ESSAYS ON STATUS RECOGNITION AND ITS CONSEQUENCES FOR TOP-TALENT MOBILITY AND PRODUCTIVITY

by

Brittany M. Bond

B.A. International Economic Relations
The American University, 2007

M.S. Public Policy and Management
Carnegie Mellon University, 2010

S.M. Management Research
Massachusetts Institute of Technology, 2018

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN MANAGEMENT
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author

_____

Sloan School of Management
April 29, 2020

Certified by

_____

Roberto M. Fernandez
William F. Pounds Professor in Management and Professor of Organization Studies
Thesis Supervisor

Accepted by

_____

Catherine Tucker
Sloan Distinguished Professor of Management Science
Professor of Marketing
Chair, MIT Sloan PhD Program

ESSAYS ON STATUS RECOGNITION AND ITS CONSEQUENCES FOR TOP-TALENT MOBILITY AND PRODUCTIVITY

by

Brittany M. Bond

Submitted to the Sloan School of Management on April 29, 2020
in Partial Fulfillment of the Requirements of the Degree of
Doctor of Philosophy in Management

Abstract

Organizations increasingly rely on status recognition to motivate members toward higher performance. Yet status recognition inevitably invites social comparisons. Although research in organization theory and strategy has focused on the returns to, antecedents of, and relative advantages of status recognition, whether, when, and to what extent bestowing status recognition outweigh the costs of social comparison remain open questions. My dissertation contributes to this scholarship through experimental field and archival research that illuminates the unexpected ways status recognition influences motivation, mobility, and productivity. This leads me to identify, in my first essay, how the preservation of self-image leads employees to make costly employer exits even when there are no material, career, or reputation concerns to nominal status under-recognition. In my second essay, I demonstrate how highly relational managers are more likely to artificially inflate employee performance evaluations, how this overvaluation leads to persistent underperformance, and how structured management can counteract this downside to close managerial relationships. My third essay (coauthored with Ethan J. Poskanzer), demonstrates how specialists' productivity improves after engaging in tasks that these professionals are recognized as being relatively inexpert in relative to teammates and their area of specialization. The settings I study in this dissertation pertain to professionals operating in high-status organizations: a highly competitive multinational pharmaceutical company and Major League Baseball. Overall, my dissertation contributes to our understanding of how status recognition influences motivation, mobility, and productivity in unexpected ways and among top-talent professionals in particular. This research has implications for organizational and strategy research on social status, motivation, and the management of performance review systems.

Thesis Supervisor: Roberto M. Fernandez
Title: William F. Pounds Professor in Management and Professor of Organization Studies

Acknowledgements

My advisor, Roberto Fernandez, made all the difference in my PhD. He made sure I was in the game for the right reasons and having fun. With Roberto in my corner, I took greater risks and tested further frontiers.

Roberto constantly reinforced the fundamentals. At the same time, he was only ever transparent about the difficulties of academia. Roberto never missed a chance to challenge me or call me out. Most important, Roberto's ethic of care overwhelmed any struggle I faced during my PhD. Roberto made Sloan the most supportive environment I have ever experienced. I am incredibly lucky and grateful to have such an immense mentor so invested and so worthy of trust. Thank you, coach.

Even (especially) when shifting my worldview, Ezra Zuckerman Sivan made sure I found steady ground. Throughout my PhD, I did not need to seek Ezra out for assistance. He constantly showed his support in unexpected, timely ways. Example: Ezra withstood the wildest weather with his daughter to cheer me on in my bucket-list Boston Marathon in 2018 ☺. Ezra's steadfast investment in my development encouraged me beyond what I thought possible. I am ever indebted to his proactive investment.

Emilio Castilla, whose scholarship I continually try to emulate, has always taken an interest in whatever I am studying and has encouraged me every step of the way (even through my admissions process, despite him being away on sabbatical).

Erin Kelly arrived at Sloan just in time so I could enlist her into my committee. I am thankful for all the times that her interventions and timely guidance have helped keep my ship sailing straight.

In addition to my four incredible committee members, I am so grateful for all that I have received from everyone in the Sloan community. Since the moment I met her, I have taken every opportunity to be in Susan Silbey's orbit. My approach to academia has benefitted dramatically from her force field. Every interaction with Ray Reagans was a mental power boost. Jerker Denrell and Hazhir Rahmandad both led me to discover new ways of thinking through dynamic modelling of organizational adaptation. Bob Gibbons, Jared Curhan, and Paul Osterman opened up their classes and ways of thinking to me well into my PhD and in doing so expanded my horizons. JoAnne Yates, John Carroll, Cat Turco, Kate Kellogg, and Tom Kochan have all been so supportive throughout my PhD journey. I continue to learn something new from Danielle Li every meeting. Hal Gregerson is the most caring office neighbor. I lucked out having Nate Wilmers and Basima Tewfik join Sloan in time to offer their support and sage advice when I was going on the job market. Mabel Abraham, Julia DiBenigno, Erik Duhaime, Jaekyung Ha, Oliver Hahl, Arvind Karunakaran, Danny Kim, Josh Krieger, Abhishek Nagaraj, Aruna Ranganathan, Christine Riordan, Ben Rissing, and Emily Truelove all continue to serve as amazing mentors.

Sam Zyontz was the first friend I had at Sloan and the person who convinced me that I could hack it. From my first semester, I started storing all the emails from pals Minjae Kim and Vanessa Conzon in my 'professors' email folder. They both helped me so much in understanding just what we were doing. Tristan Botelho was the best officemate any grad student could ask for, and I will be forever grateful that I had him looking out for me along the way. So many of my 'words of wisdom' to other students come with their Tristan acknowledgements. But all my officemates have been the best! Simon Friis was not only my officemate all 6 years, as well as my fellow Econ Soc cohort rep along with Melissa Staha, but a true friend. I learn more from Carolyn Fu than I ever impart her way and I got to commiserate with James Riley through the job market in E62-376. I am proud to have coauthored papers with my academic-sibling Tatiana Labuzova and baseball buff Ethan Poskanzer.

Sloan would have been a far less enjoyable place without Avi Collis, Caroline Fry, Mahdi Hashemian, James Houghton, Summer Jackson, Raquel Kessinger, Mahreen Khan, Soomi Kim, Will Kimball, Alex Kowalski, Claire McKenna, James Mellody, Jenna Myers, Dan Rock, Hye Jin Rho, Sebastian Steffen, Hagay Volvovsky, Michael Wahlen, Duanyi Yang, Heather Yang, Romain Boulongne, as well as all the recent additions to the Sloan and our strong alumni family.

I cannot give enough thanks to all those who keep Sloan running: Colette Boudreau, Patty Charest, Natalia Kalas, Helen Yap, Lena Ngor, Gilly Parker, Hillary Ross (who also braved the elements during the 2018 Boston Marathon for me with her dog, Ollie), Davin Schnappauf, and Sarah Massey.

Much gratitude goes to everyone I worked with at my research field site, which I refer to in my thesis as PharMed, for their appreciation of scientific approaches to better understand organizational processes.

I thank everyone I worked with in the Office of the Chief Economist at the US Department of Commerce and in the Census Bureau's Center for Economic Studies for my economist training and their support as I found the field of best fit for me.

If not for Jon Caulkins from the Heinz College at Carnegie Mellon University, I would never have found my way into academia. Thank you for giving me the chance to make this happen.

I thank my family for making me who I am. I am grateful to my parents, Pam and Julian. They inspire me in everything from my career path, research interests, and life choices. My aunts, uncles, and grandparents also helped raise me to be tougher than the circumstances. My siblings and cousins remain responsible for my flavor of imagination and for my determination to try to make something of myself. They are my motivation every day. Chrissy McLaughlin and Achilles were my solace and retreat from the craze of Cambridge.

Having Nick McLaughlin as my partner in this journey was like having air in my lungs. Thank you for knowing me. Thank you for asking me to dance.

Brittany M. Bond
April 29, 2020
Cambridge, Massachusetts

Table of Contents

INTRODUCTION

Firms increasingly rely on nonmonetary recognition to reward employees (Nelson 2005). The desire for positive recognition and status among peers can be a powerful motivator, and people are often motivated to earn recognition and visibility for their own sake (Goode 1978; Hirsch 1976; Ricoeur 1965). As public recognition is free for employers to bestow, such rewards can be cost-effective ways to motivate employees (Besley and Ghatak 2008; Moldovanu et al. 2007). Advocates of these rewards frequently argue that non-recognition will induce greater effort so as to achieve recognition in the future (Neckermann et al. 2014; Bradler et al. 2015; Ammann et al 2016).

For public recognition or awards to be meaningful, however, not everyone can earn them. The desirability of status recognition comes from bestowal on a small fraction of potential recipients (Goode 1978). The value of such rewards requires that they be scarce (Hirsch 1976). As a result, recognizing some employees as high performers necessitates labeling others as less proficient, if only by omission. While such contrasts may motivate some employees to work harder to earn the approbation of their peers and employers, they may also create excessively competitive work environments and decrease overall morale (Lazear 1989), thereby *demotivating* other employees.

Whether or not motivational returns to bestowing status recognition outweigh the costs of social comparison has important implications for how performance recognition reward systems should be designed and implemented. Additionally, such knowledge furthers our understanding of the relationship between status conferral and motivation more generally. Although research in organization theory and strategy has focused on the returns to, antecedents of, and relative advantages of status recognition, whether, when, and to what extent bestowing status recognition outweighs the costs of social comparison remain open questions. The answers promise deeper understanding of the relationship between status conferral and motivation.

Despite the importance of understanding the potential unintended consequences of status recognition schemes, identifying the effect of recognition in reward schemes is challenging. Three potential confounds are especially important to consider. First, underlying performance differences, even if

marginal, typically underlie the differentiation and selection of those receiving relative recognition (Merton 1968; Lynn et al. 2009). This means that under-recognized employees may exhibit behavior consistent with poor performance because they are actually less competent or less committed. Monetary rewards represent another confound; that is that employees may react negatively to under-recognition not because they experience status loss but because they experience monetary loss (Frey and Gallus 2017). In such cases, it is hard to tease apart the effects of pure status recognition from material gains. Finally, employees may react negatively to under-recognition because they fear it will damage their reputation inside and outside the organization (Holmstrom 1982). Without controlling for these complicating factors, it is impossible to establish the costs of under-recognition that are due to social comparison effects.

My first two chapters comprise causal investigations of the effects of status recognition on employees in a highly competitive multinational pharmaceutical company that I refer to as PharMed (a pseudonym). Over three years as an embedded field researcher, I've studied the performance evaluation process and management practices at PharMed, exploring the causes and consequences of employee status recognition. Like more than 90% of U.S. companies using formal performance appraisals, PharMed uses its year-end employee performance evaluations as a means of motivating employees (WorldatWork 2018). The way PharMed executes a strict forced distribution on the formal performance appraisals of its employees generates unique, natural experimental features that enable me to overcome the identification challenges discussed above. Moreover, this setting offers a unique opportunity for developing and testing effective motivational mechanisms in organizational performance management design schemes.

The first PharMed study uses a natural experiment to demonstrate the powerful negative effects that under-recognition of performance can have on employee motivation, even when under-recognized employees are materially better off and there are no career, reputation, or signaling concerns at stake. This study demonstrates severe costs to performance reward schemes that go unaccounted for by these specific social comparison costs. In this manner, this study also elevates the importance of self-image in the consequences of recognition reward schemes.

My second chapter from PharMed then looks at over-recognition, specifically when managers consciously overvalue under-performing employees' performance. I seek to explain what predicts this overvaluation and show the negative productivity consequences for the organization. I find that highly relational managers, whose teams report higher levels of trust and cooperation, are more likely to give a higher rating to an under-performing employee than can be objectively justified. I find this translates into productivity loss over time. I also find evidence that structured management, in the form of formalized work practices, mitigate the potential for such costly overvaluation of employee underperformance without compromising team trust and cooperation. This research illuminates how organizations can encourage more accurate performance recognition while preserving important and productive relational elements between managers and employees.

My third chapter extends my research on the consequences of professional status recognition by examining the importance of working within the area a specialist is known to be of high status versus assignments to inexpert work where they are relatively inferior relative to others. Coauthored with a fellow Sloan PhD student, Ethan J. Poskanzer, we analyze the performance and productivity of Major League Baseball pitchers. The study illuminates the consequences of transitioning from tasks that a specialist is recognized as being an exceptionally high performer to tasks which the specialist is inexpert and known as significantly inferior in performing relative to both teammates and in their specialist abilities. In particular, this study exploits the quasi-random timing of when a pitcher must perform an inexpert task (batting) prior to performing their specialization (pitching). This analysis shows pitching performance improves after transitioning between inexpert tasks and specialization tasks, controlling for pitcher, batter, game and other relevant fixed effects. These results suggest that when specialists engage in inexpert tasks, they are likely to over-perform when returning to tasks in which they excel. These findings offer implications for job design incorporating optimal challenge features for specialized professionals.

Together my dissertation contributes to our understanding of the importance of status recognition on motivation, mobility, and productivity, especially of top-talent professionals. This research holds

implications for understanding how status conferral affects motivation and how organizations should manage performance review systems.

References

Ammann, M., Horsch, P. and Oesch, D., 2016. Competing with superstars. *Management Science*, 62(10), pp.2842-2858.

Besley, T. and Ghatak, M., 2008. Status incentives. *American Economic Review*, 98(2), pp.206-11.

Bradler, C., Dur, R., Neckermann, S. and Non, A., 2016. Employee recognition and performance: A field experiment. *Management Science*, 62(11), pp.3085-3099.

Frey, B.S. and Gallus, J., 2017. Honours versus money: The economics of awards. Oxford University Press.

Goode, William, J., 1978. The Celebration of Heroes: Prestige as a Control System.

Hirsch, Fred. 1976. Social limits to growth. Harvard University Press.

Holmstrom, B. 1982. "Managerial Incentive Problems—A Dynamic Perspective." In Essays in Economics and Management in Honor of Lars Wahlbeck. Helsinki: Swedish School of Economics. Republished in Review of Economic Studies 66 (1999): 69-82.

Lazear, E. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–580.

Lynn, Freda B., Joel M. Podolny, Lin Tao. 2009. "A sociological (de)construction of the relationship between status and quality." *American Journal of Sociology*, 115(3): 755-804.

Merton, Robert K. 1968. The Matthew Effect in Science. *Science*, 159(3810): 56-63.

Moldovanu, B., Sela, A. and Shi, X., 2007. Contests for status. *Journal of political Economy*, 115(2), pp.338-363.

Neckermann, Susanne, Cueni, Reto, and Frey, Bruno S. 2014. Awards at Work. *Labour Economics*, 31: 205-17.

Nelson B. 2005. 1001 Ways to Reward Your Employees. Workman Publishing: New York

Ricoeur, P. A. 1965. Fallible Man. Chicago, Henry Regnery.

WorldatWork. 2018. "2017 Inventory of Total Rewards Programs & Practices". https://www.worldatwork.org/docs/research-and-surveys/Report_FINAL1.pdf

Chapter 1

The Burden of Under-Recognition in Organizations

INTRODUCTION

Firms increasingly rely on nonmonetary recognition to reward employees (Nelson 2005). Employees can be highly motivated by rewards that have no market value, such as purely symbolic public recognition (Frey and Neckermann 2008, Neckermann et al. 2014). The desire for positive recognition and status among peers is powerful; and people are often motivated to earn recognition and visibility for its own sake (Goode 1978; Hirsch 1976; Ricoeur 1965). As public recognition is free for employers to bestow, such rewards can be a cost-effective way to motivate employees (Besley and Ghatak 2008; Moldovanu et al. 2007). Advocates of these rewards frequently argue that non-recognition will induce greater effort so as to achieve recognition in the future (Neckermann et al. 2014; Bradler et al. 2015; Ammann et al. 2016).

For public recognition or awards to be meaningful, however, not everyone can earn them. Indeed, this is a ubiquitous problem since at the margin, there will always be a distinction between those that win recognition and those that do not. The desirability of status recognition comes from bestowal on a small fraction of potential recipients (Goode 1978). The value of such rewards requires that they be scarce (Hirsch 1976). As a result, recognizing some employees as high performers necessitates labeling others as less proficient, if only by omission. While such contrasts may motivate some employees to work harder to earn the approbation of their peers and employers, they may also create excessively competitive work environments and decrease overall morale (Lazear 1989), thereby demotivating other employees.

Evidence that nonmonetary recognition schemes reinforce prosocial behavior and motivate non-winners to work harder in the hope of earning future recognition come from settings where the recognition process is viewed as fair and where there is a sense of efficacy in ability to earn future recognition (Ammann et al. 2016; Neckermann et al. 2014; Bradler et al. 2016). Yet assessment of relative recognition is derived from salient constructed comparison sets (Stouffer et al. 1949; Merton 1968; McGraw et al. 2005). People are more likely to compare the returns to their efforts with that of proximate peers (Homans

1974, Burt 1982). Invidious comparisons, where individuals inflate their own performance and fixate on persons of higher rather than of equal or lower standing, can generate perceived inequality even in the face of either actual performance differences or equal treatment (Adams 1963; Festinger 1954; Nickerson and Zenger 2008; Larkin, Pierce, and Gino 2012). Moreover, reactions to such felt relative deprivation can lead to self-imposed absolute costs (Gurr 1970; Goode 1978; Frank 1985).

Whether or not motivational returns to bestowing status recognition outweigh the costs of social comparison has important implications for how performance recognition reward systems should be designed and implemented. Additionally, such knowledge furthers our understanding of the relationship between status conferral and motivation more generally. Although research in organization theory and strategy has focused on the returns to, antecedents of, and relative advantages of status recognition, whether, when, and to what extent bestowing status recognition outweighs the costs of social comparison remain open questions. The answers promise deeper understanding of the relationship between status conferral and motivation.

Despite the importance of understanding the potential unintended consequences of status recognition schemes, identifying the effect of recognition in reward schemes is challenging. Three potential confounds are especially important to consider. First, underlying performance differences, even if marginal, typically underlie the differentiation and selection of those receiving relative recognition (Merton 1968; Lynn et al. 2009). This means that under-recognized employees may exhibit behavior consistent with poor performance because they are actually less competent or less committed. Monetary rewards represent another confound; that is that employees may react negatively to under-recognition not because they experience status loss but because they experience monetary loss (Frey and Gallus 2017). In such cases, it is hard to tease apart the effects of pure status recognition from material gains. Finally, employees may react negatively to under-recognition because they fear it will damage their reputation inside and outside the organization (Holmstrom 1982). Without controlling for these complicating factors, it is impossible to establish the costs of under-recognition that are due to social comparison effects.

In this paper, I leverage a unique natural experiment to provide the first evidence that purely nominal status recognition carries substantial social comparison costs in reward schemes that rely on nonmonetary recognition. The study design uses counterfactual cases to status recognition where employees' performances merited recognition, but they did not receive it due to arbitrary restrictions on the supply of the official recognition. Specifically, the way that the company I study executes a strict forced distribution on the formal performance appraisals of its employees generates unique, natural experimental features that enable me to overcome the identification challenges discussed above. Importantly, all parties recognize that the status recognition is an arbitrary distinction among equally high-performing employees and is generated only by a predetermined scarcity of recognition to allocate among top performers. Under-recognition carries no promotion or other career consequences in this setting.

I demonstrate that employees are highly sensitive to this under-recognition which carries no signaling, career, or reputation concerns. I find that nominally under-recognized employees are much more likely to voluntarily exit the organization, going to work for a local competitor that offer less competitive wages overall. Strikingly, these employees are much more likely to exit *even though they are awarded greater monetary bonuses in compensation for their under-recognition*.

These findings cannot fully be captured by our current understanding of social comparison costs and relative deprivation. Especially when all parties recognize parity of performance and the under-recognition is not consequential in any tangibly negative way, neither construct fully explains costly reactions on the part of the under-recognized. Yet under-recognized employees may still take costly measures to change their professional situation even in the absence of perceived prejudice in the system. While peers and supervisors may overtly acknowledge and compensate for recognition arbitrarily denied, under-recognition can override what would seem like logical, even outsized, compensation.

To explain such costly reactions to nominal under-recognition, I discuss how changing one's external environment is explainable as a form of adaptive preference formation to reduce cognitive dissonance around one's self-image. A similar notion has been referred to as sour grapes (Elster 1985); the idea being that when performers are denied a desired form of recognition, frustration may most readily be

relieved by choosing different contexts. Employees may find relief from the frustration caused by under-recognition most readily by reassessing their current context more negatively and viewing the relative benefits of other alternatives more positively. This reweighting relieves the psychic costs of feeling undervalued in the present situation by making other options more appealing and simultaneously circumventing the difficulties of expressing the pain of under-recognition where superiors and peers negate any material or reputation costs that could be attributed the under-recognition.

This mechanism suggests that even when top performers are given equal opportunity for status recognition and those denied are given no less (indeed often more) positive treatment, there are still high costs to making public status distinctions. These costs of under-recognition, however, are internalized. I demonstrate that perceived damage to self-image is stronger than any monetary compensation leveraged to compensate for the arbitrary denial of status recognition. The sensitivity to relative recognition among peer employees is too dear to tolerate even the weakest, even moot, signal of misalignment.

These findings offer important insights regarding the power of conferring status recognition on employee motivation. Specifically, they add a strong cautionary note for the use of nonmonetary status recognition in the design and management of performance reward schemes. Sensitivity to under-recognition can compel employees to seek immediate outside options in an effort to rectify the discrepancy between their social status recognition and their self-image. Therefore, it is important for performance management and reward design schemes to heed employees' sensitivities to perceived discrepancies between their self-image and what is officially recognized by the organization beyond any material, career, or reputational concerns.

THEORY

The (Limited) Advantages of Using Nonmonetary Rewards

While monetary compensation is the dominant reward mechanism for employee performance, relying on financial incentives is prone to pitfalls (Lepper and Green 1978; Gneezy et al. 2011). Especially in settings

in which performance is difficult to monitor, pay for performance schemes will be vulnerable to gaming, manipulation, and incentive misalignment (Holmstrom and Milgrom 1987; Holmstrom 1999; Kerr 1995; Oyer 1998; Larkin 2007). Even when it is feasible to perfectly fix monetary compensation to output, workers may perform sub-optimally if money crowds out intrinsic motivation (Deci, Koestner, and Ryan 1999; Deci and Ryan 1985; Pfeffer 1998).

Employees are highly sensitive to social comparisons regarding pay. Negative social comparisons can lead to costs in the form of reduced effort (Akerlof and Yellen 1990), strategies to influence manager preference and other influence activities (Milgrom and Roberts 1988, 1990), or departure (Strauss 1995). Even after they receive absolute wage increases, Dube et al. (2017) find that retail workers will exit companies if they perceive equivalent peers to have received marginally higher raises. Negative peer comparisons have also been shown to increase absenteeism and foregone wages among manufacturing workers in developing countries (Breza et al. 2017).

Moreover, as tasks become more complex and work structures more diverse, it becomes increasingly difficult to design compensation schemes that are not undermined by social comparison costs (Larkin et al. 2012). This is why we observe university professors experiencing decreased job satisfaction and increased job search when peer wage information is revealed (Card et al. 2012). Evidence that their compensation falls below the median generates a sense of having been undervalued. Focused attention on peers in the same department concentrates this feeling of relative deprivation. These examples underscore how relative pay comparison concerns easily outweigh material satisfaction.

The nascent literature on nonmonetary rewards almost universally focuses on how such rewards and recognition avoid such pitfalls befalling monetary compensation schemes (for a review see Frey and Gallus 2017). Early evidence suggests that nonmonetary recognition is not prone to the same invidious social comparison costs besetting monetary rewards. Interview data from Neckermann et al. (2014) suggests that call center employees who do not win an award for exceptional effort rarely view the award negatively. Other evidence suggests that non-winners of awards are motivated to improve their own efforts when a competitor is rewarded prominent status recognition.

For instance, Ammann et al. (2016) find that the competitors of award-winning CEOs improve their own business activities across a variety of dimensions in subsequent years. Similarly, individuals improve their productivity in data entry tasks after witnessing employer recognition of high-performing peers (Bradler et al. 2016). Such evidence supports the theory that elevating high performers can motivate others to perform better (Ehrenberg and Bognanno 1990). This further supports the understanding that nonmonetary reward schemes can improve performance by motivating employees via status-based competition and social comparison (Moldovanu et al 2007; Kosfeld and Neckermann 2011; Neckermann et al. 2014).

Nonmonetary award programs are seen as avoiding the more prominent psychological and financial costs of monetary compensation by reinforcing positive and prosocial behavior (Frey and Neckermann 2008, Neckermann et al. 2014). Although they rely primarily on extrinsic motivation, as do monetary rewards, in the form of public recognition, nonmonetary rewards are seen as 'crowding-in' (rather than crowding-out) intrinsic motivation and prosocial behavior, expanding the overall returns to the organizations employing such rewards (Frey and Jegen 2001; Gneezy et al. 2011; Bowles and Polania-Reyes 2012). Thus, nonmonetary recognition can improve alignment between employee motivation and business goals and values through teamwork, better customer service and business improvement beyond what is possible through the use of direct payments alone (Brown and Armstrong 1999, Silverman 2004).

At a certain level however, bestowing awards on a selection of individuals means imposing distinctions on equally qualified employees. The more equally qualified the candidates, the more marginal the award distinctions will be. The performance of those receiving recognition may be indistinguishable from those that fall just on the other side of the margin. Nonmonetary recognition schemes, therefore, may in fact be vulnerable to the same invidious social comparison costs that arise from formal performance pay schemes (Nickerson and Zenger 2008; Larkin et al. 2012).

To date, only two studies provide evidence that nonmonetary reward schemes have unintended costs resembling those borne from monetary incentive schemes. Gubler et al. (2016) find strategic gaming of award eligibility criteria and the loss of internal motivation in tasks not included in an attendance award

program in US laundry plants. Their paper suggests that even purely symbolic awards can generate gaming and crowding out productivity in other important tasks. Gubler et al. (2016), therefore, demonstrate that symbolic nonmonetary rewards are not immune to the psychological costs inherent in monetary compensation schemes.

Additionally, in a field experiment in a nationwide health worker training program in Zambia, Ashraf et al. (2014) find that while employer recognition and social visibility increase performance, social comparison is detrimental. When low ability trainees learn that their performance will be ranked among peers, their performance declines even *before* the rankings are revealed. While adding employer recognition awards improves performance, the effects are nullified by the preemptory, exacerbated under-performance among low performers. In this case, however, there are no post-treatment effects. Rather, the productivity costs in Ashraf et al. (2014)'s experiment are driven by low performers self-handicapping so as to avoid having to make accurate updates regarding their own relative ability (pg 20).

Ashraf et al. (2014)'s results provide the only suggestive evidence to date that symbolic employee recognition can potentially carry social comparison costs. However, the implications for performance reward schemes are ambiguous. First, relative ranks were generated by objective differences in performance, meaning they conveyed career-relevant information to their employer. More important, social comparison effects were concentrated at bottom of the performance distribution, meaning the organization may stand to benefit if the effects translate into low performers self-selecting out of the organization. As such the influence of social comparison in this experiment is not distinctly one of *costs* per se. It remains unclear how relative status recognition comparisons will influence high performers' motivation, especially in situations in which it is understood that objective performance does not perfectly predicate employer recognition.

Self-Image and Ritual (Rather than Irrational) Moves

Importantly though, the Ashraf et al. (2014) experiment does offer a critical mechanism for explaining negative behavior in the face of arbitrary and nominal distinctions in employee recognition. Finding that

productivity declined after the announcement, but prior to the revealing of relative rank information suggests that the social comparisons behind employee recognition are driven by *self-image* concerns rather than competition (Charness and Grosskopf, 2001; Freeman and Gelber, 2010). In the absence of status recognition schemes, employees can hold whatever beliefs about their relative status as they see fit. However, the mere announcement of forthcoming employer recognition can lead to self-sabotaging behavior if such behavior serves to preserve one's self-image.

The preservation of self-image offers an explanation for costly reactions to nominal and arbitrary under-recognition that is inexplicably self-damaging from the vantage of our current understandings of social comparison costs and relative deprivation. Being denied expected status recognition can subvert an employee's self-image as a well-recognized top contributor. In such circumstances, employees must reconcile their self-image with the discrepant nominal recognition they receive. Therefore, nominal under-recognition, even that which has no basis in underlying quality or relative status, can sow the seeds for identity preserving moves. Nominally under-recognized employees may feel their 'face' has been compromised (Goffman 1967**,** pg. 5)[1]. Even when there are no tangible or even intangible costs to denied recognition, under-recognized employees may still suffer psychic costs. While others may recognize the situation merits compensation and restoration of self-image, these proffered face-salvaging efforts by others may act as a further affront one's self-image. Efforts at placation can in fact heighten the frustration.

Compounding their frustration, it becomes increasingly difficult to defend one's indignation at damage done to self-image when others may be aware of any material compensation for the nominal under-recognition. Nominal awards cannot be consumed; whereas cash can translate into the most desired external markers. Therefore, when there are no tangible or even intangible costs to their situation, employees may be entirely better off extrinsically when denied recognition. In the sense that the only costs are psychic ones, it becomes difficult to make a compelling case for one's indignation at the damage done to one's self-image.

[1] See also Goode (1978: 102-110) on reputation dynamics within membership groups in terms of gossip hazards and fears.

There are seemingly no devices to rectify this damage then. When material, career, signaling, or reputation effects are of no concern, it is difficult to explain costly negative reactions to denied status-recognition from a social comparison or relative deprivation framework. The necessary extrinsic explanations are seemingly exhausted. Yet recompense for denied recognition can further sensitize performers to the denial itself. Reactions to denied nominal recognition become increasingly more self-damaging than what extrinsic frames can predict or explain. Reacting negatively as by exiting the organization, becomes instead "a ritual move, conveying that [the employee] has a face to lose and that its loss is not to be permitted lightly" rather than "irrational expression of frustration" (Goffman 1967: 23).

This ritualistic move reduces the cognitive dissonance generated by denied recognition. Rather than be externally justifiable or validated, performers reconcile the discrepancy between internal evaluations and self-image through adaptive preference formation that alters the relative valuation of the present context to alternatives. Without recourse for alleviating the frustration of such denial, the most ready reconciliation may be one of deciding that the denied prize is not worth striving for further.[2] The denied recognition makes the current context one of sour grapes (Elster 1985), where performers adjust their desires and valuations according to being denied the recognition for which they were warranted to earn. All recourse to under-recognition is negated by over-compensation on the part of supervisors and peers. This forecloses any remedies that involve sustaining the under-recognized employees' attachment to the current context *and* rectifying the felt discrepancy regarding one's self-image and the nominal recognition received.

A classic example of this mechanism in operation comes from *The American Soldier* studies (Stouffer et al. 1949) where the better the overall promotion chances, the more frustrating one's individual promotion chances were felt. Importantly, promotion in this context was decided on universalistic criteria. The lowering of subjective well-being operated through soured anticipation in the midst of overall

[2] Note: this is distinct from the mechanism assumed in Card et al. (2012) where exposure to relative peer wages induces a sense of *past* undervaluation. Here, all signaling, monetary and otherwise, underscores the arbitrariness of the nominal evaluation relative to the high value held regarding the under-recognized.

increased objective well-being. This paradoxical finding has been referred to as the Tocqueville Effect (Tocqueville 1952 pp. 222-3; see also Elster 1985: 124). At the societal level, the Tocqueville Effect corresponds to the increased dissatisfaction with the current state of social organization amidst improving social mobility. Even as nominal conditions improve overall, the exposure to greater opportunities increase the overall incidence of frustration. The increase in potential opportunities increases the possibility that the current situation is viewed as unfavorable. The feasible set of possible outcomes has expanded positively, increasing the chances that one's current situation looks comparatively worse.

Accordingly, nominal under-recognition can favorably shape the attractiveness of other feasible opportunities even when overall conditions improve for performers arbitrarily denied recognition. Other forms of compensation are limited in preserving motivation when self-image is compromised, as extrinsic motivators cannot compensate a damaged self-image (Bewley 1999). Further, such overcompensation may exacerbate the psychic costs from the discrepant recognition and self-image. The net effect makes exit to another context more appealing than it was ex ante.

Thus, the moves taken to preserve self-image in the context of denied status-recognition can carry significant costs. Such costs reinforce the insight that employees are intensely sensitive to signals of self-validation (Ariely 2016; Pearson 2015; Pink 2009; Schwartz 2015; Sinek et al. 2017; Strack, 2014). Past research demonstrates the risks inherent in bestowing positive recognition, which may attract poaching by competitors for top talent (c.f. Groysberg 2010). Such studies have shown that firms that attract talent away from other firms may experience a modicum of performance increase (Groysberg, Lee and Nanda 2008; Younge and Marx 2016). However, the loss of star performers can represent a major loss of productivity for the source firm. Inadvertently motivating top performers to view their outside options more attractive, given self-image preservation impulses caused by under-recognition, may amount to significant talent and productivity loss.

For managers, this introduces a strong cautionary note for using nonmonetary status recognition in the design and management of performance reward schemes. Introducing rewards that may arbitrarily generate negative comparisons between top performers, even if there are other mechanisms to

counterbalance under-recognition, may cause more irrevocable damage than motivation. The distinction may sour the context for denied top contributors in a way that makes alternative options more appealing even in the face of increased compensation.

The rest of this paper proceeds as follows. First, I describe the field setting and natural experiment which provides the conditions to test for the effect of nominal status under-recognition among a group of high performing corporate employees. I provide evidence that nominal status under-recognition does not result from observable bias, nor at the cost of any material, career, signaling, or reputation concerns. I then demonstrate that under-recognized employees are much more likely to voluntarily exit the organization, even when status under-recognition is arbitrary. Importantly, this happens in a context where parity of performance among top performers, including those denied recognition is common knowledge. As such, under-recognized employees suffer no reputation or career effects from the arbitrary denial of status recognition at this margin. However, the more salient the under-recognition, as when there are top performers who receive official recognition working in the same team with the under-recognized employees, drives the effect under-recognition has on the likelihood of exiting the organization.

Further, I show that under-recognized employees are still much more likely to exit the organization despite being rewarded the highest monetary bonuses of any employees including those who receive the highest distinctions. Moreover, when the under-recognized exit the organization, they typically move on to weaker economic competitors in the labor market. This provides evidence that the under-recognized react to the arbitrary denial of status-recognition in ways that can incur high extrinsic costs for themselves.

After providing this sequence of evidence, I return to the discussion of relative deprivation and how the self-preservation mechanism, encapsulated by the mechanism captured by the notion of sour grapes, explains the ritualistic rather than rational costly moves that top performers take when denied nominal recognition. Finally, I conclude by summarizing the theoretical implications that this study holds on using recognition as an employee motivator. Even when eliminating possible confounding factors that may frustrate top performers who just fall on the denied side of the margin of recognition, organizations still risk losing top performers to otherwise less appealing competitors due to the sour grapes adaptation of

preferences. Top performers will sour on organizations they feel do not recognize them on par with their own image of their relative contributions and will update their attraction to alternative organizations. I end with a discussion regarding how organizations may better manage reward schemes with such understanding so as to forestall such costs.

EMPIRICAL SETTING AND METHODS

I test whether employees will exhibit behavior akin to classic reactions to social comparison costs and relative deprivation in a setting in which material, career, signaling and reputation costs be ruled out. I do so by leveraging a unique natural experiment that arises from the strict enforcement of a forced distribution protocol for rating employee performance at a large multinational pharmaceutical company, referred to here as PharMed (a pseudonym). The study design uses counterfactual cases to status recognition where employees' performances merited recognition, but they did not receive it due to arbitrary restrictions on the supply of the official recognition. I am able to observe the evolution of employee ratings over the course of the appraisal period through PharMed's administrative rating system. The focus of this study is on the appraisal period from November 2016 – February 2017 which generated ratings for all PharMed employees for the calendar year 2016 (between 5,000-10,000 employees).[3]

Calibrating a Forced Distribution of Talent Ratings

To motivate employees by creating scarcity and value in receiving top ratings, PharMed tightly restricts the percentage of employees who receive top ratings. At the start of yearly the performance evaluation process (early November), employees enter a self-evaluation report after which point managers enter in their written evaluation and rating into an HR administrative system (see Exhibit 1 for the sequence of performance rating stages). Once all peer managers in a group have completed their evaluations, an HR representative organizes a calibration meeting intended to equilibrate the ratings at the group level to make sure there is parity in ratings across managers. Managers typically have working knowledge of the

[3] Exact numbers are not reported due to my non-disclosure agreement with PharMed.

employees who report to their peer managers at each level of the organization. Employees do not find out their rating and performance evaluation many months later until after the entire company has finished the overall performance evaluation process and calibration has been completed across the company.

Once ratings have been finalized at the company level, managers are then required to meet with each employee who reports directly to them to deliver their final rating for the year as well as their bonus allocation. Like most companies using a similar pay-for-performance scheme, bonuses are tied to employee performance (WorldAtWork 2018). However, at PharMed bonus allocations are pooled for each manager who then has autonomy as to how to distribute her team's allotment among all of her direct reports.

The calibration process is used to enforce the tight limitation of top ratings and to add assurances that employees receive the top ratings are equally high productivity contributors to the organization. For the focal performance evaluation period of this study (for calendar year 2016), as well as for nearly a decade prior, PharMed's forced distribution scheme mandates that no more than 3% of all employees are designated 'Outstanding" and no more than 20% can be designated as 'Exceptional' (See Appendix Table A1). For simplicity, this population is referred to as "Exceptional+" as there are no strong analytical differences within the population for the purposes of this study (confirmed by robustness checks provided in the appendix).

Since PharMed does not enforce a predetermined distribution at the low-end of the performance distribution, the calibration sessions are focused on establishing and defending the designation of top performers. Identifying the low performers is largely left to individual manager discretion with support from their Human Resources (HR) representative. In this study period, only around 3% of PharMed employees are designated *less* than 'Solid' (receiving either a 'Partially Met' or 'Unsatisfactory' rating for the year's performance).[4] The majority of non-Exceptional+ employees (~97% of non-Exceptional+ employees) receive the rating 'Solid'.

[4] Based on field observations. Analysis regarding the reticence of giving low ratings is covered in detail in Bond (2019).

During these calibration sessions, the performance of top performing employees is discussed and compared with one another within teams and across business units. Interim performance ratings are determined at each level, where each group must meet their rating distribution curve before calibration sessions roll up to higher levels of the organization. Calibration proceeds in stages from 1: calibrating individual contributors among frontline managers, 2: senior managers calibrating frontline and lower level managers, 3: directors calibrating the senior manager level down, 4: vice presidents calibrating director level down, until finally, stage 5: senior vice president and chief officers calibrating vice president level down. As these calibration stages progress up the organizational hierarchy, adjustments to the proportion of Exceptional+ population are made until it meets the mandated forced distribution percentages.

Given that this system has been in place at PharMed for roughly a decade, all the HR calibration facilitators, as well as the managers and executives are well-acquainted with the process. Knowing that calibration will continue up successively higher levels of the organization, and given managers are reluctant to own the hard decision of whom to drop from Exceptional+ ratings (a common managerial inclination as reported by Pfeffer and Sutton 2006), relatively few changes to employee ratings occur until several periods after the first calibration meeting covering most top performing employee at risk of under-recognition (See Figure 1). As such, very rarely does an employee's rating change from one of Exceptional+ to Solid when their manager is in attendance in a calibration meeting. Most often, such changes occur in calibration meetings comprised of senior leaders with little or no direct familiarity with the employee in question.[5]

Ruling out Bias Concerns

The focus of this paper rests on the quasi-random assignment of employees who go from a rating of Exceptional+ by their manager to Solid through this calibration process. Once initially nominated to receive an Exceptional+ rating, employees become at risk of receiving a rating of Solid at any point in the

[5] This is further supported by direct observation of 18 calibration sessions at all stages of the process from line managers to chief executive levels.

performance rating calibration process. In the performance rating period for 2016, 14.3% of employees whose managers rated them as Exceptional+ ended up receiving a rating of Solid instead through calibration. Due to the hierarchical distance between the final decision makers in calibration from the managers of most nominated employees, as well as the role of the HR representatives facilitating all calibration meetings to ensure that no demographic biases or prejudices enter the calibration discussions, the assignment to Solid from Exceptional+ is nearly random among the nominated risk set. Attesting to this, there are no observable demographic characteristics or prior performance variables that predict whether or not an employee goes from Exceptional+ to Solid in the rating process (See Table 2; see also Table 8 for demographic representation across relevant company cross-sections).

Being well-acquainted with the calibration process and the chances that nominated Exceptional+ employees may end up being rated Solid through no action on their part, managers are reticent to nominate multiple Exceptional+ employees if they can distinguish performance differences among their potential nominees. Understanding that the decision to drop an employee from Exceptional+ to Solid is out of their hands once calibration proceeds to the next level up the organization, managers will nominate only the best of their direct reports, even if they believe that others on their team deserve an Exceptional+ rating relative to the company as a whole.[6]

For these reasons, I label employees who went from being rated Exceptional+ by their manager to receiving a rating of Solid for 2016 as ‘under-recognized’. Employees whose manager rated them Exceptional+ at the start of the performance evaluation process but who receive a Solid for the year do not receive the official recognition, in the form of their year-end performance rating, for the exceptional performance that all parties familiar with the employee’s performance agree is merited.

It is important to note that while the administrative data, predictive analysis, transcripts, notes and my general observations from the calibration process all support a lack of bias in the process of calibrating top performing employees from an Exceptional+ rating to Solid, this does not preclude any bias in the

[6] Reported in interviews by managers and HR representatives explaining the management of the calibration process from a people manager perspective.

overall performance rating system at PharMed. While the purposes of this paper rest on an absence of bias in generating under-recognition conditional on being rated as exceptional+, there may still be bias in who is nominated an Exceptional+ employee in the first place.

Indeed, I do find several demographic variables are predictive of whether or not employees will be rated Exceptional+ in the study period (in particular, Asian and Black employees are less likely to be nominated as Exceptional+, controlling for prior performance and PharMed office location. See Table AX1). This is important to note in the discussion of results comparing outcomes between the under-recognized top talent employees and the rest of the PharMed population. However, further exploration of the importance of this pre-calibration bias remains for further research, which I discuss at the conclusion of this paper. The identification methods regarding the costs of under-recognition rely on those who are under-recognized relative to the counterfactual experience of their counterpart peers who are not under-recognized. Possible heterogeneous treatment effects by demographic groups are tested in the analyses. Results support the conclusion that under-recognition is experienced similarly across demographic groups and that the increased likelihood of exit is a common response.

One non-demographic and non-performance-based factor that does influence the probability an employee becomes under-recognized is the number of total Exceptional+ employees nominated by any manager. Less than one third of nominated top performers are the only top performer their manager nominates (See Table 3). Exceptional+ employees whose managers nominate more than one of their employees for Exceptional+ are more likely to be under-recognized, although within this condition there again are no biases based on observable characteristics (See Table 4). Observations and transcripts from the calibration sessions show that this decision heuristic is based on an attempt to provide equity in the process at least at the manager level. Without much first-hand knowledge to base the decision to drop an Exceptional+ individual to Solid for the year, final decision-makers at the top of the organization attempt to make it more likely that all managers who have Exceptional+ employees on their teams will be able to reward at least one of their employees with an Exceptional+ rating for the year.

Ruling Out Material Concerns

This concentration of under-recognition among the top employees of managers with more than one nominated Exceptional+ top performing employee contributes to an important feature of this setting for ruling out potential material confounding factors. Combined with the fact that managers are able to distribute their team's collective bonus pool among their team as they see fit, this means that managers are able to compensate under-recognized employees by redistributing bonus sums earned by having Exceptional+ employees on their team. I find strong evidence that indeed managers in PharMed do compensate under-recognized top performers for under-recognition by giving them the highest possible bonuses for the year. In fact, under-recognized top performers are given the biggest year-end bonuses of any population of employees in PharMed (See Figure 2). Under-recognized employees, on average, receive 20% of their annual salary in merit bonus versus the 16% of annual salary the Exceptional+ employees receive (t-test: 6.50 $p<0.0001$).[7]

Demonstrating that under-recognized employees receive *more* money in bonus pay than their counterfactual comparison Exceptional+ peers helps rule out monetary confounds as explaining the effect that under-recognition has on reactions taken by the under-recognized. Furthermore, there is no evidence that under-recognized employees experience any reversion in future bonuses. If that were the case, then the argument could be made that there is no real monetary compensation at play as managers would instead be borrowing from the employee's future allotment to try to signal amends in the present. Instead, there is evidence that, if anything, managers continue to reward under-recognized employees for under-recognition given future opportunities (See Table 5). Finally, there is no evidence of any statistical differences in

[7] Using bonus multiplier (percent of base salary) is the most general common denominator across the two comparison groups. Managers are unable to supersede bonus multiplier ceilings within each employees rating category designation. In other words, managers may want to give under-recognized employees even *more* in bonuses as recompense for under-recognition but are limited by the predetermined multiplier ceilings which depend on the rating received. Given that the ranges for bonus multipliers overlap significantly (See Table AX2), an under-recognized employee making the same in annual salary as an exceptional+ employee can still receive a 4 percentage point greater bonus multiplier while both employees still receive bonus multipliers that fall within their respective rating ranges. Analysis below is robust to using either multiplier percent of annual salary or absolute nominal amount received.

changes to annual base salary and allowances in 2017 depending on whether an employee was Exceptional+, under-recognized, or always rated Solid, controlling for whether promoted or not (for instance, the average salary increase for Exceptional+ was about $7,000, while for under-recognized it was about $21,000, t-test: -0.337; see also Appendix Table AX3).

Ruling out Career, Signaling, or Reputation Concerns

In PharMed, promotions do not rest on the results of overall performance ratings. There is no requirement that an individual receive an Exceptional+ rating to be eligible for promotion. Rather, promotions are based on performing well the work that pertains to a higher pay grade or exhibiting the potential to be stretched into a new role. All promotion and job level adjustments are kept separate from the formal performance evaluation processes, and PharMed maintains a distinct, standalone promotion administration system. The separation of these two processes is a key feature in the calibration sessions for year-end performance ratings. Discussing promotion potential is strictly prohibited in calibration sessions. Prior to the start of calibration every year, all HR facilitators undergo formal calibration policy preparation. A key focus of this formal preparation is to reinforce the policy that promotion and job level discussions do not enter into calibration discussions. HR facilitators are also required to notify all calibration participants in advance of calibration of this policy.[8]

Furthermore, PharMed "freezes" the administrative system which processes any promotion or job level adjustment for two months during the performance evaluation period. During this period, all managers are unable to access the system to initiate any job level changes for their employees. This pertains to any internal job movement. All employees are evaluated solely based on their performance in their current role and managers are unable to adjust anything about employees' jobs until after the performance evaluation process is complete and compensation planning for the coming year has concluded.

[8] These facts are emphasized in the performance management handbook and calibration preparation material provided and reviewed with all HR facilitators prior to the start of the year-end performance evaluation process.

Still there does remain a propensity to discuss different ways of managing the need to calibrate some top performers from Exceptional+ to Solid ratings, including the use of near-term promotions (similar in motivation to the effort by managers to try to recompense the under-recognized with large bonuses). Indeed, in every calibration session (18 in total) that I observed, the matter of at least one individual's likelihood of soon being promoted was raised. In every case, however, the HR facilitator immediately stepped in to foreclose further discussion of promotion and remind the manager about the policy and material shared prior to the calibration session. In no occasion did I observe the admission of whether or not an employee merited a near-term promotion or job change as relevant information to the calibration discussions.

In addition, in all my interviews, managers, as well as employees who receive the Exceptional+ rating, describe the arbitrariness of the distinction when it comes to career progression.[9] Teams and peers discuss the common strategies that they employ for assuring employees who receive a Solid but are known to perform exceptionally well. These include offering "verbal congratulations," promises and assurances that continued effort of the current caliber will certainly merit Exceptional+ rating the following year, increased frequency of informal conversations about their exceptional level of work, and giving them more opportunities to display quality of work in internal and external forums and presentations. Nearly all discussions about assuring high performing employees who receive Solid include emphasizing the arbitrariness of not receiving an Exceptional+ rating or how it was generated by a random divide given the tight enforcement of such a strict Exceptional+ rating percentage cutoff.

Despite all such safeguards and assurances, however, there is a higher incidence of promotions among the Exceptional+ population than among the under-recognized the following year. The percent of Exceptional+ employees who receive a 2017 promotion (10.34%) is greater than that among those who are under-recognized (5.39%), but only slightly statistically significant (t-test:2.25, $p<0.05$). The difference is

[9] The majority of my interview data comes from this population. Given the high exit rates among the under-recognized and the sensitivity around the event, I was only able to conduct one formal interview with an under-recognized individual before she eventually departed PharMed.

largely driven by the fact that under-recognized employees exit before any promotion would take effect**.** Moreover, the statistical difference disappears once controls are introduced into the logit models estimating the odds of promotion (See Table 6). Over the following year, the promotion rates equalize. Excluding employees who were promoted in 2017, 11% of both Exceptional+ and under-recognized employees are promoted (t-test: 0.094)[10]. Finally, even when promoted, Exceptional+ employees did not experience an increase in salary and allowances compared with the under-recognized the following year.

The above evidence supports the conclusion that among the population of nominated top performers, under-recognition in the form of receiving a Solid rather than Exceptional+ rating is arbitrary and that there are no concordant confounds that would cause negative reactions other than that of reacting only to the discrepancy of rating recognition. In the following section, I turn to the results showing under-recognition indeed causes costly turnover among top performers. These results are generated using logistic regressions with coefficients reported in odds ratio for ease of interpretation. All findings are robust to using Ordinary Least Squares linear regression specifications.

MAIN RESULTS

Nominal Under-Recognition Increases Exit

To demonstrate how under-recognition demotivates top performers even in the absence of bias or prejudice in the recognition bestowal as well as in the absence of any material or career, signaling or reputation concerns, I examine the differences in the likelihood of exiting the organization depending on whether a top performer receives an Exceptional+ rating or whether they are instead under-recognized. *Ex ante*, there are no differences between employees who receive an Exceptional+ rating for the year and those who are designated solid due to the limitation of Exceptional+ ratings available to bestow. Yet whether a top performer is arbitrarily on the under-recognized margin for these ratings has significant consequences for their career and the organization.

[10] Including those who were promoted the previous year generates 10.1% of Exceptional+ being promoted to 10.5% of under-recognized promoted (t-test: -0.120).

Consistent with the expectation that under-recognition will lead to an increased likelihood of exiting the organization, I find that under-recognized employees are more likely to exit the organizations within 18 months of under-recognition (Table 7, models A-D). Specifically, the odds of a top performer exiting the organization within 18 months after an overall performance evaluation is about 1.72 times greater when than they are under-recognized (Model 7A; $p<0.01$), than if they receive an Exceptional+ rating for a year of high performance.

Sensitivity to Relative to Top-Performing Peers Drives Under-Recognition Effect

To test whether this effect is driven by the number of other top performers on the same team, Model 7B controls for this team factor. Interestingly, the main effect for the number of other top performers on the same team decreases the odds of exiting among the nominated population of top performers. The number of other top performers on the same team is an important factor for the likelihood a top performer is under-recognized (See Tables 2 and 4). However, for employees who receive an Exceptional+ rating, working on a team with more top performers like themselves lowers the odds of exiting the organization. For every additional top performer on the same team, an Exceptional+ employee has about 0.68 lower odds of exiting the organization within 18 months (Model 7B; $p<0.001$). Including this nominal factor also slightly reduces the odds that an under-recognized employee exits the organization to a difference of about a 1.52 times greater likelihood than other top performing employees who receive an Exceptional+ rating (Model 7B; $p<0.01$).

Considering that the impact of under-recognition is likely conditional on the proximate comparison set of other top performers and their relative recognition, I test for an interaction effect between under-recognition and the number of other top performers on the same team. Indeed, I find that it is the sensitized relative comparison to other top performers that drives the increase in exit likelihood among under-recognized top performers. For under-recognized top performers, working in a team with one additional top performer results in a 1.47 times greater likelihood of exiting the organization within 18 months (Model 7C; $p<0.01$) than an Exceptional+ peer with the same number of top performing peers.

In this model (7C), the main effect of under-recognition loses statistical significance and changes valence. This points to the decisive impact of under-recognition being the combined effect of receiving a Solid rating in a year of exceptional performance in the context of working in a team where a peer *was* recognized with an Exceptional+ rating for the same level of performance.[11] In other words, it is the stark contrast with recognized peers which makes under-recognition unbearable.

No Difference in Under-Recognition Effect Depending on Monetary Compensation

Controlling for the size of the bonus that managers reward both the under-recognized and Exceptional+ employees, the increased likelihood of exit given under-recognition still holds (Model 7D). Under-recognized employees are still 1.65 more times more likely to exit within 18 months than their Exceptional+ counterparts ($p<0.05$). The coefficient for bonus size in this model conveys no directional difference in the impact of increased monetary reward. Overall, then, the managerial strategy of giving under-recognized employees the largest possible bonuses is futile in the effort to prevent damage to their self-image and from becoming demotivated. To the under-recognized, it is an irrelevant factor.

When controlling for the size of the bonus received, the statistical significance of the explanatory power of under-recognition on exit is reduced. However, that the significance of the effect of under-recognition on exit is reduced when controlling for bonus is largely explained by the fact that this model relies on the sample of employees still employed when the bonus payout occurred. This means it excludes the employees who left right after the ratings were revealed to employees (the overall sample size reduces by roughly 11%). Financial Directors at PharMed have assured me that most of these employees who exited very soon after receiving their year-end performance ratings still receive the bonuses they earned for the prior year; the financial reports summarizing 2016 bonus payouts also reflect this. However, I am

[11] That top-performing peers are nominated for the same level of performance follows from the managerial strategy of preventing *reverse* recognition among their teams. Knowing that calibration decisions are out of their hands after their level's meeting, they report being sure to only nominate employees for the same top performer rating if they themselves are unable to distinguish their relative contributions as exceptional employees. Any factor with which to objectively make this distinction among their direct reports, helps the manager explain whatever the final outcomes are of the performance evaluation process in that they have objective criteria to point to so long as the relative distinctions submitted survive through the calibration process.

limited in my ability to match employees to their exact bonus sums from the administrative compensation records if they exit the organization before these compensation reports are generated.

Still, this reduction in sample size matters when examining the two main contributing factors for the effects of under-recognition: proximity to Exceptional+ employees and the managerial efforts to compensate the under-recognized with outsized bonuses, together in the same model. As shown by Model 7D, the coefficient for under-recognition still reflects an increased likelihood that the under-recognized are more likely to exit the organization within 18 months of receiving a Solid rating for a year of exceptional performance. However, the under-recognition coefficient is no longer significant in this full model.

The coefficient for the interaction term for under-recognition by the number of other Exceptional+ employees on the same team becomes insignificant in the model with the bonus rewards (staying in the same direction, though only slightly). Given a large factor in the differences between these two models (with and without controls for the bonus reward) is the reduction in sample size of leavers who exit very soon after performance evaluations is important for interpreting this change. Logically, the difference can be interpreted as reflecting the likelihood that those who were most sensitized to the under-recognized by being on teams with top performers who received the performance rating recognition, were the most likely to leave the fastest. Clearly, no matter what their manager was able to give them as a bonus to recompense their under-recognition, it was not enough to keep them from exiting.

Sour Grapes at Play

While the experimental nature of the under-recognition treatment clearly demonstrates the increased likelihood of exit based on the arbitrary assignment to a Solid rating from a top performer designation, the underlying explanatory mechanisms are not experimentally testable in this setting. However, supporting evidence supports the conclusion that under-recognized employees are making ritualistic rather than rational adjustments given under-recognition.

Most important is the fact that PharMed pays some of the highest wages for all levels and functions of the business across their local labor markets. Evidence for this comes from internal reports covering a

recent ‘job leveling’ program PharMed begun in 2018. Like many large companies, PharMed participates in industry compensation surveys administered by well-known business consultancies in return for market benchmarks. The persistent evidence that compensation packages were well-above market averages, recent efforts are underway to bring them more in line with the market. To this end, PharMed has begun restructuring compensation grade assignments to align wage rates and job levels to local market competitors. What is important to note, however, is that for the period covered by this paper, compensation packages for all business functions and all levels offered by PharMed far outpaced these competitors, allowing PharMed to attract some of the industry’s most qualified talent.

Despite their comparatively high pay, most leavers join less competitive employers in the same area. Indeed, only 5.65% of all top performer leavers exit the metropolitan area where they worked for PharMed. There is no difference in this rate by under-recognition (under-recognized move to another employer outside of their prior geography 7.6% of the time, t-test= 0.623, $p<0.267$). As such, all evidence suggests that these employees are not leaving to earn more money elsewhere. Rather, these other organizations must offer more attractive alternatives satisfying these employees’ newly updated preferences.

No Significant Heterogeneous Treatment Effects by Demographic Groups

While the focus of this paper is the main effect of under-recognition across the entire population that experiences this treatment in the 2016 performance evaluation process, it is important to test whether there may heterogeneous treatment effects across treated sub-groups as well. Whether or not different populations react similarly to the experience of under-recognition informs how general the self-image preservation through exit is as a response. Table 8 reports the representativeness of the treated population and exit groups by age, female, and minority status. There are no statistically significant differences across these three statuses in exit likelihood post-under-recognition in the 2016 performance evaluation period. Among the under-recognized, 23% of the minority employees left the organization compared with 21% of non-minorities (t-test: 0.229; $p<0.409$). The difference between males and females is slightly larger with a

5.6-point lower share of females who exited than males; however the difference is non-significant (t-test: 1.125; $p<0.131$). Future research will examine the strength these non-significant treatment effects by demographic groups using a larger data set including under-recognized employees from PharMed's 2017 performance evaluation process.

ROBUSTNESS CHECKS

Voluntary Turnover Only

Given the sample size limitations when including the important controls for monetary compensation in the outcomes of the performance evaluation process, the models above test the effects of under-recognition on all categories of exit (both voluntary and involuntary). In support of the main results reflecting voluntary reactions at self-image preservation, Table 9, replicates the findings for the main effect of under-recognition using only voluntary exit as the dependent variable (Models 9A and B; See also Figure 2).

The main differences between modelling all forms of top performer exit versus only voluntary exit stem from the influence of proximate Exceptional+ peers. Unlike in the main models, when examining only voluntary exit outcomes, the coefficient on the number of peer top performers on the same team loses importance. The number of other top performers has no directional impact on likelihood of voluntary exit. One interpretation of the difference in impact the number of other top performers has on voluntary versus all exits is that under-recognized employees are more likely to exhibit problematic behavior when working with peers who receive the recognition they are denied. The large difference in performance citations and position eliminations among the under-recognized with at least one other top performer on their team lends some support to this conjecture (See Table 10).

Comparing Under-Recognized to Other Solid-rated Employees

The treatment of under-recognition is applied to a relatively homogenous peer group of top performing employees. Therefore, the effects of under-recognition are likely driven by the contrast in the experience of those arbitrarily denied recognition relative to the top performers receiving the merited recognition. However, it is important to test that the mechanism motivating under-recognized employees to exit the

organization is one of deprivation costs relative to their recognized top performing and not a main effect of receiving a Solid rating generally. In other words, even though a large majority of PharMed employees receive a Solid rating every year, it is still important to rule out that the higher turnover rate among the under-recognized is simply driven by receiving Solid in general rather than the denial of a merited Exceptional+ rating.

The simplest evidence that the propensity to exit the organization is higher among under-recognized compared to other Solid employees is a comparison of means. The overall exit rate for under-recognized employees is 22 percentage points greater than employees whose ratings remained Solid throughout the 2016 performance evaluation process (t-test: 17.296; $p<0.0000$). Similarly, under-recognized performers are 13 percentage points more likely to voluntarily exit than other employees rated Solid in 2016 (t-test: 7.637; $p<0.0000$; see Figure 2).

Including all demographic and prior performance controls in a logistic regression predicting odds of exit in Table 11, under-recognized performers have 7.65 times greater odds of leaving under any exit category than comparable peer employees who were rated solid throughout the 2016 performance evaluation period (Model 11C; $p<0.001$). As in the comparison against Exceptional+ peers, this difference drops somewhat when looking only at voluntary turnover to 3.53 times greater odds of exit for under-recognized as compared to other Solid-rated employees (Model 11F; $p<0.001$). In sum, these results strongly underscore that the self-image preservation mechanism of exiting the organization post-under-recognition is not an artifact of reacting just to the experience of receiving a nominal Solid rating. What matters is that under-recognized performers delivered exceptional performance in the prior year that merited an accordingly high rating but were arbitrarily denied such recognition while some of their peers did receive the commensurate Exceptional+ rating for top performance.

Not an Unfortunate By-Product

To consider the full effects of the costs of under-recognition it is important to demonstrate that the costs are not simply byproducts to a motivational schema that actually is effective at incentivizing more

productivity overall. In other words, even though the results above demonstrate a much greater likelihood of exiting among a significant portion of PharMed's top performing employees, they do not rule out any positive overall productivity outcomes from restricting the quantity of Exceptional+ ratings for top performance.

Given that this study looks at employees across all business functions of PharMed, common productivity proxies are scarce. One way to assess whether the restriction of Exceptional+ ratings may have positive motivational effects, however, is by examining the year-over-year dynamics of rating progressions across the various rating categories. If there were noticeable differences in the pattern of future ratings between the under-recognized group and those who received Exceptional+ ratings, there may be added reason to believe there could be positive motivational byproducts to the rating system, net of the selection effects from the under-recognized performers exiting at higher rates.

In the results from multinomial logistic regressions, there is no evidence that there is a difference between employees who received Exceptional+ ratings in 2016 versus those who were under-recognized (See Appendix Table AX4). Conditioning on employees who received a rating of Solid in 2015, the experience of under-recognition in 2016 does not generate any difference in odds of receiving and Exceptional+ rating in 2017 (conditioning on the 2015 rating ensures that the only performance rating difference is the 2016 outcome and not potentially different trajectories). In fact, there are no differences in odds of attaining any 2017 rating category between the Exceptional+ population and the under-recognized. Both groups are more likely to be nominated to Exceptional+ ratings in 2017 by their manager as well as successfully receive this top performance recognition than employees who received Solid ratings in both 2015 and 2016 (see also Figures AX2 and AX3).

Of course, a more robust test of whether the restriction of Exceptional+ ratings has positive productivity implications overall requires comparing objective productivity measures of employees under such tight restrictions of Exceptional+ ratings versus a looser supply of such ratings. Future research, focusing specifically on employees in the Commercial function of PharMed (where quarterly sales metrics offer objectivity to productivity measurement) and employees in Research & Development (R&D)

function (where productivity measures by individual are less precise but where the importance to PharMed's core strategy is critical), promises progress on this dimension. This future examination will also generate insights regarding the role team dynamics play in the impact of individual top performers' experiencing relative under-recognition as Commercial personnel operate independently while most R&D work is conducted within project teams.

SUMMARY

In this paper, I provide novel evidence that recognition schemes carry serious inherent risks. I show that employees who do not receive a performance rating that is commensurate with their contributions are much more likely to exit the organization than their recognized peers. There are no differences ex-ante among the top performer population that predict the likelihood of being denied recognition in the year-end performance evaluation process. Rather, under-recognition, in this case, is but a by-product of the organization's pay-for-performance scheme which restricts Exceptional+ ratings to a predetermined small percentage of the overall company population.

When denied recognition that is commensurate with merited high achievement, top performers are more likely to exit the organization despite receiving greater performance reward bonuses than any other employee and experiencing significantly higher wage growth after under-recognition as well. After adjusting for the increased exit rate among the under-recognized top performers, their prospects of future promotion also match that of their recognized peers. Moreover, managers and peers recognize the arbitrariness of the denial of recognition in this performance rating form, accepting lower performance bonuses as recompense to the under-recognized and employing mitigation attempts to preserve the self-image of those denied official rating recognition for superior performance.

Despite such monetary over-compensation and substantive peer assurances, under-recognized top performers are likely to still feel a loss of face when receiving a formal rating that does not equate with their demonstrated high performance. In fact, such recompense all but forecloses any other option to respond to the self-image discrepancy but exiting the organization. Under-recognition causes top

performers to reassess the relative costs of staying in an organization that produced this threat to self-image versus the moving to a competing organization even if the transition comes with a loss in overall monetary compensation. That I find this self-image preservation reaction is consistent across all under-recognized performers, regardless of demographic suggestions this is a general, if not universal, response.

DISCUSSION & IMPLICATIONS

Retaining key talent is top priority for firms. Prior studies have looked at the consequences of top talent mobility for recipient firms (Groysberg, Lee and Nanda 2008) as well as the value of keeping top talent through mechanisms such as noncompete agreements (Younge and Marx 2016). Such studies have shown that firms that attract talent away from other firms may experience a modicum of performance improvement. However, the loss of star performers can represent a major joss of productivity. Furthermore, motivating employees is a critical element of competitive organizational strategy. Employees will work harder for tangible nonmonetary rewards, such as paid vacations or high-end merchandise, than they would for the commensurate market value of the reward. Tangible nonmonetary rewards are more memorable, more pleasurable to anticipate and share with others, and do not come with purchase justification costs (Jeffrey 2004; Jeffrey and Shaffer 2007). Yet, while there is growing evidence that nonmonetary recognition holds sway over organizational members' motivations, the potential for recognition to come with social comparison costs have gone largely unexamined.

Evidence suggesting that non-recognized parties in nonmonetary recognition schemes may improve their own efforts while bearing no ill-will against the winners of recognition (e.g. Ammann et al. 2016; Bradler et al. 2016; Ehrenberg and Bognanno 1990) has led to the belief that recognition schemes will inherently avoid the undesirable social comparison costs pervasive in compensation schemes (Frey and Neckermann 2008; Neckermann et al. 2014; Moldovanu et al. 2007; Kosfeld and Neckermann 2011). However, to be effective, recognition schemes must distinguish few among many as worthy of special esteem for delivering exceptional contributions. This means that for recognition to be effective, there must always be participants or members who do not receive similar esteem or recognition.

Given that, at the margin, the performance of under-recognized top performers is indistinguishable from their recognized counterparts, the design of such reward schemes has serious unintended risks regarding talent retention and motivation strategies. This study demonstrates that organizations may inadvertently encourage a significant portion of their top performers to favor working for competitors via strategies meant to motivate their workforce to perform exceptional work. Under-recognition inadvertently makes competing firms more attractive to their top performers under "the perpetual illusion that 'the grass is always greener on the other side of the fence'" (Hirschman 1970: 27).

Companies face the perennial challenge of accurately assessing talent and making sure there are substantive rewards for high productivity and real downsides for sub-par performance. In an effort to solve ratings-inflation tendencies (Pfeffer 1997; Grote 2005; Pfeffer and Sutton 2006), companies frequently make the mistake of applying a forceful rating system universally across all critical business units. Success from relying such a talent management scheme based primarily on forced distribution of rankings is highly contingent.

It is important to note the especially significant consequences of these finding for firms that rely on collaborative output for their market strategy. Because this organization applies the same calibration system across the entire company, functions that are especially team-dependent may experience more intensive treatment effects as not being able to reward teams as a unit for exceptional work generated collaboratively may have considerable spill-over effects (see Siegel 2008 for a similar case of the unintended consequences of applying a forced ranking distribution across an entire company population also in the research and development space). Newsworthy cases such as how drastically Microsoft's copying of GE's stacked ranking model backfired especially because of their mistake in applying the method on Operating System engineers building one of the largest collaborative consumer products there is attests to this fact (Impraise.com 2016). When top performers are irreplaceable and their work depends on collaborating with peers, proactively generating relative comparisons, especially when rankings are ordinal and differentiation becomes arbitrary, is sure to backfire painfully.

Finally, whether or not there are any demographic differences in the sensitivity and response to under-recognition deserves further investigation. The present study is limited by the sample size of different demographic groups among the under-recognized population to definitively test such differences. However, whether different groups, such as females, are more or less sensitive to the self-image impact of under-recognition can have important inequality implications. In the case of the employees studied here, leavers from under-recognition circumstances most likely *did not* go to *more* competitive companies, especially if they stayed within the industry, complicating the relative assessment of value of staying versus leaving. Moreover, stayers among the under-recognized experienced wage and bonus increases above their peers. To make progress on such inequality concerns, further exploration on this question would do well to expand to settings where different market dynamics can be explored.

REFERENCES

Adams J (1963) Towards an understanding of inequity. *J. Abnormal Soc. Psych.* 67(5):422–436.

Adams, J.S. 1965. Inequity in social exchange. *Advances in Experimental Social Psychology*. 62, 335-343.

Akerlof, G.A. and Yellen, J.L., 1990. The fair wage-effort hypothesis and unemployment. *The Quarterly Journal of Economics*, *105*(2), pp.255-283.

Ariely, D., 2016. Payoff: The hidden logic that shapes our motivations. Simon and Schuster.

Ashraf, N., Bandiera, O. and Lee, S.S., 2014. Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization*, 100, pp.44-63.

Baker G, Gibbs M, Holmstrom B (1994) The internal economics of the firm: Evidence from personnel data. *Quart. J. Econom.* 109(4):881–919.

Besley, T. and Ghatak, M., 2008. Status incentives. *American Economic Review*, 98(2), pp.206-11.

Bewley, T., 1999. Why don't wages fall in a recession.

Bowles, S. and Polania-Reyes, S., 2012. Economic incentives and social preferences: substitutes or complements?. *Journal of Economic Literature*, 50(2), pp.368-425.

Bradler, C., Dur, R., Neckermann, S. and Non, A., 2016. Employee recognition and performance: A field experiment. *Management Science*, 62(11), pp.3085-3099.

Brennan, Geoffrey and Pettit, Philip. 2004. The Economy of Esteem: An Essay on Civil and Political Society. Oxford, UK: Oxford University Press.

Breza, E., Kaur, S. and Shamdasani, Y., 2017. The morale effects of pay inequality. *The Quarterly Journal of Economics*, 133(2), pp.611-663.

Brown D, Armstrong M (1999), Paying for Contribution: Real Performance-Related Pay Strategies, Konan Page: London

Burt, Ron. 1982. Toward a structural theory of action: Network models of social structure, perception, and action. Academic Press.

Card, D., Mas, A., Moretti, E. and Saez, E., 2012. Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, 102(6), pp.2981-3003.

Charness, G. and Grosskopf, B. (2001). Relative payoffs and happiness: An experimental study. *Journal of Economic Behavior and Organization*, 45:301–328.

Cohen-Charash Y, Mueller J (2007) Does perceived unfairness exacerbate or mitigate interpersonal counter productive work behaviors related to envy? *J. Appl. Psych.* 92(3):666–680.

Cohen-Charash, Y. and Spector, P.E. 2002. "The role of justice in organizations: A meta-analysis" *Organizational behavior and human decision processes*, 86(2), 278-321.

Deci EL, Koestner R, Ryan RM. 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125(6): 627–668.

Deci EL, Ryan RM. 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press: New York.

Dube, Arindrajit, Laura Giuliano, and Jonathan S. Leonard. "Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior" NBER Working Paper 24906 August 2018.

Ehrenberg RG, Bognanno ML (1990) Do tournaments have incentive effects? *J. Political Econom.* 98(6):1307–1324.

Elster, John. 1985. Sour Grapes: Studies in the Subversion of Rationality. Cambridge University

Press.
Festinger L (1954) A theory of social comparison processes. *Human Relations* 7(2):117–140.
Freeman, R. and Gelber, A. (2010). Prize structure and information in tournaments: Experimental evidence. *American Economic Journal: Applied Economics*, 2(1):149–64.
Frey, B.S. 1997. Not Just for the Money. Edward Elgar Publishing: Cheltenham, UK and Brookfield, WI.
Frey, B.S. and Gallus, J., 2017. Honours versus money: The economics of awards. Oxford University Press.
Frey, B.S. and Jegen, R., 2001. Motivation crowding theory. *Journal of Economic Surveys*, 15(5), pp.589-611.
Gallus, Jana and Frey, Bruno S. 2016. Awards: A Strategic Management Perspective. *Strategic Management Journal*. 37: 1699-1714.
Gneezy, U., S. Meier, and P. Rey-Biel, 2011. "When and why incentives (don't) work to modify behavior." *Journal of Economic Perspectives*, 25: 191–210.
Goffman, E., 1967. On face-work. Interaction ritual, pp.5-45.
Goode William, J., 1978. The Celebration of Heroes: Prestige as a Control System.
Grant, A. M., 2007. "Relational job design and the motivation to make a prosocial difference." *Academy of Management Review*, 32: 393-417.
Groysberg, B., 2010. Chasing stars: The myth of talent and the portability of performance. Princeton University Press.
Groysberg, B., Lee, L.E. and Nanda, A., 2008. Can they take it with them? The portability of star knowledge workers' performance. Management Science, 54(7), pp.1213-1230.
Gubler, Timothy & Larkin, Ian & Pierce, Lamar. (2016). Motivational Spillovers from Awards: Crowding Out in a Multitasking Environment. *Organization Science,* 27(2), pp. 286-303.
Gurr, T., 1970. Why men rebel. Princeton, N.J.: Published for the Center of International Studies, Princeton University [by] Princeton University Press
Hirsch, Fred. 1976. Social limits to growth. Harvard University Press.
Holmstrom, B. and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organization* 7: 24-52.
Holmstrom, B. 1999. "The Firm as a Subeconomy." *Journal of Law, Economics and Organizations* 15: 74-102.
Holmstrom, B. 1982. "Managerial Incentive Problems—A Dynamic Perspective." In Essays in Economics and Management in Honor of Lars Wahlbeck. Helsinki: Swedish School of Economics. Republished in Review of Economic Studies 66 (1999): 169-82.
Homans, G.C., 1974. Social behavior: Its elementary forms.
Impraise.com, 2016. "Microsoft Throws Stack Ranking Out the Window" https://blog.impraise.com/360-feedback/microsoft-throws-stack-ranking-out-the-window
Jeffery, Scott. 2004. "The Benefits of Tangible Non-Monetary Incentives'. Working Paper, University of Chicago.
Jeffrey, Scott A. and Shaffer, Victoria. 2007. "The Motivational Properties of Tangible Incentives". *Compensation Benefits Review* 39:44-50.
Jurgensen, C.E., 1978. Job preferences (What makes a job good or bad?). *Journal of Applied psychology*, 63(3), p.267.

Kerr, Steven. 1995. On the folly of rewarding A, while hoping for B. *The Academy of Management Executive*. February: 9(1) 7-14.

Larkin I. 2007. The cost of high-powered incentives: employee gaming in enterprise software sales. HBS Working Paper. Harvard Business School, Boston, MA.

Larkin I, Pierce L, Gino F. 2012. The psychological costs of pay-for-performance: implications for the strategic compensation of employees. *Strategic Management Journal* 33(10): 1194–1214.

Lazear, E. (1989). Pay equality and industrial politics. Journal of Political Economy, 97(3):561–580.

Lepper, M. R., and D. Green, 1978. The Hidden Costs of Reward: New Perspectives on the Psychology of Human Behavior. Hillsdale, NY: Erlbaum

Lind EA, Tyler TR (1988) The Social Psychology of Procedural Justice (Springer Science & Business Media, Berlin).

Lynn, Freda B., Joel M. Podolny, Lin Tao. 2009. "A sociological (de)construction of the relationship between status and quality." *American Journal of Sociology.* 115(3): 755-804.

Martin J. 1981. Relative deprivation: a theory of distributive injustice for an era of shrinking resources. In Research in Organizational Behavior: an Annual Series of Analytical Essays and Critical Reviews (Vol. 3), Cummings LL, Staw BM (eds). JAI Press: Greenwich, CT; 53–107.

McGraw, A.P., Mellers, B.A. and Tetlock, P.E., 2005. Expectations and emotions of Olympic athletes. Journal of Experimental Social Psychology, 41(4), pp.438-446.

Merton, Robert K. 1968. "The Matthew Effect in Science." Science 159(3810): 56-63.

Moldovanu, B., Sela, A. and Shi, X., 2007. Contests for status. *Journal of Political Economy*, 115(2), pp.338-363.

Neckermann, Susanne, Cueni, Reto, and Frey, Bruno S. 2014. 'Awards at Work'. Labour Economics 31: 205-17.

Nelson B. 2005. 1001 Ways to Reward Your Employees. Workman Publishing: New York.

Nickerson JA, Zenger TR. 2008. Envy, comparison costs, and the economic theory of the firm. *Strategic Management Journal* 29(13): 1429–1449.

Obloj, T. and Zenger, T., 2017. Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, 28(1), pp.1-18.

Oyer P (1998) Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality. *Quart. J. Economic*. 113(1): 149–185.

Pearson, T. (2015). The end of jobs: Money, meaning and freedom without the 9-5. Lioncrest Publishing.

Pink, D. H. (2009). Drive: The surprising truth about what motivates us. New York, NY: Riverhead Hardcover.

Pfeffer, J. and Langton, N., 1993. The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly*, pp.382-407.

Pfeffer, J. (1998), "Six dangerous myths about pay", Harvard Business Review, May-June issue.

Pfeffer, J. and Sutton, R.I., 2006. Hard facts, dangerous half-truths, and total nonsense: Profiting from evidence-based management. Harvard Business Press.

Ricoeur, P. A. (1965). Fallible Man. Chicago, Henry Regnery.

Rynes, S.L., Gerhart, B. and Minette, K.A., 2004. The importance of pay in employee motivation: Discrepancies between what people say and what they do. Human

Resource Management: Published in Cooperation with the School of Business Administration, The University of Michigan and in alliance with the Society of Human Resources Management, 43(4), pp.381-394.

Sauermann, H., and W. Cohen, 2010. What makes them tick? Employee motives and firm innovation. *Management Science,* 56: 2134-2153.

Schwartz, B. 2015. Why we work. TED/Simon & Schuster.

Siegel, J., 2008. Global talent management at Novartis. Harvard Business School Pub..

Sinek, S., Mead, D., & Docker, P. (2017). Find your why. New York, NY: Portfolio/Penguin.

Silverman, M., 2004. "Non-financial recognition: The most effective of rewards." Institute for Employment Studies Research Network.

Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A. and Williams Jr, R.M., 1949. The American soldier: Adjustment during army life.(studies in social psychology in world war ii), vol. 1.

Strack, R. (2014). Rainer Strack: The workforce crisis of 2030 — and how to start solving it now. Retrieved from https://www.ted.com/talks/rainer_strack_the_surprising_workforce_crisis_of_2030_and_how_to_start_solving_it_now/up-next Tyler 2003 Psychology of Procedural Justice.

Strauss G. 1955. Group dynamics and intergroup relations. In Money and Motivation: An Analysis of Incentives in Industry, Whyte WF, (ed). Harper & Row: New York; 90–96.

Younge, K.A. and Marx, M., 2016. The value of employee retention: Evidence from a natural experiment. Journal of Economics & Management Strategy, 25(3), pp.652-677.

Exhibit 2. Full Performance Evaluation Process (OPR: Overall Performance Rating)

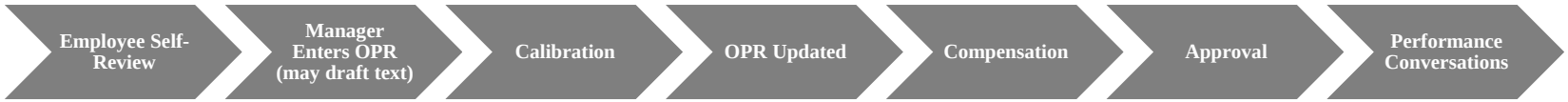


Exhibit 1. How PharMed calibration process results in 20% of highest performing employees (whose performance warrants an "Exceptional" rating) becoming arbitrarily under-recognized

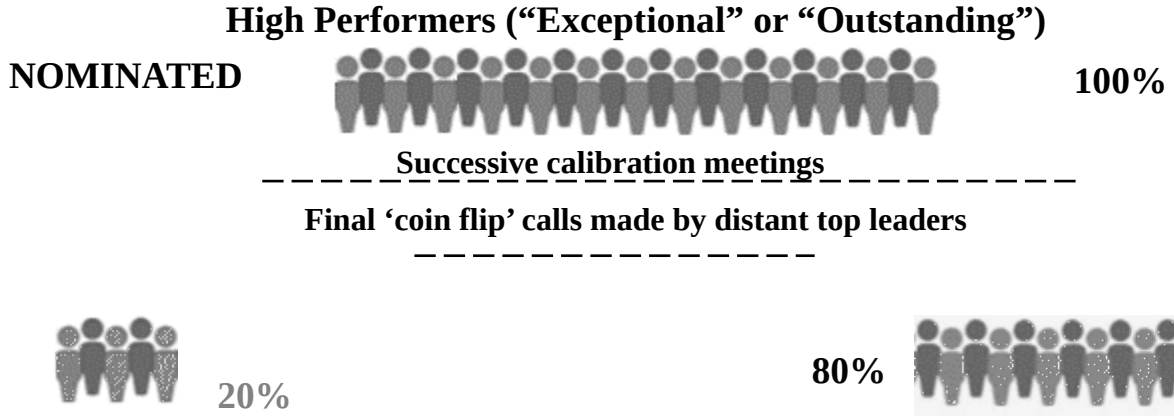
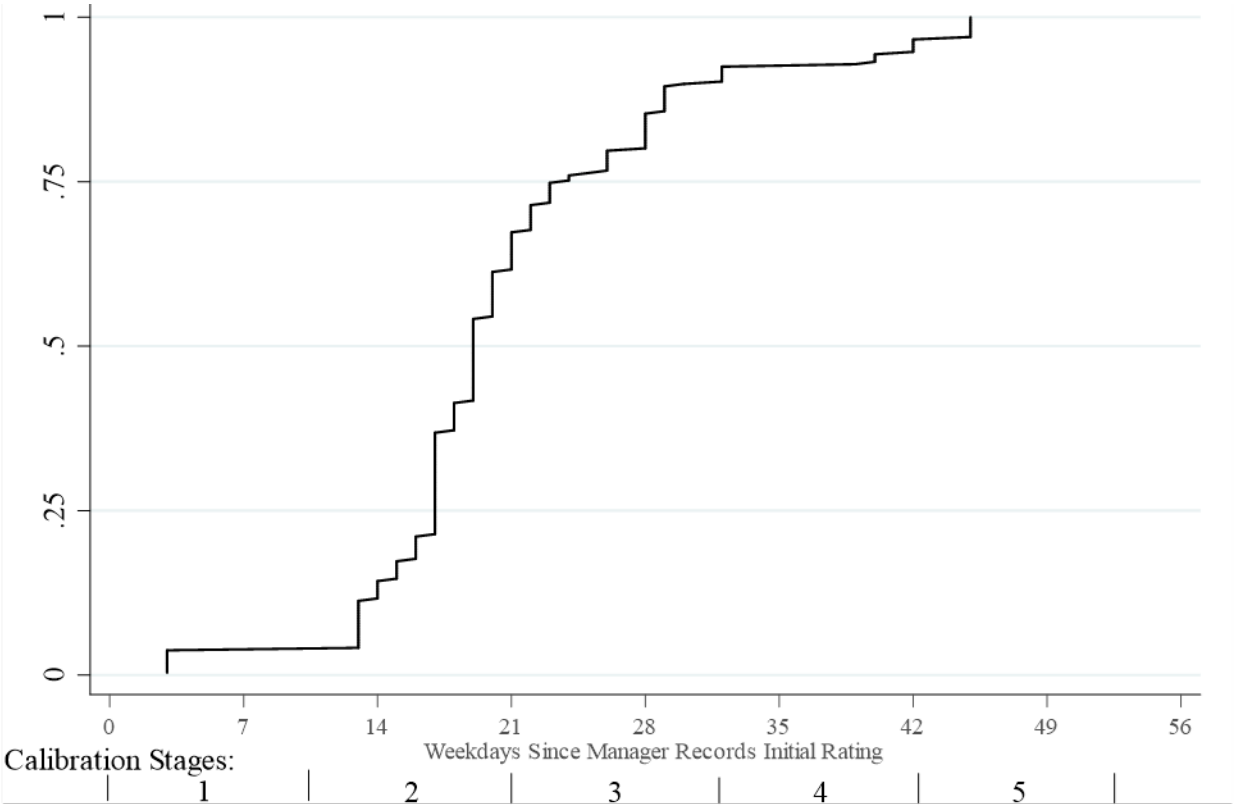


Figure 1. Cumulative distribution of the timing a Top Talent employee goes from an Exceptional+ rating to Solid, in weekdays since manager entered first rating.



Note: while all managers of director reports are required to enter ratings and evaluations for their employees by the same date a few weeks after the performance review period becomes open, the strictest requirement is that this occurs at least a day before the calibration meeting covering that employees' level (so the HR facilitator can organize the spreadsheets used for calibration). Therefore, this cumulative distribution illustration depicting the timing that employees ratings change from Exceptional+ to Solid are normalized where timing begins when their manager first enters their rating into the administrative system. The calibration period is a rough approximation based on the amount of time the total calibration process takes to proceed from calibrating individual contributor employees up through the last level where senior vice presidents and chief officers calibrate all levels down.

Table 1. Descriptive Statistics

	Top Performer Population					**Other Employees**			
	[mean]					[mean]	[std. Err.]	[mean]	[std. Err.]
	Exceptional +	**Under-Recognized**	**| Diff |**	**P-value**		**Solid**	**Solid**	**Low Performer**	**Low Performer**
Age (years)	41.2	42.8	1.611	0.002	***	42.9	(8.891)	43.4	(8.798)
Team size (n)	5.63	5.66	0.04	0.46		6.03	(3.571)	5.76	(4.194)
Tenure, years	5.41	4.39	1.02	0.0003	***	5.73	(4.959)	5.42	(4.842)
Female	0.50	0.44	0.063	0.04	*	0.49	(0.500)	0.53	(0.502)
Female Manager	0.37	0.39	0.017	0.31		0.37	(0.483)	0.29	(0.456)
Asian	0.13	19.05	0.05	0.043	*	0.15	(0.355)	0.15	(0.363)
Black	0.06	2.72	0.03	0.062		0.11	(0.318)	0.26	(0.444)
Hispanic/Latino	0.03	4.76	0.01	0.23		0.04	(0.193)	0.00	(----)
Other Race/ 2+Races	0.01	0.00	0.02	0.043	*	0.02	(0.125)	0.02	(0.147)
Number of Excpetional+s on team	1.88	2.16	0.28	0.0001	***	1.85	(1.00)	1.85	(1.057)
Bonus Percent for 2016,%base salary	15.85	19.81	3.97	0.0000	***	14.50	(8.644)	14.13	(9.525)
Bonus Amount for 2016, $USD	29,414.46	40,502.05	11,087.58	0.0007	***	22,347.44	(25972.44)	15,653.71	(14722.7 0)
Rank 2015: Outstanding	0.05	0.03	0.02	0.07		0.02	(0.134)	0.00	(----)
Rank 2015: Exceptional	0.3	0	0.03	0.16		0.19	(0.394)	0.02	(0.155)
Rank 2015: Solid	0.63	0.67	0.05	0.09		0.75	(0.430)	0.89	(0.315)
Rank 2015: Partially Met	0.003	0.004	0.001	0.41		0.01	(0.111)	0.04	(0.189)
Rank 2015: New	0.03	0.03	0.006	0.29		0.02	(0.144)	0.05	(0.217)

Table 2. Logistic Regressions predicting whether nominated Exceptional+ employees will be Under-Recognized, in odds ratios (population size excluded for non-disclosure protection)

Logit:	Under-Recognized, 2016	Std. Err.
Age (years)	1.13	(0.128)
Age, squared	1.00	(0.001)
Female	0.95	(0.219)
Asian	1.58	(0.455)
Hispanic	2.33	(1.148)
[Base: White; not enough data for Black, Other]		
Female Manager	1.46	(0.341)
Team size (n)	0.69	(0.029)
Nominated Exceptional+ on team, #	1.29***	(0.103)
Tenure[a]	0.96	(0.026)
Rank in 2015: Outstanding	2.25	(2.207)
Rank in 2015: Exceptional	2.11	(1.647)
Rank in 2015: Solid	1.81	(1.390)
[Base: Rank in 2015: Low]		
Constant	0.002*	(0.005)

***p <0.05, ** p<0.01, *** p<0.001**

*Analyses exclude employees designated as 'New' (employed less than a year in 2016)

Note: Logistic regresions are used to estimate the binary outcomes such as being under-recognized in order to report results in odds ratios. All results are robust to using OLS specifications (results available by request). **'**Age, squared' included as a demographic control to improve model accuracy for the effect of age which may have nonlinear relationship with the odds of being under-recognized.

Table 3. Number of Exceptional+ Employees on Same Team

Number of Exceptional+s on same team	Percent of Excpetional+ Popn
1	31.8%
2	32.6%
3	19.9%
4	8.4%
5	4.3%
6	2.3%
7	0.8%

Table 4. Logistic Regressions predicting whether nominated Exceptional+ employees will be Under-Recognized among managers with more than one nominated Exceptional+ employee, in odds ratios (population size excluded for non-disclosure protection)

Logit:	Under-Recognized	Std. Err.
Age (years)	1.26	(0.182)
Age, squared	1.00	(0.002)
Female	0.79	(0.222)
Asian	1.30	(0.446)
Hispanic	1.88	(1.138)
[Base: White; not enough data for Black, Other]		
Female Manager	1.63	(0.466)
Team size (n)	0.99	(0.032)
Nominated Exceptional+ on team, #	1.48***	(0.150)
Tenure[a]	0.96	(0.029)
Rank in 2015: Outstanding	6185400	(8.56e+09)
Rank in 2015: Exceptional	6353329	(8.79e+09)
Rank in 2015: Solid	5329674	(7.37e+09)
[Base: Rank in 2015: Low]		
Constant	3.71E-11	(5.13e-08)

***p <0.05, ** p<0.01, *** p<0.001**

*Analyses exclude employees designated as 'New' (employed less than a year in 2016)

Figure 2. Under-Recognized Receive Biggest Merit Bonuses for Performance when Under-Recognized

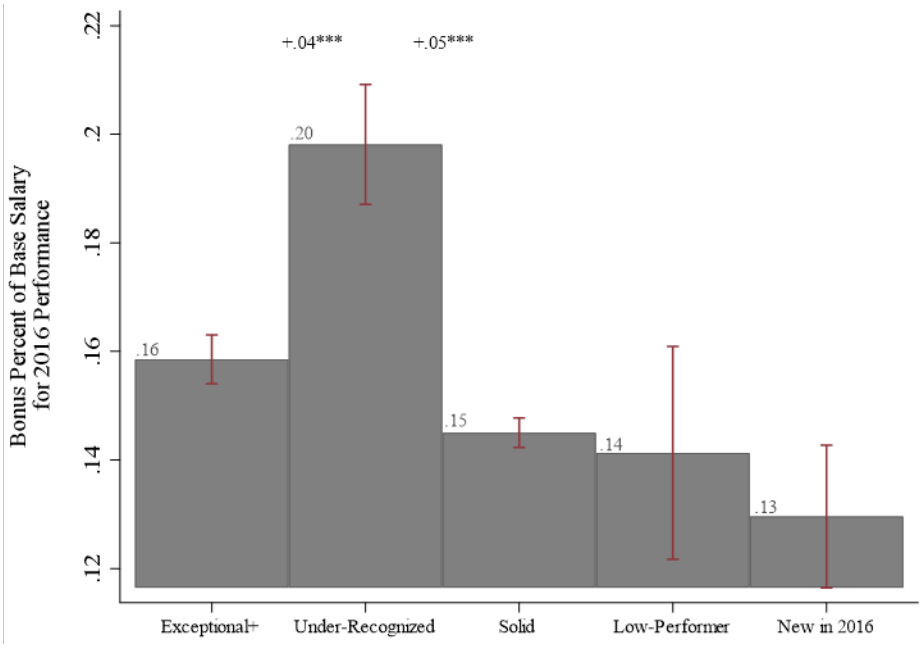


Table 5. Linear regressions predicting future bonuses, across entire company population

	Model a	**Model b**	**Model c**	**Model d**
	2017 Bonus Percent	**2017 Bonus Amount, USD**	**2018 Bonus Percent**	**2018 Bonus Amount, USD**
Under-Recognized, 2016	0.051***	14,668.24***	0.045***	14,003.91***
	(0.006)	(2011.412)	(0.007)	(3534.940)
Under-Recognized, 2017	0.029***	4,728.28*	0.033***	9,207.42***
	(0.006)	(1948.128)	(0.006)	(2826.692)
Under-Recognized, 2016 + 2017	-0.032	-13156.43*	-0.019	-4,626.93
	(0.020)	(6628.683)	(0.022)	(10634.200)
Outstanding Rating, 2017	-0.030***	-5,835.44*	-0.007	4,255.18
	(0.009)	(2882.502)	(0.011)	(5742.329)
Exceptional Rating, 2017	-0.026***	-3,274.03	-0.010	1,107.00
	(0.007)	(1986.055)	(0.009)	(478.62)
Solid Rating, 2017	-0.037***	-5,708.00**	-0.025**	-5,729.86
	(0.006)	(1872.93)	(0.008)	(4628.41)
[Base: 2017 Rank: Low or New]				
Outstanding Rating, 2018			-0.012	193.68
			(0.010)	(5141.899)
Exceptional Rating, 2018			-0.012	622.38
			(0.012)	(4480.682)
Solid Rating, 2018			-0.029*	-6,383.14
			(0.012)	(4393.591)
[Base: 2018 Rank: Low or New]				
Constant	0.182***	23,125.58***	0.203***	33,6959.97***
	(0.006)	(1819.082)	(0.016)	(5990.260)
R-squared	0.024	0.0347	0.0382	0.0343

***p <0.05, ** p<0.01, *** p<0.001**

Note: ratings in 2018 were generated using a revised performance review process that deviates from the forced distribution scheme in place in 2016 and 2017. Bonus Percent and Amount are highly correlated across years (≥ 0.908 for all pairings 2016-2018, percent or amount).

Table 6. Logistic Regressions predicting odds of promotion in 2017 by outcome of 2016 performance evaluation process across entire company population, in odds ratios

	Promoted in 2017	Promoted in 2018
Under-Recognized in 2016	0.55	0.55
	(0.268)	(0.303)
Solid in 2016	0.39***	0.73
	(0.078)	(0.139)
[Base: Exceptional+ in 2016; no info for Low Performers]		
Age (years)	0.92	1.09
	(0.084)	(0.112)
Age, squared	1.00	1.00
	(0.001)	(0.001)
Female	1.21	1.18
	(0.23)	(0.223)
Asian	1.08	1.07
	(0.292)	(0.278)
Hispanic	1.79	0.91
	(0.767)	(0.451)
Black	1.15	0.609
	(0.402)	(0.230)
Other Race	0.61	1.88
	(0.634)	(1.236)
[Base: White]		
Female Manager	1.45	1.67
	(0.283)	(0.312)
Team size (n)	1.04	0.98
	(0.026)	(0.026)
Tenure	0.98	0.99
	(0.023)	(0.022)
Bonus Amount for 2016 performance	1.00	1.00
	(4.27e-06)	(4.29e-06)
2015 Rank: Outstanding	4.03	0.99
	(3.382)	(0.724)
2015 Rank: Exceptional	4.43	1.12
	(3.317)	(0.642)
2015 Rank: Solid	2.45	1.34
	(1.815)	(0.739)
[Base: Prior Rank in 2015: Low or New]		
2017 Rank: Outstanding		5.66
		(6.19)
2017 Rank: Exceptional		2.82
		(2.957)
2017 Rank: Solid		1.54
		(1.608)
[Base: Rank in 2017: Low or New]		
Constant	0.54	0.02
	(1.057)	(0.050)
Log-likelihood	80.03	76.71

***p <0.05, ** p<0.01, *** p<0.001**

*Analyses exclude employees designated as 'New' (employed less than a year)

Table 7. Logistic Regressions predicting All Turnover by Under-Recognition Among Top Performers, in odds ratios

	Model A	Model B	Model C	Model D	Model E
Under-Recognized	1.72**	1.52**	0.75	1.65*	1.46
	(0.320)	(0.239)	(0.223)	(0.340)	(0.603)
# Nominated Exceptional+ on team		0.68***	0.61	1.06	1.04
		(0.043)	(0.048)	(0.075)	(0.088)
Under-Recognized x # Nominated Exceptional+s			1.47**		1.08
			(0.195)		(0.164)
Bonus Amount for 2016				1.00 (1.03e-07)	1.00 (1.01e-07)
Under-Recognized x Bonus Amount for 2016					1.00 (1.39e-06)
Constant	0.13***	0.48***	0.57***	0.09***	0.10
	(0.011)	(0.058)	(0.080)	(0.015)	(0.019)
Log-likelihood	7.88	46.81	54.7	7.29	8.21

***p <0.05, ** p<0.01, *** p<0.001**

*Analyses exclude employees designated as 'New' (employed less than a year in 2016)

Table 8. Demographics of Each Stage of Analysis

	All of PharMed	All Top Talent (nominated)	% Under-Recognized	% Leavers | Under-Recognized
Age (years), Ave	42.47	41.4	42.8	42.30
Female, %	49.23	49.35	43.98	36.36
Minority, %	29.87	25.2	26.53	28.13
Total, %	100	26.03	3.72	0.64

Table 9. Logistic Regressions predicting Voluntary Turnover only by Under-Recognition Among Top Performers, in odds ratios

	Model F	Model G	Model H	Model I	Model J
Under-Recognized	1.42*	1.43*	1.44	1.18	0.91
	(0.240)	(0.242)	(0.459)	(0.234)	(0.359)
# Nominated Exceptional+ on team		1.00	1.00	1.00	0.98
		(0.058)	(0.068)	(0.065)	(0.074)
Under-Recognized x # Nominated Exceptional+s			1.00		1.10
			(0.129)		(0.164)
Bonus Amount for 2016				1.00	1.00
				(9.62e-08)	(1.03e-07)
Under-Recognized x Bonus Amount for 2016					1.00
					(4.39e-07)
Constant	0.18***	0.18***	0.18***	0.15***	0.16***
	(0.013)	(0.023)	(0.129)	(0.023)	(0.026)
Log-likelihood	4.18	4.18	4.18	1.17	2.52

***p <0.05, ** p<0.01, *** p<0.001**

*Analyses exclude employees designated as 'New' (employed less than a year in 2016)

Figure 2. High Voluntary Turnover (18 months post performance review process) among Under-Recognized

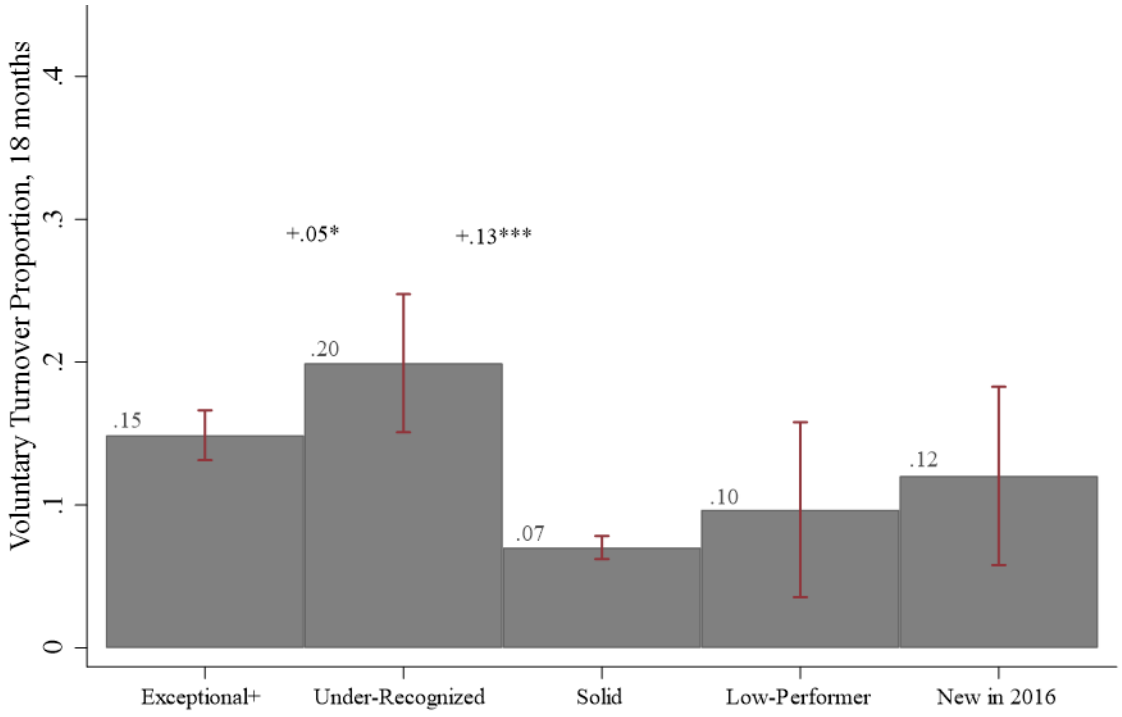


Table 10. Termination Reasons for Exits among the Under-Recognized Leavers

		Under-Recognized Leavers		Exceptional+ Leavers	
		Other Top Performers on Team?			
	Termination Reason	**No**	**Yes**	**No**	**Yes**
Voluntary :	Career Shift to New Industry and/or Different Role	19%	17%	20%	21%
	External Growth Opportunity	59%	40%	53%	66%
	Leaving Workforce	3%	0%	4%	4%
	Offered a Stronger Compensation Package	0%	3%	0%	0%
	Organization or Culture Fit	0%	6%	1%	3%
Involuntary:	Performance	5%	6%	1%	0%
	Position Elimination with Severance	14%	29%	22%	6%
	Total	100%	100%	100%	100%

Table 11. Logistic Regressions predicting Turnover (All and Voluntary) by Under-Recognition Among All Employees Rated Solid in 2016, in odds ratios

	All Turnover			Voluntary Turnover		
	Model A	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**
Under-Recognized	10.43***	7.48***	7.65***	3.30***	3.17***	3,53***
	(1.760)	(1.89)	(2.288)	(0.548)	(0.74)	(0.948)
Age (years)		0.92	0.85		0.97	0.85
		(0.079)	(0.094)		(0.062)	(0.020)
Age, squared		1.00	1.00		1.00	1.00
		(0.001)	(0.001)		(0.001)	(0.001)
Female		1.11	0.84		0.85	0.66
		(0.241)	(0.23)		(0.137)	(0.142)
Asian		1.30	1.63		1.73**	2.05**
		(0.241)	(0.528)		(0.0316)	(0.488)
Hispanic		1.89	1.73		1.17	1.22
		(0.824)	(1.00)		(0.458)	(0.617)
Black		0.33	0.43		0.33**	0.19*
		(0.199)	(0.323)		(0.142)	(0.137)
Other Race		0.84	2.28		1.35	2.04
		(0.867)	(2.42)		(0.840)	(1.596)
[Base: White]						
Female Manager		1.52	1.69		0.92	0.99
		(0.330)	(0.461)		(0.152)	(0.217)
Team size (n)		0.94	0.97		0.96	0.99
		(0.030)	(0.037)		(0.022)	(0.029)
Tenure		0.96	0.95		0.95**	0.95*
		(0.023)	(0.030)		(0.017)	(0.022)
Nominated Exceptional+ on team, #			1.10			1.15
			(0.130)			(0.109)

Rank 2015: Outstanding			0.58			0.45
			(0.745)			(0.407)
Rank 2015: Exceptional			1.02			0.72
			(0.816)			(0.0387)
Rank 2015: Solid			1.05			0.64
			(0.813)			(0.330)
[Base: Rank 2015: Low or New]						
Constant	0.032***	0.30	1.25	0.08***	0.32	5.07
	(0.003)	(0.559)	(3.116)	(0.548)	(0.439)	(9.306)
Observations	4,135	2,091	1,320	4,135	2,091	1,320
Log-likelihood	152.51	80.81	68.43	42.86	68.96	68.49

***p <0.05, ** p<0.01, *** p<0.001**

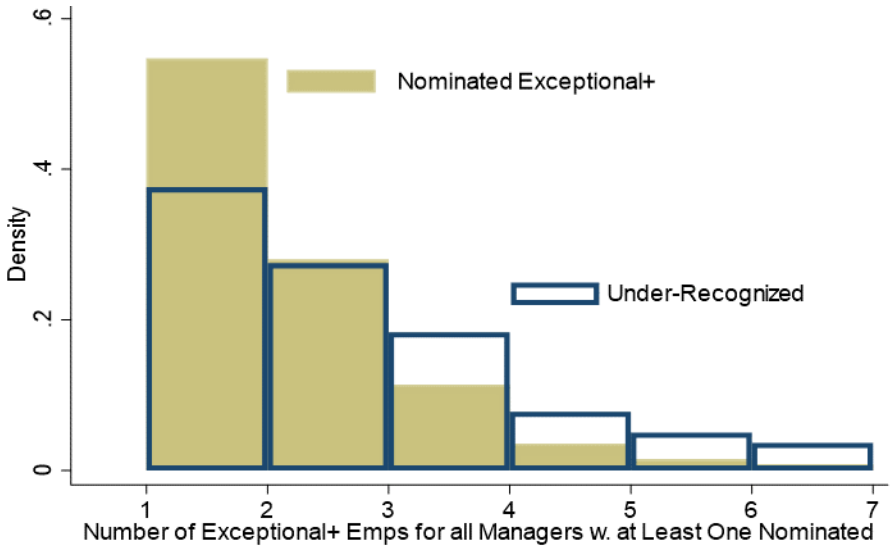
*Analyses exclude employees designated as 'New' (employed less than a year in 2016)

Note: Bonus amount for 2016 performance reward excluded in model on all Solid employees since this variable is perfectly collinear with the treatment of being under-recognized relative to being rated Solid throughout the performance evaluation process in 2016.

Table A1. Description of the Performance Ranks and the percent of employees designated for 2016

Rating	Performance Description[12]	Percent
Outstanding	Delivers excellent results in a remarkable way: - Top 3% of performers at their level in the past year - Consistently sets and exceeds ambitious goals - Universally viewed as a role model for core behaviors; has a strong positive impact on the team/department/organization	2.5 %
Exceptional	"Delivers excellent results and lives our Core Behaviors" or "Is a role model of our Core Behaviors" - Who also meets our high expectations on delivering results" and is "top 20% of performers at their level in the past year	19.8 %
Solid	Meets the company's high expectation for delivering results and living our Core Behaviors: - Meets high expectations for most goals, may exceed some goals - Consistently meets the expectations of core behaviors at their level	74.7 %
Partially Met	"Delivers in results but needs to improve in modeling our core behaviors" or "Lives our core behaviors but does not consistently accomplish their goals"	2.7 %
Unsatisfactory	Does not meet our expectations for performance (results and behaviors) at the company	0.3 %

Figure AX1. Histogram of nominated Exceptional+ distribution n versus distribution of those who are Under-Recognized



[12] From the company's internal performance management toolkit documentation, 2016.

Table AX1. Logit Models predicting nomination to Exceptional+ Rating, in odds ratios (population size excluded for non-disclosure protection)

Logit:	**Exceptional+**	**Std. Err.**
Age (years)	1.13**	(0.050)
Female	1.08	(0.1030
Asian	0.67**	(0.080)
Hispanic	0.88	(0.218)
Black	0.45***	(0.083)
Other	0.906	(0.346)
[Base: White]		
Female Manager	0.98	(0.098)
Team size (n)	1.03**	(0.013)
Tenure	0.97***	(0.010)
Prior Rank: Outstanding	1.93	(0.702)
Prior Rank: Exceptional	1.28	(0.358)
Prior Rank: Solid	0.8	(0.215)
[Base: Prior Rank: Low]		
Constant	0.058**	(0.055)

***p <0.05, ** p<0.01, *** p<0.001**

*****Analyses exclude employees designated as 'New' (employed less than a year) Controlling for Location of PharMed office.

Appendix Table AX3. Linear Regressions Predicting Total Base Pay and Total Salary + Allowances changes from 2016-2017 based on outcomes from Performance Evaluation Ratings for 2016

	Total Base Pay Change 2016 - 2017	Total Salary + Allowances Change 2016 - 2017
Under-Recognized in 2016	-4,780.40	-4,896.44
	(34044.83)	(34045.43)
Solid in 2016	9,191.97	9,210.71
	(14536.65)	(14536.90)
Low Performer in 2016	243.93	190.93
	(59866.37)	(59867.43)
[Base: New employee in 2016]		
Age (years)	3.276.67	3,286.95
	(5985.071)	(5985.176)
Age, squared	-39.86	39.972
	(67.241)	(67.242)
Female	-17,110.61	-17,211.30
	(13609.89)	(13610.13)
Asian	43035.38*	43,076.04*
	(18871.03)	(18871.36)
Hispanic	854.63	836.54
	(35634.24)	(35634.86)
Black	-36.28	-46.99
	(23043.70)	(23044.10)
Other Race	-3,486.85	-3,604.42
	(57620.36)	(57621.37)
[Base: White]		
Female Manager	22,152.35	22,078.91
	(14113.75)	(14113.99)
Team size (n)	-1,176.83	-1183.43
	(1844.039)	(1844.07)
Tenure	-943.13	-941.10
	(1346.03)	(1346.053)
Nominated Exceptional+ on team, #	-4,524.82	-4,487.62
	(6787.95)	(6788.065)
Bonus Amount for 2016 performance	0.08	0.08
	(0.225)	(0.226)
2015 Rank: Outstanding	1,948.01	1,906.61
	(50123.19)	(50124.08)
2015 Rank: Exceptional	301.11	249.35
	(36071.60)	(36072.24)
2015 Rank: Solid	7,242.81	7,301.59
	(34111.79)	(34112.39)
[Base: Prior Rank in 2015: Low or New]		
Promoted in 2017	-799.92	-821.54
	(27050.71)	(27051.19)
Constant	-48115.87	-48299.23
	(132384.7)	(132387.10)
R-squared	0.0063	0.0063

***p <0.05, ** p<0.01, *** p<0.001**

Appendix Table AX4. Multinomial Logistic Regressions predicting 2017 Performance Ratings based progression from Solid Rating in 2015 to Solid Rating in 2016 and whether Under-recognized compared to employees progressing from Solid Rating in 2015 to Exceptional+ Rating in 2016, results reported in relative risk ratios (rrr)

2017 Performance Rating	[base: Solid_15 -> Exceptional+_16]	RRR	Std. Err.	z	P>|z|	95% Conf.	Interval
Outstanding	Solid_15 -> Under Recognized _16	0.93	0.486	-0.15	0.883	0.331	2.589
	Solid_15 -> Solid _16	0.31 ***	0.063	-5.79	0	0.205	0.457
	Constant	0.06 ***	0.006	-26.06	0	0.048	0.073
Exceptional	Solid_15 -> Under Recognized _16	1.18	0.290	0.67	0.501	0.729	1.909
	Solid_15 -> Solid _16	0.56 ***	0.047	-6.91	0	0.470	0.656
	Constant	0.27 ***	0.015	-23.63	0	0.239	0.298
	[Base outcome: Solid Rating in 2017]						
Partially_Met	Solid_15 -> Under Recognized _16	0.65	0.665	-0.42	0.674	0.088	4.828
	Solid_15 -> Solid _16	1.09	0.257	0.36	0.715	0.686	1.731
	Constant	0.02 ***	0.004	-21.61	0	0.015	0.030
Unsatisfactory	Solid_15 -> Under Recognized _16	0.00	0.001	0	0.997	0.000	.
	Solid_15 -> Solid _16	0.12 *	0.124	-2.01	0.044	0.014	0.943
	Constant	0.00 ***	0.002	-14.2	0	0.002	0.010

***p <0.05, ** p<0.01, *** p<0.001**

Figure AX2. Percent of Employees Nominated by their Manager to Exceptional+ Rating in 2017 by Outcome of 2016 Performance Evaluation Process

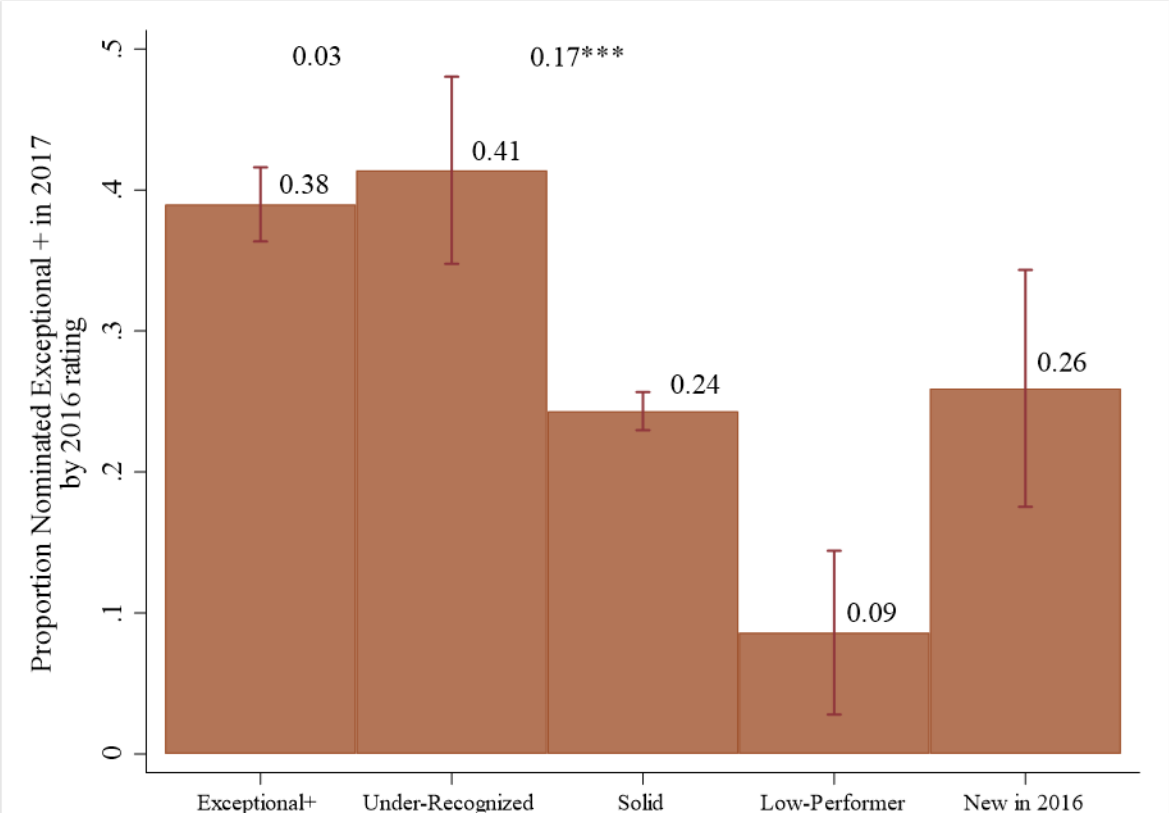
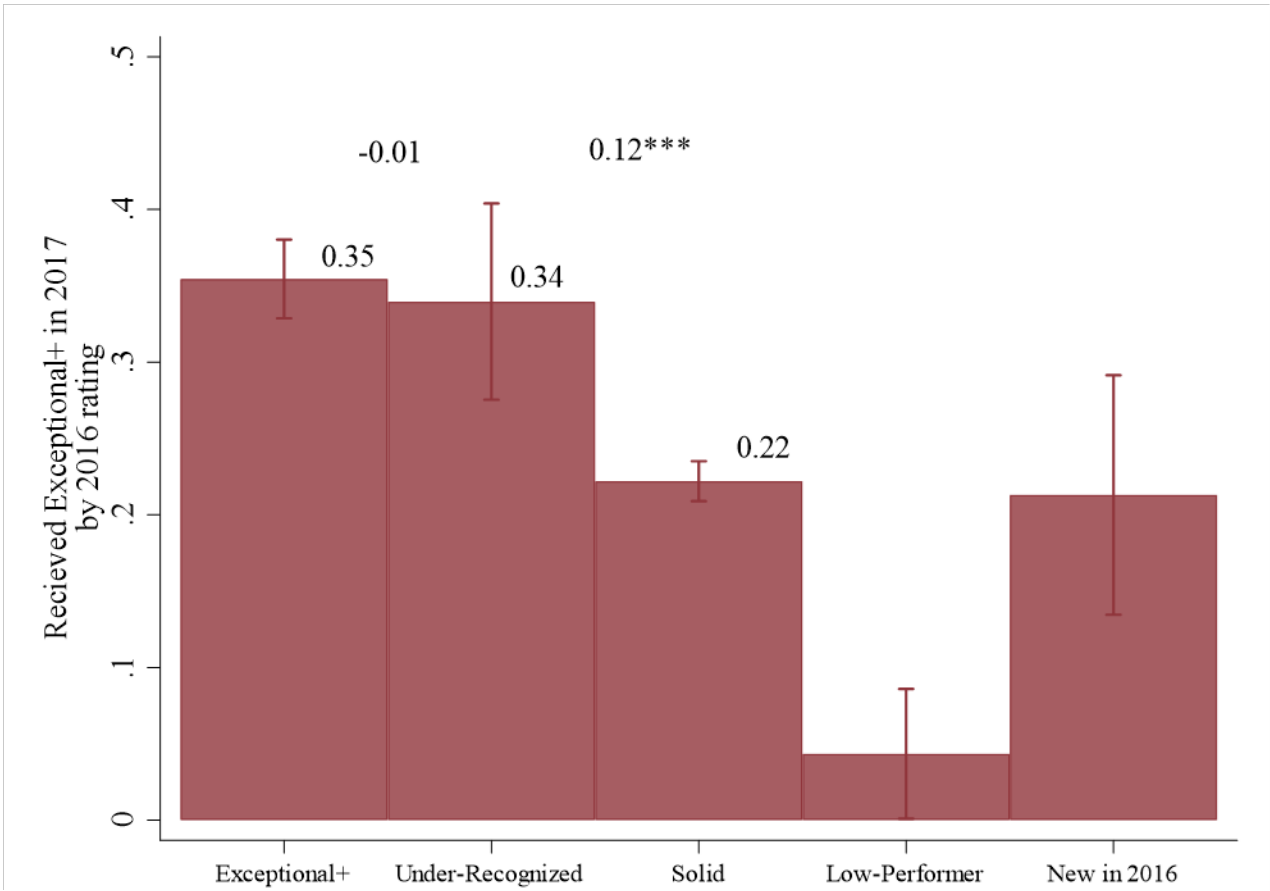


Figure AX3. Percent of Employees Receiving Exceptional+ Final Rating in 2017 by Outcome of 2016 Performance Evaluation Process



CHAPTER 2

Too Close To Call?
Relational vs. Structural Management and the Accuracy of Employee Performance Evaluations

INTRODUCTION

More than 90% of U.S. companies rely on formal performance appraisals to direct human capital investment and deployment, according to a 2018 WorldatWork survey. By tying merit bonuses to performance evaluations, companies incentivize and reward employees for reaching certain performance thresholds. Even for employees falling short of expectations, accurate assessments also enable the direction of remedial attention and resources toward performance improvement (Grote 2005). However, possible concern that accurate but poor performance assessments will reduce near-term effort and performance of under-performers (Milkovich and Newman 1996) may outweigh a manager's concern for accurate human resource allocations. If a manager believes that a low assessment will demotivate an under-performer or negatively impact their working relationship, they may decide to inflate an underperformer's evaluation.

This tradeoff marks a principal-agent problem. When managers inflate evaluations of under-performing employees, they deviate from the organization's interest in efficient information sharing and resource allocation (Prendergast and Topel 1993, 1996). This particularistic concern for under-performers compromises the general welfare of the firm by making idiosyncratic exceptions to formal organizational policy (Heimer 1992). Eventually, the overall effectiveness of the performance appraisal process will erode. Elevating the relational concerns of underperforming employees above consistent and accurate performance assessments across all employees compromises organizational efficiency (MacLeod 2003).

In addressing such principal-agent concerns, scholars have proposed the adoption of formalized work practices as a way to minimize the risk of such sources of bias. Defined as management approaches governed by clear written rules and procedures (Anderson & Tomaskovic-Devey, 1995; Huffman & Velasco, 1997; Weber, 1946), formalized work practices constrain the discretion that managers may exercise in favoring certain employees over others (Elvira & Graham, 2002; Reskin, 2000; Kalev, Dobbin

and Kelly 2006; Castilla 2008, 2011). Recent evidence suggests that, formalized management practices offer greater productivity enhancement than more relational styles of management. In this manner, the literature has largely treated the two management approaches as competing alternatives, the idea being that managers may tend toward *either* a formal respect relationship *or* an informal friendship (Fernandez 1991).

In this vein, Chatterji et al. (2019) found that new technology firms demonstrate the benefit of receiving advice from managers with a more formal management approaches than from relational managers for years. This was also particularly so when the new technology firms had little formal management training (Chatterji et al. 2019). Additional evidence from Canales (2014) and Canales and Greenberg (2016) shows that more relational loan officers generate positive returns only when structured, rule-enforcing peers provide a check on their discretionary inclinations (Canales 2014). Canales and Greenberg (2016) also show that loan officers who oscillate between relational and structured approaches from client to client generate the worst outcomes, emphasizing the importance of consistency in these lending arrangements (Canales 2014; see also Canales and Greenberg 2016). Such findings underscore the incoherence suggested from a possible dual relational-structured management approach.

Understanding whether relational and structured management approaches really are exclusionary in coherent and productive management is important for organizational effectiveness. Highly relational management can nurture competitive organizational advantages (Gittell 2016; Jordan 2015; Kotzé and Roodt 2005). At the same, time the known risks that caution close managerial relationships in organizations (Uzzi 1997; Uzzi and Lancaster 2004; Sorenson and Waguespack 2006; Doering 2018). While the preponderance of evidence shows productivity enhancement from the adoption of formalized or structured management (Bloom and Van Reenen 2006; 2010; 2011; Bloom, Sadun, and Van Reenen 2007; 2009; 2012; Bloom, Brynjolfsson, Foster, Jarmin, Sporta-Eckstein and Van Reenen 2013; 2016; Chatterji et al. 2019), there is also the caution against trying to maintain both simultaneously (Canales and Greenberg 2016).

Complicating the endeavor is the potential that individuals do oscillate between the two approaches, depending on specific situations. For instance, a manager may adhere to a strict formal approach when

evaluating employee performance in general but, in the case of a particular under-performing employee, may make an exception based on unique relational concerns or circumstances. Whether or not this represents a structured/formal managerial approach or one driven by relational concerns may depend on the specific interpretation of the particular circumstances. In other words, it remains unclear whether the two approaches really are dichotomous. Additionally, research on this topic has varied widely over different levels of analysis. This range opens up the question: whether the predicted exclusiveness of either relational or structural management approach and corresponding organizational benefit may depend on the level of analysis.

In this paper, I test whether two broad orientations toward managing - relational (i.e. informal) and structural (i.e. formal) - predict whether a manager will inflate the evaluation of an under-performing employee's performance. I address these questions empirically, using a hypothetical vignette experiment reflecting their own firm's employee performance evaluation policy as well as rich administrative and survey data on managers and the employees that they manage in a multinational pharmaceutical company. A key outcome of interest is whether either one of the two approaches dominates the evaluation decision or whether both dimensions can simultaneously influence whether a manager inflates the performance evaluation of an under-performing employee. This research illuminates the potential limits of the returns to close managerial relationships inside organizations. It also offers insight into how organizations may be able to reap the benefits of close managerial relationships while managing to avoid potentially costly pitfalls.

To begin, I first discuss how highly relational management can nurture competitive organizational advantages. Equal attention is given, as well, to the known risks that close managerial relationships may pose to competing interests, such as organizational concerns. I further articulate the implications that such inflation of employee performance evaluations can carry for the organization. Second, I review a widely discussed approach for formalizing management – *structured management practices* – and assess whether such an approach can preserve the positive aspects of relational managerial relationships while also preventing bias in performance evaluation. In particular, I assess whether the benefits promised by

structured management practices will vary depending on how relational the manager is as measured by the levels of trust and cooperation in the work unit they manage. To the degree the effectiveness of structured management does not vary depending on the closeness of the managerial relationship, organizations can benefit from adopting structured management practices, no matter the levels of trust and cooperation experienced in the team.

My findings demonstrate that relational and structural manager approaches operate independently and in opposing directions, on the likelihood that managers inflate the performance evaluations of underperforming employees. These findings support the hypothesis that managers can both be highly relational and still make accurate evaluations of employees who are underperforming through structured management. I complement this analysis using company administrative data to show that inflating performance evaluations of underperforming employees indeed does carry negative productivity consequences for the organization. Finally, I conclude by discussing the theoretical implications of the independence of these processes, as well as the practical implications for strategic management approaches.

THEORY

Relational Management and Organizational Outcomes

A highly relational management approach, specifically, one that is marked by fostering high levels of trust and cooperation within the managed work unit, is expected to generate positive operational outcomes. Mutual trust, goodwill, and commitment between managers and employees enhance internal coordination mechanisms within organizations (Gittell 2016; Jordan 2015; Kotzé and Roodt 2005; Zuckerman 2014). These attributes underpin the idea that strong relationships between managers and employees are foundational for effective organizational practices and competitive market advantages (see, e.g., Pfeffer 1997). Trust encourages employee investment in a mutually beneficial future, improving competitiveness and reducing transaction costs (Uzzi 1996; DiMaggio and Louch 1998; Sako and Helper 1998; Mizruchi and Stearns 2001; Uzzi and Lancaster 2003). Over time, employees better understand

implicit expectations which enables efficient production despite having incomplete information about their employers' overall strategy (Rousseau 2004).

Such benefits may extend to performance review practices, an often thorny and contentious aspect of manager-worker relationships (Pfeffer and Sutton 2006). Here, high trust and cooperation may facilitate delivering difficult, if accurate, evaluations, as they enable broadening the level of shared understanding between managers and their employees (Stone, Patton, Heen 2000; Stone and Heen 2015; Gittell 2016; Jordan 2015; Kotzé and Roodt 2005). When managers and employees share a common roadmap for development, evidence suggests that managers can more easily communicate areas for improvement without employees second-guessing their intentionality (Rousseau 2004). Employees are more likely to increase their efforts to help their manager reach her goals when they believe she will respond in kind (Gittel and Fields 2003; for review see Gittell and Douglass 2012). For such reasons, I offer the first of two competing hypotheses that:

> H1a: *Highly relational managers, operating on a greater basis of trust and cooperation in their work units, are <u>less likely to inflate</u> an under-performing employee's performance evaluation.*

Still, the same antecedents behind the common roadmap shared between employees and highly relational managers may also predict a greater likelihood for inflating and under-performer's performance evaluation. For instance, inflating a performance evaluation may generate relational credit for a manager, generating extra motivation on the part of their employee in the future. In fact, Bol (2011) suggests that through improved fairness perceptions on the part of the employee, such inflation may positively affect employees' incentives. In such a relationship, a manager may reasonably inflate the rating of an employee whose performance came up short, for instance, to signal faith and an expectation that the employee will reciprocate by improving productivity in the future (Axelrod 1984; Powell 1990:305; Poppo et al. 2008; Gittel and Douglass 2012). Thus, I also hypothesize:

> H1b: *Highly relational managers, operating on a greater basis of trust and cooperation in their work units, are <u>more likely to inflate</u> an under-performing employee's performance evaluation.*

Organization-Wide Costs of Inflating Employee Performance Evaluations

Under the above logic, the particularistic bias on the part of a relational manager may enhance the productiveness of the managerial relationship. However, in addition to this potential positive rationale for why relational managers may inflate the ratings of underperformance, such an approach toward particular employees with whom they share this mutual understanding may also compromise the manager's ability to balance broader organizational goals (Heimer 1992). A manager be unable to balance the particular concern for an underperforming employee with the organizational goals (Fernandez 1991; see also Bond et al. 2018 and Bond and Fernandez 2020). The exclusionary tension may be especially true when considering underperforming employees with whom the manager has developed a long-lasting embedded relationship of trust and cooperation (Uzzi 1997; Uzzi and Lancaster 2004; Sorenson and Waguespack 2006).

Preferential treatment toward some employees over others can create agency problems where the consequences of performance evaluations are not in the best interest of parties outside the focal managerial relationship (Prendergast and Topel 1993, 1996). Rewarding some employees for work incommensurate with specific standards can misallocate scarce resources that would be better deployed to motivate higher performers. Further, such a bias may potentially lead to misallocating workers away from more appropriate or better fitting jobs through unjustified promotion or retention (Prendergast and Topel 1996). Preferential treatment of favored employees also demotivates employees who are simultaneously discriminated against, which can in turn reduce employee effort (MacLeod 2003)[13].

To the extent that performance inflation translates into negative performance consequences for the organization, there should be evidence that managers who adopt a more relational, particularistic response to employee underperformance when evaluating are also less likely to terminate underperformers. In addition to testing whether a more relational manager is more likely to inflate an underperforming employee's performance evaluation, I look for additional evidence that greater embeddedness will decrease the likelihood of dismissing an employee for underperformance. This leads to my second hypothesis:

[13] For a review of further negative employee reactions to perceptions of inequity in performance evaluations from the social psychological literature on social comparisons see Larkin et al. 2012.

H2: More relational managers are less likely to dismiss an employee for underperformance.

Formalized Work Practices as a Way to Minimize Performance Inflation

One benefit of formalized work practices is that they constrain the discretion that managers may exercise in favoring certain employees over others (Elvira & Graham, 2002; Reskin, 2000; Kalev, Dobbin and Kelly 2006; Castilla 2008, 2011). Research on a particular articulation of formalized management practice – *structured management practices* – the programmatic use of monitoring, goal setting, incentives and data in decision making – strongly suggest they preserve strong relational team elements while minimizing potentially harmful interpersonal biases (Bloom and Van Reenen 2006; 2010; 2011; Bloom, Sadun, and Van Reenen 2007; 2009; 2012; Lemos and Scur 2018). At the firm level, significant productivity differences between otherwise similar firms have been explained by the variation of structured management practices (Bloom and Van Reenen 2006; 2010; 2011; Bloom, Sadun, and Van Reenen 2007; 2009; 2012; Bloom, Brynjolfsson, Foster, Jarmin, Sporta-Eckstein and Van Reenen 2013; 2016; Chatterji et al. 2019). Firms that adopt such structured management practices are found to be more efficient, more likely to persist, and more able recruit and retain workers with higher average human capital.

One open question is how much variation in structured management practices will be found among managers within a single given firm. An important implication from answering this question is whether such heterogeneity will similarly predict performance-related outcomes. At both levels of analysis, structured management practices can be analogized to collecting and analyzing key performance indicators (KPIs) and making promotion decisions based on performance rather than tenure considerations. Together they indicate greater likelihood of good management of important resources (Bloom and Van Reenen 2007). If the extent of structured management practices varies significantly within firms, by implication, a given firm may experience overall productivity gains when managers shift to more structured management practices overall. Specifically, I look at whether managers who adopt a more systematic, or structured, approach to tracking and evaluating employee performance are less likely to inflate the performance evaluation of an underperforming employee. This leads to my third hypothesis:

H3: Managers with more structured management practices are less likely to inflate an employee's performance evaluation.

Are Relational and Structured Management Approaches Independent?

Prior research supports the idea that the closeness of managerial relationships will moderate the influence of structured management practices on organizational outcomes. For one, certain levels of trust and cooperation may reflect necessary conditions for structured management practices to manifest productivity gains (Gibbons and Henderson 2013). There is also cautionary evidence that this complementarity is not strictly increasing. For instance, while Blader, Gartenberg, and Prat (2016) find evidence that "intangible relational factors" determine the optimal set of managerial practices, their evidence also implies that structured management practices may only enhance productivity under certain relational conditions. This research supports the idea that 1) a highly relational approach will not preclude a manager from adopting a more formalized approach to managing their team, and that 2) a highly relational approach may moderate productivity enhancement from structured management practices.

Despite the predictions that benefits from structured management practices are sensitive to how relational a manager is, there is still reason to expect, in the context of employee performance evaluations, that the level of trust and cooperation will not influence the effect of structured management practices in reducing the likelihood of inflating the evaluation of an under-performing employee. Importantly, such an independence between relational factors and a structured approach to managing carries important, positive implications. Namely, this independence would open a way to reconcile the many motivations for fostering trust and cooperation in organizations through highly relational management approaches, while offering a mechanism for minimizing their potential downsides in formal personnel management processes. To the extent that adopting a structured management approach does not crowd out the positive by-products of close managerial relationships, marked by levels of trust and cooperation, this approach may present a solution to the risk that such embedded relationships pose in subjective employee performance evaluation processes.

Alternatively, evidence that the effect of structured management practices does depends on the level of managerial trust and cooperation would imply that the prescriptive power of structured management

practices is less robust than otherwise. For instance, if the power of structured management practices to reduce the likelihood of inflating an underperforming employee's performance evaluation is weaker when the level of trust and cooperation between manager and employee is stronger, it is a less viable solution to the complications of managerial relational embeddedness. Absence of such evidence, however, is not the same as evidence that trust and cooperation do not ever shape critical conditions for structured management practices to persist or to be effective. Instead, such absence of evidence suggests that, in general, organizations may benefit from increasing structured management practices no matter the levels of trust and cooperation within teams. Thus, my last hypothesis is:

> *H4: relational and structured management operate <u>independently and simultaneously</u> on the likelihood that a manager inflates an underperforming employee's performance evaluation.*

DATA AND METHODS

I test whether highly relational managers, those with greater levels of trust and cooperation in their teams, will inflate the performance evaluation of under-performing employees and whether structured management practices reduce this contingency. I do so by using a hypothetical vignette case study experiment, as well as survey and administrative data on managers in a multinational pharmaceutical company headquartered in the United States, referred to here as PharMed (a pseudonym). The choice of field site for this research is motivated by the perspective of Gibbons and Henderson (2012) who "see science-driven drug discovery as a complex managerial practice that cannot be sustained without a relational contract" (16). As such, this setting offers an enticing environment to explore the limits of embedded relationships in organizational processes and their interactive effects with structured management practices.

In May 2017, I ran a hypothetical vignette case study experiment testing the predictors of inflating objective employee underperformance among a randomized sample of PharMed managers. The predictor data come from the results of survey data on reported levels of trust and cooperation from their employees and manager-reported measures of structured management practices. HR administrative data on tenure and other employee characteristics further support the analysis. The main outcome of interest is the likelihood

that a manager inflates the rating of an under-performing employee. The vignette experiment captures this objective outcome to a typically subjective process. I also test whether such inflation carries negative productivity consequences using evidence from PharMed's administrative records. Further, I assess the effectiveness of structured management practices for preserving the benefits of high levels of trust and cooperation between managers and employees while minimizing the likelihood of inflation. The unit of analysis is respondent managers to the management practices and vignette case study experimental instrument.

Dependent Variable

The outcome of interest is whether a manager will inflate the performance evaluation of an underperforming employee in a year-end review. The validity of this outcome requires an objective assessment of employee performance. I use an experimental vignette design to collect responses from PharMed managers on how they would evaluate a hypothetical, objectively, under-performing employee. The hypothetical employee's underperformance is described using guidelines in PharMed's own manager handbook articulating the identifying features of underperformance when giving employees overall performance ratings. Respondent managers are entered into this hypothetical vignette experiment, described as a case study, at the end of the survey questionnaire used to capture measures of each manager's level of structured management practices (described in detail below).

Vignettes present a hypothetical situation in which respondents are asked to consider empirically and theoretically relevant factors that can be systematically varied in the form of short descriptions (Rossi and Anderson, 1982). While critics argue that the vignette method is deficient in external validity, evidence using behavioral benchmarks demonstrates remarkable predictive power of real-world behavior (Hainmueller, Hangartner, and Yamamoto, 2015). The validity of my specific design was tested during five months of cognitive testing, focus groups, and beta testing of the final instrument conducted with PharMed managers. The method I followed in this external validity testing follows the practices outlined in Buffington et al (2016) (discussed further in the appendix; see also Hainmueller et al., 2015).

This testing ensures that not only do respondents understand the details of the hypothetical situation described in each case scenario but also that respondents know what the rating prescribed by company policy would be for the hypothetical employee in response to each case. Indeed, when presented with the cases, focus group participants expressed serious doubt as to whether the various cases would elicit anything other than the clear, objective rating that the under-performing employee had earned, as the vignette conditions intended to convey. In other words, participants doubted that there would be any deviation from respondents giving the "correct" response to the performance evaluation cases posed.

The under-performance descriptions distinctly contrast the rating that the hypothetical employee should receive this year to the rating that the hypothetical employee had received the year prior (see Methods Appendix of example vignette experimental condition). This description emphasizes that the accurate rating for the relevant year is *lower* than the rating from the prior year. Thus, giving a rating equal to or higher than the reported rating the employee earned the prior year signifies an inflation of the employee's evaluation based on explicit policy. This makes the hypothetical employee's under-performance both evident from the objective company standards as well as evident relatively given the contrast to a prior rating.

The Vignette Design

All respondent managers consider an employee Jim. The gender of the hypothetical employee, Jim, is held constant for power concerns relating to measuring the main effect of evaluating under-performance across cases with common features. Jim previously reported to a different manager so that this is the first time that the respondent manager is hypothetically evaluating Jim in a performance review setting.The intention of this construction is to remove from consideration any justification effects from their own prior evaluation history (Bazerman et al 1982). Respondent managers then decide what rating they would give Jim for the year. Anything higher than the specified drop in performance rating described is coded as inflation of under-performance.

Using this experimental vignette design allows me to control for factors in the sample design that I empirically observe to be important to managers in this organization when they are deciding what ratings

to give to their actual employees. These factors were the most frequently cited in formal interviews with managers and during official employee performance review meetings I observed first-hand as well as interviews with a sample of managers during PharMed's annual performance review process for FY 2016[14]. Controlling for the most frequently cited and empirically observed factors that influence managers when making performance evaluation decisions helps to isolate the explanatory power of the variables of interest to this study (managerial trust and cooperation on the one hand and structured management practices on the other). These factors drive the construction of the randomized conditions exploited in the experiment. To isolate the causal inference behind the relational and structured management approaches that are the focus of this study, it is important that the main variables managers raise as determining outcomes of performance evaluations are controlled for in the experimental set-up. Therefore, I randomly assign participant managers to 1 of a possible 12 cases about the hypothetical employee Jim, which varied along the three condition categories reflecting the main factors that were raised as driving performance rating deviations from managers at PharMed (see Table 1 for further details):

> 1- How large a decrease in performance is observed (4 conditions across 3 levels of performance grades that PharMed uses to classify overall performance ratings),
>
> 2- Whether the manager inherited Jim as part of a new team they've acquired or whether the underperforming employee was transferred into the manager's preexisting team (2 conditions, reflecting common occurrences in PharMed when evaluating an employee for the first time), and
>
> 3- Whether the underperformance relates to work deliverables or behavioral concerns (2 conditions, reflecting the two dimensions on which employee performance is evaluated at PharMed).

This enables estimating the effect of structured management practices net of these important and salient factors.

Although this approach constricts the variation of embeddedness (specifically, on the length of relationship dimension) within the experimental design[15], it strengthens the predictive power of the natural

[14] This sample is based on my access to managers in PharMed's global People Development organization which granted this research project, their referrals made to representative managers from various functional divisions, nominations made by the executive HR officer, as well as contacts I made through earlier interviews and meetings.
[15] While the length of relationship is not operationalized to predict outcomes in the hypothetical vignette case study, later it is used to estimate whether there is evidence that longer, embedded, managerial relationships do result in retaining under-performers.

variation along the specified trust and cooperation dimensions by minimizing the potential for confounding biases such as confirmation bias or escalation of commitment (Bazerman et al 1982). Moreover, considering a hypothetical situation in which their own experience with the under-performing employee is limited creates a conservative test regarding the predictive power of a manager's natural context along the dimensions of trust and cooperation levels. The predictive power that trust and cooperation have in this study comes from the naturally varying levels of trust and cooperation that a manager experiences in his team generally (the operationalization of which is described in detail below) as well as the effect on the likelihood that the manager will inflate the rating of a hypothetical employee new to the unit.

Data Collection

This hypothetical vignette experiment was sent to a representative, random sample of all managers below the Senior Vice President level in May 2017, across all PharMed locations globally. It followed as the second part of a company-wide assessment I constructed of current management practices (described below). The invitation to take this survey was sent by a Human Resource Vice President who sponsored the research internally. Invitees were told, correctly, that their responses to the survey would be used to inform future manager training resources. By the end of the survey window, 496 managers had completed the case study as part of the management practices survey study, for a participation rate of over 75% of those who received the survey invitation[16]. The response rate does not vary significantly by gender, function, or compensation. The response rate skewed toward recipient populations in the US and EU, likely due to language barriers for some of the targeted managers in the Asian and Latin American regions. Given the range and representativeness of functions (sales, manufacturing, research and development, etc.), the results from this sample suggest fair generalizability.

Overall, one-in-five managers gave an inflated performance rating to the hypothetical underperforming employee. Specifically, an inflated performance rating took the form of a respondent manager giving an 'Exceptional' rating (exceeded expectations for their role) to an employee described as

[16] Between 35%-65% of the company's total population of people managers under the Senior Vice President level were surveyed. The ranges here are meant to protect PharMed's identity.

either ‘Solid’ (met all expectations) or ‘Partially Met’ (failed to meet all expectations) according to PharMed’s manager handbook or giving a ‘Solid’ in the case when the employee was described as Partially Met on official objectives (See Table 1). It is important to note that PharMed fits its year-end employee performance rating distribution to a tight forced distribution scheme meaning that few employees earn an ‘Exceptional’ rating each year and many top performers receive only a ranking of ‘Solid’. Any manager with experience giving performance ratings at PharMed is aware of how hard it is to defend exceptional employees under such a tight performance curve regime. Therefore, giving an accurate ‘Solid’ rating is an easy default under these circumstances. On the other hand, when a manager must contend with whether or not to give the under-performer a ‘Partially Met’ rating, a domain of the employee performance curve distribution not subject to a predefined forced target in this setting, the manager is in more uncharted territory. Still, PharMed policy is explicit about what marks an under-performer. Further, all managers are reminded of the importance of making such accurate calls for HR resourcing, training, and role reassignment purposes before and during the annual overall performance evaluation process in company communication and in meetings with assigned HR business partners.

Likelihood of giving an inflated rating in the case did not vary substantively by personal or organizational demographics[17]. As expected, however, there is substantial variation between the 12 experimental conditions defined by the most relevant contextual influencers in the performance evaluation process at PharMed (See Table 1). For this reason, the analyses below control for these conditions when predicting case accuracy by managerial levels of trust and cooperation as well as structured management practices.

[INSERT TABLE 1 ABOUT HERE]

Independent Variables

[17] For example, the difference between male and female is non-significant t= -0.87, df=532. There was also no significant interaction between organizational function area and likelihood to inflate, F(7, 482)= 0.48, p=0.89. The predictive power of tenure on odds of inflating was also a non-significant p< 0.336.

The experimental vignette design allows me to empirically control for key conditions observed to matter when managers decide what rating to give an employee in their year-end performance appraisal. These controls enable me to more precisely estimate the effect of the key explanatory variables for this study, namely levels of reported trust and cooperation and how structured an approach to managing managers take. I measure both of these key explanatory variables using the natural variation observed in the sample population of managers from two surveys (described in detail below). Experimentally manipulating the control conditions strengthens the explanatory power from the natural variance of managerial trust and cooperation levels and structured management practices on a manager's likelihood to inflate the performance evaluation of an under-performing employee.

Team-Based Measures of Trust and Cooperation

Reported levels of trust and cooperation among employees and managers are common proxies for highly relational management approaches (Blader, Gartenberg, and Prat 2016; Moran 2005). In my setting, these indicators of strong relational management come from PharMed's March 2017 annual employee satisfaction survey. This annual employee satisfaction survey typically has a high response rate at PharMed and the one conducted in early March 2017 was no different with a response rate of nearly 90%[18]. In particular, an index capturing the overall level of trust and cooperation reported by managers is used as the main measure for the extent of relational management[19]. The six questions in this trust and cooperation index are:

- "The people I work with cooperate to get the job done."
- "At PharMed people treat one another with trust and mutual respect."
- "I trust the people I work with to put the work group's goals before their own goals."
- "At PharMed, we collaborate to more effectively meet customer needs."
- "People willingly share what they know with those from other parts of PharMed."
- "Senior leadership's actions show that they trust employees."

[18] Response rate did not substantively vary by function or gender.

[19] Analyses described below were also replicated using alternative measures of team levels of trust and cooperation, namely taking the average of the employees' under each manager's value or a combination of both the employees' and their manager's responses. Additionally, measures using 2016 survey data were also tested in various combinations including change over the prior year as predictor indicators. All results from various measures of trust and cooperation in the team are qualitatively consistent with those presented in the main analyses (and are available upon request).

Respondents answer such questions on the employee satisfaction survey using a 5-point Likert scale with response options ranging from "strongly disagree" to "strongly agree." The trust/cooperation index is a normalized average of the responses across the six questions, where 0 corresponds to choosing "strongly disagree" to all six questions and 1 corresponds to choosing "strongly agree" to all six questions. For a team-based measure, the managers' assessments are included in the average along with the assessments of their direct reports. The Cronbach's alpha coefficient for the six trust and cooperation items is 0.82. Excluding the final item concerning senior leadership only reduces the Cronbach alpha coefficient to 0.81. This slim difference minimizes the concern that including this item with a reference group distinct from the other questions will significantly influence the index's predictive properties[20]. The average favorability score across the six main trust/cooperation items for the teams covered by managers in the study population was 3.75 on the 5-point likert "agreement" scale (s.d.= 0.68). Specifically, 44% of respondent managers have teams with favorable trust/cooperation index scores (4 or above), 45% are more neutral (between 3-4), and only 11% negative (index score less than 3) (see Appendix Figure A1).

Management Practices Inventory Survey Questions

To assess the extent to which existing management practices in a manager's work environment are structured, a 'Management Practices Inventory' (MPI) survey was adapted from the U.S. Census Bureau's Manufacturing and Organizational Practices Survey (MOPS) (see Bloom et al 2013), which itself was developed from the World Management Survey (see Bloom and Van Reenen 2007). The Manager Practices Inventory (MPI) was tailored to the business context of PharMed by excluding questions that would be relevant to only one or a few business lines or that asked for information retrievable for PharMed through other means. This adaptation was further refined through five and a half months of focus groups, in-person cognitive testing, and beta testing the live instrument following the methodology specified in Buffington et al (2016) involving almost 50 PharMed people managers in early 2017[21].

[20]Excluding the question "At PharMed, we collaborate to more effectively meet customer needs" has the most impact on the Cronbach alpha coefficient. However, its exclusion reduces the coefficient only to 0.77.

[21] These managers were excluded from the randomized sample of managers targeted to take the final survey instrument.

The final instrument design entails a two-part survey. The first part has 10 questions. These questions comprise the Manager Practices Inventory (MPI), which assesses the degree and extent to which each respondent manager reports executing structured management practices. The second part contains the experimental vignette case exercise that collects manager responses to a hypothetical instance of evaluating a recognized underperforming employee (described above).

Just as the trust and cooperation index registers a general sense of the levels of these two variables within a team, the same generalized approach was taken to measure how the presence of structured management practices influences each respondent manager. Specifically, the MPI collects two measures of structured management practices from each respondent manager: 1) each manager's self-report about how he manages his team of direct reports and 2) the perception that this respondent manager has about how his own manager manages. This combination incorporates the possible influence that the perception the respondent manager has regarding his own manager's approach on the outcome of interest in this study: the likelihood the respondent manager will inflate an objectively under-performing employee's rating. In other words, it incorporates possible extensive considerations such as signaling competence to or alignment with each manager's own supervisor.

The 10 questions of the MPI questionnaire capture measurements around the management elements of goal setting, monitoring, incentivizing, and use of data in decision making, and follow closely the design used in Bloom et al. (2017), Brynjolfsson et al. (2013), Bloom and Van Reenen (2007), who build in part from the literature around the principles of lean manufacturing (e.g. Womack, Jones and Roos 1990). I aggregate results from the 10 benchmarked questions into a single measure of structured management in two steps using the same methodology as Brynjolfsson et al. (2013).

First, the responses to each of the 10 management questions are normalized on a 0-1 scale. The response option associated with the most structured management practice is normalized to 1, while the response associated with the least structured is normalized to zero. More structured management practices are defined as those that are more specific, formal, frequent or explicit. For example, when asking "Once it has been determined that an employee's under-performance is not resolvable, how long does it currently

take to reassign or dismiss that employee from the following teams?" the response "6 months or less of identifying under-performance" (the shortest timeframe offered as a response) is ranked 1 and the response "Rarely or never" is ranked 0. For this question and others with the option "not applicable…", when this option is selected the question does not factor into the final index score for the respondent manager[22]. For questions which allow for the selection of more than one answer per year, the average of the normalized answers is used as the score for the particular question[23].

Second, the structured management score is calculated as the unweighted average of the normalized responses for the 10 management questions. For the analyses that follow, I then take the z-score of the average, which normalizes this index to have a mean of 0 and a standard deviation of 1. Z-scores are used to facilitate interpreting the outcome of interest vis-á-vis one standard deviation increases in the indexed explanatory variables. Results are very similar to not using the z-score because the average z-score is extremely correlated with the unweighted normalized average.

Altogether, the MPI covers the following elements:

- How many and what type of GOALS or targets for business deliverables and other monitored performance indicators are set (3 questions)
- How frequently activity is MONITORED. For example: Checking in on Goals & Giving feedback about Behavior of employees (4 questions)
- How achievement of those Goals is INCENTIVIZED (2 questions)
- What kind of DATA is used in DECISION MAKING (1 question, parts a and b)

Again, for every question in the MPI, managers respond twice: first, reporting the management practices they use in their own teams, and second, reporting the practices they perceive their own managers as using. See the Methods Appendix for full Management Practices Inventory instrument. I construct a single variable averaging these two parallel sets of responses for each manager, thereby creating a general measure of the level of structured management practices each manager both experiences and practices. This general measure offers a more balanced construct to weigh against the generalized index of trust and

[22] This "not applicable" option departs from the MOPS/WMS protocol but was strongly desired by PharMed. All main results remain robust to inclusion/exclusion of managers who gave a "not applicable" response at some point during the questionnaire (analyses available upon request).

[23] Such "select all that apply" questions also depart from the MOPS/WMS protocol, but again were strongly desired by PharMed. Importantly, the responses to this question do not predict the likelihood of inflating an under-performing employee's rating independently from the composite index used to measure structured management practices. This supports the internal validity of using the MPI as a construct for structured management practices.

cooperation levels which managers reported in their work environment than just using a manager's own self-report alone. However, results remain qualitatively the same if either measure is used independently rather than combined[24].

The mean manager's structured management practices index (SMPi) score (the normalized average across the 10 MPI questions) is 0.63 with a standard deviation of 0.097 (see Appendix Figure A3 for a histogram of the SMPi distribution for respondent managers). The mean score constructed from the focal managers' responses concerning the practices of their supervisors is slightly, but statistically significantly, lower at 0.60 with a larger standard deviation of 0.116 (t-stat= 6.5357; d.f.=361; Appendix Figure A4). A manager's own structured management practice score is moderately correlated with the score given by their responses about their own supervisor's practices, ($r(362)=0.455$, $p<0.001$).

Averaging the two SMPi scores within manager also mitigates the issues of multicollinearity when using both measures to predict the outcome of inflating an underperforming employee's performance rating. The average of these two measures of structured management practices (Averaged SMPi) is 0.62 with a standard deviation of 0.090 (Appendix Figure A2). See Table 2 for the descriptive statistics of the operationalized variables.

[TABLE 2 ABOUT HERE]

The team-based measure of trust and cooperation for respondent managers is moderately (though statistically significantly) predictive of the averaged SMPi score (coeff.=0.05, $p< 0.008$; see Figure A6). As this study is designed to leverage the natural variation of the two main predictive approaches of management on the outcome of interest, inflation of an under-performing employee's performance evaluation, this is not problematic. Rather, the objectivity of the experimentally designed vignette enables the assessment of the relative strength of each managerial approach no matter the extent to which they

[24] In the appendix, I include a battery of robustness analyses reporting the individual effects of each of the two measures separately and discuss the implications of each measure's relative predictive power for likelihood of inflating the performance rating of an under-performing employee.

covary among the populations of managers in this study. The main models control for both relative approaches simultaneously.

Administrative Data: Underperforming employees in embedded managerial relationships

Finally, I use HR administrative data collected from PharMed to test for evidence that managerial embeddedness is costly through increased likelihood of inflating employee performance evaluations. These records document employee tenure, number of managers, and whether terminated employees left for performance reasons (i.e. involuntarily) or for non-performance reasons, in years 2014-2017 (See Table 3). I use the measured levels of trust and cooperation from PharMed's March 2017 annual employee satisfaction survey (described above) to predict whether such indicators of managerial embeddedness translate into systematic termination patterns.

[INSERT TABLE 3 ABOUT HERE]

A second analysis with PharMed's administrative data calculates the number of managers each employee has had during their tenure with PharMed to proxy employees' average length of managerial relationship as a measure for managerial embeddedness. Controlling for overall employee length of service with the company, it is expected that employees terminated for performance reasons will have had more managers than those not terminated or those who left voluntarily. In other words, underperforming employees will be more likely to be dismissed for underperformance when the length of their relationship with their manager is shorter (where they are less likely to have developed an embedded relationship).

ANALYSIS

All survey-based measures are at the individual respondent manager level. Although there is no systematic difference in the likelihood of inflating an underperforming employee by manager demographics, other scholars have reported that age and gender have positive effects on leniency bias (Bol 2011). Therefore, all analyses include controls for manager age and gender.

The main analyses are logit regressions of the two main explanatory variables: the trust and cooperation index and the average structured management practice index, on the outcome variable of

whether or not a manager gives an inflated performance rating in the hypothetical case vignette experiment concerning an under-performing employee[25]. To assess whether the effect of structured management practices on reducing a manager's likelihood of inflating a hypothetical under-performing employee depends on the level of the trust and cooperation index, I use a simple interaction term ('trust and cooperation index' x 'average structured management practice index'). The explanatory variables are entered as z-scores of their indexes and the results are reported in odds ratios, meaning the coefficients indicate the change in likelihood of accuracy for a one standard deviation increase in a given explanatory variable index score. The outcome variable is binary where 1 equals an inflated performance rating given in the case response and 0 signifies a non-inflated, or accurate, case response.

Administrative Data

To assess whether there is empirical evidence that embedded managerial relationships are associated with a cost in retaining underperformers, I test whether indicators for trust and cooperation, as well as manager attachment, are associated with a lower likelihood of terminating employees for performance reasons specifically. Such evidence is suggestive here rather than causal as underperformers may be more likely to be shifted around to multiple managers.

First, I logistically regress team trust/cooperation index (using all available data rather than only respondents to the MPI)[26] on the likelihood an employee is terminated for performance reasons rather than leaves voluntarily or remains with the company. I further test this relationship on only the subset of employees that left the organization. Second, I test whether the length of manager-employee attachment predicts the termination of underperforming employees. Specifically, I test whether there is a difference in the number of managers each employee had in each of the termination categories: involuntary, voluntary, and not terminated, controlling for length of tenure and other demographics. Given that trust and cooperation are elements developed over time between managers and employees, the longer the managerial relationship the less likely the manager will terminate an under-performing employee.

[25] All outcomes are robust to using OLS specifications as well.

[26] Data limitations limit the ability to conduct similar validity checks on the *structured management practices construct* on the likelihood of terminating employees for performance reasons using the administrative data.

RESULTS

Controlling for empirical confounds

Before testing the explanatory power of the competing managerial approaches to evaluating an underperforming employee (embedded relational versus structured managerial approach), I first use the experimentally manipulated conditions of the hypothetical case vignettes to predict whether or not a manager inflates a rating. These conditions control for the most common variables that may confound employee performance decisions. Thus they establish a robust baseline from which to test the relative predictive powers of relational and structured management approaches.

Unsurprisingly, two of these empirically motivated control conditions, a) justifying a 'Solid' rating rather than a 'Partially Met' rating and b) underperformance on the behavioral dimension rather than work deliverables, are large in magnitude and significance (Model 1 in Table 4). Almost all respondents who considered an employee whose accurate rating is one that went from 'Exceptional' in the year prior to 'Solid' accurately rate the under-performer (only 2.35% of managers considering an employee's whose accurate performance rating is 'Solid' give an inflated rating in this condition). On the other hand, when a manager must contend with whether or not to give an under-performer a 'Partially Met' rating (a domain of the employee performance curve distribution that is not subject to a predefined forced target in this company), many more (27.27% of managers in this condition) give an inflated rating (most commonly a 'Solid' rating). The simple tabulations of manager accuracy by all 12 conditions manipulated in the vignette design are reported in Table 1. Again, these 12 conditions serve to address and control for the most common reasons that managers at PharMed offer as determinants of performance ratings given.

More surprisingly, the behavioral concern conditions significantly predicts lower odds of inflating the under-performing employee's rating compared to under-performance concerns in work deliverables. One potential explanation for this is that in this organization, which has high generally reported levels of trust and cooperation (see Figure A1), managers are more willing to accurately down-rate employees who exhibit problematic behaviors which may seem to violate common norms of cooperativeness. Indeed, the official PharMed performance evaluation guidelines emphasize that the behavioral cause for concern is

dominated by the potential impact on peers and problems within the work team[27]. Finally, there is no significant predictive difference between the conditions manipulating how the employee came to the respondent manager's team in the vignette.

Relational Managerial: High Trust and Cooperation

H1a: *Highly relational managers, operating on a greater basis of trust and cooperation in their work units, are less likely to inflate an under-performing employee's performance evaluation. vs.*

H1b: Highly relational managers, operating on a greater basis of trust and cooperation in their work units, are more likely to inflate an under-performing employee's performance evaluation.

Above, I argue that in the context of employee performance evaluation, high levels of trust and cooperation may prevent managers from making accurate rating decisions regarding an underperforming employee. Specifically, I expect that managers with higher reported levels of trust and cooperation in their work environment are more likely to inflate the performance rating of an underperforming employee. As seen in Model 2 of Table 4, the main effect of a one standard deviation increase in manager's reported level of cooperation and trust in their work environment predicts a 76% increase in odds that the manager gives an underperforming employee an inflated performance rating (s.d.=0.30, $p<0.0001$; controlling for empirically observed confounds and demographic control variables [entered independently in Model 1]). Indeed, managers reporting higher levels of trust and cooperation are more likely to inflate an under-performing employee's performance.

Predicting likelihood of retaining under-performers

H2: More embedded managers are less likely to dismiss an employee for underperformance.

Given the experimental survey evidence that managers fostering greater trust and cooperation in their teams are more likely to inflate an underperforming employee's evaluation, we expect this will translate to actual employee outcomes as well. As seen in Table 5 (Model 7), a one standard deviation in trust and cooperation levels reported translates to 78% lower odds ($p<0.021$) of termination for performance

[27]The specific wording in behavioral problem cases is: "Jim's work delivery satisfactorily meets expectations. However, since Jim came to your team ten months ago, Jim's behavior has been disappointing. Despite the quality of his work, it seems his delivery comes at the expense of his peers. Relationships with Jim have been deteriorating and complaints about his behavior have been increasingly common. His behavior, generally, has become a concern for your team. You have held several one-on-one meetings with Jim to discuss expectations and what Jim needs to do in order to meet them but despite this attention, Jim's behavior has not been improving."

reasons, controlling for length of tenure and including both terminated and non-terminated employees since 2014. This translates into a 55% percentage point decrease for managers with mean levels of trust and cooperation in their teams compared to those reporting trust and cooperation levels one standard above the mean. Restricting the analysis to predicting involuntary termination only among the terminated employees, we find a similar relationship (Model 8, Table 5: 0.05, $p<0.025$).

Looking at length of managerial relationships and employee termination for performance reasons, of all employees who have left PharMed in the 2014-2017 window, those who were terminated for performance related reasons (left involuntarily) on average had 2.8 more managers over the same length of tenure with the organization than employees who were not terminated. For comparison, those who left voluntarily (not for performance related reasons) had fewer managers (-0.25 on average, n.s.) than those who stayed with the organization (controlling for employee's tenure). These findings suggest that dismissal for underperformance requires an employee to be rotated through several managers. The length of relationship with one manager is negatively correlated with the likelihood that the employee will eventually be dismissed for performance reasons. See Figure 3. While this evidence from PharMed's administrative HR data sources is suggestive rather than causal, these two analytical tests point in the direction that we would expect if employee under-performance is more likely to persist the more embedded the managerial relationship.

Structured Management Practices

H3: Managers with more structured management practices are <u>less likely to inflate</u> an employee's performance evaluation.

Literature on formalized work practices and structured management practices suggest that a more structured approach to management will lower the odds that a manager inflates the performance rating of an under-performing employee. I look not only at how structured an approach to management each respondent manager takes but also how structured an approach to management they perceive their own manager takes. Both of these approaches have similar direct effects on predicted odds of inflating the performance rating of an under-performing employee. A standard deviation increase in a respondent manager's reported level of structured management practices reduces the odds of inflating an under-

performer's performance rating by nearly 33% (s.d.=0.10, p<0.01; Model 3 in Table 4; See Figure 2). This essentially reverses the odds as predicted by high levels of trust and cooperation in a manager's team. Using respondent managers' perceptions of their own supervisor's practices produces highly similar predictions of inflating an under-performer's rating (as seen by comparing Model 3 and 4 in Table 4). This is unsurprising given the expected multicollinearity between these two measures.

With the average of the two structured management practices indexes – the respondent manager's own score and the score they give their own manager – as the main explanatory predictor of structured management practices on the likelihood that a manager inflates the performance rating of an under-performing employee generates a structured management practices indicator that is more aligned with the nature of the trust and cooperation indicator[28]. While the magnitude of this averaged index is lower than the coefficient of either component alone, its predictive power is unsurprisingly stronger since they generally move in tandem (see figure A5). This averaged indicator of structured management practices predicts that the odds that a manager inflates an underperforming employee's rating is reduced by 45% given a standard deviation increase in the level of structured management a manager reports (s.d.=0.10, p<0.001; Model 5 in Table 4). This supports the prediction that the more structured an approach to management a manager takes and experiences himself, the lower the odds are that he will inflate the rating of an underperforming employee (See Figure 2).

Are Relational and Structured Management Approaches Independent?

H4: relational and structured management operate <u>independently and simultaneously</u> on the likelihood that a manager inflates an underperforming employee's performance evaluation.

Finally, I assess whether the reduction in odds of inflating an underperforming employee's performance evaluations from increased structured management practices depends on the level of trust and cooperation in the team. Model 6 in Table 4 tests for this dependence by incorporating an interaction term of the two variables. How the two main independent drivers of a manager's willingness to inflate an underperformer's rating operate together draws attention to whether management practices may either

[28] It also addresses multicollinearity, as explained above.

moderate or mediate relational embeddedness. The interaction term for these two main independent variables ('trust and cooperation index' x 'structured management practices index') is non-significant (Model 6 in Table 4; Figure 3)[29]. Consequently, there is no evidence that higher trust and cooperation in teams reduce the power for structured management practices to lower the predicted likelihood that a manager will inflate the rating of an under-performer[30].

[TABLE 4 ABOUT HERE]
[FIGURE 2 & FIGURE 3 ABOUT HERE]
[FIGURE 4 ABOUT HERE]

DISCUSSION

Implications for Managers and Organizations

The managerial implications for this research are straightforward. How an under-performing employee is evaluated and rated in performance assessments can carry important consequences for employee and team morale, competency improvements, unit efficiency, and general productivity. Unwillingness to make accurate performance evaluations hinders performance management systems and interferes with personnel processed designed to develop employees by improving identified weaknesses. Consequently, both relational and structured management practices carry important equity and productivity implications for organizations.

In a population of managers in a multinational company, I find that managers who report higher levels of cooperation and trust are more likely to inflate the performance rating of an employee whose performance is objectively lower than last year's rating. As results from analyses on the company's HR administrative data reveal, such a relational bias in employee performance evaluations translates into retaining more under-performing employees over time. It takes being rotated among other managers for poor performers to be dismissed for under-performance. At the same time, I find that managers who take a

[29] This analysis is robust to including only the top quartile interaction of structured management practices or lowest as well as a simple multiplicative interaction term between the two variables (see Table A1).
[30] More detailed mediation and moderation analyses were also examined but unsurprisingly, given the non-significance of the interaction term between the two explanatory variables, these also did not offer any support for structured management practices predicting the lower likelihood of rating inflation by influencing the level of trust or cooperation in the team or its influence on the manager's likelihood to inflate the rating of an under-performer. Further analyses interacting the two SMPi constructs, respondent manager and supervisor scores, separately with the trust and cooperation index yield similar non-significant results.

more structured (more specific, formal, frequent or explicit) approach to managing are less likely to inflate the performance rating of an under-performing employee, regardless of trust and cooperation levels. I also find empirical evidence, using administrative records from the same company, that the more embedded the relationship between managers and their employees, the more likely underperformance will go unaddressed in the organization. This evidence also suggests that this unwillingness to accurately evaluate the poor performance translates into a lower likelihood of dismissing under-performing employees.

Moreover, I find evidence that the two approaches – trust and cooperation levels and structured management practices – operate independently on the odds that a manager will inaccurately inflate the performance rating of an underperforming employee. Such independence implies that in the case of employee evaluations, structured management practices do not necessarily interact with the development of trust and cooperation within a team nor its direct influence on performance evaluations. Rather, the results suggest that the improvement which structured management practices make on evaluation accuracy is strong no matter how trusting and cooperative a manager's relation is with employees. The positive implication for managers from this independence is that it underscores the possibility of maintaining well-functioning performance management processes, such as official performance review systems, alongside a climate of high trust and cooperation, through the practice of structured management. In other words, this independence suggests that structured management approaches can improve performance management processes even when managers are highly relational in their management approach.

Part of the advantage of taking a structured management approach is the finding that it does not necessarily crowd out the relationally embedded approach. Rather, I find that taking a structured approach to managing reduces the likelihood of inflating poor performers' evaluations without minimizing the levels of trust and cooperation held within teams.

Theoretical Implications

Both the literature on embeddedness and the value of structured management practices have heretofore been limited to inter-organizational studies. This study explores the extent to which those lessons

carry over to the intra-organizational level. Such a move also illuminates the importance of interpersonal relationships on the prescriptions made by both sets of theory.

Rather than merely extend what we know about embeddedness from more aggregate studies, using intra-organizational, team-based measures can shed light on both interpersonal and organizational consequences as well as highlight how varying managerial approaches influence general organizational agency problems. The implications of this study are that the costs to over-embedded relationships are real even at the interpersonal level of managers and their teams. The finding that more embedded managerial relationships correspond with a greater likelihood that a manager will inflate an under-performing employee's evaluation demonstrates one source of the costs of over-embeddedness. Additional costs are highlighted in the evidence that longer manager relationships seem to protect under-performers from involuntary exits for performance reasons.

Despite this, trust and cooperation are still beneficial attributes for work teams to foster. This study is careful not to claim that these elements *cause* evaluation inflation. Rather, the correlation is highlighted to caution management approaches premised solely under a relational model, without safeguards to ensure that the hidden costs of over-embeddedness do not compromise broader organizational aims. The utility of formalized work practices offers a check against possible negative consequences of embedded managerial relationships. Specifically, structured management practices – the deliberate use of monitoring, goal setting, incentives, and data in decision making – drastically reduce the tendency to inflate an under-performer's performance evaluation.

Most importantly, the lack of evidence that structured management practices come at the expense of fostering trust and cooperation in teams highlights the value of studying the consequences of embeddedness in the context of countervailing and complementary processes. Managers can both manage teams through highly cooperative and trusting relationships, as well as make objective employee evaluations by taking a structured approach to managing. This insight implies that the U-shaped return to embedded relationships' productivity is not a foregone conclusion. In contrast with Canales' (2014) findings, managers are not required to negotiate their particular approach with other parties to ensure

making accurate employee evaluations. Managers can be directed toward one managerial approach over another (Chatterji et al. 2019). Yet the potential negative consequences to over-embeddedness can also be kept in check when managers adopt more structured management approaches in addition to maintaining highly trusting and cooperative approaches within their teams.

References

Anderson, C.D. and Tomaskovic-Devey, D., 1995. Patriarchal pressures: An exploration of organizational processes that exacerbate and erode gender earnings inequality. *Work and Occupations*, *22*(3), pp.328-356.

Bazerman, M.H., Beekun, R.I. and Schoorman, F.D., 1982. Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee. *Journal of Applied Psychology*, *67*(6), p.873.

Blader, Steven, Claudine Gartenberg, and Andrea Pratt. 2016. "The Contingent Effects of Management Practices." http://www.columbia.edu/~ap3116/papers/BGP_trucks1118.pdf (accessed June 1, 2017).

Blader, Steven, Claudine Gartenberg, Rebecca Henderson, and Andrea Prat. 2015. "The Real Effects of Relational contracts" American Economic Review: Papers & Proceedings 105(5): 452-456.

Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron Jarmin, Itay Sporta-Eckstein and John Van Reenen. 2017. "What Drives Differences in Management?" *NBER Working Paper Series*: working paper 23300:

Bloom, Nicholas and John Van Reenen. 2007 "Measuring and Explaining Management Practices across Firms and Countries" *The Quarterly Journal of Economics* 122(4):1351-1408).

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. "The Organization of Firms across Countries". 2012. *The Quarterly Journal of Economics.*

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2016. "Management as a Technology?" Harvard Business Working Paper 16-133.

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

Bond, Brittany M., Tatiana Labuzova, and Roberto M. Fernandez. 2018. "At the Expense of Quality" *Sociological Science* 3: 10-32.

Bond, Brittany M. and Roberto M. Fernandez. 2020. "Networks for the Unemployed?" Social Networks at Work edited by Daniel J. Brass and Stephen P. Borgatti.

Buffington, Catherine, Herrell, Kenneth and Ohlmacher, Scott. 2016. The management and organizational practices survey (MOPS): Cognitive testing.

Canales, Rodrigo. 2014. "Weaving Straw into Gold: Managing Organizational Tensions between Standardization and Flexibility in Microfinance". *Organization Science* 25(1): 1-28.

Canales, Rodrigo and Greenberg, Jason. 2016. "A matter of (Relational) Style: Loan Officer Consistency and Exchange Continuity in Microfinance". *Management Science* 62(4): 1202-1224.

Cappelli, Peter and Anna Tavis. 2016. "The Performance Management Revolution." *Harvard Business Review.* October

Castilla, Emilio J. 2008. "Gender, Race, and Meritocracy in Organizational Careers." *American Journal of Sociology* 133: 1479-1526.

Castilla, Emilio J. 2011. "Bringing Managers Back In: Managerial Influences on Workplace Inequality". *American Sociological Review*. 76(5): 667-694.

Castilla, Emilio J. and Stephen Benard. 2010. "The Paradox of Meritocracy in Organizations." *Administrative Science Quarterly* 55: 543-76.

Chan, CSC. 2009. "Invigorating the content in social embeddedness: An ethnography of life insurance transactions in China" *American Journal of Sociology.* 115(3): 712-754.

Chatterji, A., Delecourt, S., Hasan, S. and Koning, R., 2019. When does advice impact startup performance?. *Strategic Management Journal*, 40(3), pp.331-356.

Coleman, J.S., 1988. Social capital in the creation of human capital. *American journal of sociology*, *94*, pp.S95-S120.

DiMaggio, P. and Louch, H., 1998. Socially embedded consumer transactions: For what kinds of purchases do people most often use networks?. *American sociological review*, pp.619-637.

Dobbin, F., Schrage, D. and Kalev, A., 2015. Rage against the iron cage: The varied effects of bureaucratic personnel reforms on diversity. *American Sociological Review*, 80(5), pp.1014-1044.

Doering, L., 2018. Risks, Returns, and Relational Lending: Personal Ties in Microfinance. American Journal of Sociology, 123(5), pp.1341-1381.

Duhaime, Erik P., Brittany M. Bond, Qi Yang, P. de Boer, and T.W. Malone. 2020. "Recruiting Hay to Find Needles: Recursive Incentives and Innovation in Social Networks." arXiv:1912.06922 [cs.SI] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3545311

Elfenbein, D.W. and Zenger, T., 2017. Creating and capturing value in repeated exchange relationships: The second paradox of embeddedness. Organization Science, 28(5), pp.894-914.

Elvira, M.M. and Graham, M.E., 2002. Not just a formality: Pay system formalization and sex-related earnings effects. Organization Science, 13(6), pp.601-617.

Fernandez, R.M., 1991. Structural bases of leadership in intraorganizational networks. *Social Psychology Quarterly*, pp.36-53.

Gibbons, R. and Henderson, R., 2012. Relational Contracts and organizational capabilities. *Organization Science*, 23(5), pp.1350-1364.

Gibbons, Robert and John Roberts. 2013. The Handbook of Organizational Economics. Princeton: Princeton University Press.

Gittell, Jody Hoffer. 2002. "Coordinating Mechanisms in Care Provider Groups: Relational Coordination as a Mediator and Input Uncertainty as a Moderator of Performance Effects". *Management Science* 48(11):1408-1426.

Gittell, J.H., 2016. Transforming relationships for high performance: The power of relational coordination. Stanford University Press.

Gittell, J.H. and Douglass, A., 2012. Relational bureaucracy: Structuring reciprocal relationships into roles. Academy of Management Review, 37(4), pp.709-733.

Gittell, J.H. and Fields, A., 2003. *The Southwest Airlines way: Using the power of relationships to achieve high performance*. New York: McGraw-Hill.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. *American journal of sociology*, 91(3), pp.481-510.

Grote, Dick. 2005. *Forced Ranking: Making Performance Management Work.* Harvard Business Review Press.

Hamner, W.C., 1975. How to ruin motivation with pay. *Compensation Review*, 7(3), pp.17-27.

Hainmueller, J., Hangartner, D. and Yamamoto, T., 2015. Validating vignette and conjoint survey experiments against real-world behavior. Proceedings of the National Academy of Sciences, 112(8), pp.2395-2400.

Heath, C., Knez, M. and Camerer, C. 1993. "The Strategic Management of the Entitlement Process in the Employment Relationship", *Strategic Management Journal*, 14: 75-93.

Heimer, Carol A. 1992. “Doing Your Job and Helping Your Friends: Universalistic Norms about Obligations to Particular Others in Networks” in Networks and Organizations: Structure, Form, and Action, edited by Nitin Nohria and Robert G. Eccles. Boston, Mass: Harvard Business School Press.

Helper, Susan and Rebecca Henderson. 2014. “Management Practices, Relational Contracts, and the Decline of General Motors,” *Journal of Economic Perspectives*. 28: 49-72.

Henemen, Robert L. and Jon M. Werner. 2005. *Merit pay: Linking Pay to Performance in a Changing World*, 2nd ed. Greenwich, CT: Information Age Publishing.

Huffman, M.L. and Velasco, S.C., 1997. When more is less: Sex composition, organizations, and earnings in US firms. *Work and Occupations*, *24*(2), pp.214-244.

Huselid, M.A., 1995. The impact of human resource management practices on turnover, productivity, and corporate financial performance. *Academy of management journal*, 38(3), pp.635-672.

Ichniowski, C., Kochan, T.A., Levine, D., Olson, C. and Strauss, G., 1996. What works at work: Overview and assessment. *Industrial Relations: A Journal of Economy and Society*, 35(3), pp.299-333.

Ichniowski, C., Shaw, K. and Prennushi, G., 1997. The effects of human resource management practices on productivity: A study of steel finishing lines. *The American Economic Review*, pp.291-313.

Jordan, Gretchen. 2015. “What Matters to R&D Workers” Research Technology Management. 48(3):23-32.

Kahneman, D. and Tversky, A. 1979. “Prospect Theory: An Analysis if Decision under Risk”, *Econometrica*, 47:263-92.

Kotzé, Koos, and Gert Roodt. 2005. “Factors that affect the Retention of Managerial and Specialist Staff: An Exploratory Study of an Employee Commitment Model.*” SA Journal of Human Resource Management*. 3(2):48-55.

Larkin, I., Pierce, L., & Gino, F. (2012). The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, *33*(10), 1194-1214.

Lemos, R. and Scur, D., 2018. CEP Discussion Paper No 1528 January 2018 All in the Family? CEO Choice and Firm Organization.

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

Milkovich, G., Newman, J., 1996. Compensation, 4th ed. Irwin, Homewood, IL.

Moran, Peter. 2005. “Structural vs. Relational Embeddedness: Social Capital and Managerial Performance” *Strategic Management Journal* 36(12): 1129-1151.

Pfeffer, Jeffrey. 1997. *The Human Equation: Building Profits by Putting People First. Harvard Business School Press Books.*

Pfeffer, J. and Sutton, R.I., 2006. Hard facts, dangerous half-truths, and total nonsense: Profiting from evidence-based management. Harvard Business Press.

Phillips, D.J., Turco, C.J. and Zuckerman, E.W., 2013. Betrayal as market barrier: Identity-based limits to diversification among high-status corporate law firms. *American Journal of Sociology*, 118(4), pp.1023-1054.

Poppo, L., Zhou, K.Z. and Ryu, S., 2008. Alternative origins to interorganizational trust: An interdependence perspective on the shadow of the past and the shadow of the future. *Organization Science*, 19(1), pp.39-55.

Prendergast, C. and Topel, R., 1993. Discretion and bias in performance evaluation. European Economic Review, 37(2-3), pp.355-365.

Prendergast, C. and Topel, R.H., 1996. Favoritism in organizations. Journal of Political Economy, 104(5), pp.958-978.

Reskin, B.F., 2000. Getting it right: Sex and race inequality in work organizations. Annual Review of Sociology, 26(1), pp.707-709.

Rossi, Peter H. and Andy B. Anderson. 1982. "The factorial survey approach: An introduction." Measuring social judgments: The factorial survey approach: 15-67.

Rousseau, D. M. (2004). Psychological contracts in the workplace: Understanding the ties that motivate. Academy of Management Perspectives, 18(1), 120-127.

Saal, F. E., R. G. Downey, and M. A. Lahey. 1980. Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin* 88 (2): 413–428.

Sako, M. and Helper, S., 1998. Determinants of trust in supplier relations: Evidence from the automotive industry in Japan and the United States. *Journal of Economic Behavior & Organization*, 34(3), pp.387-417.

Sorenson, Olav and David M. Waguespack. 2006. "Social Structure and Exchange: Self-Confirming Dynamics in Hollywood" *Administrative Science quarterly.* 51(4): 560-589.

Stone, D. and Heen, S., 2015. Thanks for the feedback: The science and art of receiving feedback well. Penguin.

Stone, D., Heen, S. and Patton, B., 2010. Difficult conversations: How to discuss what matters most. Penguin.

Turco, C., 2012. Difficult decoupling: Employee resistance to the commercialization of personal settings. *American Journal of Sociology*, 118(2), pp.380-419.

Uzzi, B., 1996. "The sources and consequences of embeddedness for the economic performance of organizations: The network effect." *American Sociological Review*, pp.674-698.

Uzzi, B., 1997. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative science quarterly*, pp.35-67.

Uzzi, B. and Lancaster, R., 2004. Embeddedness and price formation in the corporate law market. *American sociological review*, 69(3), pp.319-344.

Weber,M.1946. Bureaucracy. InH.H.Gerth&C.WrightMills (Eds.), From Max Weber: Essays in sociology: 196–244. Princeton, NJ: Oxford University Press.

Wright Thomas A., Douglas G. Bonnett. 2007 "Job Satisfaction and Psychological well-being as nonadditive predictors of workplace turnover" *Journal of Management* 33(2): 141-160.

Womack, J.P., Jones, D.T. and Roos, D., 1990. Machine that changed the world. Simon and Schuster.

WorldatWork. 2018. "2017 Inventory of Total Rewards Programs & Practices". https://www.worldatwork.org/docs/research-and-surveys/Report_FINAL1.pdf

Zuckerman, Ezra. 2014. "In Either Market or Hiearchy, but not in Both Simultaneously: Where Strong-Tie Networks are Found in the Economy" *Contemporary Perspectives on Organizational Social Networks*, Research in the Sociology of Organizations, Vol 40: 111-133.

TABLES & FIGURES

Table 1. Percent Inflating Rating by Randomized Control Conditions in Case Vignette

Percent Rating Inflation on Case by Condition		n	
New Employee	17.89%	246	
New Manager	20.80%	250	496
Behavioral Concern	14.74%	251	
Work Quality	24.08%	245	496
Exceptional->Solid	2.35%	173	
Exceptional ->Partially Met	22.22%	154	
Solid->Partially Met	34.32%	169	496

Table 2. Descriptive Statistics for the Operationalized Variables[31]

Variable	Obs	Mean	Std.	Min	Max
Inflate Rating	496	0.19	0.40	0	1
Coop Trust Average	432	3.82	0.59	1.33	5
SMPi Manager	444	0.63	0.10	0.31	0.88
SMPi Manager's Supervisor	403	0.60	0.12	0.12	0.87
Average	362	0.62	0.09	0.36	0.84
Ave SMPi, bottom quartile	91	0.51	0.04	0.36	0.56
Ave SMPi, second quartile	90	0.60	0.02	0.56	0.63
Ave SMPi, third quartile	91	0.66	0.02	0.63	0.68
Ave SMPi, top quartile	90	0.73	0.04	0.68	0.84

Table 3. Breakdown of Termination Outcome for All Employees Ever Employed 2014-2017[32]

Termination Class	Percent of Emp Popn
Not Terminated	69%
Voluntary Termination	16%
Involuntary Termination	16%
Rounded Total N	~ 10,000

[31] For all managers with at least one calculable structured management practice score (either for respondent manager's practice or for their supervisor). The number of observations differs by variable because of differences in response rates between the employee satisfaction survey and the MPI, as well as due to incomplete responses in the questionnaire used to score the *structured management practices* index by respondent.

[32] Percentages and rounded total are used rather than raw counts to protect company identity.

Table 4. Predicted Likelihood a Manager Inflates the Performance Rating of an Underperforming Employee by Reported Cooperation/Trust and Structure Management Practices (SMP) Levels, reported in odds ratios (z signifies z- scores for ease of interpretation)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Inflate						
Trust/Cooperation index, z		1.758***				2.442***
		(0.30)				(0.68)
Mngrs own SMPi, z			0.671**			
			(0.10)			
Mngrs' Supervisor's SMPi, z				0.708*		
				(0.10)		
Averaged SMPi score, z					0.551***	0.459***
					(0.10)	(0.09)
Ave SMPi x Trust/Coop ind, z						1.264
						(0.28)
Case: New Emp vs New Mngr	0.948	0.86	0.747	0.965	0.756	0.715
	(0.24)	(0.23)	(0.22)	(0.30)	(0.26)	(0.26)
Case: Bhvr vs Work dlvrbls	0.378***	0.380***	0.281***	0.323***	0.246***	0.235***
	(0.10)	(0.10)	(0.080)	(0.10)	(0.080)	(0.09)
Case: Solid vs Partially Met	0.052***	0.050***	0.037***	0.013***	0.010***	0.009***
	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Female	1.134	1.186	1.309	1.302	1.457	1.262
	(0.29)	(0.32)	(0.39)	(0.4)	(0.49)	(0.45)
Age	1.003	0.996	0.986	0.968	0.961	0.953
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Constant	0.573	0.687	1.531	2.914	5.134	6.113
	(0.46)	(0.57)	(1.43)	(2.88)	(5.70)	(7.29)
r2_p	0.168	0.196	0.231	0.256	0.312	0.367
chi2	80.062	90.021	91.822	92.911	102.573	113.56
Observations	488	473	405	374	332	322

Note: the unit of analysis is the manager.
* p<0.05, ** p<0.01, ***

Figure 2. Predicted Likelihood of Inflating an Underperformer's Performance Rating by Average SMPi score for a Manager and Level of Trust/ Cooperation (normalized to index between 0-1)

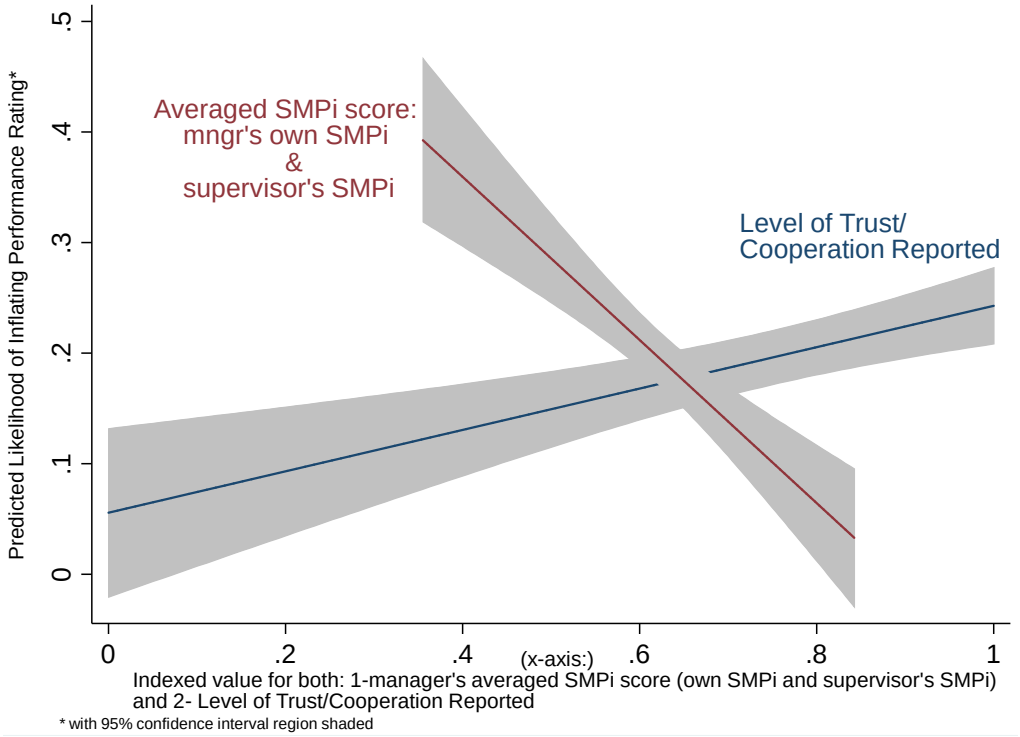


Figure 3. Predicted Likelihood of Inflating an Underperformer's Performance Rating by Interaction Terms between High(Hi) vs. Low(Lo) SMPi scores and Level of Trust/ Cooperation.

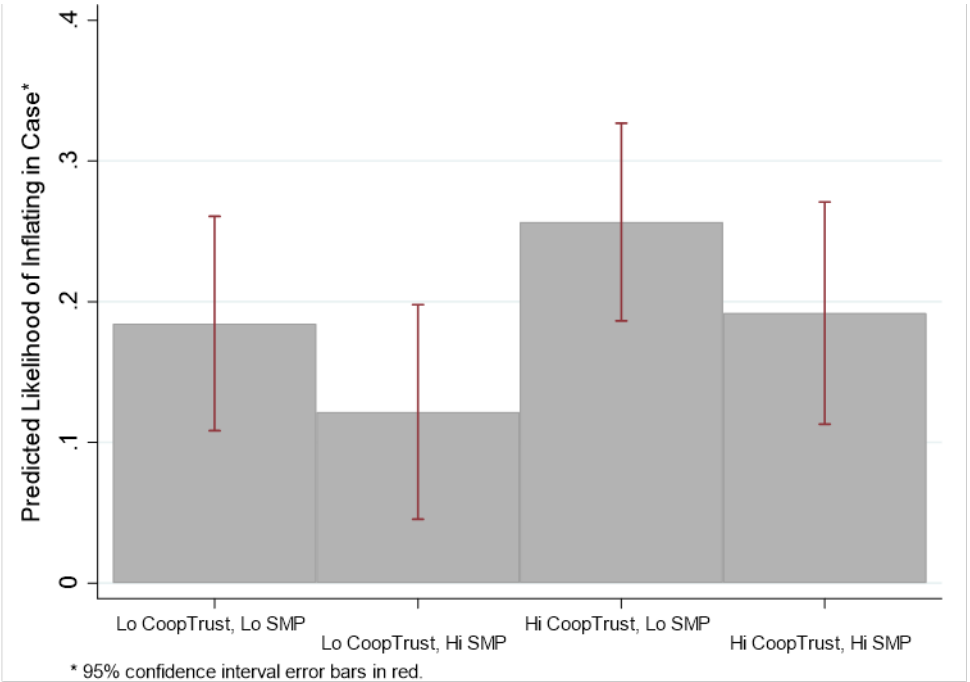
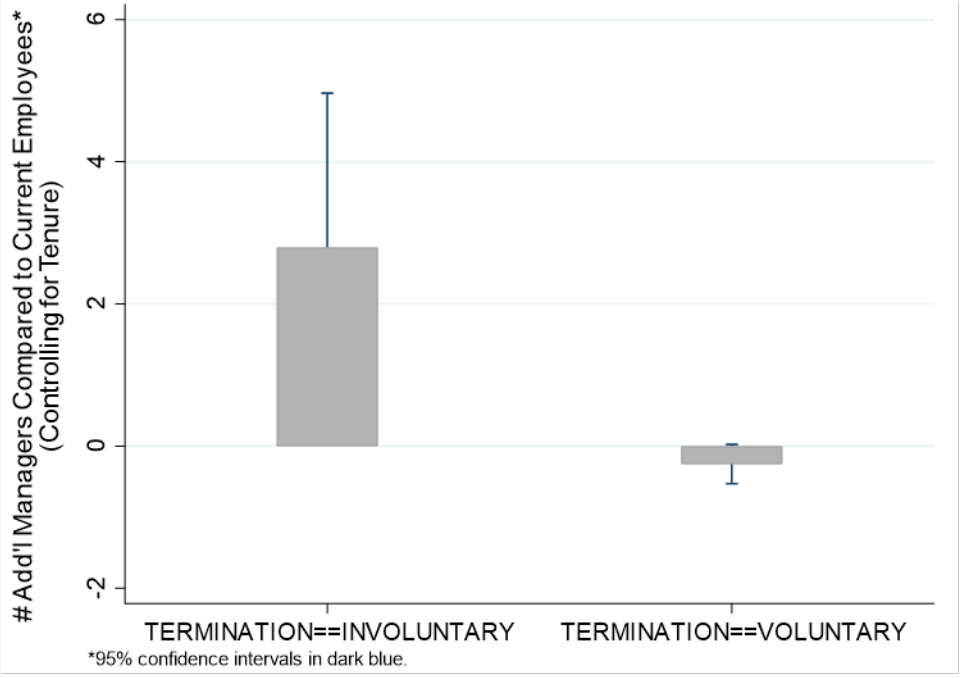


Table 5. Predicted Likelihood an Employee is Terminated for Performance Reasons (Involuntary termination by Reported Cooperation/Trust, reported in odds ratios (z signifies z- scores for ease of interpretation)

	Model 7: All Emps b/se	Model 8: Terminated Emps b/se
INVOLUNTARY TERMINATION		
Trust/Cooperation index, z	0.223*	0.051*
	(0.14)	(0.07)
Length of Service in Years	0.874	0.961
	(0.14)	(0.27)
_cons	0.000**	0.000*
	(0.00)	(0.01)
r2_p	0.217	0.569
chi2	8.633	14.633
N	844	82

* p<0.05, ** p<0.01, *** p<0.001

Figure 4. Predicted Difference in Number of Managers by Type of Termination, Compared to Non-Terminated Employees (controlling for employee tenure).



RESULTS APPENDIX

Figure A1. Distribution of Cooperation/Trust index scores for respondent manager teams

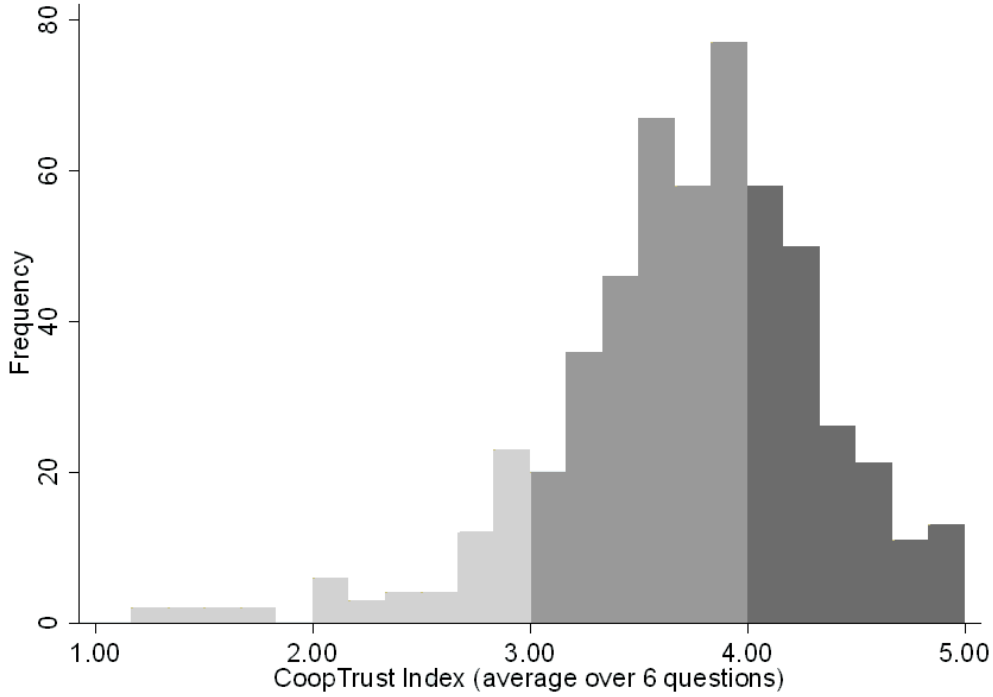


Figure A2. Distribution of Averaged Structured Management Practice Index (SMPi) scores for respondent manager (average of SMPi of resopndent rmanager and SMPi of respondent manager's supervisor)

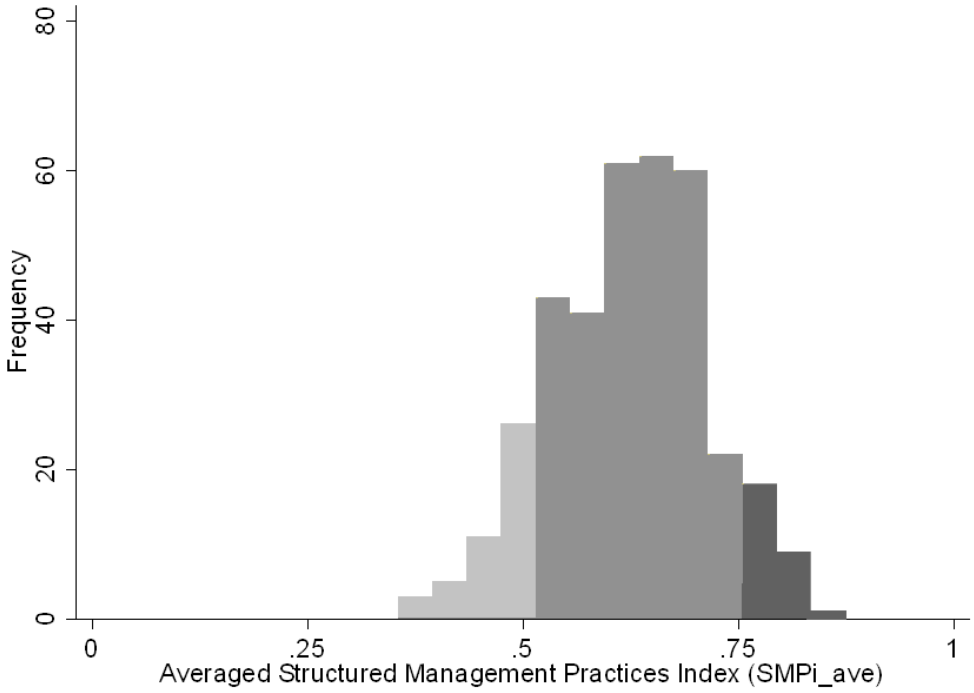


Figure A3. Distribution of Structured Management Practice Index (SMPi) scores for respondent manager's practices

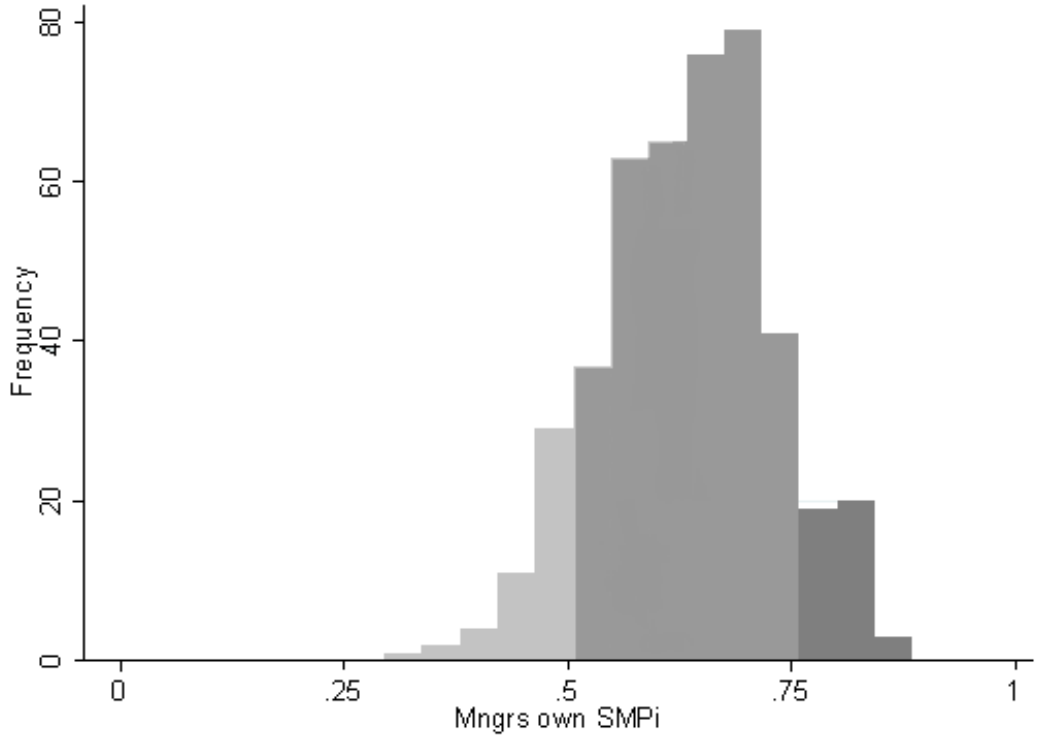


Figure A4. Distribution of Structured Management Practice Index (SMPi) scores for respondent manager's supervisor

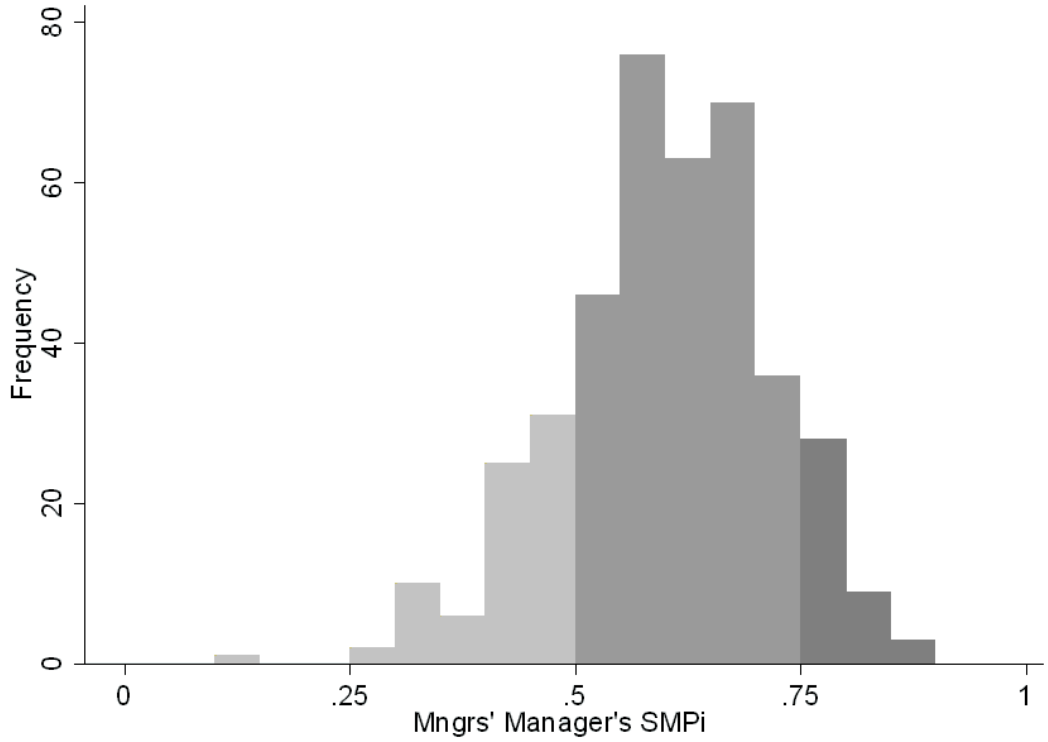


Figure A5. Manager's own Structured Management Practices (SMP) score collinear with their perception of their Supervisor's SMP score [Coeff=0.55* S.D.=0.06]**

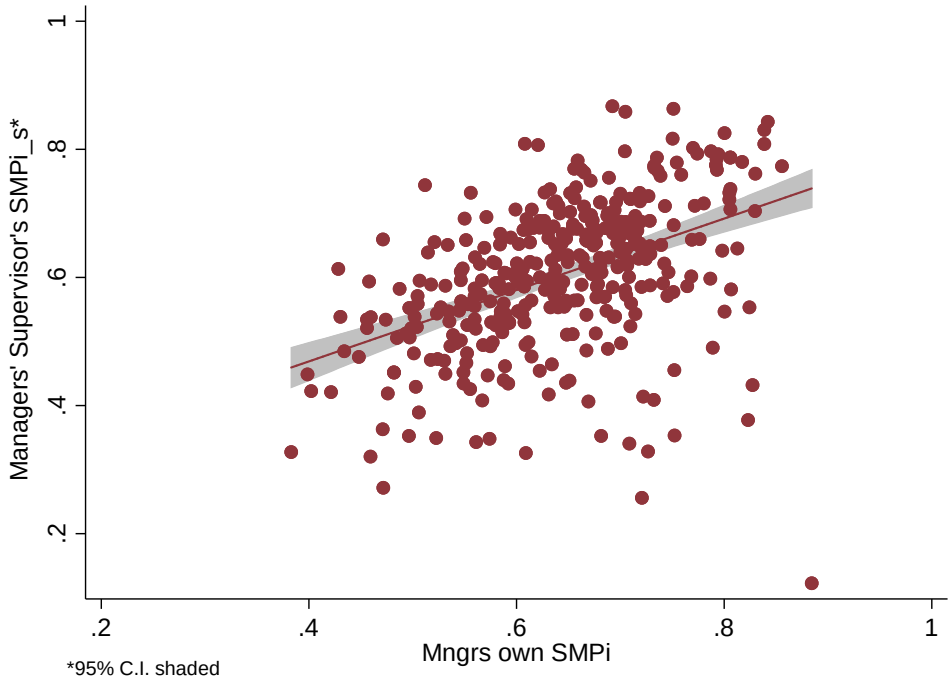


Figure A6. Correlation between Cooperation/Trust index scores and Averaged Structured Management Practice Index (SMPi) scores for respondent manager (average of SMPi of respondent manager and SMPi of respondent manager's supervisor)

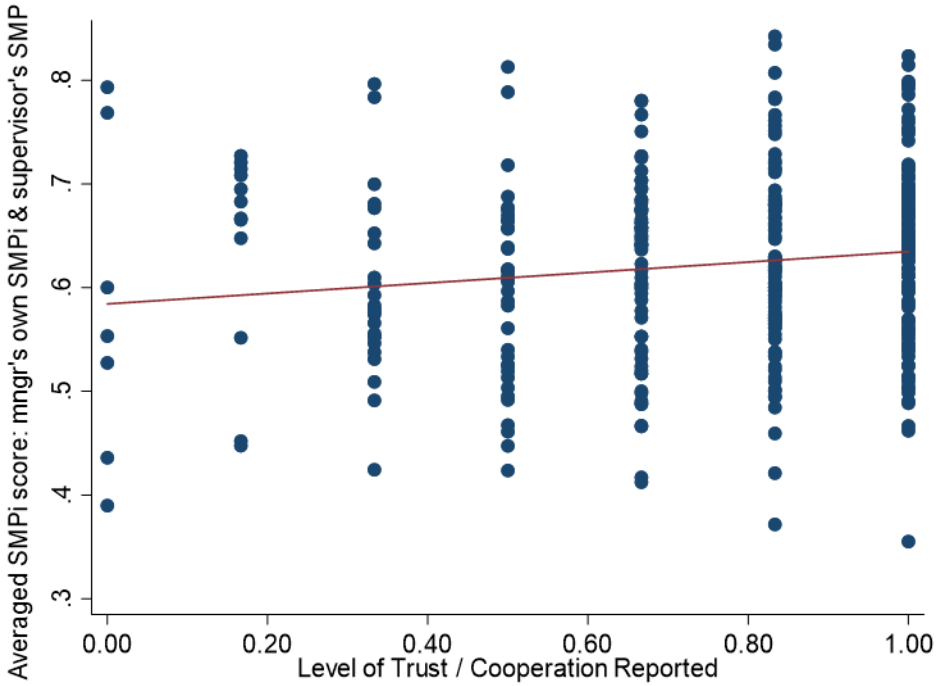


Table A1. How Structured Management Practice Index (SMPis) scores and Cooperation/Trust levels Predict Accuracy on Performance Evaluation Case with both Measures of Structured management practices included in Final Models, reported in odds ratios (z signifies z- scores for ease of interpretation)

	Model 1 b/se	Model 2 b/se	Model 3 b/se	Model 4 b/se	Model 5 b/se	Model 6 b/se
Inflate						
Cooperation/Trust index, z	1.544** (0.22)			1.604* (0.31)	1.698* (0.36)	2.522** (0.71)
Mngrs own SMPi, z		0.772* (0.09)		0.841 (0.13)	0.823 (0.13)	0.679* (0.13)
Mngrs' Manager's SMPi, z			0.763* (0.09)	0.701* (0.11)	0.691* (0.11)	0.586** (0.11)
Mngrs' Mngr SMPiX CoopTrust, z					1.132 (0.18)	1.213 (0.24)
Case: New Emp vs New Mngr						0.684 (0.24)
Case: Bhvr vs Work dlvrbls						0.259*** (0.09)
Case: Solid vs Partially Met						0.010*** (0.01)
_cons	0.219*** (0.03)	0.250*** (0.03)	0.238*** (0.03)	0.208*** (0.03)	0.198*** (0.03)	0.783 (0.23)
r2_p	0.020	0.010	0.012	0.043	0.044	0.343
chi2	10.249	4.682	4.783	14.604	15.182	109.141
N	519.000	444.000	403.000	345.000	345.000	322.000

* p<0.05, ** p<0.01, *** p<0.001

Chapter 3

Striking out Swinging: The Upside of Forced Inferiority

with Ethan J. Poskanzer

INTRODUCTION

Team settings frequently entail occasions when experts or specialists are called on to perform work outside their area of specialization. Teams in organizations are often composed of members with distinct but complementary specialties to facilitate the ability to handle a wide range of work (Faraj & Sproull, 2000, Malone et al., 2010, Reagans, Miron-Spektor & Argote, 2016). However, as workflows and team compositions fluctuate, the team member who would best perform a given task may be not always be available, requiring another teammate to 'step up' and perform that task (Sundstrom et al. 1990, Summers et al. 2012, Valentine & Edmondson 2014, Stuart 2017). These situations are likely becoming more common as organizations increasingly engage in practices designed to enable teams to handle diverse challenges or pivot quickly, such as including individuals on multiple teams at once (O'Leary et al. 2011, Cummings & Haas 2012). In these situations, the designated 'point person' for a given task may be occupied elsewhere, requiring a teammate who is less proficient to perform that task to meet pressing demands.

For example, in the case of consulting teams, Perlow (2012) found that the importance of flexibility and the need to quickly respond to new demands often lead specialized team members to be called to fill others' roles when they are indisposed. Similarly, R&D professionals must often perform work that other team members are more specialized in given various project stages (Hoegl, Weinkauf & Gemuenden 2004) and sales teams switch between lead generation and client communication based on workflow (Malone, Laubacher and Johns 2011). Employees at early stage start-ups are hired for domain expertise, such as in software engineering, but often must work in "all hands-on deck" situations and assist peers elsewhere as work demands fluctuate.

Extant theory would suggest that assigning specialists to tasks in which their peers are superior (which we refer to as "inexpert work") will hinder overall productivity. Inexpert work could break a

worker's flow in an unproductive way (Csikszentmihaly and LeFevre 1989), be distracting or tiring, or introduce switching costs such as the cognitive burden of preparing and reconfiguring one's task approach (Meiran 2000, Leroy 2009). Such work could also be demotivating, particularly if the inexpert task is low status or distinct from the areas of work from which a specialist derives work satisfaction and meaning (DiBenigno and Kellogg 2014). Having specialists perform tasks outside of their specialization, therefore, will likely lower overall productivity. Since both the specialist as well as their team's productivity would likely be negatively impacted relative to having all specialists assigned to their respective areas of expertise, assessing the magnitude and implications for managers are important concerns.

While theory suggests that inexpert work is likely to be detrimental to productivity, identifying this effect in organizations is challenging. Managers tend to allocate tasks non-randomly, such as asking specialists to perform inexpert work when the opportunity costs are minimized, when the need for their primary work is low, or when success in their expert work is unlikely. Managers may also assign inexpert work to the least productive specialists in an effort to shield the most productive team members from potentially productivity-draining assignments.

We use data from Major League Baseball that is uniquely, if unusually, suited to testing the effect of inexpert assignments on productivity. This data allows us to analyze situations in which the game's rules require specialist players to perform an inexpert task at a point in time that is independent of performance in their specialization. The structure of the game entails discrete roles and tasks, allowing us also to clearly observe when one task ends and another begins and by whom.

Counter to the likely suggestion from prior theory, we find that specialists in this context become more productive after engaging with inexpert work. We subject these results to comprehensive robustness analyses and find that players are more effective following inexpert work relative to baseline expectations over a range of econometric tests. Moreover, we find that specialists in this setting do not become inured to this effect over time and that productivity enhancement is greater when stakes are higher. Overall, these findings indicate that assigning specialists to inexpert tasks in team settings can enhance productivity and that the extent of this enhancement depends on the degree to which team outcomes are at stake.

In the following section, we review a set of contending explanations for why specialists may become more productive following engagement with inexpert work. We then describe the setting, data, research design, and results in which we validate the main finding that engaging with inexpert work improves specialists' productivity. The primary purpose of this paper is to convincingly demonstrate an unexpected, yet important, consequence to engaging specialists in inexpert work. Throughout, we underscore how the available evidence weighs in favor or disfavor of specific contending mechanisms. While this research is not designed to experimentally test potential mechanisms with the aim of definitely adjudicating between them, we submit our main results to a battery of robustness analyses to test the relative strength of the potential mechanisms to the extent that the data can offer. In addition to evaluating the following contending explanations as fully as possible, we also conclude by proposing how future research could more definitively test competing mechanisms.

THEORY

SPECIALISTS IN FLEXIBLE TEAMS

In many settings, teams are composed of members with distinct but complementary specializations (Faraj & Sproull, 2000, Malone et al., 2010, Reagans, Miron-Spektor & Argote, 2016). This allows team members to learn from one another, break up projects and distribute work to the ideal member and handle a wide range of challenges. In practice, however, teams may not always be able to rely on the team member best suited for a given task to perform that work. The ideal team member for a given task may be unavailable during particular hours, when they are occupied by work elsewhere or if they have left the organization. In such circumstances, the team is left with an inadequate array of expertise and a hole in the distribution of skills (Stuart, 2017, Perlow, 2012, O'Leary et al., 2011, Cummings & Haas, 2012). When this happens, another team member often steps up and perform someone else's expertise, work that is "inexpert" to them.

Assigning inexpert work to specialists compromises their ability to contribute to their team's output relative to if they were fully engaged in their specialization. Engaging in inexpert tasks may tire or distract specialists, undermining focus on their core contributions to the organization. Inexpert tasks may

also require specialists to reacquaint themselves with their core work while losing productivity to switching costs (Bailey 1989). Engaging specialists in inexpert work is also likely to distance them from the benefits of working in their expertise, such as sources of motivation, positive feedback and rewards and reinforcements that can improve performance (Gagné & Deci, 2005). The likely negative relationship between inexpert work and performance suggests that expert professionals should be shielded from assignments outside their core competencies where negative experiences are likely to exceed positive and satisfactory performance. For this reason, organizations often attempt to define jobs as requiring but a few or even just a single competency. This approach, often called "hyperspecialization" (Malone et al., 2010), is argued to benefit productivity because workers can spend more time on tasks that maximize performance.

However, it is often impracticable to eliminate task heterogeneity within a given job (Cohen 2016). Resource constraints limit organizations' ability to hire specialists for every task and some tasks arise infrequently or irregularly, ruling out the practicality of creating a separate job for every distinct task. Although organizations are sometimes able to apportion divergent task responses into coherent jobs by reinterpreting the purpose of a specific role (Cohen 2013), some tasks within a defined job are still experienced as frustrating or onerous, especially when they are remote to a given area of expertise (Wrzesniewski and Dutton 2001). Such arrangements may be particularly unpleasant for those who view their job as a "calling" (Berg et al. 2010) as they take such specialists away from what gives them meaning in their work. Furthermore, this task heterogeneity may end up incorporating lower status work into a specialist's portfolio, which can undermine professional motivation (DiBenigno and Kellogg 2014).

Together, these arguments suggest that engaging specialists in inexpert work will lower their productivity. However, other theories offer reasons to doubt that this is necessarily the case. We discuss a broad set of six theories which offer plausible explanations for why inexpert work can enhance specialists' productivity. We underscore, in particular, the social psychological source of a relief-based explanation for why specialists perform better after returning from inexpert work. This explanation draws on the *frustration-aggression* (FA) hypothesis (Dollard et al. 1939, Breuer and Elson 2017) to capture the idea that specialists may channel their frustration from performing tasks in which they are relatively inferior by

over-performing when they resume activities in which they are superior. We refer to instances in which specialists are required to engage in tasks that involve this relative inferiority as “forced inferiority”. We propose that the agitation from such forced inferiority creates the conditions necessary for specialists to exhibit superior performance when they transition between the two task states. After describing and testing the key features of this proposed mechanism, we conclude with a discussion of promising future research designs better suited to definitively arbitrate among possible competing mechanisms.

INCORPORATING INXPERT TASKS IN JOB DESIGN

Inexpert tasks may prime specialists for strong performance in subsequent activities. If such stimulation is greater than alternative preparatory activity prior to returning to one’s specialization, enhanced productivity may ensue. This can occur if the inexpert task provides an opportunity for the specialist to learn about their expert work. Inexpert work could facilitate learning through mental or physical stimulation to creating an opportunity to apply newly acquired knowledge. Inexpert work may also lead to better performance overall by affecting specialists’ identities. Below, we sketch five plausible theoretical explanations ranging from learning to identity theories for why specialists may become more productive following engagement with inexpert work.

First, inexpert tasks may present the specialist with an “optimal challenge” (Guadagnoli and Lee 2004). In this state, a specialist improves their repertoire by adjusting and succeeding at increasingly difficult tasks. Tasks that challenge professionals in the near term can stretch them to higher productivity over time, bringing the specialist down a learning curve in their core competency (Tucker et al. 2007, Staats and Gino 2012). If inexpert tasks help the specialist improve their abilities in their area of specialization, this would be revealed by better performance after the transition from inexpert work back to their area of specialization.

Second, inexpert tasks may break the specialist out of unproductive physical or mental monotony. Variety in a worker’s task set has been positively associated with productivity (Hackman and Oldham 1980, Fisher 1993, Ichniowski and Shaw 1999) and can lead to a more refreshed and focused orientation upon

returning to their specialization (Csikszentmihaly and LeFevre 1989, Fisher 1993; Podsakoff, et al. 2007). In these cases, the inexpert tasks prevent the specialist from becoming stultified by concentrating on only one domain. When they return to their area of specialization after an inexpert task, the specialist performs above baseline due to this refreshed perspective.

Third, inexpert tasks may provide specialists with new knowledge that they can apply towards areas of specialization. Experience in one task may improve performance in another by enhancing the understanding of common elements across the tasks (Singley and Anderson 1989). Implicit learning in one task may also produce a broader base of tacit knowledge that can be applied to problems and decisions in other situations (Reber 1989). Through successful imitation or "reverse engineering", engaging in reciprocal tasks can improve abilities in counterpart tasks (Polya 1957; Robertson 2017). In other words, related tasks may enable the specialist to see aspects of their own specialization from an illuminating alternative perspective. To the extent that specialists learn something related to their area of specialization through inexpert work, then such engagements may help them improve in their specialization via learning.

A fourth explanation stems from the power that protecting one's social identity may hold over a specialist's actions. In particular, poor performance in inexpert work may threaten the specialist's professional identity as a merited member of their organization and motivate them to seek immediate recourse (Ibarra and Petriglieri 2016). After failing at an inexpert task, the specialists may seek to protect their professional identity with excellence in their area of expertise (Arkin 1981). In this fourth model, the specialist seeks to erase associations with their failure in the inexpert domain and eliminate doubts regarding their value to the team by proving their worth through strong performance in their expertise.

Similarly, a fifth explanation is that failing in a task that other team members commonly perform can also lead to identity changes that deepen the specialists' bonds with teammates. The failure can debase the specialist's general standing with their team, leading to status degradation. The deep emotions associated with such status degradation bring the specialist to a more vulnerable state in relation to their membership with the team (Glaser and Strauss 1971), opening the potential for the specialist to experience greater solidarity with others and enhance identification as a team member (Kanter 1968, Hernandez 2015).

When returning to their area of specialization, this enhanced solidarity and identification may translate into greater momentum when executing their specialized tasks.

FORCED INFERIORITY: PRODUCTIVE RELIEF

"Striking out may or may not reduce the frustration, but it seems to be an inherently satisfying response to the tension built up through frustration" Ted Robert Gurr, *Why Men Rebel* (1970: 23)

Inexpert task engagement may also affect specialists' productivity through pathways that are distinct from learning or professional identity. We draw on the *frustration-aggression* (FA) (Dollard et al. 1939, Breuer and Elson 2017, Maier 1949) hypothesis developed in social psychology to capture the idea that specialists will channel their frustration from forced involvement in tasks in which they are relatively inferior by over-performing when they resume activities in which they are superior. If the frustration generated from performing inexpert tasks is sufficiently strong, the release created by returning to one's expertise may generate a productivity boost over a specialist's baseline (ibid). Therefore, an advantageous sequence of tasks may channel predictable frustration from inexpert work into greater productivity in specialist activities. We elaborate on one such potential pathway below before testing the relationship between inexpert work, productivity, and the viability of competing explanatory mechanisms.

Frustration-Aggression Hypothesis

Since the original formulation of the frustration-aggression (FA) hypothesis by Dollard and colleagues, frustration has been defined as an event rather than an affective state (1939:1, 7; see also Breuer and Elson 2017:2). Broadly, frustration has generally been conceived as blockage from goal-attainment (Dollard et al 1939; Breuer and Elson 2017). Importantly though, frustration-induced behavior is not an adaptive, learned-response toward discovering means to attain possible goals through trial-and-error approaches (Hamblin 1963: 193; Grimm 2008). What is required is evidence that the response behavior is arguably non-adaptive, in the sense it is not necessarily a learned response.

The available opportunities for response shape the nature and character of how frustration-instigated aggression will transpire (Meir 1949:160). Aggression may primarily be targeted directly at the

cause of frustration (Zillman and Cantor 1976) but may just as likely be directed toward available individuals not directly responsible for the frustration (Geen 1968; Grimm 2008). Frustration-instigated responses are often nonconstructive or destructive (Maier 1949). The extent to which one might expect particularly outsized responses depends on the extent to which the reaction to the frustration serves as adequate relief (Morlan 1949). This suggests that the sequences of events or tasks can be oriented so as to possibly make aggressive responses to frustration most likely result in positive, outsized results (c.f. Berkowitz 1989, see also Burnstein and Worchel 1962, Kregarman and Worchel 1961, Rothaus and Worchel 1960). Following these insights, we propose a set of conditions below that we refer to as *forced inferiority* where specialists may be more likely to over produce results in activities following a compulsory frustrating activity in team settings.

Forced Inferiority

We carry the implications from the general frustration-aggression (FA) hypothesis to a specific application for work design and task alignment. We label this specific form of frustration 'forced inferiority' to signify the act of placing specialists in work situations in which they must perform work that is remote from their expertise, they are unlikely to succeed, and are likely to perform worse than other team members. Circumstances where specialists perform a task outside their expertise and in which they are predictably worse relative to others and to their performance in their own core competency place the specialist in two types of relative inferiority. First, they are inferior relative their teammates' abilities in the focal task. Second, they are inferior relative to their ability in their own specialization. These dual inferiority prospects create the conditions for a social psychological theoretical explanations why inexpert tasks may improve specialized work.

Following this premise, we expect circumstances to condition the effects of forced inferiority. In particular, we expect any heightened attention or importance attached to the performance of the professional's inferior assignment to exacerbate the frustration felt and the resulting reactive aggression. As heightened stakes will exacerbate the deprivation felt by the specialist's inability to contribute when engaged in inferior tasks, we expect the response to forced inferiority to be greater when heighted attention

or importance is attached to the specialist's performance. At the same time, the tasks must be clearly distinct to both acting parties and audience so as not to confuse the specialist's overall value to the organization with their performance on the inferior task. Rather, the juxtaposition of forced inferiority and the specialist's expertise serves to emphasize the superiority-inferiority contrast with regards to the specialist's ability and to construct a productive channel for reactions to forced inferiority (Gurr 1970: 67). It is this cadence which directs the goalless reactions under the FA hypothesis into productive outcomes.

SETTING

"Who would people rather see, a real hitter hitting home runs or a pitcher swinging a wet newspaper?" - All-Star pitcher Max Scherzer (ESPN 2018)

We test the effect of inexpert work on overall productivity in Major League Baseball (MLB). Professional athletics is an advantageous context for management research as performance is clearly observable, participants have clearly defined incentives, and rules structure the variation in individuals' choices or actions (Day, Gordon and Fink, 2012). Accordingly, sports have been used to illuminate a range of management inquiries such as the importance of role structures on a team (Stuart and Moore, 2017), the conditional expression of racial bias (Zhang, 2017, 2019), the effect of performance expectations (Dai, Dietvorst, Tuckfield, Milkman and Schweitzer, 2018) and the demand for authenticity (Hahl 2016).

Within MLB, we study inexpert work in the form of pitchers taking turns as batters. Baseball games consist primarily, although of course not exclusively, of a series of interactions between two players: a pitcher and a batter. The pitcher tries to throw a ball past the batter who attempts to strike that ball with a wooden stick. Pitchers are highly specialized in throwing the baseball and accordingly, rarely practice batting and are not typically required to bat in developmental settings such as college (MLB.com 2018). As Washington Nationals coach Kevin Long said, "There's a reason they're [pitchers] in the big leagues, and it's not because of their hitting" (Diamond 2018).

Batting is seen as clearly distinct from the set of responsibilities that pitchers are expected to be competent in, and pitchers bat only as a result of the rules' imperative. Pitchers rarely practice batting and do not take at-bats in developmental settings such as college or the minor leagues, instead choosing to focus

on pitching as batting skill is unlikely to further their careers. Reflecting, professional pitcher Max Scherzer distinguishes between himself and “real hitters” in the headline quote of this section (ESPN 2018). As batting is so far outside pitchers’ set of competencies, many have argued that the rules should be changed to prevent pitchers from batting, making the arguments that watching pitchers fail at this task is uninteresting for spectators and that some pitchers are so underprepared to bat that they may injure themselves (Calcaterra, 2015).

The MLB represents a conservative setting to test the effect of inexpert work on productivity. Professional baseball players are the survivors of an extremely tight selection process and are at the very peak of their profession. As a result, those in our sample have played countless baseball games and switched between batting and pitching many times and as such, should be as inured as possible to any effects from this transition. Pitchers usually play once a week, or roughly every five games during a season and will bat an average of 2.25 times per game in which they play. During any single game, a professional starting pitcher is likely to pitch to between 15 to 20 batters of the opposing team. Additionally, baseball is highly influenced by quantitative analysis; players and teams are likely to have identified most conceivable advantages of various game tactics and personnel assignments, including transitions during a game. As such, pitchers’ baseline performance - their average expected performance in any given game situation - is likely near its peak. Therefore, finding potential task transitions which may move these professionals to a higher state off such a baseline should be more challenging than in less extreme settings, where individuals have yet to reach such expert levels of performance. In other words, this extreme level of athletic professionalism helps us rule out picking up spurious results attending other means of improving in their specialization. This, of course, also helps us rule out contending mechanisms which cannot explain productivity improvements that do not bring the pitchers to higher productivity baselines but only syncopated boosts following task transitions.

DATA & METHODS

Our analyses use data from Retrosheet, an online baseball statistics archive. The data include all in-game events from MLB games played between 1997 and 2018. MLB is comprised of two leagues: the National League (NL) and the American League (AL). The rules only require pitchers to bat in NL games, and as such, our analysis is limited to these games. Our sample, however, includes all interleague games between teams each in the NL and AL which were played according to NL rules. This restriction also allows us to rule out other differences between the two leagues, such as the influence of having pitchers in the batting order at all, from influencing results. The NL-only dataset consists of 2,125,351 at-bats occurring over 27,939 games.

Analytically, we exploit the timing of a pitcher's turn to take an at-bat to study the effect of engaging in inexpert work on subsequent performance in specialist work. Our primary independent variable is a binary indicator for whether or not the pitcher batted in the half-inning (the last point at which the pitcher's team batted) prior to resuming their pitching duties. By selecting the closest point in the game at which the pitcher could have batted, we can be as sure as possible that results are not affected by unobserved variation occurring between a pitcher's at-bat and returning to pitching.

The rules of baseball require pitchers to take turns ("at-bats") as batters at certain points determined before the game by the managers' posted batting line-ups. As the timing of pitcher' at-bats is determined before the game, pitchers are required to perform inexpert work at points that are exogenous to their overall performance. This is ideal for making inferences regarding productivity. In most organizational settings, managers exercise discretion over whom to assign to inexpert activity (see Bidwell 2009). However, because of the pre-determined order in which players bat, managers do not have discretion over the timing of a pitcher's turn at bat and the pitcher's turn at bat is substantively unrelated to their performance in their expert work. While managers do exercise discretion over when a pitcher is removed from the game and ceases taking part in any in-game events, we are able to demonstrate robustness to this action in the "Selection into Pitching" section. As the timing of transitions between inexpert and expert work are independent from prior productivity in this setting (with the aforementioned exception of the decision to

remove players from the game), we are able to closely identify the effect of such transitions have on overall productivity.

In subsequent models, we split our sample based on whether the pitcher was successful or unsuccessful as a batter in the prior inning. We expect the productivity effects of engaging in inexpert work on subsequent performance to be greater when the pitcher is unsuccessful as a batter, as this heightens the salience of the relative inferiority that batting creates. Splitting the sample this way also allows us to test whether such engagement with inexpert work could "backfire" or be detrimental to productivity if the professional happens to unexpectedly succeed at the inexpert task.

The primary dependent variable, labeled "Pitcher Performance", is a binary indicator of whether the pitcher was successful in getting an out in a given at-bat[33]. In a given at-bat, the pitcher's objective is to get an out, while the batter's objective is to reach a base safely. As such, outs represent successes for the pitcher, and are coded with a value of 1. Each at-bat is an important input into winning or losing the game; whether or not the first batter of an inning reaches base or makes an out changes the probability of a run[34] scoring in that inning by 26% (Fangraphs.com 2010).

Our most important control variables are fixed effects for each individual pitcher. This allows us to estimate changes in pitcher's performance after inexpert work from their baseline standard of productivity. Individual pitcher fixed effects control for variation between pitchers with regards to quality, skill, or other selective criteria, as well as unobserved personal characteristics such as emotional orientation, resilience, grit (Duckworth et al. 2007), or confidence (Deiner and Dweck 1978). To the extent that these traits evolve slowly over time, pitcher and year fixed effects should absorb their influence on pitchers' responses to engaging in inexpert tasks.

In all analyses, we account for the number of at-bats that have elapsed since the pitcher resumed pitching to measure distance from batting. For observations in which the pitcher did not just bat, this

[33] Batters can also reach base via "errors", in which the batter hits the ball in a way that is deemed by the official scorekeeper as likely to be an out, but the batter reaches base due to a mistake by one of the pitcher's teammates. These events are not counted against pitchers in official statistics and are excluded from our analyses.
[34] The team that scores more runs wins the game.

variable measures the distance from the point in the game at which the pitcher would have batted. The batters elapsed measure begins with a zero for the first batter that the pitcher pitches to after resuming pitching and increases by one for each batter the pitcher faces thereafter in a given inning. All models include an interaction between the forced inferiority indicator variable and the number of batters elapsed in the inning, as we expect the pitcher to return to their pre-established baseline as they spend more time performing in their area of specialization and the salience of the inexpert task diminishes. The sign of this interaction demonstrates this fading pattern over the series of events occurring after the pitcher resumes pitching.

We also perform a series of supplementary analyses to test potential mechanisms behind the positive relationship between inexpert work and pitchers' performance. These analyses use a group of mediator variables to outline the conditions under which inexpert work is most impactful for performance. In the first analysis, we test the "optimal challenge" mechanism by testing whether a pitchers' performance in the most recent inning mediates the relationship between inexpert work and performance. If inexpert work is beneficial by presenting a challenge over work that is too easy, the effect should be largest when the pitcher has been performing well and may be in the "flow" of their work, as this is when a new challenge would be most impactful. We use the number of batters that a pitcher got out in the prior inning divided by the total number of batters faced to measure recent performance[35]. If "optimal challenges" drive the relationship between inexpert work and performance, the effect on inexpert work will be largest when the pitcher has been performing well.

To test whether inexpert work affects performance by breaking pitchers out of unproductive monotony, we test whether the effect of inexpert work varies by length of time that a pitcher has been performing a single task by measuring how long a pitcher has been pitching without taking a turn as a better. This measure starts with a value of zero at the beginning of the game and increases by one for every batter than the pitcher pitches to. When the pitcher takes an at-bat themselves, the value returns to zero. If inexpert

[35] This can be thought as the number of successes divided by the number of opportunities in the pitcher's last turn at pitching.

work affects pitchers by breaking monotony, the effect of inexpert work should be largest when the pitcher has been pitching for a long time without performing a different task, as this is when monotony is likely to be greatest.

We then estimate whether the effect is larger for pitchers on American League (AL) teams to test whether inexpert work affects productivity by allowing pitchers to develop new knowledge. As pitchers are only required to bat in National League (NL) ballparks, American League pitchers bat with much less frequency than NL peers, as they only bat in interleague games played in NL ballparks. If the effect of inexpert work was due to pitchers developing new knowledge, inexpert work should be more impactful for AL pitchers, as those who bat less frequently should have more new information to learn.

In a secondary test of this mechanism, we estimate whether pitchers become more likely to use strategies that were used against them during their turn batting. Baseball pitchers throw two types of pitches: fastballs, which are thrown as fast as possible past the batter, and breaking balls, which utilize spin and drag to take nonlinear paths through the air and trick the batter. We measure whether pitchers become more likely to use pitches that were used against them to test whether the pitcher gained new knowledge from their turn as a batter and tried to incorporate it into their own pitching strategy. In particular, we identify the type of the last pitch used against the pitcher during their turn at-bat (fastball or breaking ball) and estimate whether they become more likely to use this type of pitch than would otherwise be expected[36].

To test whether inexpert work affects performance by threatening the pitcher's professional identity, we test whether the effect of inexpert work varies by the stakes at that point in the game. Stakes are measures as the difference in score between the two teams. The stakes are highest when the score is close, as events are more determinative of the team's probability of winning the game. If failing at an inexpert task affects performance by threatening the pitcher's sense of professional identity as an athlete, the effect of inexpert work should be unaffected by the stakes for the pitcher's team at the time of a given interaction. The pitcher's identity as a strong performer is at stake whenever they compete and regardless

[36] This analysis is preliminary but is something we intend to do with a new supplementary dataset.

of the team's situation. For some, professional identity may be all that is left to fight for in situations in which the team's outcome has largely been decided.

We test whether inexpert work affects performance through status degradation by testing whether inexpert work has a greater effect on performance for pitchers whose status is likely to be vulnerable. We use status as a "rookie" pitcher to measure status vulnerability. Rookies are players that are in their first season as a professional baseball player, and accordingly, have the lowest status on the team. Reflecting their low status, rookie players have often been subjected to hazing rituals (New York Times 2016). If inexpert work affects performance through status degradation, instances of inexpert work will be most impactful for players whose are likely to feel that their status is vulnerable. Rookies also have the shortest track records of performance by the team, and as such, may be more likely to respond to small failures than those with longer track records of strong performance.

To test whether pitchers' performance improves as a result of relief from released frustration, we test whether pitchers adopt more aggressive strategies after batting. In this mechanism, failure as a batter frustrates the pitcher. That frustration is then released in a productive, aggressive way when they return to an area of relative dominance. We use the probability that a pitcher throws a given pitch in the "strike zone" to indicate an aggressive strategy. The "strike zone" is the area deemed by the umpire[37] as a reasonable location for the batter to make contact with the pitch[38]. If the pitcher throws too many pitches outside this area, the batter is automatically awarded a base, in what is referred to as a "walk"[39].

Throwing pitches in the strike zone is a generally more aggressive strategy – the pitcher is attempting to beat the batter with high-quality pitches that are in an area in which they could possibly hit the ball. In support of this, throwing in the strike zone is often referred to as "challenging" the batter and pitchers often doing so against particularly fearsome batters. To test whether pitchers adopt a more

[37] Umpires are on-field officials or referees and enforce rules during the course play.
[38] The strike zone is an area that is 17 inches horizontally and extends roughly from the batter's knees to the middle of their chest.
[39] This rule prevents pitchers from throwing totally unhittable pitches, such as those that bounce or are way over the batter's head.

aggressive approach after batting, we estimate whether they become more likely to throw pitches in the strike zone. We then test whether this more aggressive strategy translates to a performance increase by estimating whether walks become less likely after the pitcher takes an at-bat.

SUMMARY STATISTICS

Table 1 shows key summary statistics. Pitchers engage in inexpert activity through batting with moderate frequency. Of all at-bats[40] in which the pitcher is engaged in their specialization, pitching, 23.9% occur in innings after the pitcher himself took an at-bat. Reflecting pitchers' low probability of success at batting, 84.4 % of these occur after the pitcher had been unsuccessful as a batter.

In support of the argument that batting is outside the expertise and specialization of professional pitchers, Figure 1 shows density plots of career batting performance for every player in our sample split by position. Batting performance is measured as the proportion of at-bats in which that player reaches base, or the inverse of the variable used for pitching performance in other models. Pitchers' mean career performance is .148, while the next lower position, catchers, is .293. The gap between pitchers and catchers is four times as large as that between catchers and the best performing position, designated hitters. The clear distance between pitchers and all other players underlines the distinction between those who specialize in pitching or in batting on a baseball team.

RESULTS

The descriptive relationship between the pitcher having batted in the prior inning and performance is shown in Figure 2. The Y axis indicates the probability that a given at-bat ends in an out (success for the pitcher), while the X axis shows the number of batters elapsed since the pitcher returned to their specialization of pitching. The solid line represents at-bats in which the pitcher had just engaged with inexpert work as a batter, while the dotted line represents all other at-bats. Pitchers are more productive in innings after their turn to bat came up. As the pitcher spends more time performing their specialization, the

[40] An at-bat is equivalent to one event in the game.

positive relationship between having batted and performance fades and pitchers return to the performance baseline.

Table 2 shows results from multivariate linear probability models that test this relationship[41]. As the outcome is binary, we also estimate logistic regressions using the same equation (results in appendix). Model 1 estimates that the pitcher having batted in the prior inning leads to a .35% increase in the probability of the first at-bat of the inning ending in an out, which represents a 1.30% increase in performance, providing evidence that having just engaged in inexpert work has a positive effect on specialist performance in this setting.

Model 2 splits the independent variable by whether the pitcher was successful or unsuccessful as a batter in the prior inning. We find that the positive effect of having batted on performance only occurs if the pitcher is unsuccessful as a batter, and in doing so, underperforms relative to others. The pitcher having been unsuccessful as a batter is associated with a .46% increase in the probability of the pitcher getting an out immediately after returning to their specialist work. The marginal effects of this regression are shown in Figure 3. Performance is unaffected if the pitcher is successful as a batter, and as such, we find no evidence that engagements with inexpert work risks a "backfiring" effect, as only positive or neutral responses to batting are observed. In other words, we observe no evidence of risk for productivity when engaging in inexpert tasks unexpectedly leads to success rather than underperformance.

ADDRESSING SELECTION INTO PITCHING

While the timing of a pitcher's at-bat is exogenous to pitching performance, managerial discretion may however play a role in whether a pitcher is left in the game to bat at all. In MLB, when a player is removed from a game they are disqualified from taking part in all further game events, whether pitching or batting. As such, managers are more likely to leave pitchers in the game on days when they are performing well at pitching, and in doing so, pitchers are more likely to take at-bats on such occasions. This could potentially bias the results if the positive relationship between taking an at-bat and pitching performance is

[41] Results from logit models with the same equation are in Appendix X. Results are qualitatively similar.

the result of within-pitcher, between-day variation in performance and ensuing changes in likelihood of batting.

Models 3-5 shows results from tests that are designed to show robustness to this selection process. Model 3 controls for pitcher-game level fixed effects (necessarily relaxing the use of individual fixed effects and controls that do not vary by game, such as whether the pitcher is at their home ballpark). In this model, if a pitcher is performing well on a given day and is left in the game for a longer period, this will establish a higher productivity baseline than the pitcher's baseline across days. To observe a positive effect, performance after engaging in inexpert work must exceed this relatively higher baseline. Within the pitcher's performance in a single day, we estimate that the first batter of an inning after the pitcher batted is 2.51% more likely to be out than would otherwise be expected.

Lastly, Model 7 shows models that are conditional on selection into batting and estimates the effect of the outcome of the pitcher's at-bat on subsequent pitching performance. The pitcher's performance as a batter is likely to be independent of the decision to allow the pitcher to bat. As pitchers are generally very poor batters, the decision whether to leave them in is likely to be totally driven by their likelihood of future success as a pitcher. Reflecting this, the median pitcher in our sample by career batting performance would rank below the 1st percentile of all other players and only 6.2% of pitchers exceeded the 5th percentile. In the vast majority of situations, the team manager likely had better options to substitute for a pitcher if they were concerned about batting success. As such, the outcome of the pitcher's at-bat is likely to be unrelated to any process that could bias the relationship between having batted and pitching performance. We find that being unsuccessful as a batter is associated with a positive effect on pitching performance conditional on the manager's decision to allow the pitcher to bat. This aligns with model 2 and suggests that underperformance in the inexpert task more strongly drives the positive effect on performance.

Model 5 uses coarsened exact matching (CEM) to match innings in which pitchers were left in the game to bat with a set of "control" innings that are comparable on factors likely to affect a manager's decision to keep a pitcher in the game (Iacus, King & Porro, 2011). This decision is likely to be driven by fatigue, how successful the pitcher has been that day and their likelihood of continued success. The

matching criteria are the number of pitches the pitcher has thrown that day to measure fatigue, how well the pitcher has performed that day, as measured by the mean number of at-bats in that game in which the pitcher successfully got an out, and whether the next upcoming batter is of the same handedness as the pitcher[42]. All matching is done within individual pitchers, as the inclusion of these factors make within-game matches unlikely. Within the matched innings, we find the first batter of an inning is .489% more likely to be out if the pitcher just batted than would otherwise be expected. This estimate similar to that of Model 1.

TESTING POTENTIAL MECHANISMS

Models 6-12 in Table 3 show tests of mechanisms that could potentially support the positive relationship between inexpert work and performance. Model 6 tests the "optimal challenge" explanation by testing whether inexpert work is more impactful when the pitcher is already performing well. We estimate that the pitcher's recent performance does not mediate the relationship between inexpert work and performance, and accordingly, find no support for the optimal challenge explanation.

Model 7 tests whether inexpert work breaks pitchers out of an unproductive monotony by testing whether the effect of inexpert work is greater when pitchers have been performing one task (pitching) for a long time. We find that inexpert work is actually less impactful for performance in situations when the pitcher is likely to have settled into monotony, and accordingly, do not find evidence that the effect is driven by inexpert work breaking pitchers out of monotony.

Model 8 estimates whether the effect of batting on pitching performance is more important for AL pitchers, who experience batting with much less frequency than NL pitchers. If inexpert work helps pitchers perform better by providing them with new knowledge that they can use while pitching, the effect is likely to be greater for pitchers who do not bat frequently, as there will be more new knowledge available to them

[42] Pitchers are argued to have an advantage when throwing to batters of the same handedness due to difficulty seeing the ball. For example, right handed pitchers are argued to have an advantage over right handed batters.

when they bat. We find there is no difference in the effect of having batted between AL and NL pitchers, and as such, find no evidence that inexpert work helps pitchers develop new knowledge.

Model 9 tests whether the stakes of the situation mediate the relationship between inexpert work and performance. Stakes are measured by the difference in the score of the two teams playing at the time of a given in-game event. If inexpert work improved productivity solely through professional identity, the team's stakes at a given moment will not mediate the relationship between inexpert work and productivity, as pitchers' professional identities should be on the line at all times when they are performing their expert professional work. We estimate that engaging in inexpert work affects performance until the two teams are four runs apart [.00635 > (3*.00180)]. This is a substantial difference, as 78.03% of at-bats in our sample take place with a score difference of three runs or less. Figure 4 shows the marginal effects of this regression. Overall, we do not find evidence that the relationship between inexpert work and performance is driven by professional identity.

Model 10 tests whether inexpert work affects performance through status degradation. We estimate whether the effect of inexpert work is largest for rookie pitchers, whose status is likely to be most vulnerable. We find that inexpert work is actually less impactful for rookie pitchers' performance, and accordingly, do not find evidence that status degradation is the mechanism at play.

Models 11-13 test whether inexpert work leads pitchers to adopt a more aggressive strategy that leads to stronger performance, which would be consistent with a response to frustration. In Model 11, we estimate that pitchers are .616% more likely to throw a given pitch in the strike zone to the first batter of an inning after the pitcher batted than would otherwise be expected. Throwing pitches in the strike zone is an aggressive strategy, as these pitches are in an area in which the batter could possibly hit the ball, and are seen as "challenges" to the batter.

Figure 5 breaks down the change in pitches' locations after the pitcher bats. Red areas indicate that a pitch is more likely to be in that space after the pitcher bats, while blue areas indicate that a pitch is less likely to be in that space after the pitcher bats. The thick black rectangle outlines the border of the strike zone. The number in each space is the point estimate of the relationship between the pitcher having batted

on the pitch being in that space from a regression using the same specification as Model 1. After the pitcher bats, pitches are generally more likely to be in the strike zone. Additionally, pitches are more likely to be in a high location near the physical top of the strike zone. This is consistent with a more aggressive strategy, as high pitches are generally more difficult for the batter to make contact with, but are more likely to be hit in the air, and thus travel far than physically lower pitches.

Models 12 and 13 test whether the more aggressive strategies adopted by pitchers translate into better performance. Batters can reach base through two pathways: walks and hits. If the pitcher throws too many (four) pitches outside the strike zone, the batter is permitted to "walk" to first base, and if the batter hits the ball with their bat is able to run to first base, this is called a "hit". In Model 12, we find that batters are .536% less likely to walk after the pitcher took an at-bat, which is likely related to pitchers throwing more pitches inside the strike zone. In Model 13, we estimate no difference in the batter's probability of getting a hit after the pitcher batted. Overall, these results are consistent with a pattern in which pitchers adopt the strategy of aggressively throwing pitches in the area where batters could make contact with the ball, but with high-quality pitches that the batter is unable to hit well.

Models 13 and 14 in Table 4 show tests of whether the increase in pitchers' performance after batting contributes to better outcomes for the team. To do so, we test whether the batting team is more or less likely to score "runs" in situations after which the pitcher batted. When a player on the batting team is able to touch all four bases[43], their team is awarded a run, and the team who score more runs wins the game. Pitchers try to get three outs before the batting team is able to score a run, and accordingly, better pitcher performance as measured by our primary dependent variable (whether a given individual batter is out) is likely to contribute strongly to run prevention for the team. In Model 13, we estimate that the batting team is .698% less likely to score at least one run in situations in which the pitcher just batted, and on average,

[43] This measure is similar to an aggregation of our primary dependent variable, whether the batter is out. Each batter is either out or reaches a base. Players on the batting team need to touch all four bases before three outs are recorded to score a run. It generally requires multiple players to reach base for one player to touch all four bases and score a run.

.018 less runs score in these situations. Together, these models indicate that pitchers' better performance after inexpert work has a positive effect on the overall team's outcomes.

SUMMARY AND IMPLICATIONS

We test the effect of engaging in inexpert task on overall performance by exploiting the quasi-random timing of when a pitcher must pitch immediately following their own at-bat in Major League Baseball. We find that having batted prior to pitching has a positive effect on performance than would otherwise be expected controlling for pitcher-game level fixed effects. Importantly, the effect disappears - but does not reverse - when pitchers *succeed* as batters. This suggests that when specialists are forced to engage in inexpert activity, they are likely to over-perform when returning to tasks in which they excel. This finding implies that organizations should provide productive avenues for professionals to turn to following situations when they must perform inexpert work. Further, we find evidence that professionals do not become inured to this stimulus over time.

Taken together, these findings imply that even highly specialized professionals should not necessarily be shielded from tasks outside their specialization. This may be particularly useful when tasks arise that are necessary but unlinked to a job's core responsibilities. Since the expectation for performance is low, these activities should not come at a cost to the organization (Williamson 1975). However, several benefits may accrue. First, our evidence suggests that expanding into new tasks can lead to improved performance in concurrent 'bread and butter' tasks. Second, in terms of training and development, engagement on core tasks may become extra generative whenever a professional is also concurrently coming down a learning curve in parallel work, and time spent outside of core specialization, in general, may prove useful (Tucker et al. 2007).

Third, because we find evidence that repeated exposure to inexpert task engagement does not eliminate the effect, managers can strategically syncopate tasks so that core tasks follow potentially frustrating assignments. Fourth, our findings suggest that some bureaucratic tedium may be a feature rather

than a bug of organizational reality. If working in a large bureaucratic organization entails some mix of unpleasant tasks, it may not always pay to outsource these activities, even when they can be executed to greater effect by others and perhaps at a marginally lower cost.

This leads to our final strategic implication. At the core of our analysis is the idea that even (or especially) top performers are highly sensitive to contrasts between tasks in which they are inferior to others and their own expertise and specialization, and that these professionals may overcompensate in reaction. For example, it may be worth requiring consulting directors to put their own slide decks together, at least for internal presentations, ahead of important external meetings. The results may not be pretty for the internal audiences (like watching a pitcher swing 480 degrees around home plate) but might just clinch a home run business deal. Perhaps there is something to be gained by reminding superstars that they are not always the all-star in certain managed situations, and that top performers can improve in their own domain by coming up short outside their own specialty.

LIMITATIONS AND FUTURE RESEARCH

The contribution of this study is first and foremost providing convincing evidence that engaging specialists in inexpert work has surprising and unexpected positive consequences for productivity among elite professionals. Specifically, we find that engaging specialists in tasks that are remote from where they make their primary contribution to the organization has a positive effect on performance in the area of specialization.

The forced inferiority mechanism we propose is limited to tasks and activities which follow regular cadence. This enables intentionality in the design of work where tasks may be scheduled or routinized so as to channel the effects of FA productively. Therefore, we make no claims as to whether the mechanism would pertain to more irregular work design considerations such as those at the project-level. However, rather than limiting the power of forced inferiority as a productivity inducement, we believe this is a feature to the extent it allows for routinizing and constructing predictable productivity channels for professionals

to respond to frustrating tasks. To the extent that professional work entails such unavoidable tasks within in a given role, this mechanism offers a means to preserve the role so defined while counteracting the negative consequences of frustrating elements with outsized productivity in specialized tasks.

As this is the first strong evidence of such an unexpected result[44], we believe our findings merit attention from a broad management audience. We justify this belief by subjecting our main findings to stringent robustness checks. We pay special attention to ruling out any potential selection issues that could bias our results. The persistently strong relationship between engaging elite professionals in inexpert work and enhanced productivity when they return to their area of specialization offers convincing evidence of this novel phenomenon.

Given the empirical design necessary to substantiate such claims, we are unfortunately unable to use this framework to isolate the specific underlying mechanism that is the exact cause of this relationship. As discussed above, however, we believe that such a finding may have been obscured from focus on purely motivational mechanisms regarding productivity-enhancement channels. For this reason, we suggest a more relief-based mechanism we call forced inferiority, summarized above.

Still, we believe the implications of our productivity findings merit further investigation into the specific underlying mechanisms producing this effect. Specifically, future research testing the causal linkages between engaging specialists in inexpert work would do well to apply lab-based, controlled settings. Such an approach would enable isolating competing elements that could more definitively adjudicate between motivation-based and relief-based causal linkages. In future research, we plan to test such competing mechanisms in team-based settings of specialized professionals working in a competitive

[44] In a non-professional setting, Ranganathan (2018) finds similar results where tea pickers who must perform peripheral tasks like clearing debris before picking feel more connected to their output and are in turn more productive. The mechanism Ranganathan invokes in her study though rests on increased identification with the final product through such scut work, which is distinct from the forced inferiority of performing a task one is much inferior in relative to other professionals and relative to one's own area of specialization.

corporate environment. We are excited about the potential of such research for underscoring the generalizability of the insights developed here.

References

Arkin, R. M. (1981). Self-presentation styles. In J. T. Tedeschi (Ed.), *Impression management theory and social psychological research* (pp. 311-333). New York, NY: Academic Press.

Averages, N. 2019. National League Batting Year-by-Year Averages | *Baseball-Reference.com*. [online] Baseball-Reference.com. Available at: https://www.baseball-reference.com/leagues/NL/bat.shtml [Accessed 9 Jan. 2019].

Bailey, C. "Forgetting and the learning curve: A laboratory study." *Management Science* 35.3 (1989): 340-352.

Barrick, M,, Mount, M. and Li, N. "The theory of purposeful work behavior: The role of personality, higher-order goals, and job characteristics." *Academy of Management Review* 38.1 (2013): 132-153.

Ballparks of Baseball - Your Guide to Major League Baseball Stadiums. (2018). *Ballparks of Baseball - Your Guide to Major League Baseball Stadiums.* [online] Available at: https://www.ballparksofbaseball.com/ [Accessed 30 Dec. 2018].

Bateson,G. 1941. "The Frustration-Aggression Hypothesis and Culture," *Psychological Review*, LXVIII (July), 354-355.

Berg, J., Grant A. and Johnson, V. "When callings are calling: Crafting work and leisure in pursuit of unanswered occupational callings." *Organization Science* 21:5 (2010): 973-994.

Berkowitz, L., 1989. "Frustration-aggression hypothesis: Examination and reformulation." *Psychological Bulletin*, 106(1), p.59.

Berman, S., Down, J., and Hill, C. 2002. Tacit knowledge as a source of competitive advantage in the National Basketball Association. *Academy of Management Journal*, 45: 13-31.

Bertolotti F, Mattarelli E, Vignoli M, Macrì DM (2015) Exploring the relationship between multiple team membership and team performance: The role of social networks and collaborative technology. Res. Policy 44(4):911–924

Bidwell, M., 2009. Do peripheral workers do peripheral work? Comparing the use of highly skilled contractors and regular employees. ILR Review, 62(2), pp.200-225

Breuer, J. and Elson, M. 2017. "Frustration–Aggression Theory." *The Wiley Handbook of Violence and Aggression*, pp.1-12.

Burnstein, E. and Worchel, P. 1962. "Arbitrariness of frustration and its consequences for aggression in a social situation." *Journal of Personality*, 30(4), pp.528-540.

Buss, A. H. (1963). "Physical aggression in relation to different frustrations." *Journal of Abnormal and Social Psychology*, 67(1), 1–7. doi:10.1037/h0040505.

Buss, A. H. (1966). "Instrumentality of aggression, feedback, and frustration as determinants of physical aggression." *Journal of Personality and Social Psychology*, 3(2), 153–162. doi:10.1037/h0022826.

Carton, A. 2018. ""I'm Not Mopping the Floors, I'm Putting a Man on the Moon": How NASA Leaders Enhanced the Meaningfulness of Work by Changing the Meaning of Work." *Administrative Science Quarterly* 63(2): 323-369.

Csikszentmihalyi, M. and LeFevre, J. "Optimal experience in work and leisure." *Journal of Personality and Social Psychology* 56.5 (1989): 815.

Csikszentmihalyi, M. and Rathunde, K., 1993. The measurement of flow in everyday life: toward a theory of emergent motivation . In J. E. Jacobs (Ed.), *Current theory and research in motivation, Vol. 40. Nebraska Symposium on Motivation, 1992: Developmental perspectives on motivation* (pp. 57-97). Lincoln, NE, US: University of Nebraska Press.

Cohen, L. 2013. "Assembling Jobs: A Model of How Tasks are Bundled into and Across Jobs" *Organizational Science*, 24(2): 432-454.

--- . 2016. Jobs as Gordian Knots: A New Perspective Linking Individuals, Tasks, Organizations and Institutions" in *The Structuring of Work in Organizations* (pp 25-59). Emerald Group Publishing Limited.

Cummings JN, Haas MR (2012) So many teams, so little time: Time allocation matters in geographically dispersed teams. J. Organ. Behav. 33(3):316–341

Dai, H., Milkman, K. L., Hofmann, D. A., and Staats, B. R. "The impact of time at work and time off From work on rule compliance: The case of hand hygiene in health care." *Journal of Applied Psychology* 100.3 (2015): 846.

Dai, H., Dietvorst, B., Tuckfield, B., Milkman, K., and Schweitzer, M. (2018). Quitting When the Going Gets Tough: A Downside of High Performance Expectations. *Academy of Management Journal*, *61*(5), 1667-1691.

Day, D., Gordon, S., and Fink, C. 2012. The sporting life: Exploring organizations through the lens of sport. *Academy of Management Annals*, 6: 397-433.

Diamond, J. (2018). Major League Baseball Is Moving Closer to a Universal Designated Hitter. *Wall Street Journal*. [online] Available at: https://www.wsj.com/articles/major-league-baseball-is-moving-closer-to-a-universal-designated-hitter-1531914518 [Accessed 30 Dec. 2018].

DiBenigno, J. and Kellogg, K. 2014. "Beyond occupational differences: The importance of cross- cutting demographics and dyadic toolkits for collaboration in a US hospital." *Administrative Science Quarterly*, *59*(3), pp.375-408.

Dollard, J., Miller, N.E., Doob, L.W., Mowrer, O.H. and Sears, R.R. 1939. *Frustration and aggressio*n. New Haven, CT, US: Yale University Press

Doeringer, P. and M. Piore. 1971. *Internal Labor Markets and Manpower Analysis.* Boston: D.C. Heath and Company

Drucker, P.F., 1988. The coming of the new organization. *Harvard Business Review*, 66, 45-53.

Duckworth, A., Peterson, C., Matthews, M. and Kelly, D. 2007. "Grit: perseverance and passion for long-term goals." *Journal of personality and social psychology* 92(6): 1087.

Durkheim, E., 1893. 1984. *The division of labor in society*. New York: Macmillan

The Economist. (2016). Is it ever a good idea to let a hurler hit?. [online] | *The Economist.* Available at: https://www.economist.com/blogs/gametheory/2016/06/pitchers-plate-baseball [Accessed 30 Dec. 2018].

ESPN, (2018). Max Scherzer: No one wants to see pitchers bat. *ABC News* [online]. Available at: https://abcnews.go.com/Sports/max-scherzer-pitchers-bat/story?id=30597790 [Accessed 30 Dec. 2018].

Fangraphs.com. (2010). Win Expectancy | *FanGraphs Baseball*. [online] Available at: https://library.fangraphs.com/misc/we/ [Accessed 12 Jan. 2019].

Fangraphs.com. (2017). The Recent History of Free Agent Pricing | *FanGraphs Baseball*. [online] Available at: https://blogs.fangraphs.com/the-recent-history-of-free-agent-pricing/ [Accessed 14 Jan. 2019].

Fangraphs.com. (2018). The Recent History of Free-Agent Pricing | *FanGraphs Baseball*. [online] Available at: https://www.fangraphs.com/blogs/the-recent-history-of-free-agent-pricing/ [Accessed 30 Dec. 2018].

Fangraphs.com. (2019). The State of Pitchers Hitting| *FanGraphs Baseball*. [online] Available at: https://tht.fangraphs.com/the-state-of-pitchers-hitting/[Accessed 19 Nov. 2019].

Faraj, Samer, and Lee Sproull. "Coordinating expertise in software development teams." Management science 46.12 (2000): 1554-1568.

Fisher, C. "Boredom at work: A neglected concept." *Human Relations* 46.3 (1993): 395-417.

Fox, S., and Spector, P. E. (1999). A model of work frustration–aggression. *Journal of Organizational Behavior*, 20(6), 915–931. doi:10.1002/(SICI)1099-1379(199911)20:6 < 915::AID-JOB918>3.3.CO;2-Y

Fox, S., Spector, P.E. and Miles, D., 2001. "Counterproductive work behavior (CWB) in response to job stressors and organizational justice: Some mediator and moderator tests for autonomy and emotions." *Journal of Vocational Behavior*, 59(3), pp.291-309.

Gagné, M. and Deci, E.L. 2005. "Self-determination theory and work motivation." *Journal of Organizational Behavior*, 26(4), pp.331-362.

Geen, R., 1968. “Effects of frustration, attack, and prior training in aggressiveness upon aggressive behavior.” *Journal of Personality and Social Psychology*, 9(4), p.316.

Glaser, Barney and Anselm Strauss (1971) Status Passage. Chicago: Aldine.

Goode, W.J., 1978. *The celebration of heroes: Prestige as a control system*. Berkeley. University of California Press

Green, B,, and Zwiebel, J. "The hot-hand fallacy: Cognitive mistakes or equilibrium adjustments? Evidence from Major League baseball." *Management Science* 64.11 (2017): 5315-5348.

Grimm, J., 2008. *Frustration Aggression Theory*. *The International Encyclopedia of Communication* Malden, MA : Blackwell Pub.

Guadagnoli, M.A. and Lee, T.D., 2004. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. Journal of motor behavior, 36(2), pp.212-224.

Gurr, T., 1970. *Why men rebel*. Princeton, N.J.: Published for the Center of International Studies, Princeton University [by] Princeton University Press

Hackman, J. and Oldham, G. 1980. *Work Redesign* . Reading, Mass: Addison-Wesley.

Hahl, O., 2016. Turning back the clock in baseball: The increased prominence of extrinsic rewards and demand for authenticity. Organization Science, 27(4), pp.929-953.

Hamblin, R., Bridger, D., Day, C. and Yancey, W. 1963 “The Interference-Aggression Law?” *Sociometry*, XXVI (June).

Haner, C. and P. Brown. 1955. “Clarification of the Instigation to Action Concept in the Frustration-Aggression Hypothesis,” *Journal of Abnormal and Social Psychology*, LI (September), 204-206, discussed in Berkowitz, 53-54.

Hernandez, S.M., 2015. A better understanding of bullying and hazing in the military. Mil. L. Rev., 223, p.415.

Hoegl, Martin, Katharina Weinkauf, and Hans Georg Gemuenden. "Interteam coordination, project commitment, and teamwork in multiteam R&D projects: A longitudinal study." *Organization science* 15.1 (2004): 38-55.

Hoffer, E., 1951. *The True Believer*, New York, NY.Harper, 21-28

Hovland, C. and Sears, R. 1940. “Minor studies of aggression: VI. Correlation of lynchings with economic indices.” *The Journal of Psychology*, 9(2), pp.301-310.

Huckman, R., Staats, B., and Upton, D. "Team familiarity, role experience, and performance: Evidence from Indian software services." *Management Science* 55.1 (2009): 85-100.

Hulin, C., and Blood, M. "Job enlargement, individual differences, and worker responses." *Psychological Bulletin* 69.1 (1968): 41.

Ibarra, H. and Petriglieri, J., 2016. Impossible selves: Image strategies and identity threat in professional women's career transitions.

Ichheiser, G. 1950. “Frustration and aggression or frustration and defence: A counter-hypothesis. *Journal of General Psychology*, 43(1), 125–129. doi:10.1080/00221309.1950.9710609

Ichniowski, C. and Shaw, K. "The effects of human resource management systems on economic performance: An international comparison of US and Japanese plants." *Management Science* 45.5 (1999): 704-721

Jung, H., Vissa, B. and Pich, M. "How do entrepreneurial founding teams allocate task positions?." *Academy of Management Journal* 60.1 (2017): 264-294.

Kahn, W., 1990. “Psychological conditions of personal engagement and disengagement at work.” *Academy of Management Journal*, 33(4), pp.692-724.

Kanter, R. M. 1968. “Commitment and Social Organization: A Study of Commitment Mechanisms in Utopian Communities,” *American Sociological Review*, (33), pp. 499-517.

Kim, J. and King, B.. "Seeing stars: Matthew effects and status bias in major league baseball umpiring." *Management Science* 60.11 (2014): 2619-2644.

Kregarman, J. and Worchel, P. 1961. “Arbitrariness of frustration and aggression.” *The Journal of Abnormal and Social Psychology*, 63(1), p.183.

Lazear, E., 2004. “Balanced skills and entrepreneurship.” *American Economic Review*, 94(2), pp.208-211.

Leifer, E., 2009. *Making the majors: The transformation of team sports in America*. Cambridge, MA. Harvard University Press.

Lepisto, D., Crosina, E., and Pratt, M. 2015. Identity work within and beyond the professions: Toward a theoretical integration and extension. In A. Desilva and M. Aparicio (Eds.), *International Handbook of Professional Identities* (pp. 11-37). Rosemead, CA: Scientific and Academic Publishing.

Locke, E. and Latham, G.P., 1990. *A theory of goal setting and task performance.* Englewood Cliffs, N.J.: Prentice Hall.

Maier, N., 1949. *Frustration, the study of behavior without a goal*. New York, NY, US: McGraw-Hill.

Malone T., Crowston, K., Lee J., Pentland, B., Dellarocas, C., Wyner, G., Quimby, J., Osborn, C.,
Bernstein, A., Herman, G., Klein, M. and O’Donnell E. "Tools for inventing organizations: Toward a handbook of organizational processes." *Management Science* 45.3 (1999): 425-443.

O’Leary MB, Mortensen M, Woolley AW (2011) Multiple team membership: A theoretical model of its effects on productivity and learning for individuals and teams. Acad. Management Rev. 36(3):461–478.

Malone, T., Laubacher, R. and Johns, T. 2011. "The age of hyperspecialization." *Harvard Business Review* 89(7/8): 56-65.

Marx, K., 1867. *Das Kapital*, Band I, English translation by Ben Fowkes of the 4th edn (1894).

McCelland, D. and Apicella, F.. 1945. “A Functional Classification of Verbal Reactions to Experimentally Induced Failure,” *Journal of Abnormal and Social Psychology,* XL (July), 376-390

Meiran, N., 2000. Reconfiguration of stimulus task sets and response task sets during task switching. Control of cognitive processes: Attention and performance XVIII, pp.377-400.

Merton, R. K. (1968). *Social theory and social structure*. Glencoe, IL. Free Press.

Mintzberg, H., 1983. *Structures in fives: Designing effective organizations*. Englewood Cliffs, NJ. Prentice-Hall

MLB.com. 2018. National League could add DH before long. MLB.com [online] Available at: https://www.mlb.com/news/national-league-could-add-dh-before-long/c-162404308 [Accessed 30 Dec. 2018].

Morlan, G., 1949. “A note on the frustration-aggression theories of Dollard and his associates.” *Psychological Review*, 56(1), p.1.

Perlow, L., 2012. *Sleeping with your smartphone: How to break the 24/7 habit and change the way you work*. Cambridge, MA. Harvard Business Press.

Podsakoff, N., LePine, J. and LePine, M., 2007. Differential challenge stressor-hindrance stressor relationships with job attitudes, turnover intentions, turnover, and withdrawal behavior: a meta-analysis. *Journal of applied psychology*, *92*(2), p.438.

Polya, G. (1957). *How to Solve It: A New Aspect of Mathematical Method*. Princeton, NJ: Princeton University Press.

Ranganathan, A. 2018. “Sticking to Picking Tea: Task Variety, Alignment and Productivity in a Tea Plantation in India” Presentation at Wharton’s People and Organizations Conference September 2018 – and correspondence with author.

Retrosheet.com. (2018). Play-by-Play Data Files (Event Files) | *Retrosheet*. [online] Available at: https://www.retrosheet.org/game.htm [Accessed 15 Oct. 2018].

The Ringer. 2018. Let’s Stop Pretending That Pitchers Can Hit. *The Ringer* [online] Available at: https://www.theringer.com/mlb/2018/6/7/17437016/national-league-pitcher-hitting-dh [Accessed 30 Dec. 2018].

Robertson, S.I., 2016. *Problem solving: Perspectives from cognition and neuroscience*. Psychology Press.

Rothaus, P. and Worchel, P., 1960. “The inhibition of aggression under nonarbitrary frustration.” *Journal of Personality*, 28(1), pp.108-117.

Ross, J., and Sharapov, D. (2015). When the leader follows: Avoiding dethronement through imitation. *Academy of Management Journal*, *58*(3), 658-679.

Scott, W. and Mitchell, T. 1976. *Organizational theory* (3rd ed.), Homewood, IL: Irwin.

Singley, M.K. and Anderson, J.R. (1989). *The Transfer of cognitive skill.* Cambridge, MA: Harvard Univ. Press.

Smith, A., 1776. *An inquiry into the nature and causes of the wealth of nations: Volume One*. London: printed for W. Strahan; and T. Cadell.

Spreitzer, G., 1995. “Psychological empowerment in the workplace: Dimensions, measurement, and validation.” *Academy of Management Journal*, 38(5), pp.1442-1465.

Staats, B., and Gino, F. 2012. "Specialization and variety in repetitive tasks: Evidence from a Japanese bank." Management Science 58(6): 1141-1159.

Stouffer, S., Suchman, E., DeVinney, L., Star, S. and Williams Jr, R., 1949. *The American soldier: Adjustment during army life: Studies in social psychology in World War II, Vol. 1* .Oxford, England: Princeton Univ. Press.

Stuart, C., and Moore, C. (2017). Shady characters: The implications of illicit organizational roles for resilient team performance. *Academy of Management Journal*, *60*(5), 1963-1985.

Summers JK, Humphrey SE, Ferris GR (2012) Team member change, flux in coordination, and performance: Effects of strategic core roles, information transfer, and cognitive ability. Acad. Management J. 55(2):314–338.

Sundstrom E, De Meuse KP, Futrell D (1990) Work teams. Amer. Psychologist 45(2):120–133. Taylor, F. Principles of scientific management. New York: Harper, 1911.

Thompson, R., and Kolstoe, R. 1974. “Physical aggression as a function of strength of frustration and instrumentality of aggression.” *Journal of Research in Personality*, 7(4), 314–323. doi:10.1016/0092-6566(74)90053-1

Tucker, A., Nembhard, I. and Edmondson, A. "Implementing new practices: An empirical study of organizational learning in hospital intensive care units." *Management Science* 53.6 (2007): 894-907.

Reagans, Ray, Ella Miron-Spektor, and Linda Argote. "Knowledge utilization, coordination, and team performance." *Organization Science* 27.5 (2016): 1108-1124.

Reber, A.S., 1989. Implicit learning and tacit knowledge. Journal of experimental psychology: General, 118(3), p.219.

Valentine MA, Edmondson AC (2014) Team scaffolds: How mesolevel structures enable role-based coordination in temporary groups. Organ. Sci. 26(2):405–422

Williamson, O. (1975). *Markets and hierarchies, analysis and antitrust implications: A study in the economics of internal organization*. New York: Free Press.

Wrzesniewski, A., and Dutton, J. 2001. "Crafting a job: Revisioning employees as active crafters of their work." Academy of Management Review 26(2): 179-201.

Zillman, D. and Cantor, J. 1977. Affective responses to the emotions of a protagonist. Journal of Experimental Social Psychology, 13(2), pp.155-165.

TABLES and FIGURES

Table 1: Summary Statistics

	N	Mean	St. Dev.	Min	Max
Pitcher Batted in Prior Inning	2,125,351	0.239	0.427	0	1
Pitcher was Successful as Batter	2,125,351	0.202	0.401	0	1
Pitcher was Unsuccessful as Batter	2,125,351	0.037	0.189	0	1
Batter is Out	2,125,351	0.672	0.469	0	1
Batters Elapsed After Return to Pitching	2,125,351	1.67	1.49	0	13
Pitcher at Home Stadium	2,125,351	0.510	0.500	0	1
Mediators Used to Test Potential Mechanisms					
Pitcher Performance in Prior Inning	1,332,462	0.758	0.209	0	1
Batters Pitched to Without Batting (Monotony)	2,094,863	6.638	5.745	0	48
Batters Pitched to Without Batting (Monotony)	2,125,351	0.095	0.293	0	1
Pitcher Plays for AL Team	2,125,351	0.095	0.293	0	1
Score Difference	2,125,351	2.107	2.245	0	24
Pitcher is a Rookie	2,032,951	0.131	0.337	0	1

Figure 1: Batting Performance by Position

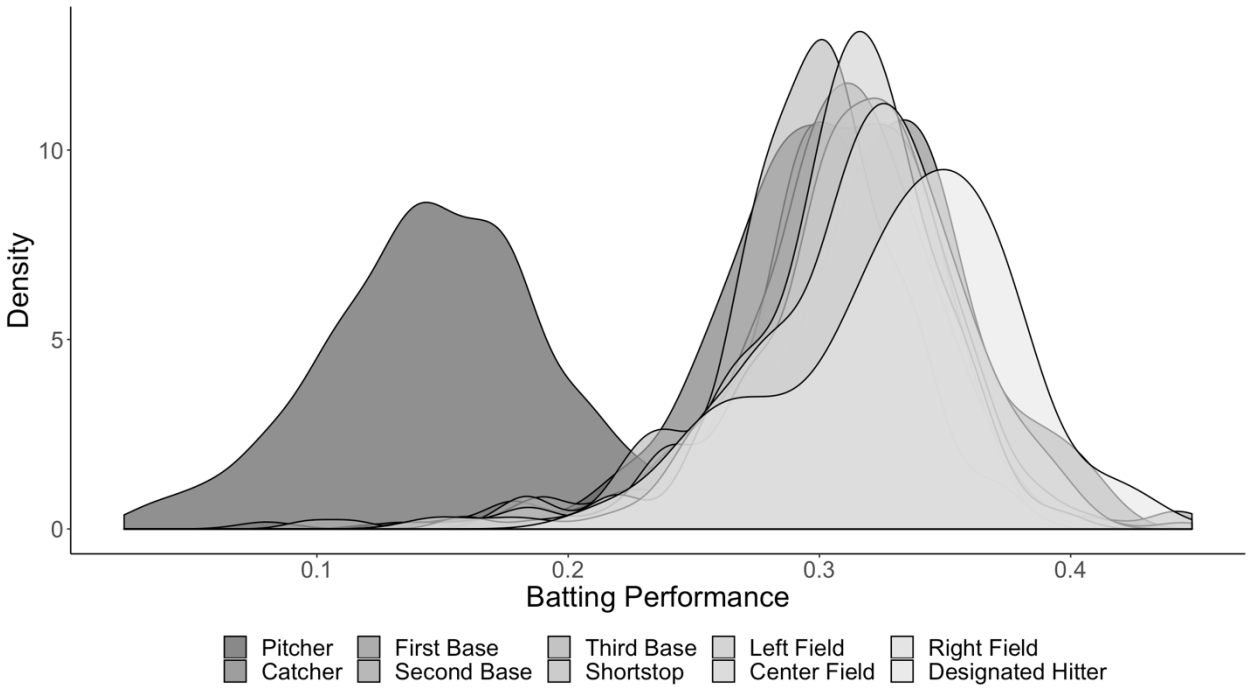


Figure 2: Relationship Between Pitcher At-Bats and Performance

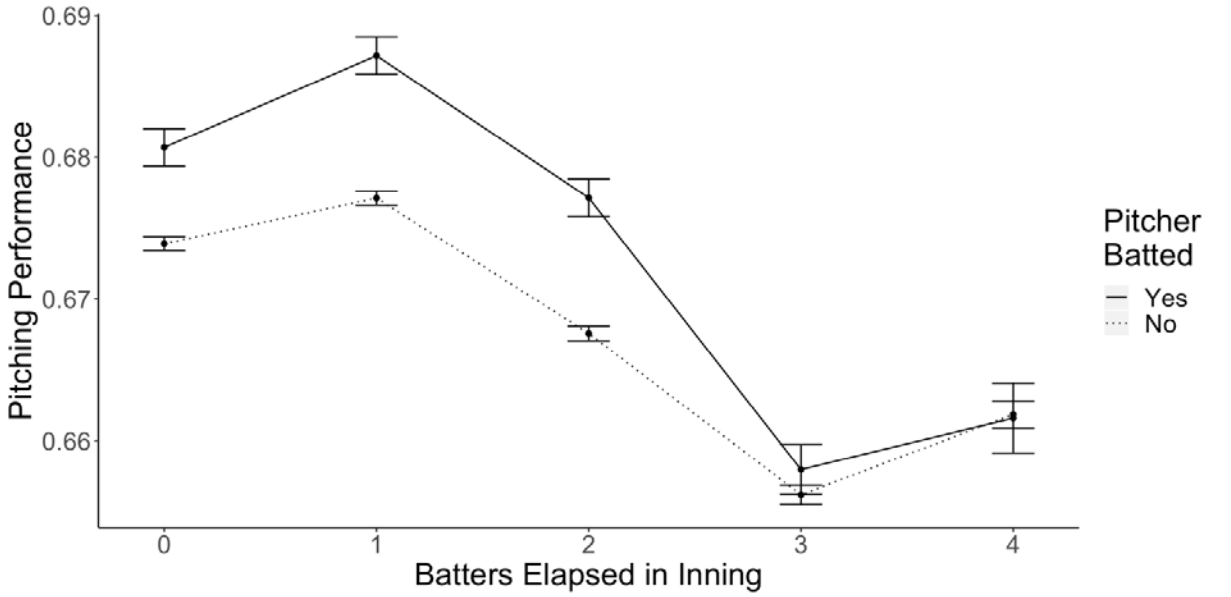


Table 2: The Effect of Having Batted on Pitcher Performance

	Dependent Variable:				
	Batter is Out				
	Model 1	Model 2	Model 3	Model 4	Model 5
Pitcher Batted in Prior Inning (batted = 1)	0.00353***		0.0251***		0.00489**
	(0.00)		(0.00)		(0.00)
Pitcher was Unsuccessful Batter		0.00461***		.00688**	
		(0.00)		(0.00)	
Pitcher was Successful Batter		-0.00232			
		(0.00)			
Batters Elapsed in Inning (first batter = 0)	-0.00316***	-0.00316***	0.0245***	-0.00171	-0.00110***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pitcher at Home Ballpark	0.0121***	0.0121***		0.0105***	0.00917***
	(0.00)	(0.00)		(0.00)	(0.00)
Pitcher Batted in Prior Inning × Batters Elapsed	-0.00142***		-0.0160***		-0.00059
	(0.00)		(0.00)		(0.00)
Pitcher was Unsuccessful Batter × Batters Elapsed		-0.00183***		0.00688***	
		(0.00)		(0.00)	
Pitcher was Successful Batter × Batters Elapsed		0.000779			
		(0.00)			
Constant	0.672***	0.672***	0.688***	0.677***	0.673***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Individual Pitcher Fixed Effects	Yes	Yes	No	Yes	Yes
Pitcher-Game Fixed Effects	No	No	Yes	No	No
Opponent Place in Batting Order Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Inning Fixed Effects	Yes	Yes	Yes	Yes	Yes
Matched Sample	No	No	No	No	Yes
Observations	2,125,337	2,125,337	2,111,515	508,298	1,003,704
Adjusted R^2	0.010	0.010	0.024	0.010	0.014

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All models OLS Regression. Observations in which the pitcher faced only one batter in a game are dropped in Model 3 to include pitcher-game fixed effects. Sample size varies in Model 4 because the data is limited to only situations in which the pitcher batted in the prior inning and in Model 5 because the data is limited to a matched sample

Figure 3: Predicted Effect of Having Batted on Pitching Performance

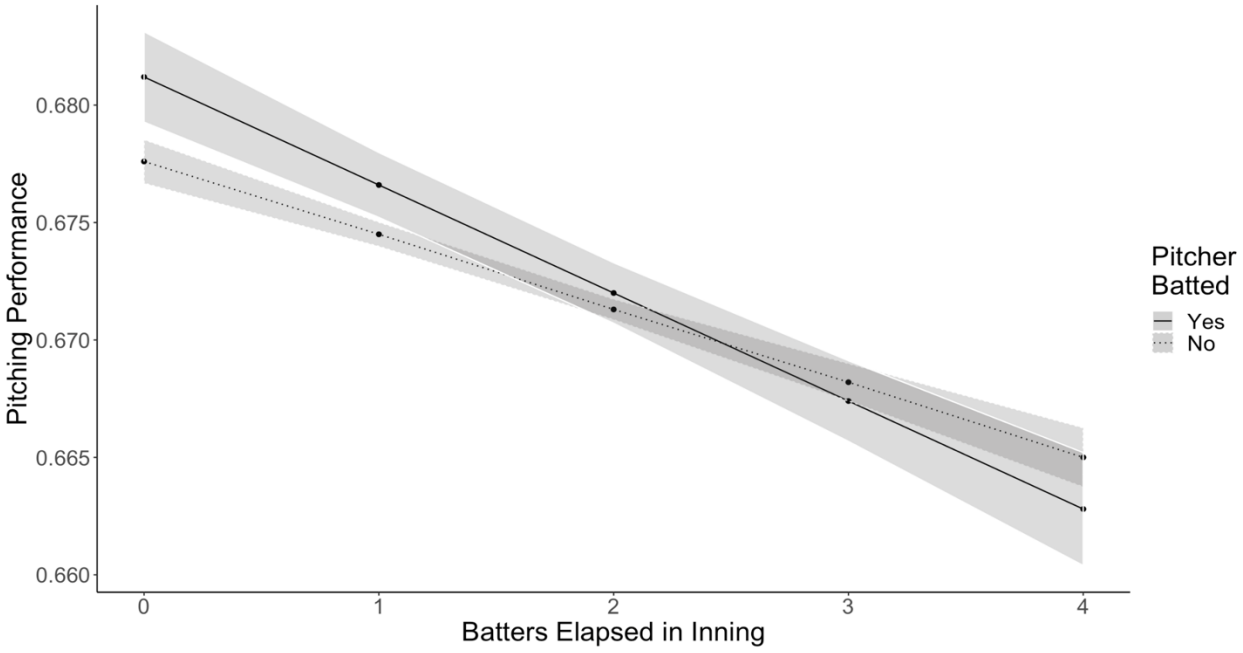


Figure 4: Predicted Effect of Having Batted by Situational Stakes

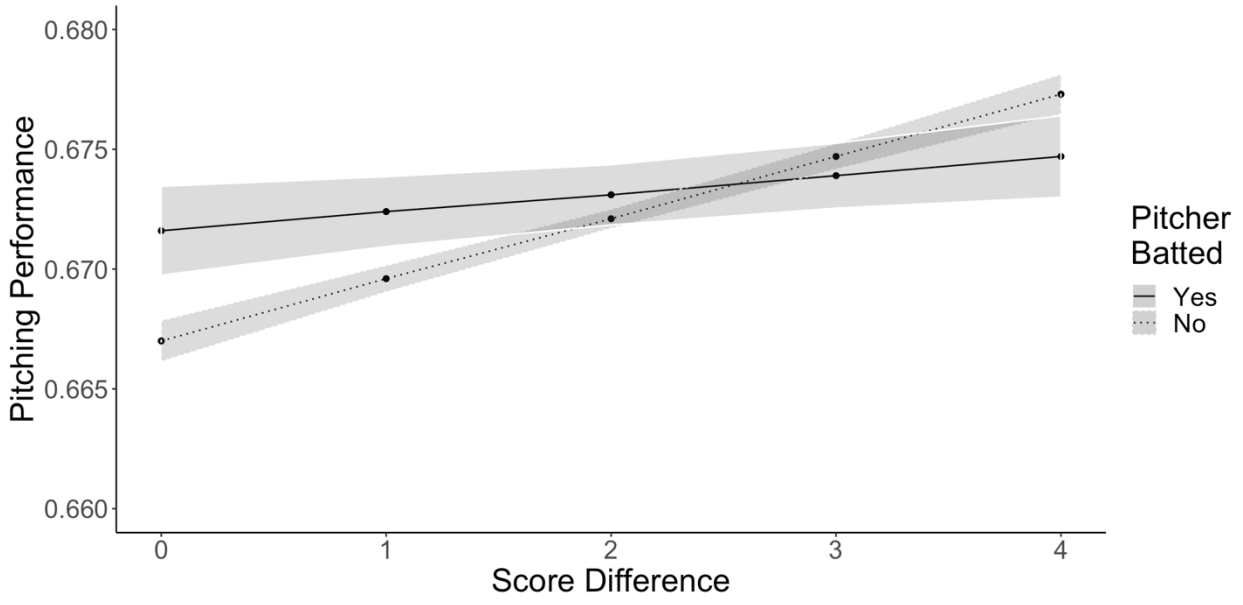


Table 3: Testing Potential Mechanisms Underlying the Positive Relationship Between Having Batted and Performance

	Dependent Variable:							
	Batter is Out				Pitch in Strike Zone	Batter Walks	Batter Gets Hit	
	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11	Model 12	Model 12
Pitcher Batted in Prior Inning (batted = 1)	0.00381*	0.00473*	0.00356**	0.00635***	0.00491***	0.00616***	-0.00536***	0.00192***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Batters Elapsed in Inning (first batter = 0)	-0.00356***	-0.00269***	-0.00316***	-0.00358***	-0.00318***	-.00466***	0.00730***	-0.00434***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pitcher Batted × Batters Elapsed	-0.00101	-0.000399	-0.00142**	-0.00105*	-0.00150**	-.00168**	0.00089**	0.00043
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pitcher at Home Ballpark	0.0107***	0.0123***	0.0124***	0.0125***	0.0119***	.00173	-0.00617***	-0.00560***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Performance in Last Inning	0.00537*							
	(0.00)							
Pitcher Batted × Performance Last Inning	0.00151							
	(0.00)							
Time Since At-Bat		-0.000381***						
		(0.00)						
Pitcher Batted × Time Since At-Bat		-0.00106*						
		(0.00)						
American League Pitcher			0.00606**					
			(0.00)					
Pitcher Batted × AL Pitcher			-0.000378					
			(0.00)					
Score Difference				0.00257***				
				(0.00)				
Pitcher Batted × Score Difference				-0.00180***				
				(0.00)				
Rookie Pitcher					0.00051			
					(0.00)			
Pitcher Batted × Rookie Pitcher					-0.01091***			
					(0.00)			
Constant	0.680***	0.671***	0.671***	0.672***	0.675***	.50514***	0.0849***	0.23340***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Individual Pitcher Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inning Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Opponent Place in Batting Order Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,332,450	2,094,854	2,125,337	2,125,337	2,032,937	1,405,910	2,125,337	2,125,337
Adjusted R^2	0.014	0.010	0.010	0.010	0.012	0.005	0.011	

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All models OLS Regression. Sample Size varies because the data is limited to innings in which the pitcher was in the game in the prior inning in Model 6 and Model 11 is estimated within a dataset of all pitches thrown during a subset of the sample

Figure 5: Changes in Pitches' Locations in Innings After the Pitcher Bats

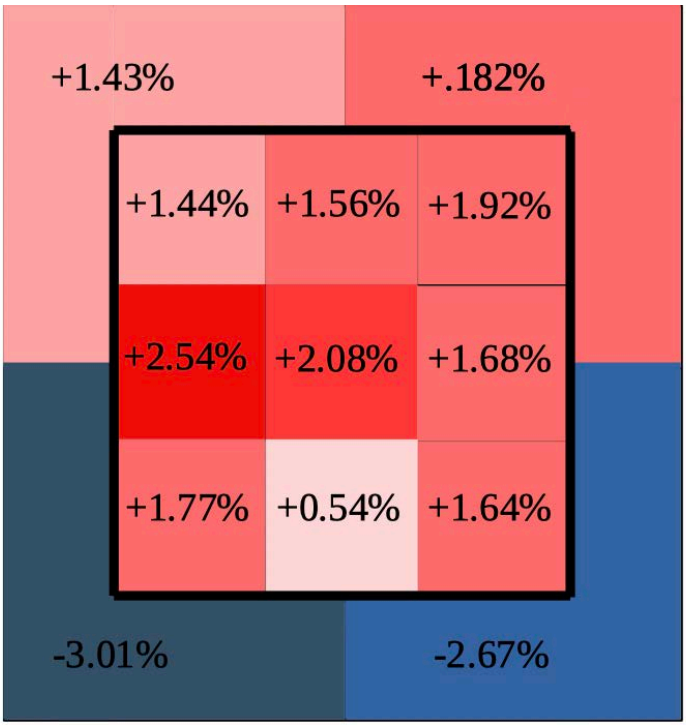


Table 4: The Effect of Inexpert Work for Pitchers on Overall Team Outcomes

	Dependent Variable:	
	Any Run Scores	Runs Scored
	Model 14	Model 15
Pitcher Batter in Prior Inning (batted = 1)	-0.00698***	-0.0175***
	(0.00)	(0.00)
Batters Elapsed in Inning (first batter = 0)	0.0204***	0.0468***
	(0.00)	(0.00)
Pitcher Batted in Prior Inning × Batters Elapsed	0.00298***	0.00724***
	(0.00)	(0.00)
Pitcher at Home Ballpark	-0.00324***	-0.00613***
	(0.00)	(0.00)
Constant	0.0181***	0.0176***
	(0.00)	(0.00)
Individual Pitcher Fixed Effects	Yes	Yes
Year Fixed Effects	Yes	Yes
Inning Fixed Effects	Yes	Yes
Opponent Place in Batting Order Fixed Effects	Yes	Yes
Observations	2,125,337	2,125,337
Adjusted R^2	0.022	0.031

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All models OLS Regression.

APPENDIX A: Additional Statistics on Batting Performance by Position

A1: Bunting Rates by Position

Fielding Position	Proportion of Bunts	Proportion of Swinging At-Bats
Pitchers	14.4%	85.6%
Non-Pitchers	1.1%	98.9%

A2: Mean Batting Performance by Fielding Position

Fielding Position	Batting Average	On-Base Percentage	Slugging Percentage	OBP + Slugging
Pitchers	0.131	0.148	0.165	0.314
Catchers	0.232	0.293	0.356	0.650
First Base	0.247	0.317	0.414	0.731
Second Base	0.245	0.305	0.356	0.661
Third Base	0.245	0.308	0.383	0.691
Shortstop	0.241	0.296	0.350	0.646
Left Field	0.243	0.309	0.387	0.696
Center Field	0.248	0.308	0.373	0.681
Right Field	0.249	0.314	0.402	0.715
Designated Hitter	0.254	0.329	0.439	0.768

CONCLUSION

Status under-recognition may induce negative responses by damaging an individual's self-image, even when the under-recognition carries no material, career, or reputation concerns. Nominally under-recognized employees may feel their 'face' has been compromised (Goffman 1967**,** pg. 5).[45] Even when there are no tangible or even intangible costs to denied recognition, under-recognized employees may still suffer psychic costs. Moreover, it becomes increasingly difficult to defend one's indignation at damage done to self-image when others may be aware of any material compensation for the nominal under-recognition. Reacting negatively by exiting the organization becomes "a ritual move, conveying that [the employee] has a face to lose and that its loss is not to be permitted lightly" rather than "irrational expression of frustration" (Goffman 1967: 23).

Using a unique natural experiment, my first chapter essay provides the first evidence that *purely nominal* status recognition carries substantial social comparison costs in reward schemes that rely on nonmonetary recognition. The study design uses counterfactual cases to status recognition where employees' performances merited recognition, but they did not receive it due to arbitrary restrictions on the supply of the official recognition. Importantly, all parties recognize that the status recognition is an arbitrary distinction among equally high-performing employees and is generated only by a predetermined scarcity of recognition to allocate among top performers. Under-recognition carries no future promotion or other career consequences in this setting.

I demonstrate that employees are highly sensitive to under-recognition even when it carries no monetary, signaling, or career reputation concerns. I find that nominally under-recognized employees are much more likely to voluntarily exit the organization, often going to work for a local competitor. Strikingly, these employees are much more likely to exit *even though they are awarded greater monetary bonuses in compensation for their under-recognition*. Moreover, I find that under-recognized employees who exit

[45] See also Goode (1978: 102-110) on reputation dynamics within membership groups in terms of gossip hazards and fears.

often join weaker economic competitors. Perceived damage to self-image thus weighs more heavily than material considerations.

These findings add a strong cautionary note for the use of nonmonetary status recognition in the design and management of performance reward schemes. Sensitivity to under-recognition can compel employees to seek outside options in an effort to rectify the discrepancy between their social status recognition and their self-image. Therefore, it is important for performance management and reward design schemes to heed employees' sensitivities to perceived discrepancies between their self-image and what is officially recognized by the organization beyond any material, career, or reputational concerns. Building on these insights regarding the effects of discrepant status recognition, my second essay turns to investigating the causes and consequences of *overvaluing* performance.

In my second chapter, I investigate the determinants and consequences of *overvaluing* employee performance. Overvaluing the performance of some employees over others can reflect agency problems where the consequences of performance evaluations are not in the best interest of the organization (Prendergast and Topel 1993, 1996). Such a bias may prevent under-performing workers from finding more appropriate or better fitting jobs or lead to unjustified promotion or retention (Prendergast and Topel 1996). Preferential treatment of favored employees also demotivates slighted employees, reducing employee effort (MacLeod 2003).[46]

However, providing necessary negative feedback can be challenging for a manager who takes a relational approach, one which prioritizes the building of trust with subordinates, as they may be more likely to elevate the motivational concerns of the under-performer. The longer their relationship and the more it is marked by a strong sense of trust and cooperation, the more easily managers can adopt their perspective (Granovetter 1985; Coleman 1988; Montgomery 1998). Such mutual trust, goodwill, and commitment between managers and employees can enhance internal coordination mechanisms within organizations (Gittell 2016; Jordan 2015; Kotzé and Roodt 2005).

[46] For a review of further negative employee reactions to perceptions of inequity in performance evaluations from the social psychological literature on social comparisons see Larkin et al. 2012.

Managers thus face a tension in evaluating the performance of under-performing employees: relax official evaluation standards, thereby avoiding potential conflict and prioritizing the relationship with their underperforming employees, or enforce official evaluation standards as an agent of their principle, the organization. In such an embedded relationship, a manager may inflate the rating of an employee, signaling faith and expectations that the employee will reciprocate by improving productivity in the future (Gittel and Douglass 2012). Such a particularistic "ethic of care" approach toward particular employees with whom they share this mutual understanding may compromise the manager's ability to balance broader organizational goals (Heimer 1992).

Embeddedness scholarship leads us to expect limits to and eventual costs from such highly trusting and cooperative relationships should they become over-embedded – a state where the relationship is so close as to forestall alternative relationships or flexibility in strategy (Uzzi 1997; Sorenson and Waguespack 2006; Doering 2018). Rather than being able to balance the particular concern for an underperforming employee with organizational goals, a manager may find it only possible to take one of two approaches when evaluating an under-performing employee's performance: either that of the trusting, relational manager with an ongoing cooperative relationship or that of an executor of disinterested organizational policy. Managers cannot always reconcile the two roles simultaneously when evaluating their employees (Fernandez 1991). The exclusionary tension may be especially difficult when considering underperforming employees with whom the manager has developed a long-lasting embedded relationship of trust and cooperation.

One benefit of formalized work practices, or management approaches governed by clear written rules and procedures (Anderson and Tomaskovic-Devey 1995; Huffman and Velasco 1997; Weber 1946), is that they constrain the discretion that managers may exercise in favoring certain employees over others (Elvira and Graham 2002; Reskin 2000). To the extent that adopting such a formal management approach does not crowd out the positive by-products of high levels of trust and cooperation within teams, this approach may present a solution to the risk such embedded relationships pose in performance evaluation processes.

In this vein, Canales (2014) finds that the effectiveness of relational styles is contingent on the influence of rule-enforcing peers who take a formalized approach and provide a check on the discretionary inclinations of the actors with relational styles in a financial services setting. Notably though, individual loan officers who have a style defined by oscillating between relational and formal from client-to-client generate the worst organizational outcomes, emphasizing the importance of consistency in these lending arrangements (see also Canales and Greenberg 2016). These insights, based on external lending arrangements, are limited, however, in their prescription for managers of employees within a firm which by-and-large would benefit from high trust and cooperation within teams.

My second dissertation chapter advances this line of research by testing the extent to which the embeddedness and formalized work practice literatures predict managerial overvaluation of employees within an organization rather than in client or inter-firm relationships. By studying managers in their own organizational context, I am able to make progress on theory underpinning both professional relational embeddedness and formalized work practices.

I find that highly relational managers are more likely to overvalue under-performing employee performance beyond what is justifiably objective. I further find that managers taking a more structured approach to managing are less likely to overvalue the performance of under-performing employees, regardless of the degree of relational embeddedness. Complementary analysis using PharMed administrative data provides evidence that overvaluing performance of underperforming employees indeed carries negative productivity consequences for the organization.

This research highlights how different approaches to managing people can predict overvaluation of employee performance. On the level of practice, it further illuminates how organizations can tip the balance toward one desired outcome over another by fostering formalized management practices, while reconciling various welfare concerns at the same time.

Status recognition can also have subtle performance implications, influencing day-to-day productivity. Many approaches to job design, such as self-determination theory, emphasize the importance of worker motivation for productivity (Gagné and Deci 2005). However, jobs often come to encompass

tasks that are necessary for the function of the larger organization, but that are distant from the worker's superior competencies ("inexpert" tasks) or the aspects of work from which they derive meaning (Wrzesniewski and Dutton 2001). As an example, the importance of flexibility and the need to respond quickly to clients means that sometimes a specialized member of a consultancy team will be called to fill in for another's role when the other is indisposed (Perlow 2012). Such task heterogeneity often implies incorporating lower status or inexpert work, which may be demotivating for professionals when such tasks are not clearly assigned across collaborating occupations or ranges of expertise (DiBenigno and Kellogg 2014).

Recent research has proposed that performing inexpert tasks can positively affect productivity in workers' core activities by fostering identification with the final product (Ranganathan 2018). However, inexpert tasks are not always easily connected to the meaning of the final focal product. Therefore, understanding how frustrating, peripheral, or inexpert tasks influence overall productivity when a meaningful connection to the final product is tenuous remains unclear.

My third chapter (coauthored with a fellow student coauthor) tests whether switching between inexpert and specialist tasks has positive, negative, or neutral effects on overall productivity. This essay draws on the *frustration-aggression* (FA) hypothesis (Dollard et al. 1939; see also Breuer and Elson 2017) to capture the idea that performing frustrating inexpert tasks can positively affect productivity. Specialists can react to having to engage in tasks in which they are relatively inexpert by over-performing when they resume activities in which they are superior. We call this theoretical mechanism through which these peripheral tasks can affect productivity, "forced inferiority." We theorize that such over-performance following compulsory tasks, in which professional are recognized as inferior, will emerge even relative to their own baseline performance in their superior tasks.

While prior theory suggests that inexpert tasks are likely to be detrimental to productivity, either by being distracting, tiring, or breaking the worker's flow (Csikszentmihalyi & LeFevre 1989), we find that engaging in inexpert tasks can affect specialist performance positively upon return to their specialization, regardless of whether meaning is derived. Pitchers perform better upon returning to their superior

competency (pitching) following forced inferiority (batting) than would otherwise be expected controlling for pitcher, game, year and other relevant fixed effects. Additionally, forced inferiority's effect on individual performances has a tangible impact on organizational-level outcomes in the form of the team's probability of winning a game. The effect of forced inferiority on performance is greatest when the situational stakes are high. Furthermore, pitchers do not become inured to this effect over repeated instances of forced inferiority. These results suggest that when professionals engage in inexpert tasks in which they are inferior, they are likely to over-perform when returning to tasks in which they excel.

Together, these findings carry important implications for how organizations should handle the need for team members to work on tasks *outside* their superior competencies. We find that such tasks have a positive effect on overall productivity, and this suggests that organizations can benefit from engaging specialists in activities outside their core expert role, or at least not seek to eliminate such activities out of hand. Additionally, organizations should provide productive avenues for professionals to channel reactions to the frustration that comes from performing a task that professionals are inferior in performing. Lastly, we find that *even elite* performers are sensitive to tasks that accentuate their weaknesses and will productively overcompensate when given the opportunity to excel. This job design insight offers a potential productivity enhancement mechanism even for those at the peak of their specialization.

References

Anderson, C.D. and Tomaskovic-Devey, D., 1995. Patriarchal pressures: An exploration of organizational processes that exacerbate and erode gender earnings inequality. *Work and Occupations*, 22(3), pp.328-356.

Canales, Rodrigo. 2014. Weaving Straw into Gold: Managing Organizational Tensions between Standardization and Flexibility in Microfinance. *Organization Science*, 25(1): 1-28.

Canales, Rodrigo and Greenberg, Jason. 2016. "A matter of (Relational) Style: Loan Officer Consistency and Exchange Continuity in Microfinance". *Management Science*, 62(4): 1202-1224.

Coleman, J.S., 1988. Social capital in the creation of human capital. *American Journal of Sociology*, 94, pp.S95-S120.

Doering, L., 2018. Risks, Returns, and Relational Lending: Personal Ties in Microfinance. *American Journal of Sociology*, 123(5), pp.1341-1381.

Elvira, M.M. and Graham, M.E., 2002. Not just a formality: Pay system formalization and sex-related earnings effects. *Organization Science*, 13(6), pp.601-617.

Fernandez, R.M., 1991. Structural bases of leadership in intraorganizational networks. *Social Psychology Quarterly*, pp.36-53.

Gittell, J.H., 2016. Transforming relationships for high performance: The power of relational coordination. Stanford University Press.

Gittell, J.H. and Douglass, A., 2012. Relational bureaucracy: Structuring reciprocal relationships into roles. *Academy of Management Review*, 37(4), pp.709-733.

Goffman, E., 1967. On face-work. Interaction ritual, pp.5-45.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3), pp.481-510.

Heimer, Carol A. 1992. "Doing Your Job and Helping Your Friends: Universalistic Norms about Obligations to Particular Others in Networks" in Networks and Organizations: Structure, Form, and Action, edited by Nitin Nohria and Robert G. Eccles. Boston, Mass: Harvard Business School Press.

Huffman, M.L. and Velasco, S.C., 1997. When more is less: Sex composition, organizations, and earnings in US firms. *Work and Occupations*, 24(2), pp.214-244.

Jordan, Gretchen. 2015. "What Matters to R&D Workers" *Research Technology Management*, 48(3):23-32.

Kotzé, Koos, and Gert Roodt. 2005. Factors that affect the Retention of Managerial and Specialist Staff: An Exploratory Study of an Employee Commitment Model. *SA Journal of Human Resource Management*, 3(2):48-55.

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

Montgomery, J.D., 1998. Toward a role-theoretic conception of embeddedness. *American Journal of Sociology*, 104(1), pp.AJSv104p92-125.

Prendergast, C. and Topel, R., 1993. Discretion and bias in performance evaluation. *European Economic Review*, 37(2-3), pp.355-365.

Prendergast, C. and Topel, R.H., 1996. Favoritism in organizations. *Journal of Political Economy*, 104(5), pp.958-978.

Reskin, B.F., 2000. Getting it right: Sex and race inequality in work organizations. *Annual Review of Sociology*, 26(1), pp.707-709.

Sorenson, Olav and David M. Waguespack. 2006. Social Structure and Exchange: Self-Confirming Dynamics in Hollywood. *Administrative Science Quarterly*. 51(4): 560-589.

Uzzi, B., 1997. Social structure and competition in interfirm networks: The paradox of

embeddedness. *Administrative Science Quarterly*, pp.35-67.
Weber,M.1946. Bureaucracy. InH.H.Gerth&C.WrightMills (Eds.), From Max Weber: Essays in sociology: 196–244. Princeton, NJ: Oxford University Press.