 7-point scale from (1) 'strongly disagree' to (7) 'strongly agree' with all points labelled. First, I ask an effort heuristic measure: "I believe I will

---

grade Grammarly would have awarded participants (χ^2 (4, N = 136) = 138.10, $p < 0.001$). Two participants did not have a Grammarly score because their responses were too short for Grammarly to output a score.

receive a higher evaluation if the evaluator knows I spent more time on my section of the report" (*Effort Heuristic*). I developed this measure based on theory (e.g., Kruger et al., 2004). It is mainly intended as a process measure, to better support why subjectivity may impact overwork, but can also be used as an individual difference measure (see section 5.2.4 for more details on this). I also ask measures related to participants' perceptions of their writing task evaluations: "In this writing task, I believe it is difficult for the evaluator to evaluate writing quality" (*Difficult*) and "I feel certain about **how** I can get a good evaluation on my section of the report" (*Certain;* **bolded** in experiment). I also ask a social comparison question: "While I was completing the writing task, I thought about how my performance on the writing task compared to my team members' performance" (*State_Social_Comparison*; adapted from Tafkov, 2013). These measures are intended as process measures, to gain a better understanding of the underlying theoretical mechanism behind subjectivity in performance evaluation and group identity as antecedents to overwork.

I then ask questions related to participants' perceptions of the task: "I was motivated to do well on this task" (*Motivated*), "It was important for me to be evaluated well on this task" (*Important*), and "I understood the task instructions" (*Understood*). These measures allow me to control for participants who are not invested in the task or did not feel they understood the task, and so can help increase the power of my tests.[29]

Finally, I ask three trait measures verified in the literature: "I do not like to work on a problem unless there is a possibility of finding a clear-cut and unambiguous solution" (*Ambiguity*

[29] There may also be differences in these measures between conditions (e.g., stronger group identity may lead to higher scores on the *Motivated* measure). However, my main purpose in including the measures is to test if participants are on average invested in the task and understand it enough to care about how they are evaluated. As noted in Chapter 3, this is an important feature in my setting of interest: employees care about their work and about how they are evaluated. I expand on how I use these measures in Chapter 5.

Tolerance; Hartmann and Slapnicar, 2012); "If I want to find out how well I have done something, I compare what I have done with how others have done" and "I always like to know what others in a similar situation would do" (*Social Comparison Orientation*; Buunk & Gibbons, 2007). I use these trait measures as they may impact individuals' responses to my manipulations. Specifically, individuals that are low in tolerance for ambiguity (i.e., those who need unambiguous information in ambiguous settings) may feel demotivated by the uncertainty that comes with greater subjectivity in performance evaluation (Hartmann and Slapnicar, 2012). Similarly, individuals who are low in social comparison orientation may not be motivated to socially compare even when they identify strongly with their group members (Buunk & Gibbons, 2007). Therefore, measuring these traits allows me to control for these differences if needed to increase the power of my tests.

4.8 Pilot Testing of Independent Variable Manipulations

4.8.1 Overview of Pilot Studies

I anticipated potential design challenges in my main experiment, related to the manipulations of my two independent variables. First, my group identity vignette manipulation had not been tested in prior research with my proposed group of participants. Although I follow Estep (2021) in using a vignette that participants read separately on their computers, Estep (2021) uses senior audit participants that have professional experience working in teams and are likely to relate to the team vignette described in her experiment. Therefore, her participants are likely to have experienced emotions from circumstances like that in the vignette, and in turn, this could allow the vignette manipulation to be more powerful with that participant group. In contrast, my participants are MTurk workers who may not relate to the vignette manipulation to the same extent that the participants in Estep (2021) do. Besides Estep (2021), prior research that has manipulated group identity has generally done so in a lab with students (e.g., Kelly & Presslee, 2021; Towry,

2003), or has measured it in a field study (e.g., Shang et al., 2020; Van der Vegt & Bunderson, 2005). Therefore, I pilot tested my group identity manipulation to ensure that it could be done effectively online using vignettes. Note that, in contrast to prior studies where the group identity manipulation allows participants to interact with each other, my manipulation, which does not allow participant interaction, biases against my finding results as it potentially makes the group identity manipulation weaker. My design choice thus results in a more conservative test of my theory.

My second potential design challenge related to my manipulation of subjectivity in performance evaluation. My manipulation is somewhat based on Gorenflo and Crano (1989); however, the main difference is that I manipulate how participants perceive how they will be evaluated, rather than manipulating how they should make an evaluation, as is done in Gorenflo and Crano (1989). My manipulation had not been tested previously. Further, because both my *Higher* and *Lower Subjectivity* conditions are told the objective writing criteria, I needed to ensure my vignette was perceived differently between the conditions and proved a strong manipulation.

Thus, I ran one pilot study testing the efficacy of my group identity manipulation, and another pilot testing the efficacy of my subjectivity manipulation. The results of my pilot studies are as follows (all p-values two-tailed).[30]

[30] These pilot studies do not have a writing task because such a task would not help test the efficacy of my manipulations, which is the main purpose of the pilots.

4.8.2 Group Identity Pilot Study

My group identity pilot has 179 participants recruited from Amazon Mechanical Turk.[31,32] I test for differences between the *Stronger* and *Weaker Identity* conditions in my group identity manipulation check measures (the same as those detailed in Section 4.5.1 above). I average my three identity variables of *Happy*, *Belong*, and *Like* and obtain a single measure for *Group_Identity_Manipulation_Check.* Principal components analysis (untabulated) indicates that these variables represent one factor with an eigenvalue of 2.88, which together explain 96% of the variance and have a Cronbach alpha of 0.98, indicating high reliability (all factor loadings equal 0.98; Pituch & Stevens, 2016; Kline, 2005). I additionally use the *IOS Overlap* measure. I also test for differences in two social comparison questions (both measured on a 7-point scale from (1) 'not at all' to (7) 'a great extent', with an unlabeled midpoint): *Think About* (the extent participants would think about their performance on team-related work compared to their team members' performance) and *Concerned About* (the extent participants would be concerned about their performance compared to their team members' performance on team-related work; adapted from Tafkov, 2013). I average these two measures to obtain a single measure for *Social Comparison.*[33] As with the original component variables, *Group_Identity_Manipulation_Check, IOS Overlap* and *Social Comparison* are anchored at 1 and 7.

---

[31] I recruited 181 participants, two of whom attempted to take the experiment twice and so were removed, leaving 179 responses. One participant answered the attention check question wrong, but the participant emailed after completing the study to say she meant to answer the question differently, which would have been the correct answer to the attention check question. A closer look at her answer to the open response question showed that she was in fact paying attention, and so she was retained.

[32] I tested three different levels of my group identity conditions, *Stronger Identity*, *Moderate Identity*, and *Weaker Identity*. However, I only use two conditions in my main experiment: *Stronger Identity* and *Weaker Identity* (n = 179 together). This is because my social comparison measures differed significantly between these two conditions in the pilot, but not between the *Stronger Identity* and *Moderate Identity* conditions. Here, I only report results related to the *Stronger Identity* and *Weaker Identity* conditions.

[33] Principal components analysis (untabulated) indicates that these variables represent one factor with an eigenvalue of 1.73, which together explain 86% of the variance and have a Cronbach alpha of 0.84, which indicates high reliability (Pituch & Stevens, 2016; Kline, 2005).

As shown in Table 1 (Panels A and B), I find that the measure for *Group_Identity_Manipulation_Check* is significantly higher in the *Stronger Identity* condition (M = 6.56) than the *Weaker Identity* condition (M = 2.39, t = 26.49, $p < 0.001$). Further, *IOS Overlap* is greater in the *Stronger Identity* condition (M = 5.72) than the *Weaker Identity* condition (M = 2.11, t = 18.90, $p < 0.001$). I also find that *Social Comparison* differs significantly between the *Stronger Identity* (M = 5.05) and *Weaker Identity* (M = 4.42, t = 2.77, p = 0.006) conditions. Finally, because group identity should impact the desire to socially compare, I test whether there is an indirect effect of my group identity manipulation on *Social Comparison,* through the *Group_Identity_Manipulation_Check* measure. I run a PROCESS model (Model 4) in SPSS, a path analysis modelling tool used to estimate direct and indirect effects in mediation models (Hayes, 2017). I use my group identity manipulation as my independent variable, the *Social Comparison* measure as my dependent variable, and the *Group_Identity_Manipulation_Check* measure as my mediator (Hayes & Preecher, 2014). The model outputs the indirect effect of the *Weaker Identity* condition relative to the *Stronger Identity* condition. As shown in Figure 4, the model indicates that *Group_Identity_Manipulation_Check* mediates the relationship between the group identity manipulation and *Social Comparison.* My group identity manipulation significantly influences the *Group_Identity_Manipulation_Check* (a = -2.08), and the *Group_Identity_Manipulation_Check* significantly influences *Social Comparison* (b = 0.262). A bootstrap confidence interval for the indirect effect (ab = -0.545) based on 5,000 bootstrap samples was all below zero, 95% CI = [-0.9386, -0.1341]).

4.8.3 Subjectivity Pilot Study

My subjectivity pilot study has 174 participants from Amazon Mechanical Turk.[34] My dependent variables are subjectivity manipulation checks: a) whether participants feel like their evaluation is subjective (*Subjective*), and b) whether they feel like their evaluation is objective (*Objective*). I also ask three additional questions that I expect to differ between the *Higher* and *Lower* subjectivity conditions: a) whether they feel certain about how they can get a good evaluation (*Certain*), b) whether they feel like their writing will be evaluated better if their evaluator knows they spent more time on it (*Effort Heuristic*), and c) whether they feel like they can control how they are evaluated (*Control*). All dependent variables are on 7-point Likert scales ranging from *Strongly Disagree* (1) to *Strongly Agree* (7), with all points labelled.

As shown in Table 2, I find that, as anticipated, the *Subjective* measure is significantly higher in the *Higher Subjectivity* condition (M = 5.92) compared to the *Lower Subjectivity* condition (M = 3.01, t = 10.76, $p < 0.001$). In contrast, the *Objective* measure is significantly lower in the *Higher Subjectivity* condition (M = 3.40) compared to the *Lower Subjectivity* condition (M = 5.99, t = -10.13, $p < 0.001$). Furthermore, as expected, *Certain* is significantly lower in the *Higher Subjectivity* condition (M = 4.49) compared to the *Lower Subjectivity* condition (M = 5.76, t = -6.40, $p < 0.001$). As expected, the *Effort Heuristic* measure is marginally higher in the *Higher Subjectivity* condition (M = 4.59) compared to the *Lower Subjectivity* condition (M = 4.12, t = 1.92, p = 0.056). Finally, as expected, *Control* is significantly lower in the *Higher Subjectivity* condition (M = 3.94) compared to the *Lower Subjectivity* condition (M = 5.08, t = -5.21, $p < 0.001$).

[34] I recruited 188 participants, 11 of whom attempted the experiment twice and so were removed, leaving 177 responses used. Three participants answered the attention check question incorrectly and so were also removed, leaving a final sample of 174 usable responses.

Thus, my group identity and subjectivity pilot studies support the efficacy of my group identity and subjectivity manipulations.

4.9 Conclusion

I employ my main experiment, designed as a 2 x 2 between-subjects online experiment, to test Hypotheses 1, 2, and 3, and to explore Research Question 1. I manipulate the strength of group identity as either *Stronger* or *Weaker*, and I manipulate subjectivity in performance evaluation as either *Higher* or *Lower*. My dependent variable is how long participants spend on completing a writing task beyond how long I contract them to spend on the writing task. A summary of the experimental procedures for my main experiment is presented in Appendix A. The experimental instrument for my main experiment is presented in Appendix C. The next chapter provides the results.

CHAPTER 5: MAIN EXPERIMENT RESULTS

5.1 Introduction

This chapter reports the results of my main experiment. Section 5.2.1 reports demographic information about my participants, section 5.2.2 reports the results of my manipulation checks, and section 5.2.3 reports the tests of hypothesis 1, 2 and 3, as well as RQ1. Section 5.2.4 reports the supplemental analysis related to process. Section 5.2.5 reports the other measured variables in the post-experiment questionnaire. Section 5.2.6 discusses my results including my excluded participants. I summarize in Section 5.3.

5.2 Results

5.2.1 Participants and Demographics

My final sample consists of 290 MTurkers. I start by collecting 404 participant responses. I remove one response where the participant spent more than the maximum time of one hour that I allowed on the writing task: the respondent spent more than 2.2 hours on the study and 1.35 hours on the writing task, indicating inattentiveness. I also remove two respondents who had a reCAPTCHA score lower than 0.5, as this indicates bot responses, per Qualtrics guidelines.[35] I also remove 100 respondents who indicated inattentiveness on my three attention check questions. In the experiment, if participants responded incorrectly to these questions, they were asked to try again. Thus, in my final sample, I retain participants who answered these questions correctly the first time, or who may have erred once. This is indicated by participants who only clicked three or

[35] The reCAPTCHA score is based on a technology that detects whether a response is likely to be a bot or a human and is automatically calculated by Qualtrics.

four times on the page with all three questions, a metric automatically captured by Qualtrics. This results in 301 valid responses. This means that 26% of participants were removed due to inattention, which is in line with previous research that removes MTurk participants that indicate inattention (Bentley, 2021).

Finally, I remove 11 outlier respondents across all conditions who are more than 2.5 standard deviations away from the mean of my dependent variable – *DV_Overwork* – bringing my final sample to 290 participants.[36] Removing outlier respondents is often a necessary step in MTurk studies because of the limited experimental control over how participants complete the study (e.g., Haesebrouck, 2021; Lauck et al., 2020; Buchheit et al., 2018; Tan & Yu, 2018; Panero et al., 2016). This is especially necessary considering the open-ended nature of my dependent variable (i.e., there is no upper limit on how long participants can work on the writing task), making it easily susceptible to noise from extreme outliers (Litman et al., 2017). A cut-off value of 2.5 standard deviations is an oft-used standard for outliers in studies, such as reaction time studies (e.g., Rule et al., 2009; Van Selst & Jolicoeur, 1994; Ratcliff, 1993). Section 5.2.6 outlines my main results if I include my excluded participants.

My participants are paid $2.67 for completing the study, and they spend an average of 15.32 minutes in the study.[37] See Table 3 for demographics about these participants. On average, participants are 41.54 years old, and 67.25% of participants completed college or university (including those who completed or did some graduate school). All but one participant notes that English is the language they feel most confident writing in, and 62.76% of participants indicate

[36] I find that *DV_Overwork* is normally distributed per Kline (2005), i.e., |skewness| < 3 at 1.56 and |kurtosis| < 10 at 1.79.

[37] This is a reasonable pay for MTurk workers as it amounts to $10.46/hr, more than the United States federal minimum wage of $7.25 (U.S. Department of Labor, 2023). It is also in line with, or more than, accounting studies pay for Amazon Mechanical Turk workers (Buchheit et al., 2018).

that the highest education level for which they have written essays in English is for undergraduate classes. Random assignment appears to have been successful since there are no significant differences in age or education between my conditions (untabulated).[38]

5.2.2 Manipulation Checks

I first examine whether my subjectivity in performance evaluation manipulation check differs as expected by condition. I ask participants whether they feel like their evaluation is subjective, based on the evaluator's personal views and opinions on each section of the report (*Subjective*). Results are reported in Table 4 (Panels A and B). As expected, *Subjective* is significantly higher in the *Higher Subjectivity* condition (M = 6.53) compared to the *Lower Subjectivity* condition (M = 2.32, t = 23.74, $p < 0.001$). Thus, I effectively manipulated subjectivity in performance evaluation.

I next test the effectiveness of my group identity manipulation. I average my identity manipulation check variables of *Happy* ('To what extent would you be happy to be a part of this team?'), *Belong* ('To what extent would you feel like a member of your team?'), and *Like* ('To what extent would you like your team members?'), and I obtain a single measure for *Group_Identity_Manipulation_Check.* As shown in Table 5, principal components analysis indicates that these variables represent one factor with an eigenvalue of 2.91, which together explain 97% of the variance and have a Cronbach alpha of 0.98, indicating high reliability (Pituch & Stevens, 2016; Kline, 2005). As with the original component variables, *Group_Identity_Manipulation_Check* is anchored at 1 and 7. As shown in Table 6 (Panels A and B), *Group_Identity_Manipulation_Check* is significantly higher in the *Stronger Identity* condition

[38] Age is correlated with my dependent variable (*DV_Overwork;* untabulated). Including age as a covariate in my main tests does not make a qualitative difference to my results, but only makes them more significant. For expositional ease and a more conservative reporting of the results, I do not include age in my analysis.

(*M* = 6.59) than the *Weaker Identity* condition (*M* = 2.14) (*t* = 39.72, *p* < 0.001). Furthermore, the *IOS Overlap* measure is significantly higher in the *Stronger Identity* condition (*M* = 5.76) than the *Weaker Identity* condition (*M* = 1.80, *t* = 30.35, *p* < 0.001). Thus, I effectively manipulated group identity.

5.2.3 Tests of Hypotheses and Research Question

H1 predicts a positive effect of subjectivity in performance evaluation (*Subjectivity*) on *DV_Overwork*, while H2 predicts an ordinal interaction of *Subjectivity* and *Group Identity*, such that *Subjectivity* will have a stronger effect on *DV_Overwork* in the *Stronger* (versus *Weaker*) *Identity* condition. H3 predicts a positive effect of *Group Identity* on *DV_Overwork*. I also test Research Question 1 (RQ1), which examines the interactive effect of *Subjectivity* and *Group Identity* on participants' performance in the writing task (*Grammarly_Performance* and *Rater_Performance*). Table 7 presents the descriptive statistics for *DV_Overwork* (Panels A and B), *Grammarly_Performance* (Panels A and C), and *Rater_Performance* (Panels A and D), across conditions.

I first report the descriptive statistics for *DV_Overwork* as shown in Table 7, Panel B. When *Group Identity* is *Stronger*, the mean level of *DV_Overwork* for the *Higher Subjectivity* condition (*M* = 137.05, *SD* = 145.27) is higher than the mean for the *Lower Subjectivity* condition (*M* = 96.91, *SD* = 128.04). When *Group Identity* is *Weaker*, the mean level of *DV_Overwork* for the *Higher Subjectivity* condition (*M* = 89.08, *SD* = 109.55) is higher than the mean for the *Lower Subjectivity* condition (*M* = 85.83, *SD* = 97.85).

To formally test my hypotheses, I start by running an ANOVA as a general overview of the results, with *Subjectivity* and *Group Identity* as my independent factors, and *DV_Overwork* as my dependent factor. I follow my ANOVA with simple effects analysis. As presented in Table 8,

Panel A, I find that *Subjectivity* has a marginally significant positive effect on *DV_Overwork* (F = 2.34, p = 0.064, one-tailed). I also find that *Group Identity* has a significant positive effect on *DV_Overwork* (F = 4.34, p = 0.019, one-tailed). Finally, my main effects are qualified by the marginally significant interaction of *Subjectivity* and *Group Identity* on *DV_Overwork* (F = 1.69, p = 0.097, one-tailed), which provides initial support for H2. Because my ANOVA shows a marginally significant interaction, I follow up with simple effects to further examine H1 and H3.

I find that *Subjectivity* has a positive effect on *DV_Overwork* in the *Stronger Identity* condition (Table 8, Panel B: F = 3.00, p = 0.043, one-tailed) but not in the *Weaker Identity* condition (F = 0.04, p = 0.424, one-tailed). Thus, H1 is supported in the *Stronger Identity* condition, but not in the *Weaker Identity* condition. This also provides initial support for H2, which I expand on below. An analysis of simple effects also shows that *Group Identity* has a significant positive effect on *DV_Overwork* in the *Higher Subjectivity* condition (Table 8, Panel C: F = 4.85, p = 0.015, one-tailed), but not in the *Lower Subjectivity* condition (F = 0.36, p = 0.274, one-tailed). Thus, H3 is supported in the *Higher Subjectivity* condition, but not in the *Lower Subjectivity* condition.

I perform further analyses to test H2. Whereas H2 predicts an ordinal interaction, an ANOVA assigns contrast weights that assume a disordinal interaction, and thus it does not provide a statistically powerful test for my theory (Buckless & Ravenscroft, 1990); as indicated above, it provides only marginally statistically significant support for H2. Therefore, following previous research, I use a planned contrast to test the predicted pattern of results in H2, as a planned contrast is the most appropriate and statistically powerful test of my hypotheses (e.g., Lambert & Agoglia, 2011; Hirst et al., 2007; Kadous et al., 2003; Rosnow and Rosenthal, 1995). With *DV_Overwork* as my dependent variable, I use contrast weights of +3 [*Higher Subjectivity/ Stronger Identity*], -1

[*Higher Subjectivity/ Weaker Identity*], -1 [*Lower Subjectivity/ Stronger Identity*] and -1 *[Lower Subjectivity/ Weaker Identity*]. These weights test an ordinal interaction, such that the *Higher Subjectivity/ Stronger Identity* condition has the highest weight compared to the rest of the conditions.

I follow the three-part test suggested by Guggenmos et al. (2018) when analyzing my planned contrast. First, as outlined in Table 8, Panel D, the planned contrast is significant (F = 7.57, p = 0.003, one-tailed), and the between-cells residual variance not captured by the planned contrast is insignificant (p = 0.828; Buckless & Ravenscroft, 1990). Second, the evaluation of the relative contrast variance residual, q^2, is 0.048, indicating that only 4.8% of the systematic variance is not explained by the contrast (Guggenmos et al., 2018). Guggenmos et al. (2018) do not indicate a specific cut-off value for q^2 but suggest it as an indicator to be combined with the significance test and a visual inspection of the results. Thus, I examine the visual evaluation of fit of the results (Figure 5) and the predicted pattern (see Figure 3). The pattern of results approximates that in the predicted figure, supporting the ordinal interaction in H2. Thus, using the three-part test by Guggenmos et al. (2018), H2 is supported.

For robustness, I also test H2 using a more specified contrast that incorporates my predicted main effects – the main effect of *Subjectivity* and the main effect of *Group Identity* – within my predicted ordinal interaction effect. Thus, with *DV_Overwork* as my dependent variable, I use contrast weights of +4 [*Higher Subjectivity/ Stronger Identity*], -2 [*Higher Subjectivity/ Weaker Identity*], +1 [*Lower Subjectivity/ Stronger Identity*] and -3 *[Lower Subjectivity/ Weaker Identity*]. This planned contrast, compared to my main ordinal interaction contrast of {+3, -1, -1, -1}, has the cost of biasing the main effect of group identity upward and so is less conservative than the main ordinal interaction contrast. First, as outlined in Table 8, Panel E, the planned contrast is

significant (F = 7.10, p = 0.004, one-tailed), and the between-cells residual variance not captured by the planned contrast is insignificant (p = 0.655; Buckless & Ravenscroft, 1990). Second, the evaluation of the relative contrast variance residual, q^2, is 0.107, indicating that only 10.7% of the systematic variance is not explained by the contrast (Guggenmos et al., 2018). The pattern of results approximates that in the predicted figure, supporting the ordinal interaction in H2. Thus, using the alternative planned contrast as a robustness test, H2 is supported.

Therefore, overall, using the less powerful disordinal ANOVA contrast test for H2, I find marginally significant support that *Subjectivity* and *Group Identity* interact to influence *DV_Overwork*. Further, using the more appropriate but less conservative planned ordinal contrast test (Buckless & Ravenscroft, 1990), I find significant support for H2. Finally, for robustness, using an even less conservative ordinal contrast that incorporates both my predicted main effects of *Subjectivity* and *Group Identity* on *DV_Overwork* and the interaction effect, I find similar support for H2. Thus, results support that the effect of subjectivity in performance evaluation on the level of overwork is increasing in group identity strength.

Finally, I test RQ1, which examines whether there is an interaction effect of *Subjectivity* and *Group Identity* on participants' performance on the writing task, first as scored by the Grammarly writing software (*Grammarly_Performance*), then as scored by an independent rater (*Rater_Performance*).

I first describe the descriptive statistics for *Grammarly_Performance* as shown in Table 7, Panel C. When *Group Identity* is *Stronger*, the mean *Grammarly_Performance* for the *Higher Subjectivity* condition (M = 74.26, SD = 14.78) is lower than the mean for the *Lower Subjectivity* condition (M = 74.67, SD = 13.82). When *Group Identity* is *Weaker*, the mean

Grammarly_Performance for the *Higher Subjectivity* condition (*M* = 73.47, *SD* = 15.62) is lower than the mean for the *Lower Subjectivity* condition (*M* = 78.40, *SD* = 14.89).

Next, I run an ANOVA as a general overview of the results, with *Subjectivity* and *Group Identity* as my independent factors, and *Grammarly_Performance* as my dependent factor. As presented in Table 9, Panel A, the ANOVA revealed no significant interaction between *Subjectivity* and *Group Identity* (*F* = 1.67, *p* = 0.197, two-tailed). There is also no main effect of either *Subjectivity* or *Group Identity* on *Grammarly_Performance* (all $p > 0.120$).

I next report the descriptive statistics for *Rater_Performance* as shown in Table 7, Panel D. When *Group Identity* is *Stronger*, the mean *Rater_Performance* for the *Higher Subjectivity* condition (*M* = 3.67, *SD* = 1.30) is higher than the mean for the *Lower Subjectivity* condition (*M* = 3.53, *SD* = 1.13). When *Group Identity* is *Weaker*, the mean *Rater_Performance* for the *Higher Subjectivity* condition (*M* = 3.72, *SD* = 1.28) is lower than the mean for the *Lower Subjectivity* condition (*M* = 3.73, *SD* = 1.28).

I then run an ANOVA with *Rater_Performance* as my dependent factor. As is presented in Table 9, Panel B, the ANOVA revealed no significant interaction between *Subjectivity* and *Group Identity* (*F* = 0.24, *p* = 0.624, two-tailed). There is also no main effect of either *Subjectivity* or *Group Identity* on *Rater_Performance* (all $p > 0.300$).

Thus, in answer to RQ1, there appears to be no main or interactive effects of *Subjectivity* and *Group Identity* on participants' performance, with performance either scored by Grammarly or by an independent rater.

Interestingly, as is apparent in Figure 6, Panel A, the pattern of results for *Grammarly_Performance* mirror those for *DV_Overwork*, but in reverse. Thus, I test the relationship between *DV_Overwork* and *Grammarly_Performance*; I find that they are

significantly negatively correlated ($r(284) = -0.12$, $p = 0.050$; Table 9, Panel C). There are two possible explanations for this result. The first is that participants who have a higher level of overwork prioritize overworking over real performance outcomes, such that their goal is to increase their level of overwork at the expense of improving their writing quality. The second is that participants who overwork are those that are worse at writing and so need more time to create what they perceive to be 'good enough' writing quality.

For completeness, I also test the relationship between *DV_Overwork* and *Rater_Performance*. I find that they are not significantly correlated ($p = 0.754$; Table 9, Panel C).

5.2.4 Supplemental analyses: Process evidence

I examine participants' post-experimental questionnaire (PEQ) responses in my main experiment to further examine the mechanisms underlying my results. Specifically, I first examine my effort heuristic measure (*Effort Heuristic*): "I believe I will receive a higher evaluation if the evaluator knows I spent more time on my section of the report". Although it does not differ by subjectivity condition ($p = 0.640$; Table 10, Panel A), I also examine if this measure captures individual differences in the use of the effort heuristic, as research finds that individuals exhibit differences in their use of heuristics (Jackson et al., 2016).[39]

If the effort heuristic is a main process by which subjectivity impacts the level of overwork, then I expect individuals who have a higher (lower) tendency to use the effort heuristic to (not) show variation in overwork as subjectivity and group identity vary, such that support for H1 and H2 is increasing in the effort heuristic measure. Thus, as presented in Table 10 (Panels B and C), I perform a median split on *Effort Heuristic*, such that the *High Effort Heuristic* and the *Low Effort*

[39] As previously discussed, my measure of effort heuristic differed by subjectivity condition in my manipulation pilot tests, but a limitation of my measure is that it has not been previously validated in the literature (Asay et al., 2022).

Heuristic groups have a mean (standard deviation) of 4.77 (0.91) and 2.28 (0.70), respectively (Panel B). The means of the two groups are statistically different from each other (t = 26.08, $p < 0.001$, Table 10, Panel C).

I next run the same ANOVA I ran for my main analyses for both the *High Effort Heuristic* and the *Low Effort Heuristic* groups, with *Subjectivity* and *Group Identity* as my independent factors, and *DV_Overwork* as my dependent factor (Table 10, Panel E). For *High Effort Heuristic* participants, I find that *Subjectivity* has a marginally significant effect on *DV_Overwork* ($F = 2.18$, $p = 0.071$, one-tailed), partially supporting H1. The interaction term in the ANOVA indicates no significant interaction between *Subjectivity* and *Group Identity* ($p = 0.355$). However, using the planned contrast of {+3, -1, -1, -1} to test H2, I find support for the predicted pattern ($F = 6.86$, $p = 0.005$, one-tailed) and the between-cells residual variance not captured by the planned contrast is insignificant ($p = 0.243$; Table 10, Panel F). Further, the evaluation of the relative contrast variance residual, q^2, is 0.29, indicating that only 29% of the systematic variance is not explained by the contrast (Guggenmos et al., 2018). Finally, I examine the visual evaluation of fit of the results and the predicted pattern (see Figure 7, Panel A). The pattern of results reasonably approximates that in the predicted figure, providing some additional support for the ordinal interaction in H2. I also report simple effects of *Subjectivity* on *DV_Overwork*, within the *Stronger* and *Weaker Identity* conditions (Table 10, Panel G). I find that *Subjectivity* does not have a significant effect on *DV_Overwork* within either the *Stronger* or *Weaker Identity* conditions ($p > 0.280$). Thus, support for H2 is mixed but largely consistent with expectations for the *High Effort Heuristic* participants.

For *Low Effort Heuristic* participants, I find that *Subjectivity* has no main effect on *DV_Overwork* ($p = 0.526$) and the ANOVA shows no significant interaction ($p = 0.175$); the

planned contrast results also do not support the predicted pattern (F = 1.49, p = 0.224, two-tailed; Table 10, Panels E and F). Thus, consistent with theory underlying my hypotheses, I find stronger support for H1 and H2 for those participants who are naturally higher (versus lower) in their use of the effort heuristic.

I also ask two trait social comparison questions in the PEQ: "If I want to find out how well I have done something, I compare what I have done with how others have done" and "I always like to know what others in a similar situation would do" (Gibbons & Buunk, 1999). If social comparison is the mechanism by which group identity moderates the relationship between subjectivity and the level of overwork, then I expect strong support for the predicted H2 relationship for individuals who are high in trait social comparison, but not those who are low in trait social comparison. Thus, I create an *Average Trait Social Comparison* measure consisting of both trait social comparison questions. Principal components analysis (untabulated) indicates that these variables represent one factor with an eigenvalue of 1.63, which together explain 82% of the variance and have a Cronbach alpha of 0.78, indicating adequate reliability (Pituch & Stevens, 2016; Kline, 2005). *Average Trait Social Comparison* does not differ by group identity condition (p = 0.257; Table 11, Panel A). I then median split this measure, such that the *High Trait Social Comparison* and the *Low Trait Social Comparison* groups have a mean (standard deviation) of 5.50 (0.59) and 3.34 (1.01), respectively (Table 11, Panel B). The means of the two groups are statistically different from each other (Table 11, Panel C: t = 22.44, p < 0.001).

I next run the same ANOVA I ran for my main analyses for both the *High Trait Social Comparison* and the *Low Trait Social Comparison* groups, with *Subjectivity* and *Group Identity* as my independent factors, and *DV_Overwork* as my dependent factor. I find that there is a statistically insignificant interaction between *Subjectivity* and *Group Identity* (p = 0.603; Table 11,

Panel E). I follow this up with the same planned contrast that I ran for my main test of H2 {+3, -1, -1, -1}. I find that for the *High Trait Social Comparison* participants, planned contrast results support the predicted pattern (Table 11, Panel F: $F = 4.10$, $p = 0.023$, one-tailed) and the between-cells residual variance not captured by the planned contrast is insignificant ($p = 0.663$). Further, the evaluation of the relative contrast variance residual, q^2, is 0.17, indicating that only 17% of the systematic variance is not explained by the contrast (Guggenmos et al., 2018). I also examine the visual evaluation of fit of the results and the predicted pattern (see Figure 8, Panel A). The pattern of results approximates that in the predicted figure, supporting the ordinal interaction in H2. I also report simple effects of *Subjectivity* on *DV_Overwork*, within the *Stronger* and *Weaker Identity* conditions (Table 11, Panel G). I find that *Subjectivity* does not have a significant effect on *DV_Overwork* within either the *Stronger* or *Weaker Identity* conditions ($p > 0.239$). Thus, support for H2 is mixed but largely consistent with expectations for the *High Trait Social Comparison* participants.

Next, for the *Low Trait Social Comparison* participants, the ANOVA shows a statistically insignificant interaction between *Subjectivity* and *Group Identity* ($p = 0.223$; Table 11, Panel E). I follow this up with the same planned contrast that I ran for my main test of H2 {+3, -1, -1, -1}; results show only marginal support for the predicted pattern (Table 11, Panel F: $F = 3.16$, $p = 0.077$, two-tailed). The between-cells residual variance not captured by the planned contrast is insignificant ($p = 0.954$), and the evaluation of the relative contrast variance residual, q^2, is 0.03, indicating that only 3% of the systematic variance is not explained by the contrast (Guggenmos et al., 2018). I also examine the visual evaluation of fit of the results and the predicted pattern (see Figure 8, Panel B). The pattern of results approximates that in the predicted figure, supporting the ordinal interaction in H2. However, I find that *Subjectivity* does not have a significant effect on

DV_Overwork (Table 11, Panel G) within either the *Stronger* or *Weaker Identity* conditions ($p > 0.226$). Overall, I find somewhat stronger support for H2 for those participants who are naturally higher (versus lower) in trait social comparison, consistent with theory underlying my hypotheses.

5.2.5 Other Measured Variables

As discussed in section 4.2.7, I ask participants to respond to additional questions in the post-experiment questions. All measures in this section are on a scale from (1) to (7), where (1) is *Strongly Disagree* and (7) is *Strongly Agree*. The effort heuristic and trait social comparison measures are discussed in section 5.2.4. I also ask two questions related to participants' perceptions of their writing task evaluations: "In this writing task, I believe it is difficult for the evaluator to evaluate writing quality" (*Difficult*) and "I feel certain about how I can get a good evaluation on my section of the report" (*Certain*). These measures are intended as process measures and may differ by *Subjectivity* condition according to my theory, such that *Difficult* should be higher in the *Higher Subjectivity* condition, and *Certain* should be higher in the *Lower Subjectivity* condition. As shown in Table 12 (Panels A and B), I find that *Difficult* is indeed significantly higher in the *Higher Subjectivity* condition ($M = 3.72$) than the *Lower Subjectivity* condition ($M = 3.11$, $t = 3.35$, $p < 0.001$). As expected, I also find that *Certain* is significantly higher in the *Lower Subjectivity* condition ($M = 4.74$) than the *Higher Subjectivity* condition ($M = 4.00$, $t = 4.65$, $p < 0.001$).

I also ask a social comparison question: "While I was completing the writing task, I thought about how my performance on the writing task compared to my team members' performance" (*State_Social_Comparison;* adapted from Tafkov, 2013). According to my theory, this should be higher in the *Stronger Identity* condition than the *Weaker Identity* condition. However, I find that this measure does not differ significantly between my two conditions (Table 12, Panel D: $t = 0.23$, $p = 0.815$).

I also ask questions related to participants' perceptions of the task: "I was motivated to do well on this task" (*Motivated*), "It was important for me to be evaluated well on this task" (*Important*), and "I understood the task instructions" (*Understood*). As shown in Table 12, Panels E to F, I find that on average, participants are between (5) *Somewhat Agree* and (7) *Strongly Agree* on all three measures: *Motivated* ($M = 6.04$), *Important* ($M = 5.84$), and *Understood* ($M = 6.49$). I also find that *Motivated* and *Important* differ by *Identity* condition ($p = 0.030$ and $p = 0.095$, respectively). This can be explained by group identity acting as a social control, such that stronger identity leads to increased alignment between participants' objectives and that of the team (Akerlof & Kranton, 2005, see footnote 18). However, these measures do not correlate with my dependent variable, *DV_Overwork* (untabulated). Including these two variables in an ANCOVA, with *Subjectivity* and *Group Identity* as my independent variables, and *DV_Overwork* as my dependent variable does not lead to qualitatively different results from my main ANOVA reported in section 5.2.3 (untabulated). *Understood* does not differ significantly between conditions ($p > 0.600$).

Finally, I use the tolerance for ambiguity trait measure from Hartmann and Slapnicar (2012; *Ambiguity Tolerance*): "I do not like to work on a problem unless there is a possibility of finding a clear-cut and unambiguous solution". I find that this measure does not differ between conditions ($p > 0.700$; Table 12, Panel G and H).

5.2.6 Robustness Tests with Excluded Participants (untabulated)

In this section, I discuss my results for H1, H2, H3 and RQ1 if I include my excluded participants. As noted in section 5.2.1, I collect 404 participant responses. I first remove 26% of participants due to inattention, resulting in 301 valid responses. As expected, retaining these participants in my sample results in insignificant differences between conditions in the tests of H1, H2, and H3, due to noise from the inattentive participants.

I next additionally remove 11 outlier respondents across all conditions who are more than 2.5 standard deviations away from the mean of my dependent variable - *DV_Overwork* - bringing my final sample to 290. If I include my 11 outliers in the analyses, my contrast results in support of H2 remain significant ($p = 0.034$, one-tailed). H3, predicting a positive effect of *Group Identity* on *DV_Overwork*, is also significantly supported ($F = 4.05$, $p = 0.023$, one-tailed). However, H1 is no longer supported: *Subjectivity* does not have a significant positive effect on *DV_Overwork* ($p = 0.264$, two-tailed). Results for RQ1 remain qualitatively similar.

5.3 Summary

This chapter provides the results for Hypotheses 1, 2, and 3, RQ1, as well as supporting process evidence. I find partial support for Hypothesis 1: subjectivity in performance evaluation has a significant positive effect on the level of overwork, but only when group identity is stronger and not when it is weaker. I find support for Hypothesis 2: consistent with my prediction, I find that the positive effect of subjectivity in performance evaluation on the level of overwork is increasing in group identity strength. I find partial support for Hypothesis 3: group identity has a significant positive effect on the level of overwork, but only at higher levels of subjectivity in performance evaluation and not at lower levels.

I also report supportive process evidence in the form of cross-sectional analysis: supporting my process mechanism, I find that participants who are more inclined to use the effort heuristic show stronger support for Hypotheses 1 and 2. Similarly, those who are more inclined to socially compare show somewhat stronger support for Hypothesis 2. I provide further process support in Chapter 6 in the form of a secondary, supplemental experiment.

In terms of RQ1, I find that there is an insignificant interaction effect between subjectivity in performance evaluation and group identity on participants' performance on the writing task.

CHAPTER 6: SECONDARY EXPERIMENT DESIGN AND RESULTS

6.1 Introduction

This chapter reports the design and results of my secondary experiment. Section 6.2 provides an overview of the experiment. Section 6.3 describes the research method. Section 6.3.1 provides details about participant recruitment. Section 6.3.2 describes the experimental procedures and task details. Section 6.4 reports the results of the secondary experiment. Section 6.4.1 reports demographic information about my participants, and section 6.4.2 reports the manipulation checks. Section 6.4.3 reports the tests of process, and section 6.4.4 reports the other measured variables in the post-experiment questionnaire. I summarize in Section 6.5.

6.2 Overview

I employ my secondary experiment to provide further evidence that subjectivity in performance evaluation impacts the level of overwork through the effort heuristic mechanism and through feelings of uncertainty. Although I use a measure in the post-experimental questionnaire of my main experiment to capture the effort heuristic, this has the disadvantage of creating carryover effects from my dependent variable (*DV_Overwork*) to my process measure (Asay et al., 2022). These carryover effects could result in participants hypothesis guessing or attempting to answer the process measures consistently with their work on the writing task or attempting to justify their work on the writing task (Asay et al., 2022). These threats in combination can lead to noise in the measurement of process, reducing the power in testing a statistical mediation model in my main experiment.

Thus, the experimental literature stresses the benefits of using separate experiments to test theoretical process (Asay et al., 2022; Spencer et al., 2005). I follow previous research in running

an experiment to test the direct impact of my independent variable on my theoretical process of interest (e.g., Kelly & Presslee, 2017). Thus, I test the direct impact of subjectivity in performance evaluation on participants' use of the effort heuristic and on their perceptions of whether they feel certain about how they can get a good evaluation, using a 2 x 1 between-subjects online experiment.[40] The design of my secondary experiment largely follows the design of my main experiment but with no manipulation of group identity and with no writing task. Other notable differences in design are discussed in section 6.3.

6.3 Research Method

6.3.1 Participant Recruitment

I recruit participants from Amazon Mechanical Turk (MTurk). Consistent with what I do in my main experiment, in my secondary experiment I screen for CloudResearch approved MTurk workers (i.e., MTurkers who have been vetted by CloudResearch for their evidence of attention), who have an approval rating higher than 95% and who have completed more than 100 Human Intelligence Tasks (HITs) on the platform (Bentley, 2021; Eyal et al., 2021). I use these screens to

[40] I also employed another experiment testing the effect of group identity on my theoretical process of interest for group identity as a moderator, i.e., state social comparison. I manipulated group identity in the same way I manipulated it in my main experiment. However, I find that my state social comparison measures (the same as the state social comparison measures used in my group identity pilot study [see section 4.8.2]) do not differ between my group identity conditions ($p > 0.350$, two-tailed). I also find that the average of my two trait social comparison measures (the same as those used in my main experiment [see section 5.2.4]) unexpectedly differ by group identity condition ($p = 0.095$, two-tailed). More specifically, one of the trait measures ("I always like to know what others in a similar situation would do") differs significantly between conditions ($t = 2.28$, $p = 0.024$, two-tailed). The other trait measure ("If I want to find out how well I have done something, I compare what I have done with how others have done") does not ($p = 0.394$). These trait measures are derived from the literature (Gibbons & Buunk, 1999) and should therefore be reliable measures of individual differences. Therefore, they should not differ by condition if random assignment is successful in the experiment. This leads me to believe random assignment was not achieved, and I do not report further results related to this experiment. It is worth noting that in my group identity pilot study, state social comparison differed as expected by group identity condition (see section 4.8.2).

reduce inattentive participants in my study. Only participants in the United States can participate. I do not allow participants from my main experiment to participate in my secondary experiment.

6.3.2 Experimental Procedures and Task Details

To begin, participants consent to participate in the study and answer a reCAPTCHA question to ensure they are not bots. Participants are shown vignettes of scenarios that they are told to assume that they are a part of. These vignettes are the same as those used in my main experiment, where participants are told that they and their team members will be working on a report for a client and that each team member will be working on a section of the report, and each section will be evaluated separately. Participants then receive either the *Higher* or *Lower Subjectivity* manipulation, which is the same manipulation as in my main experiment. They then proceed to answer the process measure dependent variables (see below). As noted above, unlike my main experiment, they do not complete a writing task. They are given a comprehension check question to ensure their understanding of the task. Finally, participants complete the post-experimental questionnaire that comprises demographic questions and other measured variables (described in section 6.4.4).

My dependent variables are two measures of the effort heuristic ("I believe the evaluator will evaluate my writing better if they know I spent more time on it"; "I believe the evaluator will evaluate my writing better if they know I put more effort into it"). Compared to my main experiment, I added an additional effort heuristic measure to my secondary experiment, to allow for more points of discrimination in the measurement of the effort heuristic and to allow for a measure of reliability (Cronbach's alpha). Compared to my main experiment, I also changed the first effort heuristic item to be more consistent with my measure in the pilot study (see section 4.8.3). I also have a measure of certainty as my other dependent variable (the same measure used

in my main experiment). All three measures use 7-point Likert scales from *Strongly Disagree (1)* to *Strongly Agree (7)* with all points labelled.

6.4 Secondary Experiment Results

6.4.1 Participants and Demographics

My final sample consists of 181 participants.[41] Participants are paid $1.60 for completing the study, and they spend an average of 4.68 minutes in the study.[42] As shown in Table 13, on average participants are 39.89 years old and 67.96% of participants completed college or university (including those who completed or did some graduate school).[43] Of all participants, 100% indicate that English is the language in which they feel most confident writing. These participant demographics are comparable to those in my main experiment. Further, 92.82% of participants indicate that they have previous work experience working as part of a team. Random assignment appears to have been successful since there are no differences in age or education between my conditions ($p > 0.170$; untabulated). Age and education are not correlated with my dependent variables ($p > 0.200$; untabulated; see details on dependent variables below).

6.4.2 Manipulation Checks

To test the effectiveness of my subjectivity manipulation, I examine whether my subjectivity manipulation checks differ as expected by condition. As I do in my main experiment,

[41] I recruit 190 participants and remove 9 participants who answer the comprehension check question incorrectly. Retaining these participants does not make a qualitative difference to the results reported below. Unlike in my main experiment, I do not have participants who had a reCAPTCHA score lower than 0.5, so none are removed from my final sample for this reason in my secondary experiment. Unlike my main experiment, I also do not remove respondents based on time, since my dependent variable in my secondary experiment is not open-ended or time-based.
[42] This is a reasonable pay for MTurk workers as it amounts to $20.51 per hour, more than the United States federal minimum wage of $7.25 (U.S. Department of Labor, 2023). It is also in line with, or more than, accounting studies pay for Amazon Mechanical Turk workers (Buchheit et al., 2018).
[43] One participant typed an age of 2000 as their response and was removed from the age analysis.

I employ the measure *Subjective*. I additionally ask participants whether they feel their section of the report will be evaluated objectively, based on objective writing criteria such as the number of spelling, grammar, and punctuation mistakes in the section of the report being evaluated (*Objective*). Results are reported in Table 14. As expected, the *Subjective* measure is significantly higher in the *Higher Subjectivity* condition (M = 5.76) compared to the *Lower Subjectivity* condition (M = 2.41, t = 14.13, $p < 0.001$). Also as expected, *Objective* is significantly lower in the *Higher Subjectivity* condition (M = 3.91) compared to the *Lower Subjectivity* condition (M = 6.18, t = -10.23, $p < 0.001$). Thus, I effectively manipulated subjectivity in performance evaluation.

6.4.3 Tests of Process

As a reminder, the purpose of my secondary experiment is to provide additional process evidence that subjectivity in performance evaluation impacts employees' level of overwork through the effort heuristic and through feelings of uncertainty. Thus, I manipulate subjectivity in performance evaluation, and my dependent variables are my two measures of the effort heuristic and my measure of certainty (*Certainty*; see section 6.3.2). As shown in Table 15, principal components analysis indicates that my two effort heuristic variables represent one factor with an eigenvalue of 1.77, which together explain 88% of the variance and have a Cronbach alpha of 0.87, indicating high reliability (Pituch & Stevens, 2016; Kline, 2005). I average my two effort heuristic variables and obtain a single measure for *Average_Effort_Heuristic*.

As shown in Table 16 (Panels A and B), I find, as expected, that the *Average_Effort_Heuristic* measure is significantly higher in the *Higher Subjectivity* condition (M = 4.90) compared to the *Lower Subjectivity* condition (M = 3.99; t = 4.16, $p < 0.001$). I also find that the *Certainty* measure is significantly lower in the *Higher Subjectivity* condition (M = 4.56) compared to the *Lower Subjectivity* condition (M = 5.85; t = -6.93, $p < 0.001$). This evidence

supports my theory that subjectivity in performance evaluation leads individuals to feel less certain about what they need to do to be evaluated well, and that it increases their tendency to use the effort heuristic.

6.4.4 Other Measured Variables

Prior psychology research indicates that individual trait differences influence the use of heuristics such as the effort heuristic (Cheng et al., 2017; Jackson et al., 2016). Thus, I ask participants to answer two trait measures capturing the Protestant Work Ethic to ensure that differences in the use of the effort heuristic between conditions is not due to individual trait differences (PWE; Cheng et al., 2017; see section 3.4.2). Both measures use 7-point Likert scales from *Strongly Disagree* (1) to *Strongly Agree* (7) with all points labelled and are as follows: "Any person who is able and willing to work hard has a good chance of succeeding" (*Work_Hard*), and "Life would be more meaningful if we had more leisure time" (*Leisure_Time*, reverse coded). As shown in Table 17 (Panels A and B), I find that both variables do not differ between *Subjectivity* conditions ($p > .100$, two-tailed). Thus, I conduct no follow-up analyses with these variables.

Finally, I ask participants about their perception of how much they can control their evaluation on their section of the report ("I feel like I can control how I am evaluated on my section of the report"; *Control*). Theoretically, participants should feel more control over their evaluation in the *Lower Subjectivity* compared to the *Higher Subjectivity* condition, because there is less uncertainty associated with the *Lower Subjectivity* condition. I find support for this, as shown in Table 17 (Panels A and B): I find that the *Control* measure is significantly higher in the *Lower Subjectivity* ($M = 5.25$) than the *Higher Subjectivity* condition ($M = 4.29$, $t = 4.44$, $p < 0.001$).

6.5 Summary

This chapter reports the results of my secondary experiment, where I collect additional process evidence underlying the relationship between subjectivity in performance evaluation and the level of overwork. I find support that subjectivity in performance evaluation impacts the level of overwork through the effort heuristic and through feelings of uncertainty, such that there is a statistically significant difference in measures for the effort heuristic and uncertainty between my subjectivity conditions.

CHAPTER 7: CONCLUSION

7.1 Introduction

This chapter provides the conclusion to my dissertation. In Section 7.2, I discuss the results of my hypotheses and research question testing. In Section 7.3, I discuss the limitations of my study and identify opportunities for future research. Finally, in Section 7.4, I conclude my dissertation.

7.2 Discussion of Hypotheses Testing and Research Question Results

This dissertation examines subjectivity in performance evaluation and employee group identity as antecedents of employee overwork. I hypothesize that subjectivity in performance evaluation and group identity each have separate, positive main effects on employees' level of overwork (H1 and H3, respectively), and that they interact such that the positive effect of subjectivity in performance evaluation on employees' level of overwork is higher when employee group identity is stronger compared to weaker (H2). My experimental results largely support my hypotheses. In my main experiment, I find that participants in the *Higher Subjectivity* condition have a higher level of overwork than those in the *Lower Subjectivity* condition, but only in the *Stronger Identity* condition and not in the *Weaker Identity* condition. This partially supports H1 and is consistent with H2. I also find that participants in the *Stronger Identity* condition have a higher level of overwork than those in the *Weaker Identity* condition, but only when *Subjectivity* is *Higher* and not when it is *Lower*. This partially supports H3. Additionally, I find that there is a significant interaction effect between *Subjectivity* and *Group Identity*, such that the positive effect of subjectivity in performance evaluation on overwork is increasing in group identity strength. This supports H2. Finally, Research Question 1 considers the interactive effect of my two

independent variables on participants' performance on the writing task, via overwork. I do not find evidence that *Subjectivity* and *Group Identity* interact to impact performance.

Taken together, my results provide evidence of the following: (1) higher levels of subjectivity in performance evaluation lead to higher levels of employee overwork when group identity is stronger, (2) stronger group identity leads to higher levels of employee overwork at higher levels of subjectivity in performance evaluation, and (3) the positive effect of subjectivity in performance evaluation on overwork is higher when group identity is stronger compared to weaker.

Besides my main results, in my secondary experiment, I find that *Subjectivity* has a positive main effect on measures of effort heuristic and feelings of uncertainty. This provides support that the effort heuristic and feelings of uncertainty form the underlying theoretical mechanism for subjectivity in performance evaluation impacting the level of overwork.

7.3 Limitations and Opportunities for Future Research

My study has limitations that provide opportunity for future research. First, I take an initial step in understanding the antecedents of employee overwork, and as such only examine a single period setting where a history of employee overwork has yet to develop. Although I expect my theory to strengthen over repeated periods of time as employee overwork becomes a norm, future research can examine the impact of repeated exposure to my antecedents. Relatedly, an exception that may lead to reduced overwork in a multi-period setting would be if there are signals in the setting, after the first performance period, that indicate that overwork is not desired and that there are other antecedents to good performance evaluation (e.g., better efficiency or creativity). I expect this may act as an intervention in my setting which decreases the level of overwork.

Second, limitations exist in my tests for process evidence. I use data from my main experiment to perform cross-sectional analysis that helps support that the effort heuristic, and to some extent social comparison, form theoretical processes underlying my predictions. I also use an additional experiment that separately examines the effect of subjectivity in performance evaluation on the effort heuristic and on feelings of uncertainty. However, I do not provide evidence using statistical mediation analysis which could more completely support a mediation model. This is because my process measures in the main experiment are measured after the dependent variable of overwork. This has the disadvantage of creating carryover effects from my dependent variable (*DV_Overwork*) to my process measures, reducing the power in testing a statistical mediation model (Asay et al., 2022). Thus, future research can design an experiment that is more targeted at supporting a complete mediation model.

Third, my setting of interest is one in which employees have a neutral level of trust in their evaluator, because the evaluator has no explicit incentive to be untrustworthy. Similarly, I make no assumptions about how much employees identify with their evaluator. However, trust in and shared values with the evaluator can impact how employees respond to subjective performance evaluation (Gibbs et al., 2004; MacLeod, 2003). Thus, future research might examine how a setting with employee distrust in or a lack of shared values with their evaluators impact how subjectivity in performance evaluation influences employees' level of overwork.

Finally, the overwork literature often discusses how individual-level employee factors are predictors of overwork, such that employees who are more intrinsically motivated or conscientious may be more inclined to overwork (e.g., Ng and Feldman, 2008; Brett & Stroh, 2003). I do not consider individual-level factors in my theoretical model; however, this may be important to

consider in future research, to the extent settings with high levels of overwork tend to attract self-motivated employees.

7.4 Conclusion

Employees often overwork by working longer than contractual or statutory standard working time for no immediate additional monetary gain. Despite its prevalence and its repeated scrutiny in the popular press, little is understood about why employees overwork (Lupu & Empson, 2015; Surowiecki, 2014; Pozen, 2012; Cha, 2010). A better understanding of the antecedents to employee overwork within a firm is important so that management can guide employee overwork to a desired level. When employees overwork, it suggests increased alignment between employee and firm objectives (if overworking has positive returns to firm performance); however, management may also wish to decrease employee overwork due to its potential negative consequences and the impact of these consequences on firm value. I predict and find partial support that greater subjectivity in performance evaluation increases employees' level of overwork. I further predict and find partial support that stronger group identity increases the level of overwork. I also predict and find that the effect of subjectivity on overwork occurs to a greater extent when group identity is stronger versus weaker. Supplemental analyses and an additional experiment provide support that the effort heuristic, and some support that social comparison, act as mechanisms underlying my findings. Finally, I pose a research question examining the interactive effect of subjectivity and group identity on performance; I find no evidence that they interact to affect performance.

My research contributes to theory and practice. First, my study contributes to the overwork literature that examines the consequences and possible determinants of employees' long working hours. Importantly, there is little empirical evidence that examines management controls that lead

to overwork and that examines a direct causal relationship between management controls and overwork. I contribute to filling this gap. I show how two key management controls, subjectivity in performance evaluation and group identity, lead employees to work longer and contribute to the propagation of firm overwork cultures (e.g., Ladva and Andrew, 2014; Michel, 2011; Feldman, 2002). This is consequential for firms that want to discourage employees from overworking, due to its negative consequences, but it is also informative to firms designing their control systems as it allows them to have a more complete understanding of what motivates their employees.

My study also contributes to the subjective performance evaluation literature (e.g., Cai et al., 2022; Kelly et al., 2015; Bol, 2011; 2008). I develop behavioural theory that suggests the uncertainty that comes with subjective performance evaluation can result in judgment heuristics that lead employees to believe working longer will result in better evaluation. This is unlike previous research that has theorized that the uncertainty that comes with subjective performance evaluation can result in contracting issues: rational employees may in fact reduce effort because they do not see how their effort translates to compensation due to a lack of clear, objective measurement criteria (Grabner et al., 2020; Luft et al., 2016; Bol, 2008). Further, in studying employee overwork, my study contributes to filling a gap in the literature, noted by Wick (2021): there is little research that looks at the effects of subjectivity in performance evaluation on outcomes other than employee bonuses and performance.

Finally, my study also extends the identity literature that examines the impact of identity on employee behaviour (e.g., Brown et al., 2022; Estep, 2021). My theoretical model predicts that stronger group identity can result in behaviour that is not group-aligned, whereby each employee wishes to avoid having the lowest level of overwork in their group. Paradoxically, this creates a negative externality in the group because each group member then feels compelled to keep

increasing their overwork levels to match those of their peers. Whereas current research in management accounting shows how stronger group identity can lead to group aligned behavior (Shang et al., 2020; Kelly & Presslee, 2017; Towry, 2003), my study examines a context where stronger identity can result in behavior that may not align with the interest of the group.

REFERENCES

Abernethy, M. A., Bouwens, J., & Kroos, P. (2017). Organization identity and earnings manipulation. *Accounting, Organizations and Society, 58*, 1-14.

Afota, M. C., Ollier-Malaterre, A., & Vandenberghe, C. (2019). How supervisors set the tone for long hours: Vicarious learning, subordinates' self-motives and the contagion of working hours. *Human Resource Management Review, 29*(4), 100673.

Agoglia, C. P., Brazel, J. F., Hatfield, R. C., & Jackson, S. B. (2010). How do audit workpaper reviewers cope with the conflicting pressures of detecting misstatements and balancing client workloads?. *Auditing: A Journal of Practice & Theory, 29*(2), 27-43.

Akerlof, R. (2016). "We thinking" and its consequences. *American Economic Review*, *106*(5), 415-419.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, *115*(3), 715-753.

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, *19*(1), 9-32.

Akerlof, G. A., & Kranton, R. E. (2008). Identity, supervision, and work groups. *American Economic Review*, *98*(2), 212-217.

Agoglia, C. P., Hatfield, R. C., & Lambert, T. A. (2015). Audit team time reporting: An agency theory perspective. *Accounting, Organizations and Society*, *44*, 1-14.

Ahn, T. S., Hwang, I., & Kim, M. I. (2010). The impact of performance measure discriminability on ratee incentives. *The Accounting Review, 85*(2), 389-417.

Anderson, S. W., Dekker, H. C., Sedatole, K. L., & Wiersma, E. (2020). When one size does not fit all: Using ex post subjective ratings to provide parity in risk-adjusted compensation. *Management Accounting Research*, *49*, 100706.

Anderson-Gough, F., Grey, C., & Robson, K. (2000). In the name of the client: The service ethic in two professional services firms. *Human Relations*, *53*(9), 1151-1174.

Anderson-Gough, F., Grey, C., & Robson, K. (2001). Tests of time: organizational time-reckoning and the making of accountants in two multi-national accounting firms. *Accounting, Organizations and Society*, *26*(2), 99-122.

Alvesson, M., & Kärreman, D. (2004). Interfaces of control. Technocratic and socio-ideological control in a global management consultancy firm. *Accounting, Organizations and Society, 29*(3-4), 423-444.

Aranda, C., Arellano, J., & Davila, A. (2019). Subjective bonuses and target setting in budget-based incentive contracts. *Management Accounting Research*, *43*, 45-60.

Arnold, M. C., & Tafkov, I. D. (2019). Managerial discretion and task interdependence in teams. *Contemporary Accounting Research*, *36*(4), 2467-2493.

Arnold, M. C., Hannan, R. L., & Tafkov, I. D. (2020). Mutual monitoring and team member communication in teams. *The Accounting Review*, *95*(5), 1-21.

Arnold, M. C., Hannan, R. L., & Tafkov, I. D. (2018). Team member subjective communication in homogeneous and heterogeneous teams. *The Accounting Review 93*(5): 1-22.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596.

Asay, H. S., Guggenmos R., Kadous K., Koonce L., and Libby R. (2022). Theory testing and process evidence in accounting experiments. *The Accounting Review*, *97*(6), 23-43.

Ashforth, B. E., Harrison, S. H., & Corley, K. G. (2008). Identification in organizations: An examination of four fundamental questions. *Journal of Management*, *34*(3), 325-374.

Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review*, *14*(1), 20-39.

Awasthi, V., & Pratt, J. (1990). The effects of monetary incentives on effort and decision performance: The role of cognitive characteristics. *Accounting Review*, 797-811.

Babbar, S., & Aspelin, D. J. (1998). The overtime rebellion: symptom of a bigger problem?. *Academy of Management Perspectives*, *12*(1), 68-76.

Bachmann, R. L., Loyeung, A., Matolcsy, Z. P., & Spiropoulos, H. (2020). Powerful CEOs, cash bonus contracts and firm performance. *Journal of Business Finance & Accounting*, *47*(1-2), 100-131.

Baker, G., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics*, *109*(4), 1125-1156.

Banker, R. D., & Mashruwala, R. (2007). The moderating role of competition in the relationship between nonfinancial measures and future financial performance. *Contemporary Accounting Research*, *24*(3), 763-793.

Bauch, K. A., Kotzian, P., & Weißenberger, B. E. (2021). Likeability in subjective performance evaluations: does it bias managers' weighting of performance measures?. *Journal of Business Economics*, *91*(1), 35-59.

Bauer, T. D. (2015). The effects of client identity strength and professional identity salience on auditor judgments. *The Accounting Review*, *90*(1), 95-114.

Bauer, T. D. & Estep, C. (2019). One team or two? Investigating relationship quality between auditors and IT specialists: Implications for audit team identity and the audit process. *Contemporary Accounting Research*, *36*(4), 2142-2177.

Beheshti, N. (2021, August 18). *We worked longer hours during the pandemic-research says we need to work smarter, not harder*. Forbes. https://www.forbes.com/sites/nazbeheshti/2021/08/18/we-worked-longer-hours-during-the-pandemic-research-says-we-need-to-work-smarter-not-harder/.

Beach, S. R., & Tesser, A. (1995). Self-esteem and the extended self-evaluation maintenance model. In *Efficacy, agency, and self-esteem* (pp. 145-170). Springer, Boston, MA.

Bentley, J. W. (2021). Improving the statistical power and reliability of research using Amazon Mechanical Turk. *Accounting Horizons*, *35*(4), 45-62.

Blagoev, B., Muhr, S. L., Ortlieb, R., & Schreyögg, G. (2018). Organizational working time regimes: Drivers, consequences and attempts to change patterns of excessive working hours. *German Journal of Human Resource Management*, *32*(3-4), 155-167.

Bloomfield, R. J., Rennekamp, K., Steenhoven, B., & Stewart, S. (2021). Penalties for unexpected behavior: Double standards for women in finance. *The Accounting Review*, *96*(2), 107-125.

Bol, J. C. (2008). Subjectivity in Compensation Contracting. *Journal of Accounting Literature*, *27*, 1 -24.

Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, *86*(5), 1549-1575.

Bol, J. C., Keune, T. M., Matsumura, E. M., & Shin, J. Y. (2010). Supervisor discretion in target setting: An empirical investigation. *The Accounting Review*, *85*(6), 1861-1886.

Bol, J. C., & Leiby, J. (2018). Subjectivity in Professionals' Incentive Systems: Differences between Promotion-and Performance-Based Assessments. *Contemporary Accounting Research*, *35*(1), 31-57.

Bol, J. C., & Smith, S. D. (2011). Spillover effects in subjective performance evaluation: Bias and the asymmetric influence of controllability. *The Accounting Review*, *86*(4), 1213-1230.

Bonner, S. E., & Sprinkle, G. B. (2002). The effects of monetary incentives on effort and task performance: theories, evidence, and a framework for research. *Accounting, Organizations and Society*, *27*(4-5), 303-345.

Brandon-Jones, A., Lewis, M., Verma, R., & Walsman, M. C. (2016). Examining the characteristics and managerial challenges of professional services: An empirical study of management consultancy in the travel, tourism, and hospitality sector. *Journal of Operations Management*, *42*, 9-24.

Brett, J. M., & Stroh, L. K. (2003). Working 61 plus hours a week: why do managers do it?. *Journal of Applied Psychology*, *88*(1), 67.

Brewer, M. B., & Weber, J. G. (1994). Self-evaluation effects of interpersonal versus intergroup social comparison. *Journal of Personality and Social Psychology*, *66*(2), 268.

Brown, J. D., Novick, N. J., Lord, K. A., & Richards, J. M. (1992). When Gulliver travels: Social context, psychological closeness, and self-appraisals. *Journal of Personality and Social Psychology*, *62*(5), 717.

Brown, J. L., Sprinkle, G. B., & Way, D. (2022). The effects of multi-level group identification on intergroup helping behavior. *Journal of Management Accounting Research*, *34*(1), 97-116.

Buchheit, S., Doxey, M. M., Pollard, T., & Stinson, S. R. (2018). A technical guide to using Amazon's Mechanical Turk in behavioral accounting research. *Behavioral Research in Accounting*, *30*(1), 111-122.

Buchheit, S., Pasewark, W. R., & Strawser, J. R. (2003). No need to compromise: Evidence of public accounting's changing culture regarding budgetary performance. *Journal of Business Ethics*, *42*(2), 151-163.

Buckless, F. A., & Ravenscroft, S. P. (1990). Contrast coding: A refinement of ANOVA in behavioral analysis. *Accounting Review*, 933-945.

Buell, R. W., & Norton, M. I. (2011). The labour illusion: How operational transparency increases perceived value. *Management Science*, *57*(9), 1564-1579.

Burke, R. J. (2009). Working to live or living to work: should individuals and organizations care?. *Journal of Business Ethics*, *84*(2), 167-172.

Buunk, A. P., Dijkstra, P., Bosch, Z. A., Dijkstra, A., & Barelds, D. P. (2012). Social comparison orientation as related to two types of closeness. *Journal of Research in Personality*, *46*(3), 279-285.

Buunk, A. P., & Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, *102*(1), 3-21.

Cabrera, A., & Cabrera, E. F. (2002). Knowledge-sharing dilemmas. *Organization Studies*, *23*(5), 687-710.

Cai, W., Gallani, S., & Shin, J. E. (2022). Incentive Effects of Subjective Allocations of Rewards and Penalties. *Management Science*.

Campbell, I., & Charlesworth, S. (2012). Salaried lawyers and billable hours: a new perspective from the sociology of work. *International Journal of the Legal Profession*, *19*(1), 89-122.

Canada Labour Code, RSC 1985, c L-2.

Caprino, K. (2016, January 14). How companies like Uber, Facebook and Salesforce engage in team-building (It's not what you think). *Forbes*. Retrieved from https://www.forbes.com/sites/kathycaprino/2016/01/14/how-companies-like-uber-facebook-and-salesforce-engage-in-team-building-its-not-what-you-think/?sh=6261ffea3cc1.

Carmichael, S. (2015, December 28). *The Research Is Clear: Long Hours Backfire for People and for Companies*. Harvard Business Review. https://bit.ly/3j6ZmRI

Cameron, J. E. (2004). A three-factor model of social identity. *Self and Identity*, *3*(3), 239-262.

Cha, Y. (2010). Reinforcing separate spheres: The effect of spousal overwork on men's and women's employment in dual-earner households. *American Sociological Review*, *75*(2), 303-329.

Cha, Y., & Weeden, K. A. (2014). Overwork and the slow convergence in the gender gap in wages. *American Sociological Review*, *79*(3), 457-484.

Chaiken, S., & Maheswaran, D. (1994). Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology, 66*(3), 460–473.

Chan, E. W., Kachelmeier, S. J., & Zhang, X. (2021). Working longer but not harder: the effects of incentivizing inputs versus outputs in a heterogeneous workforce. *The Accounting Review*, *96*(5), 133-156.

Chang, C. (2004). Country of origin as a heuristic cue: The effects of message ambiguity and product involvement. *Media Psychology, 6*(2), 169-192.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review, 99*(1), 431-57.

Cheng, M. M., & Coyte, R. (2014). The effects of incentive subjectivity and strategy communication on knowledge-sharing and extra-role behaviours. *Management Accounting Research*, *25*(2), 119-130.

Cheng, Y., Mukhopadhyay, A., & Schrift, R. Y. (2017). Do Costly Options Lead to Better Outcomes? How the Protestant Work Ethic Influences the Cost–Benefit Heuristic in Goal Pursuit. *Journal of Marketing Research, 54*(4), 636–649. https://doi.org/10.1509/jmr.15.0105

Christensen, B. E., Newton, N. J., & Wilkins, M. S. (2021). How do team workloads and team staffing affect the audit? Archival evidence from US audits. *Accounting, Organizations and Society*, *92*, 101225.

Cloyd, C. B. (1997). Performance in tax research tasks: The joint effects of knowledge and accountability. *The Accounting Review*, *72*(1), 111-131.

Coffey, A. J. (1994). Timing is everything'; graduate accountants, time and organizational commitment. *Sociology*, *28*(4), 943-956.

Cortes, P., & Pan, J. (2017). Cross-country evidence on the relationship between overwork and skilled women's job choices. *American Economic Review*, *107*(5), 105-09.

Dembe, A. E. (2009). Ethical issues relating to the health effects of long working hours. *Journal of Business Ethics, 84*(2), 195-208.

Demeré, B. W., Sedatole, K. L., & Woods, A. (2019). The role of calibration committees in subjective performance evaluation systems. *Management Science, 65*(4), 1562-1585.

Dhaliwal, D. S., Radhakrishnan, S., Tsang, A., & Yang, Y. G. (2012). Nonfinancial disclosure and analyst forecast accuracy: International evidence on corporate social responsibility disclosure. *The Accounting Review, 87*(3), 723-759.

Eastman, W. (1998). Working for position: Women, men, and managerial work hours. *Industrial Relations: A Journal of Economy and Society, 37*(1), 51-66.

Eckel, C. C., & Grossman, P. J. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization, 58*(3),

Empson, L. (2018, July 20). *If You're So Successful, Why Are You Still Working 70 Hours a Week?* Harvard Business Review. https://hbr.org/2018/02/if-youre-so-successful-why-are-you-still-working-70-hours-a-week

Reid, E. (2015). Embracing, passing, revealing, and the ideal worker image: How people navigate expected and experienced professional identities. *Organization Science, 26*(4), 997-1017.

Estep, C. (2021). Auditor integration of IT specialist input on internal control issues: How a weaker team identity can be beneficial. *The Accounting Review, 96*(5), 263-289.

Eyal, P., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 1-20.

Farrell, A.M., Grenier, J.H., and Leiby, J. (2017). Scoundrels or stars? Theory and evidence on the quality of workers in online labor markets. *The Accounting Review, 92*(1): 93-114.

Feldman, D. C. (2002). Managers' propensity to work longer hours: A multilevel analysis. *Human Resource Management Review, 12*(3), 339-357.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Festinger, L., Torrey, J., & Willerman, B. (1954). Self-evaluation as a function of attraction to the group. *Human Relations, 7*(2), 161-174.

Fisher, J. G., Maines, L. A., Peffer, S. A., & Sprinkle, G. B. (2005). An experimental investigation of employer discretion in employee performance evaluation and compensation. *The Accounting Review, 80*(2), 563-583.

Forehand, M. R., Deshpandé, R., & Reed, A. II. (2002). Identity salience and the influence of differential activation of the social self-schema on advertising response. *Journal of Applied Psychology, 87*(6), 1086–1099.

Garcia, S. M., Tor, A., & Schiff, T. M. (2013). The psychology of competition: A social comparison perspective. *Perspectives on Psychological Science*, *8*(6), 634-650.

Gastorf, J. W., & Suls, J. (1978). Performance evaluation via social comparison: Performance similarity versus related-attribute similarity. *Social Psychology, 41*(4), 297-305.

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: development of a scale of social comparison orientation. *Journal of Personality and Social Psychology, 76*(1), 129.

Gibbs, M., Merchant, K. A., Stede, W. A. V. D., & Vargus, M. E. (2004). Determinants and effects of subjectivity in incentives. *The Accounting Review*, *79*(2), 409-436.

Gicheva, D. (2013). Working long hours and early career outcomes in the high-end labor market. *Journal of Labor Economics*, *31*(4), 785-824.

Goette, L., Huffman, D., & Meier, S. (2006). The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups. *American Economic Review*, *96*(2), 212-216.

Golden, L. (2012). The effects of working time on productivity and firm performance, research synthesis paper. *International Labor Organization (ILO) Conditions of Work and Employment Series*, (33).

Golden, L. (2009). A brief history of long work time and the contemporary sources of overwork. *Journal of Business Ethics*, *84*(2), 217-227.

Golden, L., & Altman, M. (2008). Why do people overwork? Over-supply of hours of labor, labor market forces and adaptive preferences. The long work hours culture: Cause, consequences and choices, 61-83.

Gonzalez, A. (14, November 2014). *Don't Eat Hours, You Idiots*. Going Concern. https://www.goingconcern.com/dont-eat-hours-you-idiots/.

Gorenflo, D. W., & Crano, W. D. (1989). Judgmental subjectivity/objectivity and locus of choice in social comparison. *Journal of Personality and Social Psychology, 57*(4), 605.

Grabner, I. (2014). Incentive system design in creativity-dependent firms. *The Accounting Review*, *89*(5), 1729-1750.

Grabner, I., Künneke, J., & Moers, F. (2020). How calibration committees can mitigate performance evaluation bias: An analysis of implicit incentives. *The Accounting Review*, *95*(6), 213-233.

Guggenmos, R. D., Piercey, M. D., & Agoglia, C. P. (2018). Custom contrast testing: Current trends and a new approach. *The Accounting Review*, *93*(5), 223-244.

Haesebrouck, K. (2021). The effects of information acquisition effort, psychological ownership, and reporting context on opportunistic managerial reporting. *Contemporary Accounting Research, 38*(4), 3085-3112.

Hartmann, F., & Slapničar, S. (2012). The perceived fairness of performance evaluation: The role of uncertainty. *Management Accounting Research, 23*(1), 17-33.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*, 451-470.

Hecht, G., Hobson, J. L., & Wang, L. W. (2020). The effect of performance reporting frequency on employee performance. *The Accounting Review, 95*(4), 199-218.

Hecht, G., Newman, A. H., & Tafkov, I. D. (2019). Managers' strategic use of discretion over relative performance information provision and implications for team-members' effort. *Management Accounting Research*, *45*, 100638.

Heinle, M. S., Hofmann, C., & Kunz, A. H. (2012). Identity, incentives, and the value of information. *The Accounting Review, 87*(4), 1309-1334.

Hensley, V., & Duval, S. (1976). Some perceptual determinants of perceived similarity, liking, and correctness. *Journal of Personality and Social Psychology*, *34*(2), 159.

Hewlett, S. A., & Luce, C. B. (2006). Extreme jobs: the dangerous allure of the 70-hour workweek. *Harvard Business Review*, *84*(12), 49-59.

Hirst, D. E., Koonce, L., & Venkataraman, S. (2007). How disaggregation enhances the credibility of management earnings forecasts. *Journal of Accounting Research, 45*(4), 811-837.

Hoffman, P. J., Festinger, L., & Lawrence, D. H. (1954). Tendencies toward group comparability in competitive bargaining. *Human Relations*, *7*(2), 141-159.

Hogg, M. A., & Terry, D. I. (2000). Social identity and self-categorization processes in organizational contexts. *Academy of Management Review, 25*(1), 121-140.

Hogg, M. A., & Turner, J. C. (1987). Intergroup behaviour, self-stereotyping and the salience of social categories. *British Journal of Social Psychology*, *26*(4), 325-340.

Hogg, M. A., & Turner, J. C. (1985). Interpersonal attraction, social identification and psychological group formation. *European Journal of Social Psychology*, *15*(1), 51-66.

Holmström, B. (2017). Pay for performance and beyond. *American Economic Review*, *107*(7), 1753-1777.

Holmström, B., & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, *7*, 24.

International Labour Office Geneva (ILO). 2004. *Overtime*. https://www.ilo.org/wcmsp5/groups/public/---ed_protect/---protrav/---travail/documents/publication/wcms_170708.pdf

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*(4), 337-349.

Jackson, S. A., Kleitman, S., Howie, P., & Stankov, L. (2016). Cognitive abilities, monitoring confidence, and control thresholds explain individual differences in heuristics and biases. *Frontiers in Psychology*, *7*, 1559.

James, J. (2013, Nov 5). What Long Hours Really Mean. *Medium*. https://medium.com/we-are-mammoth/what-long-hours-really-mean-6107a3fe60b9

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology*, *50*(4), 905-925.

Kadous, K., Kennedy, S. J., & Peecher, M. E. (2003). The effect of quality assessment and directional goal commitment on auditors' acceptance of client-preferred accounting methods. *The Accounting Review*, *78*(3), 759-778.

Kahneman. (2011). *Thinking, Fast and Slow.* Doubleday Canada.

Karaian, J., & Sorkin, A. R. (2021, March 19). *'I'm in a really dark place': Complaints at Goldman Sachs set off a workplace debate.* The New York Times. https://www.nytimes.com/2021/03/19/business/goldman-sachs-analysts-workplace-complaint.html.

Kärreman, D., & Alvesson, M. (2004). Cages in tandem: Management control, social identity, and identification in a knowledge-intensive firm. *Organization*, *11*(1), 149-175.

Kellogg, K. C. (2009). Operating room: Relational spaces and microinstitutional change in surgery. *American Journal of Sociology*, *115*(3), 657-711.

Kelly, K., Dinovitzer, R., Gunz, H., and Gunz, S. P. (2020). The interaction of perceived subjectivity and pay transparency on professional judgment in a profit pool setting: The case of large law firms. *The Accounting Review*, *95*(5), 227–46.

Kelly, K., & Presslee, A. (2017). Tournament group identity and performance: The moderating effect of winner proportion. *Accounting, Organizations and Society*, *56*, 21-34.

Kelly, K. O., Webb, R. A., & Vance, T. (2015). The interactive effects of ex post goal adjustment and goal difficulty on performance. *Journal of Management Accounting Research*, *27*(1), 1-25.

Keynes, J. M. (2010). Economic possibilities for our grandchildren. In *Essays in persuasion* (pp. 321-332). Palgrave Macmillan, London.

Kilduff, G. J., Elfenbein, H. A., & Staw, B. M. (2010). The psychology of rivalry: A relationally-dependent analysis of competition. *Academy of Management Journal, 53*, 943– 969.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. Guilford publications.

Kodz, J. Kersley, B. Strebler, M. T, & O'Regan, S. (1998). *Breaking the Long Hours Culture*. (Report No. IES-R-352). Sussex University., Brighton (England). Institute for Employment Studies.

Konnikova, M. (2014, January 7). *The Open-Office Trap*. The New Yorker. https://www.newyorker.com/business/currency/the-open-office-trap

Kunda, G. (1995). Engineering culture: Control and commitment in a high-tech corporation. *Organization Science*, *6*(2), 228-230.

Kunz, J. (2015). Objectivity and subjectivity in performance evaluation and autonomous motivation: An exploratory study. *Management Accounting Research, 27*, 27-46.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, *40*(1), 91-98.

Kruglanski, A. W., & Mayseless, O. (1990). Classic and current social comparison research: Expanding the perspective. *Psychological Bulletin*, *108*(2), 195.

Kuroda, S., & Yamamoto, I. (2019). Why do people overwork at the risk of impairing mental health?. *Journal of Happiness Studies, 20*, 1519-1538.

Ladva, P., & Andrew, J. (2014). Weaving a web of control: "The Promise of Opportunity" and work-life balance in multinational accounting firms. *Accounting, Auditing & Accountability Journal*, *27*(4), pp. 634-654.

Lambert, T. A., & Agoglia, C. P. (2011). Closing the loop: Review process factors affecting audit staff follow-through. *Journal of Accounting Research*, *49*(5), 1275-1306.

Lapierre, L. M., Spector, P. E., Allen, T. D., Poelmans, S., Cooper, C. L., O'Driscoll, M. P., Sanchez, J. I., Brough, P., & Kinnunen, U. (2008). Family-supportive organization perceptions, multiple dimensions of work–family conflict, and employee satisfaction: A test of model across five samples. *Journal of Vocational Behavior*, *73*(1), 92-106.

Lauck, J. R., Perreault, S. J., Rakestraw, J. R., & Wainberg, J. S. (2020). Strategic audit inquiry: The impact of timing and the promotion of statutory protections on client fraud disclosures. *Accounting Horizons*, *34*(3), 153-167.

Lembke, S., & Wilson, M. G. (1998). Putting the" team" into teamwork: Alternative theoretical contributions for contemporary management practice. *Human Relations*, *51*(7), 927-944.

Lewis, S. (2003). The integration of paid work and the rest of life. Is post-industrial work the new leisure? *Leisure Studies*, *22*(4), 343e345.

Lewis, S. (2007). Working time, client time and family time: accounting for time in the accountancy profession. (van der Lippe, T. and Peters, P. Eds). Edward Elgar, Cheltenham.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, *49*(2), 433-442. https://doi.org/10.3758/s13428-016-0727-z

Liu, W. (2017). *The effect of group identity on sabotage induced by relative performance information*. [Doctoral dissertation, University of Waterloo]. UWSpace. http://hdl.handle.net/10012/12270

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

López, D. M., & Peters, G. F. (2012). The effect of workload compression on audit quality. *Auditing: A Journal of Practice & Theory*, *31*(4), 139-165.

Luft, J., Shields, M. D., & Thomas, T. F. (2016). Additional information in accounting reports: Effects on management decisions and subjective performance evaluations under causal ambiguity. *Contemporary Accounting Research*, *33*(2), 526-550.

Lupu, I., & Empson, L. (2015). Illusio and overwork: playing the game in the accounting field. *Accounting, Auditing & Accountability Journal*, *28(*8), 1310-1340

Lupu, I. & Ruiz-Castro, M. (2021, January 29). *Work-Life Balance Is a Cycle, Not an Achievement*. Harvard Business Review. https://hbr.org/2021/01/work-life-balance-is-a-cycle-not-an-achievement

Lupu, I., Ruiz-Castro, M., & Leca, B. (2020). Role Distancing and the Persistence of Long Work Hours in Professional Service Firms. *Organization Studies*, *43*(1), 7–33.

Maas, V. S., Van Rinsum, M., & Towry, K. L. (2012). In search of informed discretion: An experimental investigation of fairness and trust reciprocity. *The Accounting Review*, *87*(2), 617-644.

MacLeod, W. B. (2003). Optimal contracting with subjective evaluation. *American Economic Review*, *93*(1), 216-240.

Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, *13*(2), 103-123.

Major, V. S., Klein, K. J., & Ehrhart, M. G. (2002). Work time, work interference with family, and psychological distress. *Journal of Applied Psychology, 87*(3), 427–436.

Maume Jr, D. J., & Bellas, M. L. (2001). The overworked American or the time bind? Assessing competing explanations for time spent in paid labor. *American Behavioral Scientist, 44*(7), 1137-1156.

Mazmanian, M., Orlikowski, W. J., & Yates, J. (2013). The autonomy paradox: The implications of mobile email devices for knowledge professionals. *Organization Science*, *24*(5), 1337-1357.

Michel, A. (2011). Transcending socialization: A nine-year ethnography of the body's role in organizational control and knowledge workers' transformation. *Administrative Science Quarterly*, *56*(3), 325-368.

Mitchell, R., Boyle, B., Parker, V., Giles, M., Chiang, V., & Joyce, P. (2015). Managing inclusiveness and diversity in teams: How leader inclusiveness affects performance through status and team identity. *Human Resource Management, 54*(2), 217-239.

Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society*, *30*(1), 67-80.

Morales, A. C. (2005). Giving firms an "E" for effort: Consumer responses to high-effort firms. *Journal of Consumer Research*, *31*(4), 806-812.

Morgan McKinley. (2019, October 31). *Watching the clock: Are UK professionals working too long for too little?* https://www.morganmckinley.com/uk/article/watching-clock-are-uk-professionals-working-too-long-too-little

Mousavi, S., & Gigerenzer, G. (2017). Heuristics are tools for uncertainty. *Homo Oeconomicus*, *34*(4), 361-379.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives*. Sage.

Nelson, M., & Tan, H. T. (2005). Judgment and decision-making research in auditing: A task, person, and interpersonal interaction perspective. *Auditing: A Journal of Practice & Theory*, *24*(s-1), 41-71.

Neth, H., & Gigerenzer, G. (2015). Heuristics: Tools for an uncertain world. In R. Scott & S. Kosslyn. *Emerging trends in the social and behavioral sciences* (pp. 1-18). Wiley Online Library.

Ng, T. W., & Feldman, D. C. (2008). Long work hours: A social identity perspective on meta-analysis data. Journal of Organizational Behavior: *The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *29*(7), 853-880.

Panero, M. E., Weisberg, D. S., Black, J., Goldstein, T. R., Barnes, J. L., Brownell, H., & Winner, E. (2016). Does reading a single passage of literary fiction really improve theory of mind? An attempt at replication. *Journal of Personality and Social Psychology, 111*(5), e46–e54.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184-188.

Pencavel, J. (2015). The productivity of working hours. *The Economic Journal, 125*(589), 2052-2076.

Perlow, L. A. (1999). The time famine: Toward a sociology of work time. *Administrative Science Quarterly*, *44*(1), 57-81.

Perlow, L. A., & Porter, J. L. (2009, October 2009). *Making Time Off Predictable-and Required*. Harvard Business Review. https://hbr.org/2009/10/making-time-off-predictable-and-required.

Persellin, J. S., Schmidt, J. J., Vandervelde, S. D., & Wilkins, M. S. (2019). Auditor perceptions of audit workloads, audit quality, and job satisfaction. *Accounting Horizons*, *33*(4), 95-117.

Peticca-Harris, A., Weststar, J., & McKenna, S. (2015). The perils of project-based work: Attempting resistance to extreme work practices in video game development. *Organization*, *22*(4), 570-587.

Pituch, K. A., & Stevens, J. P. (2016). *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS*. Routledge.

Pleban, R., & Tesser, A. (1981). The effects of relevance and quality of another's performance on interpersonal closeness. *Social Psychology Quarterly*, *44* (3), 278-285.

Ponemon, L. A. (1992). Auditor underreporting of time and moral reasoning: An experimental lab study. *Contemporary Accounting Research*, *9*(1), 171-189.

Pozen, R. C. (2012, October 6). *They Work Long Hours, but What About Results?* The New York Times. https://www.nytimes.com/2012/10/07/business/measure-results-not-hours-to-improve-work-efficiency.html?pagewanted=all&_r=0

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*(1), 7-63.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.

Reuters Staff. (2021, March 18). *Goldman Sachs analysts complain of long hours, unrealistic deadlines*. Reuters. https://www.reuters.com/article/us-goldman-sachs-analyst-survey-idUSKBN2BA2PI

Rosnow, R. L., & Rosenthal, R. (1995). “Some things you learn aren't so”: Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science*, *6*(1), 3-9.

Rousseau, D. M. (1998). Why workers still identify with organizations. *Journal of Organizational Behavior*, 217-233.

Rule, N. O., Macrae, C. N., & Ambady, N. (2009). Ambiguous group membership is extracted automatically from faces. *Psychological Science*, *20*(4), 441-443.

Shang, R., Abernethy, M. A., & Hung, C. Y. (2020). Group identity, performance transparency, and employee performance. *The Accounting Review*, *95*(5), 373-397.

Schrift, R. Y., Kivetz, R., & Netzer, O. (2016). Complicating decisions: The work ethic heuristic and the construction of effortful decisions. *Journal of Experimental Psychology: General*, *145*(7), 807-829.

Smith, R. H. (2000). Assimilative and contrastive emotional reactions to upward and downward social comparisons. In *Handbook of social comparison* (pp. 173-200). Springer, Boston, MA.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*(6), 845.

Sprinkle, G. B. (2000). The effect of incentive contracts on learning and performance. *The Accounting Review*, *75*(3), 299-326.

Sprinkle, G. B., & Williamson, M. G. (2006). Experimental research in managerial accounting. *Handbooks of Management Accounting Research*, *1*, 415-444.

Sprinkle, G. B. (2003). Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society*, *28*(2-3), 287-318.

Steinmetz, S., de Vries, D. H., & Tijdens, K. G. (2014). Should I stay or should I go? The impact of working time and wages on retention in the health workforce. *Human Resources for Health*, *12*(1), 1-12.

Sullivan, T. A. (2014). Greedy institutions, overwork, and work-life balance. *Sociological Inquiry*, *84*(1), 1-15.

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect?. *Current Directions in Psychological Science*, *11*(5), 159-163.

Surowiecki, J. (2014, January 20). *The Cult of Overwork*. The New Yorker. https://www.newyorker.com/magazine/2014/01/27/the-cult-of-overwork.

Sweeney, J. T., & Summers, S. L. (2002). The effect of the busy season workload on public accountants' job burnout. *Behavioral Research in Accounting, 14*(1), 223-245.

Tafkov, I. D. (2013). Private and public relative performance information under different compensation contracts. *The Accounting Review*, *88*(1), 327-350.

Tajfel. H. & Turner. J. C. (1979). An integrative theory of intergroup conflict. In W. C. Austin & S. Worchcl (eds). *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole.

Tan, H. T., & Yu, Y. (2018). Management's responsibility acceptance, locus of breach, and investors' reactions to internal control reports. *The Accounting Review, 93*(6), 331-355.

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. In *Advances in experimental social psychology* (Vol. 21, pp. 181-227). Academic Press.

Tesser, A., & Campbell, J. (1980). Self-definition: The impact of the relative performance and similarity of others. *Social Psychology Quarterly*, *43*(3), 341-347.

Tesser, A., & Campbell, J. (1982). Self-evaluation maintenance and the perception of friends and strangers. *Journal of Personality, 50*(3), 261-279.

Tesser, A., Campbell, J., & Smith, M. (1984). Friendship choice and performance: Self-evaluation maintenance in children. *Journal of Personality and Social Psychology*, *46*(3), 561.

Tesser, A., & Smith, J. (1980). Some effects of task relevance and friendship on helping: You don't always help the one you like. *Journal of Experimental Social Psychology, 16*(6), 582-590. https://doi.org/10.1016/0022-1031(80)90060-8

Towry, K. L. (2003). Control in a teamwork environment—The impact of social ties on the effectiveness of mutual monitoring contracts. *The Accounting Review*, *78*(4), 1069-1095.

Tropp, L. R., & Wright, S. C. (2001). Ingroup identification as the inclusion of ingroup in the self. *Personality and Social Psychology Bulletin*, *27*(5), 585-600.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

United States., & United States. (2011). *The Fair Labor Standards Act of 1938, as amended.* Washington, D.C.: U.S. Dept. of Labor, Wage and Hour Division.

U.S. Department of Labor. (2023). State Minimum Wage Laws. Retrieved April 10, 2023, from https://www.dol.gov/agencies/whd/minimum-wage/state#:~:text=Employers%20covered%20by%20the%20FLSA,wage%20of%20%247.25%20per%20hour.

Van den Berg, P., Dewitte, S., & Wenseleers, T. (2021). Uncertainty causes humans to use social heuristics and to cooperate more: An experiment among Belgian university students. *Evolution and Human Behavior, 4*2(3), 223-229.

Van der Stede, W. A. (2015). Management accounting: Where from, where now, where to?. *Journal of Management Accounting Research, 27*(1), 171-176.

Van Der Vegt, G. S., & Bunderson, J. S. (2005). Learning and performance in multidisciplinary teams: The importance of collective team identification. *Academy of Management Journal, 48*(3), 532-547.

Van Rinsum, M., & Verbeeten, F. H. (2012). The impact of subjectivity in performance evaluation practices on public sector managers' motivation. *Accounting and Business Research, 42*(4), 377-396.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology Section A, 47*(3), 631-650.

Vershbow, S. (2021, September 28). *Workers aren't burnt out from the pandemic. they're burnt out from overwork*. Newsweek. https://www.newsweek.com/workers-arent-burnt-out-pandemic-were-burnt-out-overwork-opinion-1632819

Voußem, L., Kramer, S., & Schäffer, U. (2016). Fairness perceptions of annual bonus payments: The effects of subjective performance measures and the achievement of bonus targets. *Management Accounting Research, 30*, 32-46.

Waddoups, N. (2022). Feedback-Driven Time Segmenting: The Effect of Feedback Frequency on Employee Behavior. *Contemporary Accounting Research*, *39*(3), 1516-1541.

Wick, S. (2020). *Subjective Evaluation of Professional Employees: Work-Day Duration as a Heuristic to Evaluate Output*. [Doctoral dissertation, Wilfrid Laurier University]. Scholars Commons @ Laurier.

Wick, S. (2021). Subjectivity in Performance Evaluations: A Review of the Literature. *Accounting Perspectives*, *20*(4), 653-685.

Wilder, D. A. (1984). Empirical contributions: Predictions of belief homogeneity and similarity following social categorization. *British Journal of Social Psychology, 23*(4), 323-333.

Work-life balance. Consultancy.eu. (n.d.). https://www.consultancy.eu/career/work-life-balance.

Xu, Y., Pace, S., Kim, J., Iachini, A., King, L.B., Harrison, T., DeHart, D., Levkoff, S.E., Browne, T.A., Lewis, A.A. and Kunz, G.M., Reitmeier, M., Utter, R.kK., & Simone, M. (2022). Threats to Online Surveys: Recognizing, Detecting, and Preventing Survey Bots. *Social Work Research*, *46*(4), 343-350.

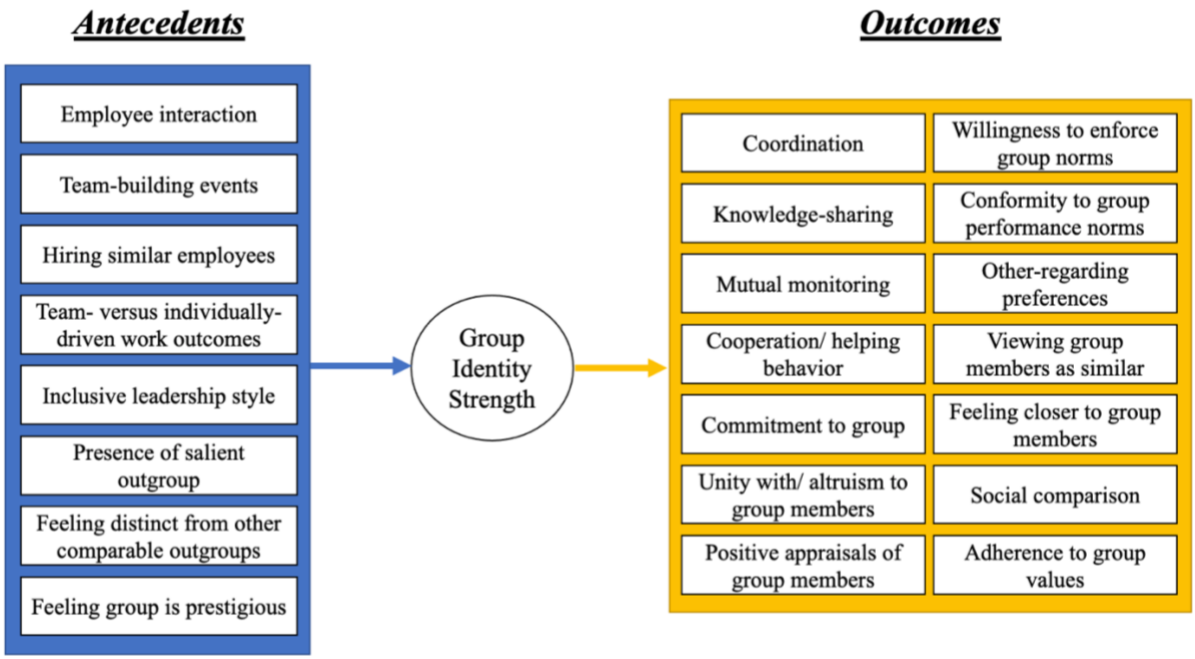
Yatsenko, D. (2022). Productivity effects of shared peer effort and relative performance information. *Management Accounting Research, 56*, 100779.

Yeung, C. W., & Soman, D. (2007). The duration heuristic. *Journal of Consumer Research*, *34*(3), 315-326.

APPENDIX A: SUPPLEMENTARY FIGURES

FIGURE 1

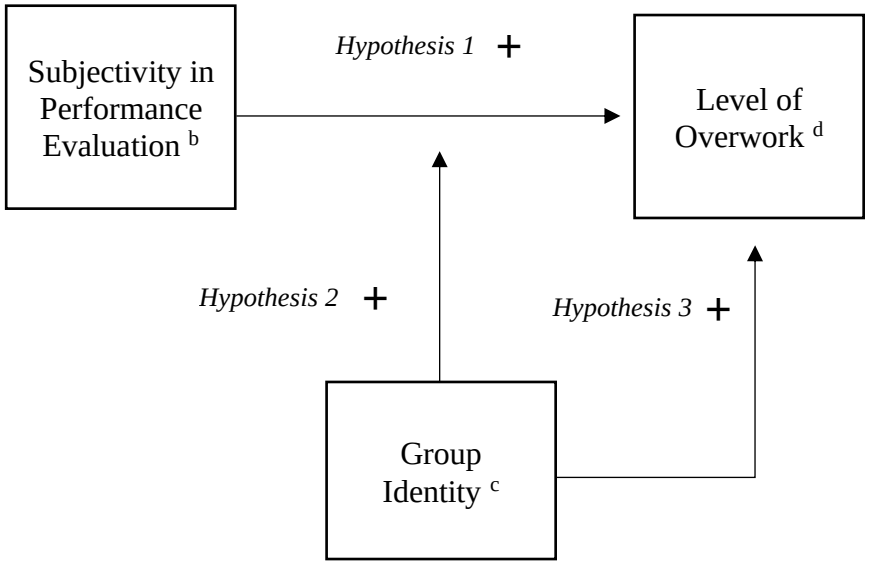
Summary of Antecedents and Outcomes of Group Identity Strength [a]



[a] Figure 1 represents a summary of the antecedents and outcomes of group identity strength. See Section 2.4 for details and relevant studies.

FIGURE 2

Theoretical Model [a]



[a] Figure 2 depicts my theoretical model. My first prediction posits a positive main effect of subjectivity in performance evaluation on employee overwork. My second prediction posits that group identity moderates the relationship between subjectivity and overwork, such that the effect of subjectivity on an employee's level of overwork will be greater under stronger, versus weaker, group identity. My final prediction posits a main effect of group identity on overwork.

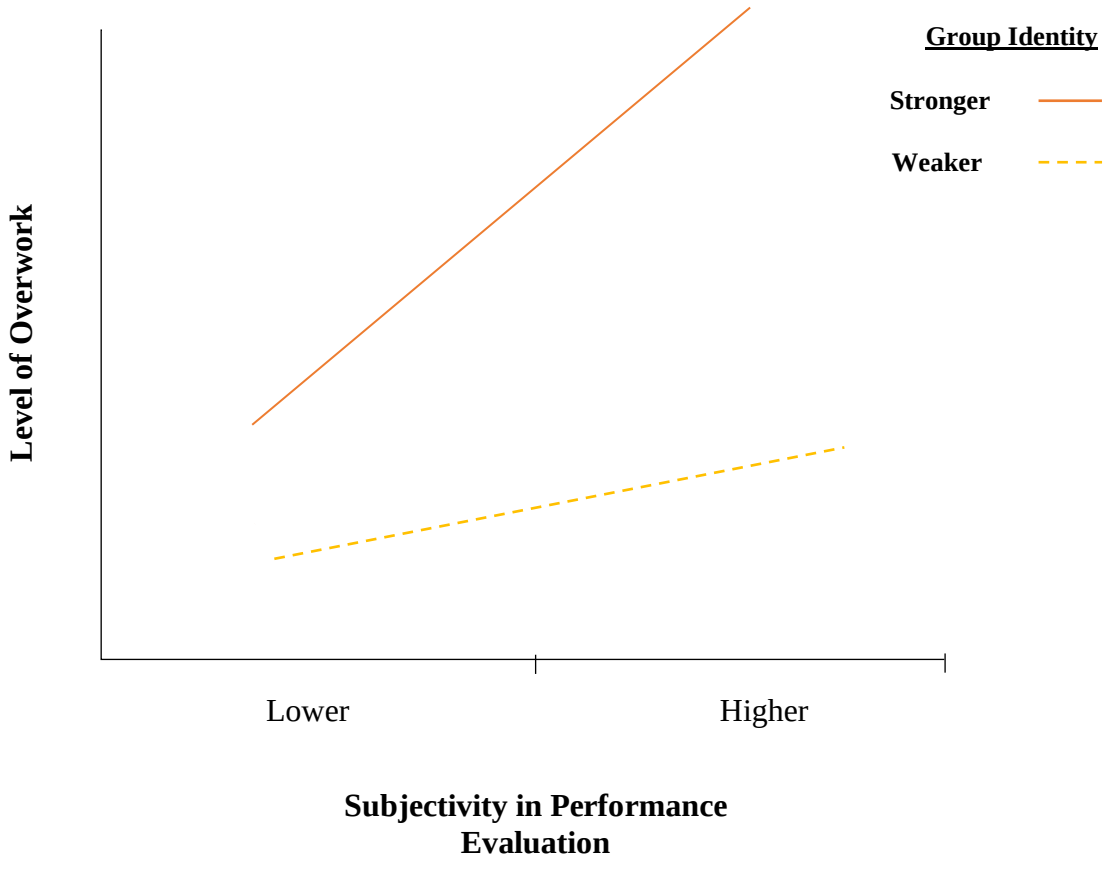
[b] Subjectivity in performance evaluation is the extent to which "judgment based on personal impressions, feelings, and opinions" is used in evaluating an employee's performance (Bol, 2008, p. 2).

[c] Group identity is the extent to which a person thinks of herself as a member of a group, such that she derives her sense of self from being a member of that group (Chen & Li, 2009; Towry, 2003; Hogg & Turner, 1987). A person with a stronger (versus weaker) group identity derives her sense of self to a larger extent from being a member of the group (Mael & Ashforth, 1992).

[d] I define overwork as employees working beyond contractual or statutory standard working time for no immediate additional monetary gain ("Work-life balance", n.d.).

FIGURE 3

Hypothesized Results (H1, H2, and H3)[a]

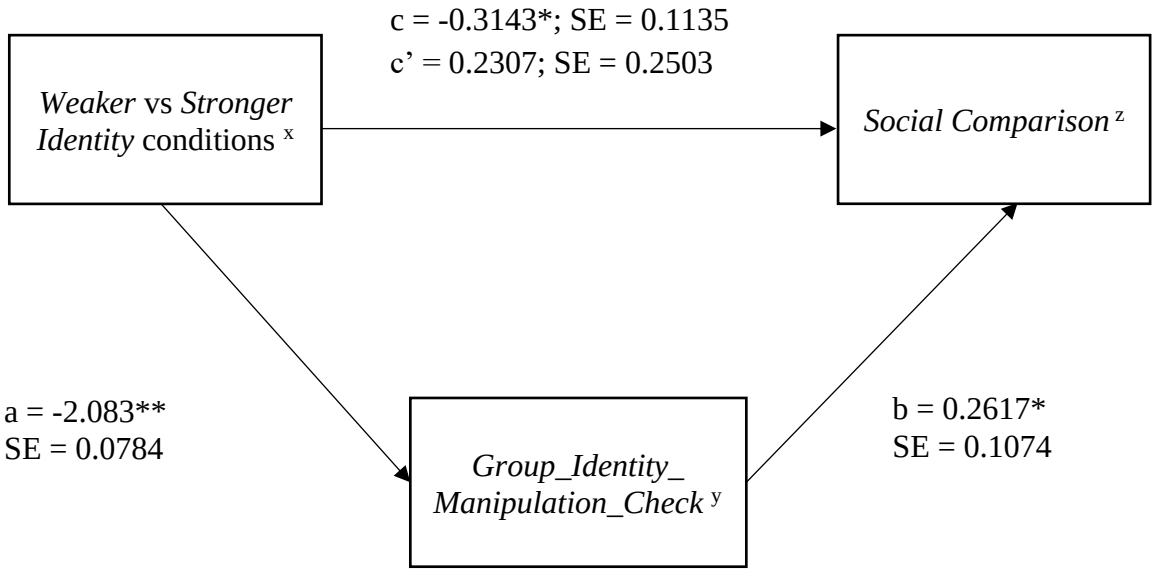


[a] Figure 3 depicts the predicted relationship between subjectivity in performance evaluation, group identity, and the level of overwork. See Figure 2 for variable definitions.

FIGURE 4

Group Identity Pilot Study Path Analysis [w]

Significant mediation: *ab* = -0.5451, 95% CI [-0.9386, -0.1341]



[w] Figure 4 captures a PROCESS model (Model 4), a path analysis modelling tool used to estimate direct and indirect effects in mediation models (Hayes, 2017). I run the model in SPSS and use my group identity manipulation as my independent variable, the *Social Comparison* measure as my dependent variable, and the *Group_Identity_Manipulation_Check* measure as my mediator (Hayes & Preecher, 2014). The model outputs the indirect effect of the *Weaker Identity* condition relative to the *Stronger Identity* condition. All dependent variables are on 7-point scales. *Group_Identity_Manipulation_Check* and *Social Comparison* have neutral midpoints and are anchored at 1 and 7. The total effect (c) is the sum of the direct effect (c') and the indirect effect (*ab*).

[x] Group identity was manipulated in vignettes at two levels: *Stronger* vs *Weaker*. The vignettes touched on three factors of group identity: (1) ingroup ties or attitude similarity, (2) liking of ingroup members or ingroup affect, *and* (3) belonging to the group or the group's centrality to the self. The *Stronger* (*Weaker*) *Identity* vignette described the participant as having high (low) levels of these factors in relation to a hypothetical team.

[y] *Group_Identity_Manipulation_Check* is the average of my three identity variables of *Happy* (the extent participants would be happy to be part of the team), *Belong* (the extent participants would feel like a member of the team), and *Like* (the extent participants would like their team members; adopted from Kelly and Presslee, 2017).

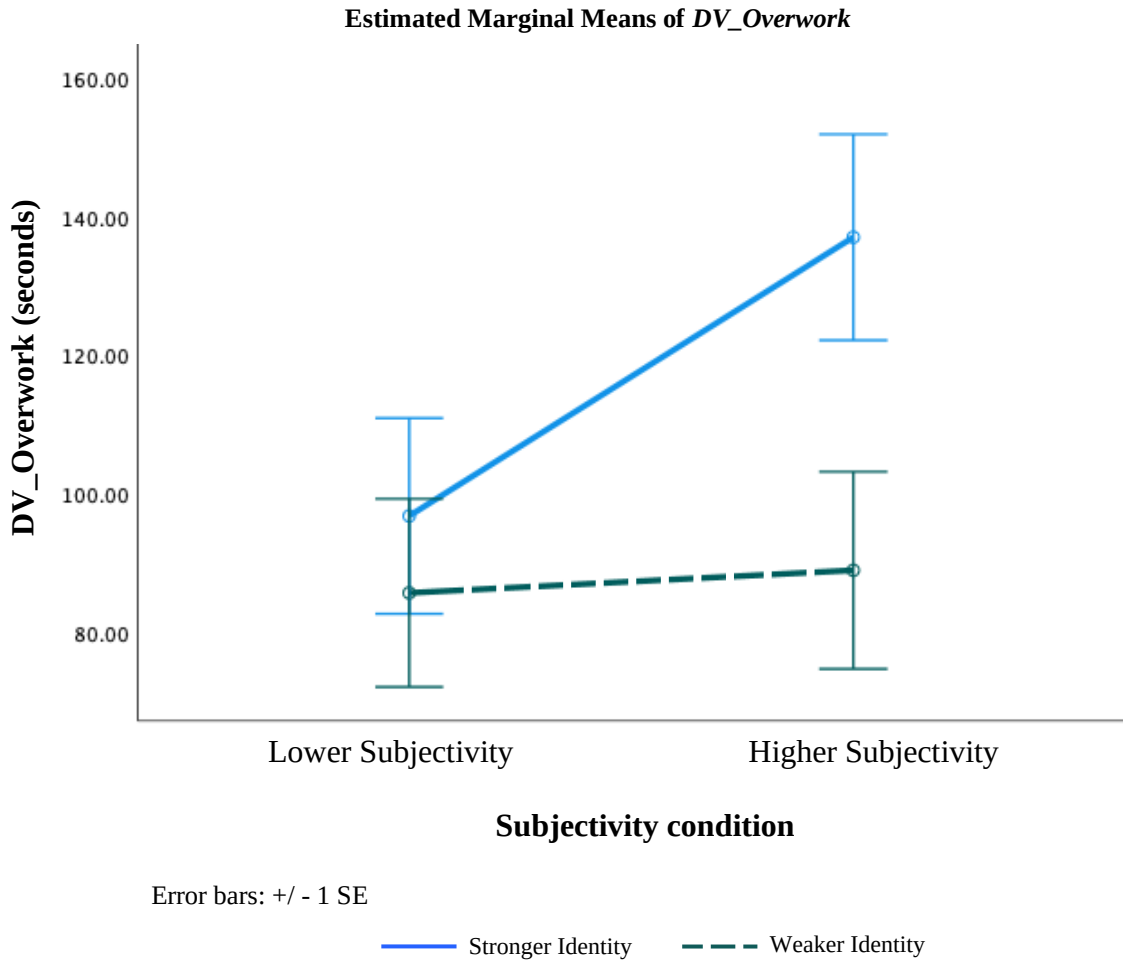
[z] *Social Comparison* is the average of my two social comparison variables: *Think About* (the extent participants would think about their performance on team-related work compared to their team members' performance) and *Concerned About* (the extent participants would be concerned about their performance compared to their team members' performance on team-related work; adapted from Tafkov, 2013).

[**] $p < 0.001$.

[*] $p < 0.050$. All p-values are two-tailed.

FIGURE 5

Actual Results (H1, H2, and H3) [a]

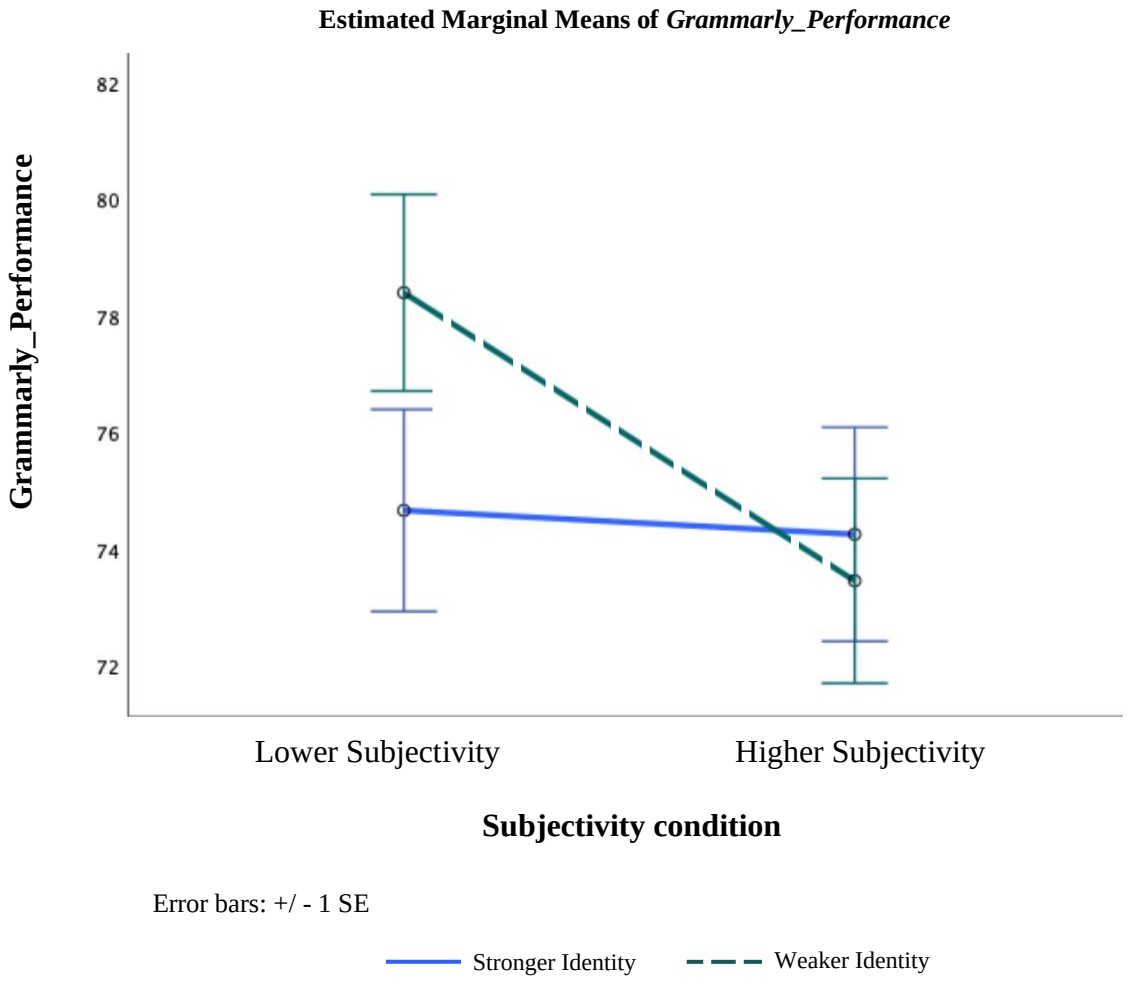


[a] Figure 5 represents the relationship between subjectivity in performance evaluation, group identity, and the level of overwork as found in my experiment. See Figure 2 for definitions of these constructs. *DV_Overwork* represents the amount of time participants worked above the contracted time of 5 minutes (300 seconds). Group identity was manipulated in vignettes at two levels: *Stronger* vs *Weaker*. The vignettes touched on three factors of group identity: (1) ingroup ties or attitude similarity, (2) liking of ingroup members or ingroup affect, and (3) belonging to the group or the group's centrality to the self. The *Stronger* (*Weaker*) *Identity* vignette described the participant as having high (low) levels of these factors in relation to a hypothetical team. Subjectivity in performance evaluation (subjectivity) was manipulated in vignettes at two levels, *Lower* vs *Higher*, such that in the *Lower Subjectivity* condition, participants were told that the evaluator will be able to objectively evaluate their output from the writing task, whereas in the *Higher Subjectivity* condition, participants were told that the evaluator will not be able to objectively evaluate their output from the writing task but would use their personal views and opinions to do so.

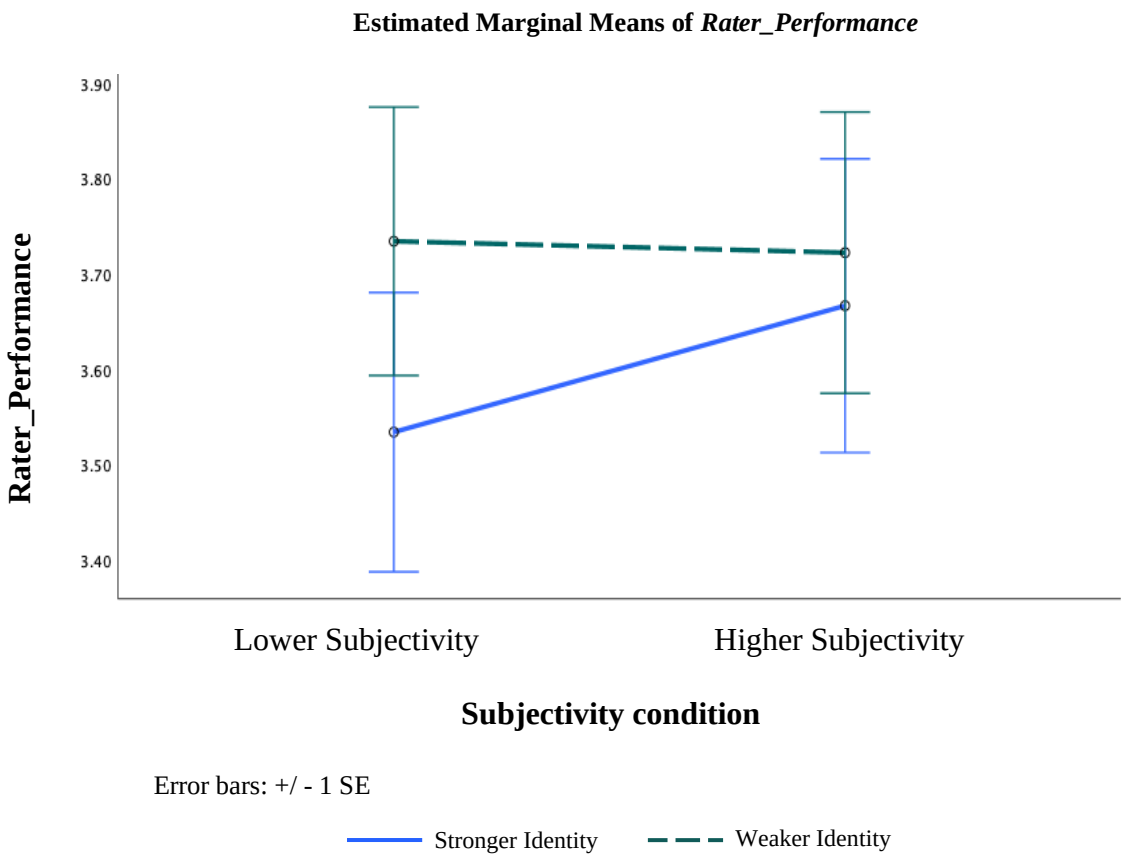
FIGURE 6

RQ1 Results [a]

Panel A: *Grammarly_Performance*



Panel B: *Rater_Performance*

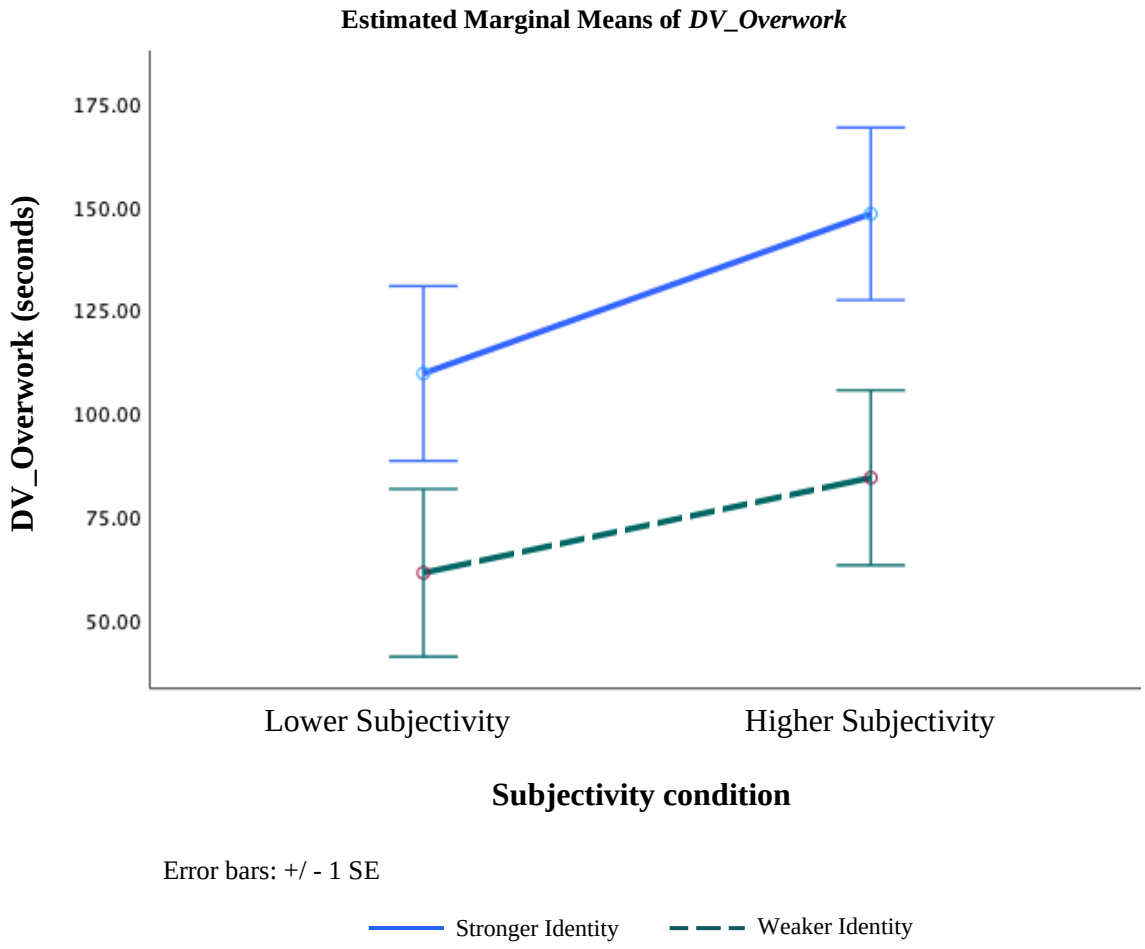


[a] Figure 6 represents the relationship between subjectivity in performance evaluation, group identity, and performance as found in my experiment (RQ1). See Figure 5 for descriptions of independent variables. Panel A uses *Grammarly_Performance* as the measure of performance. *Grammarly_Performance* is a score out of 100, and represents the score rewarded by Grammarly software on participants' writing output. Panel B uses *Rater_Performance* as the measure of performance. *Rater_Performance* is the score rewarded by an independent rater for all participants in all conditions and is on a scale from 1 (weak) to 7 (excellent).

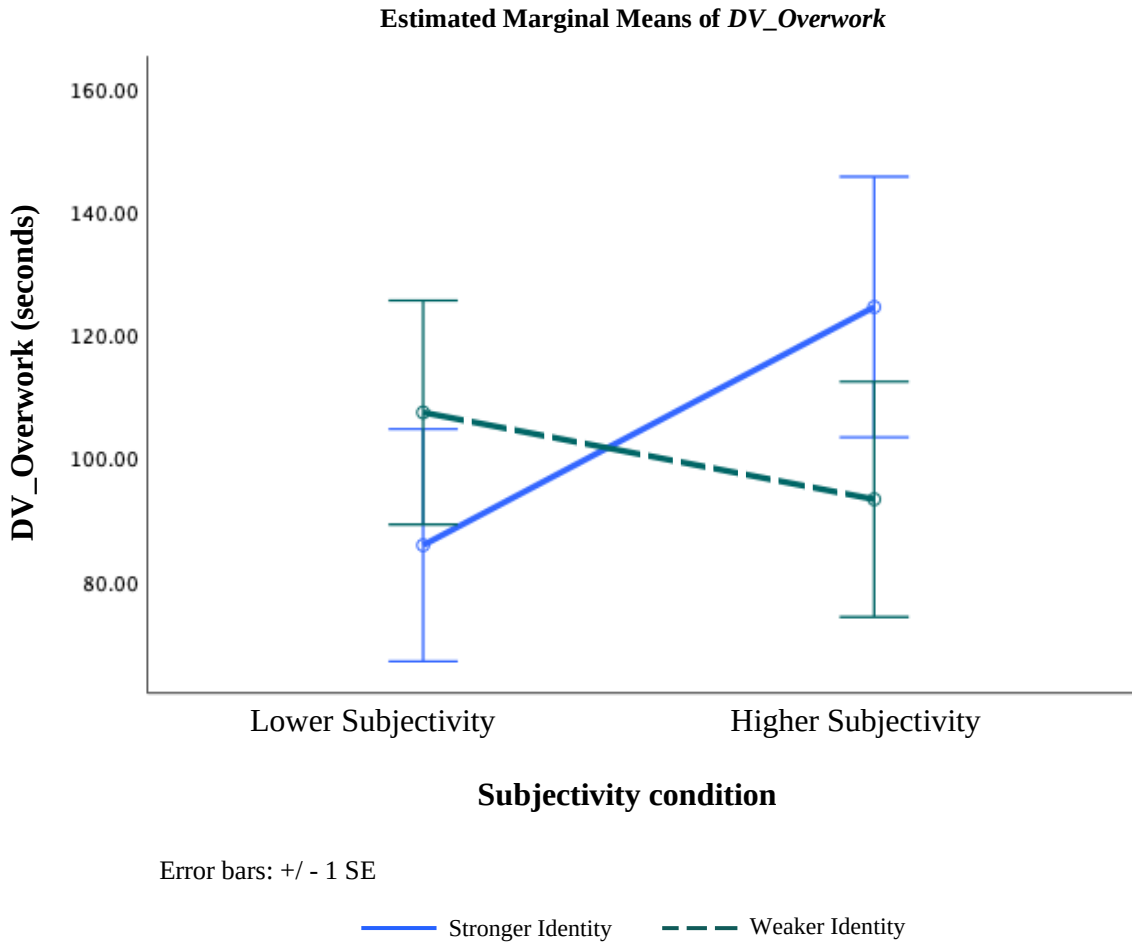
FIGURE 7

DV_Overwork *for* High *and* Low Effort Heuristic *Participants* [a]

Panel A: *High Effort Heuristic* Participants



Panel B: *Low Effort Heuristic* Participants

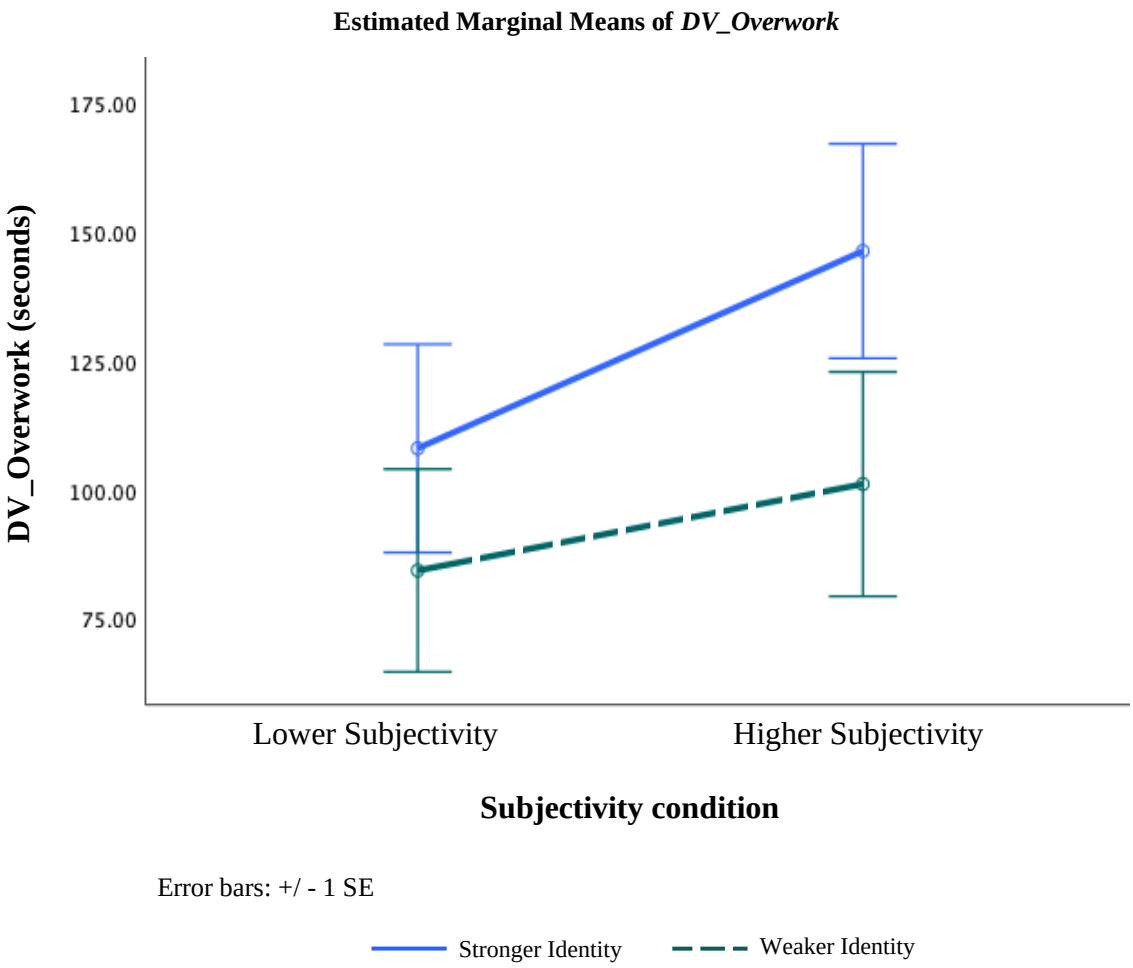


[a] Figure 7 represents the relationship between subjectivity in performance evaluation, group identity, and *DV_Overwork* as found in my experiment, for *High* and *Low Effort Heuristic* participants. See Figure 5 for descriptions of independent and dependent variables. I perform a median split on my effort heuristic measure ("I believe I will receive a higher evaluation if the evaluator knows I spent more time on my section of the report"), such that the *High Effort Heuristic* and the *Low Effort Heuristic* groups have a mean (standard deviation) of 4.77 (0.91) and 2.28 (0.70).

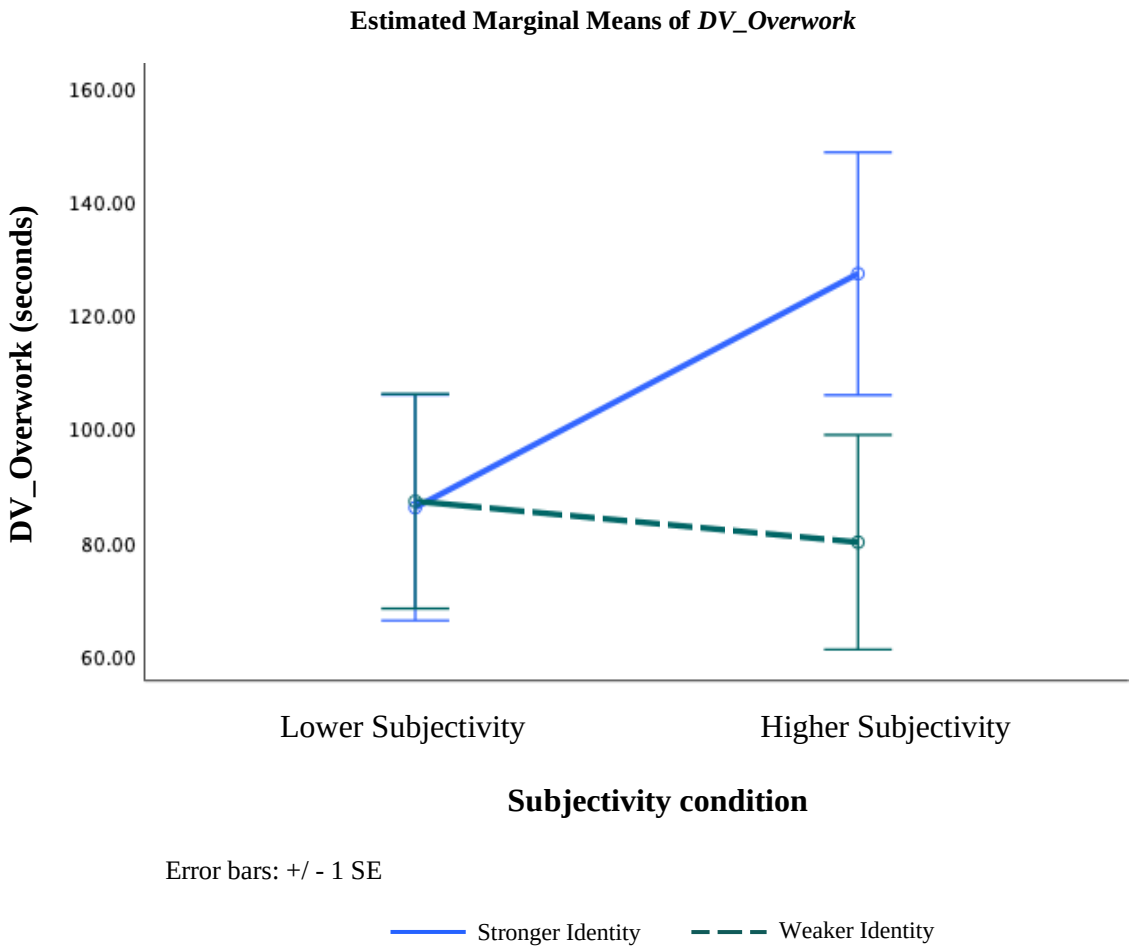
Figure 8

DV_Overwork *for* High *and* Low Trait Social Comparison *Participants* [a]

Panel A: *High Trait Social Comparison* Participants



Panel B: *Low Trait Social Comparison* Participants



[a] Figure 8 represents the relationship between subjectivity in performance evaluation, group identity, and *DV_Overwork* as found in my experiment, for *High* and *Low Trait Social Comparison* participants. See Figure 5 for descriptions of independent and dependent variables. I create an *Average Trait Social Comparison* measure consisting of two trait social comparison questions ("If I want to find out how well I have done something, I compare what I have done with how others have done" and "I always like to know what others in a similar situation would do" (Gibbons & Buunk, 1999)). I median split this measure, such that the *High Trait Social Comparison* and the *Low Trait Social Comparison* groups have a mean (standard deviation) of 5.50 (0.59) and 3.34 (1.01), respectively.

APPENDIX B: SUPPLEMENTARY TABLES

TABLE 1. Pilot Study: Group Identity Manipulation

Panel A: Mean and Standard Deviation by Condition [a]

Condition		*Group_Identity_ Manipulation_ Check*	*IOS Overlap*	*Social Comparison*
Stronger Identity	Mean	6.56	5.72	5.05
	n	90	90	90
	Std. Deviation	0.63	1.27	1.45
Weaker Identity	Mean	2.39	2.11	4.42
	n	89	89	89
	Std. Deviation	1.35	1.28	1.59

Panel B: Independent Samples T-test

Contrast: *Stronger Identity* vs *Weaker Identity*

Dependent Variable	t	df	Sig. (two-tailed)	Mean Difference	Std. Error Difference
Group_Identity_ Manipulation_Check [*]	26.49	124.29	<0.001	4.17	0.16
IOS Overlap	18.90	177	<0.001	3.61	0.19
Social Comparison	2.77	177	0.006	0.63	0.23

[a] See Figure 5 for the description of my group identity manipulation. All dependent variables are on 7-point scales. *Group_Identity_Manipulation_Check* and *Social Comparison* have neutral midpoints and are anchored at 1 and 7. *IOS Overlap* is anchored at 1 and 7. *Group_Identity_Manipulation_Check* is the average of my three identity variables of *Happy* (the extent participants would be happy to be part of the team), *Belong* (the extent participants would feel like a member of the team), and *Like* (the extent participants would like their team members; adapted from Kelly & Presslee, 2017). *IOS Overlap* is the Inclusion of Other in the Self Scale (Aron et al., 1992): I show participants seven images of two interconnecting circles, one representing the participant and one the team described in the vignette: the first image is coded 1 (no overlap between circles / weakest group identity) and the last coded 7 (near-complete overlap / strongest group identity). *Social Comparison* is the average of my two social comparison variables: *Think About* (the extent participants would think about their performance on team-related work compared to their team members' performance) and *Concerned About* (the extent participants would be concerned about their performance compared to their team members' performance on team-related work; adapted from Tafkov, 2013).

[*] Levene's test was significant for the *Group_Identity_Manipulation_Check* measure ($p < 0.001$), indicating a violation of the homogeneity of variance assumption. Thus, equal variances are not assumed for the t-test.

TABLE 2. Pilot Study: Subjectivity in Performance Evaluation Manipulation [a]

Panel A: Mean and Standard Deviation by Condition[*]

Condition		***Subjective***	***Objective***	***Certain***	***Effort Heuristic***	***Control***
Higher Subjectivity	Mean	5.92	3.40	4.49	4.59	3.94
	n	88	88	88	88	88
	Std. Dev.	1.35	2.04	1.56	1.51	1.47
Lower Subjectivity	Mean	3.01	5.99	5.76	4.12	5.08
	n	86	86	86	86	86
	Std. Dev.	2.12	1.25	0.99	1.74	1.41

Panel B: Independent Samples Test – *Higher Subjectivity* vs *Lower Subjectivity* Condition

	T-test for Equality of Means[*]				
	t	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**
Subjective	10.76	143.53	<0.001	2.91	0.27
Objective	-10.13	144.98	<0.001	-2.59	0.26
Certain	-6.40	148.02	<0.001	-1.27	0.20
Effort Heuristic	1.92	167.42	0.056	0.48	0.25
Control	-5.21	172	<0.001	-1.14	0.22

[a] See Figure 5 for a description of my subjectivity manipulation. All dependent variables are on 7-point scales from *Strongly Disagree* (1) to *Strongly Agree* (7). *Subjective* is a measure of whether participants feel like their evaluation is subjective; *Objective* is a measure of whether participants feel like their evaluation is objective; *Certain* is a measure of whether participants feel certain about how they can get a good evaluation; *Effort Heuristic* is a measure of whether participants feel like their writing will be evaluated better if their evaluator knows they spent more time on it; *Control* is a measure of whether participants feel like they can control how they are evaluated.

[*] Levene's test for equality of variances is significant for the first 4 variables ($p < 0.05$). Thus, equal variances are not assumed for the t-tests of these variables.

TABLE 3. Main Experiment: Demographic Information about Participants

Age

n	Minimum	Maximum	Mean	Std. Deviation
290	19	76	41.54	12.27

Highest Education Level

	Frequency	Percent
Some High School	1	0.34
Completed High School (or Equivalent)	25	8.62
Completed Trade or Professional School	8	2.76
Some College/ University	61	21.03
Completed College/ University	134	46.21
Some Graduate School	6	2.07
Completed Graduate School	55	18.97
Total	290	100.00

Language in Which Feel Most Confident Writing ("What is the language you feel most confident writing in?")

	Frequency	Percent
English	289	99.66
Other	1	0.34
Total	290	100.00

Highest Education Level for Writing ("What is the highest education level for which you have written essays in English?")

	Frequency	Percent
For High School Classes	50	17.24
For Undergraduate Classes	182	62.76
For Graduate Classes	52	17.93
None of Above	6	2.07
Total	290	100.00

TABLE 4. Main Experiment: Subjectivity Manipulation Check [a]

Panel A: Descriptives for Subjectivity Manipulation Check

Dependent Variable: *Subjective*

Condition	n	Mean	Std. Deviation
Higher Subjectivity	138	6.53	0.82
Lower Subjectivity	152	2.32	2.01

Panel B: Independent Samples Test: T-test for Equality of Means

	t	df	Sig. (two-tailed)	Mean Difference	Std. Error Difference
Subjective[*]	23.74	203.96	< 0.001	4.21	0.18

[a] This is my manipulation check for my subjectivity manipulation. I asked participants whether they feel like their evaluation is subjective, based on the evaluator's personal views and opinions on each section of the report (*Subjective*). *Subjective* should be higher in the *Higher Subjectivity* condition. See Figure 5 for a description of my subjectivity manipulation.

[*] Levene's test for equality of variances is significant ($p < .001$). Thus, equal variances are not assumed for the t-test of this variable.

TABLE 5. Main experiment: Group Identity Manipulation Check Variables: Principal Components Analysis

Group_Identity_Manipulation_Check **Items.**[a]

Item*	***Factor Loadings***
1. Happy: To what extent would you be happy to be a part of this team?	0.99
2. Belong: To what extent would you feel like a member of your team?	0.98
3. Like: To what extent would you like your team members?	0.99
Eigenvalue [b]	2.91
% of Variance explained	96.90

[a] *Group_Identity_Manipulation_Check* is the average of all three items in the table. All items adapted from Kelly & Presslee (2017). Responses are collected using a 7-point Likert scale with a range from (1) 'not at all' to (7) 'a great extent' with an unlabeled midpoint.

[b] This is the only component with an eigenvalue larger than 1.

Table 6. Main Experiment: Group Identity Manipulation Checks [a]

Panel A: Descriptives for Group Identity Manipulation Checks

	Condition	n	Mean	Std. Deviation
Group_Identity_ Manipulation_Check [b]	*Stronger Identity*	139	6.59	0.61
	Weaker Identity	151	2.14	1.22
IOS Overlap [c]	*Stronger Identity*	139	5.76	1.16
	Weaker Identity	151	1.80	1.06

Panel B: Independent Samples Test: T-test for Equality of Means

	t	df	Sig. (two-tailed)	Mean Difference	Std. Error Difference
Group_Identity_ Manipulation_Check[*]	39.72	224.89	< 0.001	4.45	0.11
IOS Overlap	30.35	288	< 0.001	3.96	0.13

[a] See Figure 5 for a description of my group identity manipulation.

[b] I average my three identity variables of *Happy*, *Belong*, and *Like* and obtain a single measure for *Group_Identity_Manipulation_Check*. Principal components analysis is shown in Table 5. As with the original component items, *Group_Identity_Manipulation_Check* is anchored at (1) and (7).

[c] This measure uses the Aron et al. (1992) Inclusion of Other in the Self Scale (*IOS Overlap* measure), a validated measure of identity. I show participants seven images of two interconnecting circles, one representing the participant and one the team described in the vignette: the first image is coded 1 (no overlap between circles / weakest group identity) and the last coded 7 (near-complete overlap / strongest group identity).

[*] Levene's test for equality of variances is significant ($p < 0.001$). Thus, equal variances are not assumed for the t-tests of this variable.

TABLE 7. Main Experiment: Descriptives [a]

Panel A: Total Number of Participants [b]

Group Identity Condition	Subjectivity Condition	Dependent variable		
		DV_Overwork	***Grammarly_Performance***	***Independent_Rating_Performance***
Stronger Identity	*Higher Subjectivity*	66	65	66
	Lower Subjectivity	73	73	73
Weaker Identity	*Higher Subjectivity*	72	71	72
	Lower Subjectivity	79	77	79

Panel B: Time spent on writing task above contracted time (*DV_Overwork,* in seconds)

Group Identity Condition	Subjectivity Condition	Mean	Std. Deviation
Stronger Identity	*Higher Subjectivity*	137.05	145.27
	Lower Subjectivity	96.91	128.04
Weaker Identity	*Higher Subjectivity*	89.08	109.55
	Lower Subjectivity	85.83	97.85

Panel C: Grammarly Score on Writing Task (*Grammarly_Performance,* score out of 100)

Group Identity Condition	Subjectivity Condition	Mean	Std. Deviation
Stronger Identity	*Higher Subjectivity*	74.26	14.78
	Lower Subjectivity	74.67	13.82
Weaker Identity	*Higher Subjectivity*	73.46	15.62
	Lower Subjectivity	78.40	14.89

Panel D: Independent Rater Score on Writing Task (*Rater_Performance,* score from 1 to 7)

Group Identity Condition	Subjectivity Condition	Mean	Std. Deviation
Stronger Identity	*Higher Subjectivity*	3.67	1.30
	Lower Subjectivity	3.53	1.13
Weaker Identity	*Higher Subjectivity*	3.72	1.28
	Lower Subjectivity	3.73	1.28

[a] See Figures 5 and 6 for descriptions of my independent and dependent variables.

[b] There are fewer participants for my dependent variable of *Grammarly_Performance*, as four participant responses were too short for Grammarly to output a score.

TABLE 8. Main Experiment: H1, H2, and H3 – Time spent on Writing Task above Contracted Time (*DV_Overwork*) [a]

Panel A: Conventional ANOVA Results – Tests of H1 and H3

Source	df	Mean Square	F	Sig.[b]
Subjectivity	1	33988.84	2.34	0.064
Group Identity	1	62943.68	4.34	0.019
*Subjectivity * Group Identity Interaction*	1	24565.03	1.69	0.097
Error	286	14514.69		

Panel B: Simple effect of *Subjectivity* on *DV_Overwork*

Source	df	Mean Square	F	Sig. [b]
Higher vs *Lower Subjectivity* in *Stronger Identity* condition	1	55850.35	3.00	0.043
Higher vs *Lower Subjectivity* in *Weaker Identity* condition	1	398.22	0.04	0.424

Panel C: Simple effect of *Group Identity* on *DV_Overwork*

Source	df	Mean Square	F	Sig. [b]
Stronger vs *Weaker Identity* in *Higher Subjectivity* condition	1	79237.85	4.85	0.015
Stronger vs *Weaker Identity* in *Lower Subjectivity* condition	1	4658.08	0.36	0.274

Panel D: Planned contrast – Test of H2

Time spent on the writing task will be highest in the *Higher Subjectivity/ Stronger Identity* condition, and lower in the other three conditions (i.e., the *Higher Subjectivity/ Weaker Identity* condition, the *Lower Subjectivity/ Stronger Identity* and the *Lower Subjectivity/ Weaker Identity* condition). Contrast weights are +3, -1, -1, -1, respectively.

Source of variance	Sums of Squares	Degrees of Freedom	Mean Square	F	Sig. [b]
Contrast	109922.55	1	109922.55	7.57	0.003
Residual	5483.70	2	2741.85	0.19	0.828
Total between-cells variance	115406.25	3	38468.75	2.65	0.049
Error	4151201.12	286	14514.69		
Total	4266607.37	289			
Contrast Variance Residual, q2	0.048				

Panel E: Planned contrast – Robustness Test of H2

For robustness, I test H2 using a more specified contrast that incorporates my predicted main effects – the main effect of *Subjectivity* and the main effect of *Group Identity* – within my predicted ordinal interaction effect. Thus, with *DV_Overwork* as my dependent variable, I use contrast weights of +4 [*Higher Subjectivity/ Stronger Identity*], -2 [*Higher Subjectivity/ Weaker Identity*], +1 [*Lower Subjectivity/ Stronger Identity*] and -3 *[Lower Subjectivity/ Weaker Identity*].

Source of variance	**Sums of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Sig.** [b]
Contrast	103087.70	1	103087.70	7.10	0.004
Residual	12318.55	2	6159.28	0.42	0.655
Total between-cells variance	115406.25	3	38468.75	2.65	0.049
Error	4151201.12	286	14514.69		
Total	4266607.37	289			
Contrast Variance Residual, q2	0.107				

[a] See Figure 5 for descriptions of my independent and dependent variables.

[b] p-values are one-tailed where there are directional predictions (i.e., the main effect of *Subjectivity*, the ANOVA interaction effect, the main effect of *Group Identity*, the contrast effect, and the simple effects).

TABLE 9. Main Experiment: RQ1 – Performance on Writing Task as Measured by Grammarly Writing Software (*Grammarly_Performance*) and Independent Rater (*Rater_Performance*) [a]

Panel A: Conventional ANOVA Results

Dependent Variable: *Grammarly_Performance*

Source	**df**	**Mean Square**	**F**	**Sig.**[b]
Subjectivity	1	509.23	2.33	0.128
Group Identity	1	153.36	0.70	0.403
*Subjectivity * Group Identity Interaction*	1	365.13	1.67	0.197
Error	282	218.67		

Panel B: Conventional ANOVA Results

Dependent Variable: *Rater_Performance*

Source	**df**	**Mean Square**	**F**	**Sig.**[b]
Subjectivity	1	0.26	0.17	0.682
Group Identity	1	1.18	0.75	0.386
*Subjectivity * Group Identity Interaction*	1	0.38	0.24	0.624
Error	286	1.56		

Panel C: Correlations

		Grammarly_ Performance	*Rater_ Performance*	*DV_Overwork*
Grammarly_ Performance	Pearson Correlation	1	0.47	-0.12
	Sig. (2-tailed)		< 0.001	0.050
	n	286	286	286
Rater_ Performance	Pearson Correlation	0.47	1	0.02
	Sig. (2-tailed)	< 0.001		0.754
	n	286	290	290
DV_Overwork	Pearson Correlation	-0.12	0.02	1
	Sig. (2-tailed)	0.050	0.754	
	n	286	290	290

[a] See Figures 5 and 6 for descriptions of my independent and dependent variables.

[b] p-values are two-tailed as there are no directional predictions.

TABLE 10. Main Experiment: Supplemental Analysis – *Effort Heuristic*

Panel A: Conventional ANOVA Results

Dependent Variable: *Effort Heuristic*

Source	**df**	**Mean Square**	**F**	**Sig.**
Subjectivity	1	0.48	0.22	0.640
Group Identity	1	4.71	2.14	0.144
*Subjectivity * Group Identity Interaction*	1	2.02	0.92	0.338
Error	286	2.20		

Panel B: Descriptive Statistics for *High* and *Low Effort Heuristic* Groups

Dependent Variable: *Effort Heuristic*

Effort Heuristic Group	**n**	**Mean**	**Std. Deviation**
High Effort Heuristic	140	4.77	0.91
Low Effort Heuristic	150	2.28	0.70

Panel C: Independent Samples Test – *High Effort Heuristic* vs *Low Effort Heuristic* Participants

Dependent variable	**t**	**df**	**Sig. (two-tailed)**	**Mean Difference**	**Std. Error Difference**
Effort Heuristic[*]	26.08	260.25	< 0.001	2.49	0.10

Panel D: *DV_Overwork* for *High* versus *Low Effort Heuristic* Participants [a]

	Group Identity Condition	Subjectivity Condition	n	Mean	Std. Deviation
High Effort Heuristic Participants	*Stronger Identity*	*Higher Subjectivity*	35	148.18	167.18
		Lower Subjectivity	34	109.57	129.75
	Weaker Identity	*Higher Subjectivity*	34	84.34	110.16
		Lower Subjectivity	37	61.35	67.32
Low Effort Heuristic Participants	*Stronger Identity*	*Higher Subjectivity*	31	124.48	117.22
		Lower Subjectivity	39	85.87	127.18
	Weaker Identity	*Higher Subjectivity*	38	93.31	110.31
		Lower Subjectivity	42	107.39	114.98

Panel E: Conventional ANOVA Results

Dependent Variable: *DV_Overwork*

	Source	df	Mean Square	F	Sig. [b]
High Effort Heuristic Participants	*Subjectivity*	1	33163.29	2.18	0.071
	Group Identity	1	109731.32	7.21	0.008
	*Subjectivity * Group Identity Interaction*	1	2130.62	0.14	0.355
	Error	136	15216.799		
Low Effort Heuristic Participants	*Subjectivity*	1	5574.75	0.40	0.526
	Group Identity	1	861.58	0.06	0.803
	*Subjectivity * Group Identity Interaction*	1	25700.18	1.86	0.175
	Error	146	13829.78		

Panel F: Planned contrast - *High* versus *Low Effort Heuristic* Participants

Time spent on the writing task will be highest in the *Higher Subjectivity/ Stronger Identity* condition, and lower in the other three conditions (i.e., the *Higher Subjectivity/ Weaker Identity* condition, the *Lower Subjectivity/ Stronger Identity* and the *Lower Subjectivity/ Weaker Identity* condition). Contrast weights are +3, -1, -1, -1, respectively.

High Effort Heuristic Participants	**Source of variance**	**Sums of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Sig.[b]**
	Contrast	104438.54	1	104438.54	6.86	0.005
	Residual	43496.78	2	21748.39	1.43	0.243
	Total between-cells variance	147935.32	3	49311.78	3.24	0.024
	Error	2069483.33	136	15216.79		
	Total	2217418.653	139			
	Contrast Variance Residual, q^2	0.29				

Low Effort Heuristic Participants	**Source of variance**	**Sums of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Sig.[b]**
	Contrast	20617.09	1	20617.09	1.49	0.224
	Residual	9280.50	2	4640.25	0.34	0.716
	Total between-cells variance	29897.59	3	9965.86	0.72	0.541
	Error	2019147.72	146	13829.78		
	Total	2049045.314	149			
	Contrast Variance Residual, q^2	0.31				

Panel G: Simple effect of *Subjectivity* on *DV_Overwork* for *High Effort Heuristic* Participants

Source	df	Mean Square	F	Sig.
Higher vs *Lower Subjectivity* in *Stronger Identity* condition	1	25705.86	1.14	0.289
Higher vs *Lower Subjectivity* in *Weaker Identity* condition	1	9367.53	1.15	0.288

[*] Levene's test for equality of variances is significant ($p < 0.001$). Thus, equal variances are not assumed for the t-tests of this variable.

[a] I perform a median split on my *Effort Heuristic* measure ("I believe I will receive a higher evaluation if the evaluator knows I spent more time on my section of the report"), such that the *High Effort Heuristic* and the *Low Effort Heuristic* groups have a mean (standard deviation) of 4.77 (0.91) and 2.28 (0.70). See Figure 5 for descriptions of independent and dependent variables.

[b] p-values are one-tailed where there are directional predictions (i.e., for the *High Effort Heuristic* participants, the main effect of *Subjectivity*, the ANOVA interaction effects, and the contrast effect).

TABLE 11. Main Experiment: Supplemental Analysis – *Average Trait Social Comparison*

Panel A: Conventional ANOVA

Dependent Variable: *Average Trait Social Comparison*				
Source	**df**	**Mean Square**	**F**	**Sig.**
Subjectivity	1	3.27	1.75	0.187
Group Identity	1	2.41	1.29	0.257
*Subjectivity * Group Identity Interaction*	1	0.38	0.20	0.653
Error	286	1.87		

Panel B: Descriptive Statistics for *High* and *Low Trait Social Comparison* Participants

Dependent Variable: *Average Trait Social Comparison*		
Trait Social Comparison Group	**Mean**	**Std. Dev**
High Trait Social Comparison	5.50	0.59
Low Trait Social Comparison	3.34	1.01

Panel C: Independent Samples Test – *High* vs *Low Trait Social Comparison* Participants

Dependent variable	**t**	**df**	**Sig. (two-tailed)**	**Mean Difference**	**Std. Error Difference**
Average Trait Social Comparison *	22.44	243.86	< 0.001	2.16	0.100

Panel D: Time spent on writing task above contracted time (*DV_Overwork*) [a]

	Group Identity Condition	Subjectivity Condition	n	Mean	Std. Deviation
High Trait Social Comparison Participants	*Stronger Identity*	*Higher Subjectivity*	34	146.30	131.51
		Lower Subjectivity	36	108.06	137.34
	Weaker Identity	*Higher Subjectivity*	31	101.12	117.11
		Lower Subjectivity	38	84.37	95.62
Low Trait Social Comparison Participants	*Stronger Identity*	*Higher Subjectivity*	32	127.21	160.15
		Lower Subjectivity	37	86.06	119.19
	Weaker Identity	*Higher Subjectivity*	41	79.97	104.01
		Lower Subjectivity	41	87.17	101.05

Panel E: ANOVA - *High* versus *Low Trait Social Comparison* Individuals

Dependent Variable: *DV_Overwork*

	Source	df	Mean Square	F	Sig.
High Trait Social Comparison Participants	*Subjectivity*	1	26128.52	1.78	0.184
	Group Identity	1	40963.88	2.79	0.097
	*Subjectivity * Group Identity Interaction*	1	3991.12	0.27	0.603
	Error	135	14671.35		
Low Trait Social Comparison Participants	*Subjectivity*	1	10767.28	0.74	0.392
	Group Identity	1	19875.31	1.36	0.245
	*Subjectivity * Group Identity Interaction*	1	21845.04	1.50	0.223
	Error	147	14610.04		

Panel F: Planned contrast – *High* versus *Low Trait Social Comparison* Individuals

Time spent on the writing task will be highest in the *Higher Subjectivity/ Stronger Identity* condition, and lower in the other three conditions (i.e., the *Higher Subjectivity/ Weaker Identity* condition, the *Lower Subjectivity/ Stronger Identity* and the *Lower Subjectivity/ Weaker Identity* condition). Contrast weights are +3, -1, -1, -1, respectively.

High Trait Social Comparison Participants	**Source of variance**	**Sums of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Sig.**
	Contrast	60184.36	1	60184.36	4.10	0.023[b]
	Residual	12107.90	2	6053.95	0.41	0.663
	Total between-cells variance	72292.26	3	24097.42	1.64	0.183
	Error	1980631.57	135	14671.35		
	Total	2052923.84	138			
	Contrast Variance Residual, q2	0.17				

Low Trait Social Comparison Participants	**Source of variance**	**Sums of Squares**	**Degrees of Freedom**	**Mean Square**	**F**	**Sig.**
	Contrast	46201.93	1	46201.93	3.16	0.077[c]
	Residual	1364.84	2	682.42	0.05	0.954
	Total between-cells variance	47566.77	3	15855.59	1.09	0.357
	Error	2147676.37	147	14610.04		
	Total	2195243.140	150			
	Contrast Variance Residual, q2	0.03				

Panel G: Simple effect of *Subjectivity* on *DV_Overwork*

	Source	df	Mean Square	F	Sig.
High Trait Social Comparison Participants	*Higher* vs *Lower Subjectivity* in *Stronger Identity* condition	1	25577.54	1.41	0.239
	Higher vs *Lower Subjectivity* in *Weaker Identity* condition	1	4790.68	0.43	0.515
Low Trait Social Comparison Participants	*Higher* vs *Lower Subjectivity* in *Stronger Identity* condition	1	29064.60	1.49	0.226
	Higher vs *Lower Subjectivity* in *Weaker Identity* condition	1	1063.92	0.10	0.751

[*] Levene's test for equality of variances is significant ($p < 0.001$). Thus, equal variances are not assumed for the t-tests of this variable.

[a] I create an *Average Trait Social Comparison* measure consisting of two trait social comparison questions ("If I want to find out how well I have done something, I compare what I have done with how others have done" and "I always like to know what others in a similar situation would do" (Gibbons & Buunk, 1999)). I median split this measure, such that the *High Trait Social Comparison* and the *Low Trait Social Comparison* groups have a mean (standard deviation) of 5.50 (0.59) and 3.34 (1.01), respectively. See Figure 5 for descriptions of independent and dependent variables.

[b] *one-tailed*

[c] *two-tailed*

TABLE 12. Main experiment: Other Measured Variables [a]

Panel A: Descriptive Statistics for *Difficult* and *Certain*

Dependent variable	Subjectivity Condition	n	Mean	Std. Deviation
Difficult	*Higher Subjectivity*	138	3.72	1.56
	Lower Subjectivity	152	3.11	1.55
Certain	*Higher Subjectivity*	138	4.00	1.41
	Lower Subjectivity	152	4.74	1.29

Panel B: Independent Samples T-test – *Higher Subjectivity* vs *Lower Subjectivity* conditions

Dependent variable	t	df	Sig. (two-sided)	Mean Difference
Difficult	3.35	288	< 0.001	0.61
Certain	-4.65	288	< 0.001	-0.74

Panel C: Descriptive Statistics for *State_Social_Comparison*

Subjectivity Condition	n	Mean	Std. Deviation
Stronger Identity	139	3.71	1.80
Weaker Identity	151	3.66	1.79

Panel D: Independent Samples Test – *Stronger Identity* vs *Weaker Identity* Conditions

Dependent variable	t	df	Sig. (two-sided)	Mean Difference	Std. Error Difference
State_Social_Comparison	0.23	288	0.815	0.05	0.21

Panel E: Descriptive Statistics

Dependent variable	n	Mean	Std. Deviation
Motivated	290	6.04	0.90
Important	290	5.84	1.21
Understood	290	6.49	0.67

Panel F: MANOVA Results

Source	Dependent Variable	df	Mean Square	F	Sig.
Subjectivity	*Motivated*	1	0.45	0.56	0.453
	Important	1	0.20	0.14	0.709
	Understood	1	0.05	0.12	0.731
Group Identity	*Motivated*	1	3.81	4.78	0.030
	Important	1	4.10	2.81	0.095
	Understood	1	0.09	0.19	0.664
*Subjectivity * Group Identity Interaction*	*Motivated*	1	0.16	0.21	0.650
	Important	1	0.47	0.32	0.572
	Understood	1	0.04	0.09	0.768
Error	*Motivated*	286	0.80		
	Important	286	1.46		
	Understood	286	0.46		

Panel G: Descriptive Statistics

Dependent Variable: *Ambiguity Tolerance*			
Group Identity Condition	**Subjectivity Condition**	**Mean**	**Std. Deviation**
Stronger Identity	*Higher Subjectivity*	3.92	1.62
	Lower Subjectivity	3.89	1.74
Weaker Identity	*Higher Subjectivity*	3.92	1.46
	Lower Subjectivity	3.84	1.55

Panel H: Conventional ANOVA Results

Dependent Variable: *Ambiguity Tolerance*				
Source	**df**	**Mean Square**	**F**	**Sig.**
Subjectivity	1	0.24	0.09	0.759
Group Identity	1	0.07	0.03	0.868
*Subjectivity * Group Identity Interaction*	1	0.04	0.02	0.900
Error	286	2.54		

[a] I ask participants to respond to additional questions in the post-experiment questions. I ask two questions related to participants' perceptions of their writing task evaluations: "In this writing task, I believe it is difficult for the evaluator to evaluate writing quality" (*Difficult*) and "I feel certain about how I can get a good evaluation on my section of the report" (*Certain*). I ask a social comparison question: "While I was completing the writing task, I thought about how my performance on the writing task compared to my team members' performance" (*State_Social_Comparison*). I ask questions related to participants' perceptions of the task: "I was motivated to do well on this task" (*Motivated*), "It was important for me to be evaluated well on this task" (*Important*), and "I understood the task instructions" (*Understood*). I ask the tolerance for ambiguity trait measure (*Ambiguity Tolerance*) from Hartmann and Slapnicar (2012): "I do not like to work on a problem unless there is a possibility of finding a clear-cut and unambiguous solution". All these measures are on a scale from (1) to (7), where (1) is *Strongly Disagree* and (7) is *Strongly Agree*. See Figure 5 for descriptions of independent variables.

TABLE 13. Secondary Experiment: Demographic Information about Participants

Age [a]

n	Minimum	Maximum	Mean	Std. Deviation
180	20	74	39.89	11.22

Highest education level

	Frequency	Percent
Some High School	2	1.10
Completed High School (or Equivalent)	9	4.97
Completed Trade or Professional School	5	2.76
Some College/ University	42	23.20
Completed College/ University	89	49.17
Some Graduate School	7	3.87
Completed Graduate School	27	14.92
Total	181	100.00

Language in which feel most confident writing ("What is the language you feel most confident writing in?")

	Frequency	Percent
English	181	100.00
Other	0	0
Total	181	

Team Experience ("Do you have previous work experience working as part of a team?")

	Frequency	Percent
Yes	168	92.82
No	13	7.18
Total	181	100.00

[a] One participant typed an age of 2000 as their response and was removed from the age analysis.

TABLE 14. Secondary Experiment: Manipulation Checks [a]

Panel A: Descriptives for Subjectivity Manipulation Checks

Dependent variable	Subjectivity condition	n	Mean	Std. Deviation
Subjective	*Higher Subjectivity*	90	5.76	1.34
	Lower Subjectivity	91	2.41	1.81
Objective	*Higher Subjectivity*	90	3.91	1.82
	Lower Subjectivity	91	6.18	1.05

Panel B: Independent Samples Test: T-test for Equality of Means

Dependent variable	t	df	Sig. (two-sided)	Mean Difference	Std. Error Difference
Subjective[*]	14.13	165.93	< 0.001	3.35	0.24
Objective[*]	-10.23	141.93	< 0.001	-2.27	0.22

[a] I ask participants whether they feel like their evaluation is subjective, based on the evaluator's personal views and opinions on each section of the report (*Subjective*). I also ask participants whether they feel their section of the report will be evaluated objectively, based on objective writing criteria such as number of spelling mistakes, grammar and punctuation mistakes in the section of the report being evaluated (*Objective*). See Figure 5 for description of the subjectivity manipulation.

[*] Levene's test for equality of variances is significant ($p < 0.001$). Thus, equal variances are not assumed for the t-tests of this variable.

TABLE 15. Secondary Experiment: Principal Components Analysis – Effort Heuristic Variables [a]

***Average_Effort_Heuristic* Items** [b]

Item*	**Factor Loadings**
1. I believe the evaluator will evaluate my writing better if they know I spent more time on it	0.94
2. I believe the evaluator will evaluate my writing better if they know I put more effort into it	0.94
Eigenvalue	1.77
% of Variance explained	88.38

[a] Measures use 7-point Likert scales from *Strongly Disagree* (1) to *Strongly Agree* (7) with all points labelled.

[b] *Average_Effort_Heuristic* is the average of the two items in the table.

TABLE 16. Secondary Experiment: Difference between Conditions

Panel A: Mean and Standard Error by Condition [a]

Dependent variable	Condition	n	Mean	Std. Deviation
Average_Effort_Heuristic	*Higher Subjectivity*	90	4.90	1.22
	Lower Subjectivity	91	3.99	1.67
Certain	*Higher Subjectivity*	90	4.56	1.43
	Lower Subjectivity	91	5.85	1.04

Panel B: Independent Samples T-Test – *Higher Subjectivity* vs *Lower Subjectivity* conditions

Dependent variable	t	df	Sig. (two-sided)	Mean Difference	Std. Error Difference
Average_Effort_Heuristic[*]	4.16	164.80	< 0.001	0.91	0.22
Certain[*]	-6.93	162.66	< 0.001	-1.29	0.19

[a] I manipulate subjectivity in performance evaluation between conditions (see Figure 5 for description of subjectivity manipulation), and my dependent variables are two measures of effort heuristic ("I believe the evaluator will evaluate my writing better if they know I spent more time on it"; "I believe the evaluator will evaluate my writing better if they know I put more effort into it") and a measure of *Certainty* ("I feel certain about **how** I can get a good evaluation on my section of the report"). Both measures use 7-point Likert scales from *Strongly Disagree* (1) to *Strongly Agree* (7) with all points labelled. I average my two effort heuristic variables and obtain a single measure for *Average_Effort_Heuristic*.

[*] Levene's test for equality of variances is significant ($p < .001$). Thus, equal variances are not assumed for the t-tests of these variables.

TABLE 17. Secondary Experiment: Other Measured Variables

Panel A: Descriptives for Other Measured Variables [a]

Dependent variable	Subjectivity condition	n	Mean	Std. Deviation
Work_Hard	*Higher Subjectivity*	90	5.01	1.44
	Lower Subjectivity	91	5.33	1.39
Leisure_Time	*Higher Subjectivity*	89	5.39	1.19
	Lower Subjectivity	91	5.30	1.27
Control	*Higher Subjectivity*	90	4.29	1.51
	Lower Subjectivity	91	5.25	1.41

Panel B: Independent Samples Test: T-test for Equality of Means

Dependent variable	t	df	Sig. (two-sided)	Mean Difference	Std. Error Difference
Work_Hard	-1.51	179	0.132	-0.32	0.21
Leisure_Time	0.53	178	0.600	0.10	0.18
Control	-4.44	179	< 0.001	-0.96	0.22

[a] I ask participants to answer two trait measures capturing the Protestant Work Ethic to ensure that differences in the use of the effort heuristic between conditions is not due to individual trait differences (PWE; Cheng et al., 2017). Both measures use 7-point Likert scales from *Strongly Disagree* (1) to *Strongly Agree* (7) with all points labelled and are as follows: "Any person who is able and willing to work hard has a good chance of succeeding" (*Work_Hard*), and "Life would be more meaningful if we had more leisure time" (*Leisure_Time*). I also ask participants about their perception of how much they can control their evaluation on their section of the report ("I feel like I can control how I am evaluated on my section of the report"; *Control*). See Figure 5 for the description of the subjectivity manipulation.

APPENDIX C: MAIN EXPERIMENT - EXPERIMENTAL PROCEDURES

1. Participants consent to participate in the study and answer a reCAPTCHA question.
2. Participants are told that they will be asked to complete a writing task as part of the study, and that their writing output evaluation (Outstanding, Acceptable or Poor) will be sent to them after they complete the study. Participants are told they may find evaluation useful as writing tasks provide an indication of analytical and critical thinking skills.
3. Participants are told to imagine they are in the scenario to be outlined on screens that follow while they complete writing task.
4. Participants are randomly assigned to the *Stronger* or *Weaker Identity* condition and provided a vignette with a description of a hypothetical team for each condition.
5. Participants are asked to briefly describe what they think it would be like to work with the team described in the vignette.
6. Participants answer group identity manipulation check questions.
7. Participants are told that they and their hypothetical team members will be working on a report for a client and that each team member will be working on a section of the report, and each section will be evaluated separately. They are told they will receive fixed payment for their work.
8. Participants are randomly assigned to the *Higher* or *Lower Subjectivity* condition and provided a vignette describing the evaluation method for each condition.
9. Participants answer a subjectivity manipulation check question.
10. Participants are told that they are contracted to work on their section of the report for 5 minutes. They are informed that they can choose to work longer than 5 minutes on their task.
11. Participants answer 3 comprehension check questions to ensure understanding of the task instructions.
12. Participants are taken to a screen with the writing task and are provided with a writing prompt. The same screen includes a timer showing participants how long they have been working, and a box showing them whether hypothetical team members are still working on task.
13. Participants can submit the writing task after 5 minutes on the screen.
14. Participants complete post-experiment questions and demographic questions.
15. Participants receive a feedback and appreciation letter with a debriefing statement.
16. Participants are provided a random identification number to insert into Amazon Mechanical Turk to get paid.

APPENDIX D: SECONDARY EXPERIMENT - EXPERIMENTAL PROCEDURES

1. Participants consent to participate in the study and answer a reCAPTCHA question.
2. Participants are told that they and their hypothetical team will be working on a report for a client and that each team member will be working on a section of the report, and each section will be evaluated separately.
3. Participants are randomly assigned to the *Higher* or *Lower Subjectivity* condition and provided a vignette describing the evaluation method for each condition.
4. Participants answer dependent variable measures and manipulation check measures.
5. Participants answer a comprehension check question to ensure understanding of the task instructions.
6. Participants answer demographic questions.
7. Participants receive a feedback and appreciation letter with a debriefing statement.
8. Participants are provided a random identification number to insert into Amazon Mechanical Turk to get paid.

APPENDIX E: MAIN EXPERIMENT INSTRUMENT

Writing in square brackets is not shown to participants.

[Consent form displayed]

[Instructions]

You will be asked to complete a writing task as part of this study. The writing task will be on a screen entitled "Instructions for writing your report".

You will be sent an evaluation on your writing in the writing task following submission of the HIT.

The writing evaluation grade that will be sent to you in a few days through Amazon Mechanical Turk will be either **Outstanding**, **Acceptable**, or **Poor**. You may find this evaluation useful, as writing tasks provide an indication of analytical and critical thinking skills.

While completing the writing task, please imagine that you are in the setting and situation outlined on the following screens.

[Stronger Identity Condition]

Assume that you are a member of a work team that you have a lot in common with. Your team members are named Kai, Aly & Fin. You very much like and trust your team members and feel very close to them. Being a part of this team is very important to you and you are very happy to be a part of it.

Before proceeding to the writing task, take a minute to consider what it would be like to work with this team and put down any thoughts that come to mind.

→

[Weaker Identity Condition]

Assume that you are a member of a work team that you do not have a lot in common with. Your team members are named Kai, Aly & Fin. You don't particularly like or trust your team members and you don't feel close to them. Being a part of this team is not very important to you and you are not very happy to be a part of it.

Before proceeding to the writing task, take a minute to consider what it would be like to work with this team and put down any thoughts that come to mind.

→

[Group Identity Manipulation Check Questions]

To what extent would you be happy to be a part of this team?

Not at all 1	2	3	4	5	6	To a great extent 7

To what extent would you feel like a member of your team?

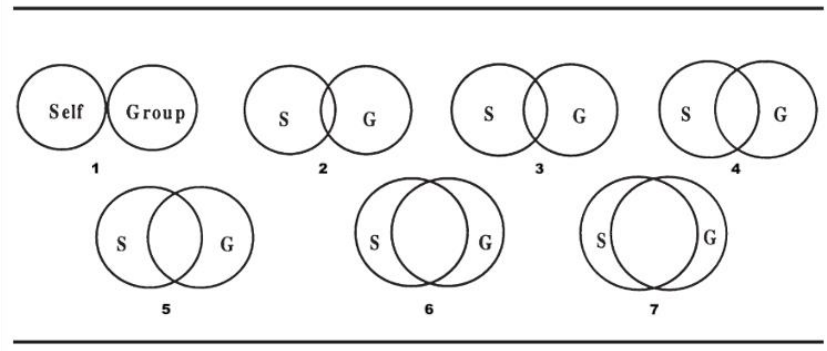
Not at all 1	2	3	4	5	6	To a great extent 7

To what extent would you like your team members?

Not at all 1	2	3	4	5	6	To a great extent 7

→

Choose the pair of circles that you feel best describes your relationship with your team members. S = self, G = group.



→

[Instructions]

(Part 1/3)

Assume that you and your team work for a consulting company that writes reports for clients. The company you work for expects its employees to write very well-written reports. Therefore, it evaluates its employees' reports before it sends them to the client.

Each of you and your team members have been asked to write a section of a report for your client, and each section of the report will then be evaluated by someone from your company. You will each be paid a fixed payment for your work.

All sections of the report are similar in length and difficulty.

The quality of **your section** of the report, which depends only on your own work, will be evaluated. Each section of the report worked on by your team members will also be evaluated separately.

→

[Higher Subjectivity Condition]

(Part 2/3)

The evaluator **will not** be able to form an objective evaluation of writing quality, even though they possess some objective criteria for quality. The objective criteria the evaluator **might refer to** in their evaluation are the number of spelling mistakes, grammar and punctuation mistakes in the report section, and the conciseness and formality of the writing. However, overall, their evaluation **will be subjective and <u>not</u> objective**: it **will** be based on the evaluator's personal views and opinions on each section of the report.

The evaluator will be able to see the sections of the report worked on by you and your team members, and how long you and your team worked on your sections of the report.

Click here to see the information on the previous page

Before proceeding to the next screen, indicate the extent you agree with the following statement:

[Subjectivity Manipulation Check]

"I feel that my section of the report will be evaluated subjectively, based on the evaluator's personal views and opinions on each section of the report."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

[Lower Subjectivity Condition]

(Part 2/3)

The evaluator **will not** be able to form an objective evaluation of writing quality, even though they possess some objective criteria for quality. The objective criteria the evaluator **might refer to** in their evaluation are the number of spelling mistakes, grammar and punctuation mistakes in the report section, and the conciseness and formality of the writing. However, overall, their evaluation **will be subjective and not objective**: it **will** be based on the evaluator's personal views and opinions on each section of the report.

The evaluator will be able to see the sections of the report worked on by you and your team members, and how long you and your team worked on your sections of the report.

Click here to see the information on the previous page

Before proceeding to the next screen, indicate the extent you agree with the following statement:

[Subjectivity Manipulation Check]

"I feel that my section of the report will be evaluated subjectively, based on the evaluator's personal views and opinions on each section of the report."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

→

(Part 3/3)

A section of the report can be completed in 5 minutes.

Therefore, each of you and your team members is contracted to work on each of your sections of the report for 5 minutes. However, each of you and your team members may work for longer than 5 minutes, if you choose to.

You can start to work on writing the report in a couple of screens.

You will have a timer on your screen that will show you how long you have been working on your section of the report.

Click here to see the information on the previous pages

→

[Comprehension Check Questions]

Please respond to the following questions:

In this hypothetical scenario, I am asked to assume that I am a member of a team in a consulting company.

- ○ True
- ○ False

[Correct response: True]

My actual writing evaluation grade (Outstanding, Acceptable, or Poor) will be sent to me in a few days through Amazon Mechanical Turk.

- ○ True
- ○ False

[Correct response: True]

I am contracted to work on my section of the report for 5 minutes, for which I will be paid a fixed payment. However, I may work on my section for longer than 5 minutes, if I choose to.

- ○ True
- ○ False

[Correct response: True]

→

[Writing Task]

Instructions for writing your report:

Your client is creating a marketing campaign that is based on the assumption that people have stopped communicating properly, due to the popularity of social media. Before they run this campaign, they would like your opinion on whether this assumption is true or false. Your client would like to see all the different points of view from all team members before making a decision on their campaign.

In the box below, please provide your opinion on the following statement: "**People communicate with each other less effectively now than in the past because of social media**." Your client would like you to use specific reasons and examples to support your opinion, from your personal or professional experience, without checking the internet.

00:00:07

Here is information about whether your team members are still working:

Team member 1: Kai ■
Team member 2: Aly ■
Team member 3: Fin ■

■ Indicates still working on the writing task
☒ Indicates no longer working on the writing task

→

[Post-experiment Questionnaire]

On this screen and the screens that follow, please indicate the extent you agree with the provided statements:

“I believe I will receive a higher evaluation if the evaluator knows I spent more time on my section of the report.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

→

"In this writing task, I believe it is difficult for the evaluator to evaluate writing quality."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

“I feel certain about **how** I can get a good evaluation on my section of the report.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

→

“While I was completing the writing task, I thought about how my performance on the writing task compared to my team members' performance.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

→

For each of the following, please indicate the extent to which you agree that the provided statement characterizes your thinking while reading the scenario.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
"I was motivated to do well on this task."	○	○	○	○	○	○	○
“It was important for me to be evaluated well on this task.”	○	○	○	○	○	○	○
"I understood the task instructions."	○	○	○	○	○	○	○

→

Below, you will find a series of statements about you. Please read each statement and decide how much you agree or disagree with that statement.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
"I do not like to work on a problem unless there is a possibility of finding a clear-cut and unambiguous solution."	○	○	○	○	○	○	○
"If I want to find out how well I have done something, I compare what I have done with how others have done."	○	○	○	○	○	○	○
"I always like to know what others in a similar situation would do."	○	○	○	○	○	○	○

→

[Demographics]

What is your age in years? ☐

What is the language you feel most confident writing in?

- ○ English
- ○ Other (please specify) _____

What is the highest education level for which you have written essays in English?

- ○ I wrote essays in English for high school classes
- ○ I wrote essays in English for undergraduate classes
- ○ I wrote essays in English for graduate classes
- ○ I have not written essays in English for any of the classes above

What is your highest education level?

- ○ Some high school
- ○ Completed high school (or equivalent)
- ○ Completed Trade or Professional School
- ○ Some College/ University
- ○ Completed College/ University
- ○ Some Graduate School
- ○ Completed Graduate School

→

[Feedback and Appreciation form displayed]

[Random Identification for Payment provided]

APPENDIX F: SECONDARY EXPERIMENT INSTRUMENT

Writing in square brackets is not shown to participants

[Consent form displayed]

[Instructions]

(Part 1/2)

Assume that you and your team work for a consulting company that writes reports for clients. The company you work for expects its employees to write very well-written reports. Therefore, it evaluates its employees' reports before it sends them to the client.

Each of you and your team members have been asked to write a section of a report for your client, and each section of the report will then be evaluated by someone from your company. You will each be paid a fixed payment for your work.

All sections of the report are similar in length and difficulty.

The quality of **your section** of the report, which depends only on your own work, will be evaluated. Each section of the report worked on by your team members will also be evaluated separately.

→

[Lower Subjectivity Condition]

(Part 2/2)

The evaluator **will** be able to form an objective evaluation of writing quality, because they possess objective criteria for quality. The objective criteria the evaluator **will use** in their evaluation are the number of spelling mistakes, grammar and punctuation mistakes in the report section, and the conciseness and formality of the writing. Therefore, overall, their evaluation **will be objective and <u>not</u> subjective**: it **will not** be based on the evaluator's personal views or opinions on each section of the report.

The evaluator will be able to see the sections of the report worked on by you and your team members, and how long you and your team worked on your sections of the report.

Click here to see the information on the previous page

→

[Higher Subjectivity Condition]

(Part 2/2)

The evaluator **will not** be able to form an objective evaluation of writing quality, even though they possess some objective criteria for quality. The objective criteria the evaluator **might refer to** in their evaluation are the number of spelling mistakes, grammar and punctuation mistakes in the report section, and the conciseness and formality of the writing. However, overall, their evaluation **will be subjective and not objective**: it **will** be based on the evaluator's personal views and opinions on each section of the report.

The evaluator will be able to see the sections of the report worked on by you and your team members, and how long you and your team worked on your sections of the report.

Click here to see the information on the previous page

→

[Dependent Variables]

For each of the following, please indicate the extent to which you agree that the provided statement characterizes your thinking while reading the scenario.

"I feel certain about **how** I can get a good evaluation on my section of the report."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

"I feel that my section of the report will be evaluated **subjectively**, based on the evaluator's personal views and opinions on each section of the report."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

"I feel that my section of the report will be evaluated **objectively**, based on objective writing criteria such as number of spelling mistakes, grammar and punctuation mistakes in the section of the report being evaluated."

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

→

“I believe the evaluator will evaluate my writing better if they know I spent more time on it.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

“I believe the evaluator will evaluate my writing better if they know I put more effort into it.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

“I feel like I can control how I am evaluated on my section of the report.”

Strongly disagree	Disagree	Somewhat Disagree	Neither agree not disagree	Somewhat agree	Agree	Strongly Agree

[Post-experiment Questionnaire]

Below, you will find a series of statements. Please read each statement and decide how much you agree or disagree with that statement.

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Any person who is able and willing to work hard has a good chance of succeeding.	○	○	○	○	○	○	○
Life would be more meaningful if we had more leisure time.	○	○	○	○	○	○	○

→

[Comprehension Check]

Please answer the following question:

In this hypothetical scenario, you were asked to assume that you are a member of a team in a consulting company.

- ○ True
- ○ False

[Correct response: True]

→

[Demographics]

What is your age in years? ☐

What is the language you feel most confident writing in?

- ○ English
- ○ Other (please specify) _____

Do you have previous work experience working as part of a team?

- ○ Yes
- ○ No

What is your highest education level?

- ○ Some high school
- ○ Completed high school (or equivalent)
- ○ Completed Trade or Professional School
- ○ Some College/ University
- ○ Completed College/ University
- ○ Some Graduate School
- ○ Completed Graduate School

[Feedback and Appreciation form displayed]

[Random Identification for Payment provided]