TRUST VERSUS REWARDS: REVISITING MANAGERIAL DISCRETION IN INCOMPLETE CONTRACTS

A Dissertation
Presented to
The Academic Faculty

by

Wenqian Hu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Accounting in the
Scheller College of Business

Georgia Institute of Technology
May 2022

TRUST VERSUS REWARDS: REVISITING MANAGERIAL DISCRETION IN INCOMPLETE CONTRACTS

Approved by:

Dr. Jeffrey Hales, Advisor
McCombs School of Business
University of Texas at Austin

Dr. Xi (Jason) Kuang, Advisor
Scheller College of Business
Georgia Institute of Technology

Dr. Nikki Mackenzie
Scheller College of Business
Georgia Institute of Technology

Dr. Kathy Rupar-Wang
Scheller College of Business
Georgia Institute of Technology

Dr. Ivo Tafkov
J. Mack Robinson College of Business
Georgia State University

Date Approved: [April 20, 2022]

ACKNOWLEDGEMENTS

I sincerely thank my dissertation committee, Jeffrey Hales (co-chair), Jason Kuang (co-chair), Nikki Mackenzie, Kathy Rupar-Wang, and Ivo Tafkov, for their guidance and tremendous support throughout the development of this paper. I appreciate the helpful comments and suggestions from Jeremy Bentley, Angelo Ditillo, Hans Frimor, Lan Guo, Alex Lyubimov, Andreas Ostermaier, Adam Presslee, Timothy Shields, Alan Webb, participants at 2022 Hawaii Accounting Research Conference, 2022 MAS Midyear Meeting, 2021 JMAR Online Rookie Camp, 2021 ABO Section Research Conference, 2021 AAA Annual Meeting, and workshop participants at Bocconi University, Concordia University, Georgia Institute of Technology, HEC Lausanne, HEC Paris, Iowa State University, University of Southern Denmark, University of Waterloo, and Wilfrid Laurier University. I thank Todd Swarthout for his help with the access to student participants at the Experimental Economics Lab of Georgia State University. I also thank Philipp Chapkovski for his guidance with Otree programming. Finally, I gratefully acknowledge the financial support provided by the Institute of Management Accountants Research Foundation Doctoral Scholars Stage III Grant.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Incentive compensation is often characterized by incomplete contracts. While managerial opportunism has been documented as one of the most pronounced problems with managerial discretion in incomplete contracts, prior work has not investigated the underlying mechanisms driving a loss of productivity. In this study, I experimentally investigate whether replacing human managers' decision making with algorithm-generated bonus schemes that mimic managers' decision making improves employee productivity. I find that compensation determined by algorithms generate higher productivity without sacrificing the residual profits. Further, the productivity-inducing effect of algorithms is stronger when the rewards are not contingent on the performance signal. These results are consistent with the idea that it is hard for managers to establish credibility for rewarding employees for their performance in incomplete contracts. Employee productivity can be improved by enhancing their trust in the rewarding mechanism, even when they are not paid a more generous bonus scheme. This study advances our understanding of the behavioral factors influencing employee productivity in incomplete contracts and the potential ways algorithm-based evaluations can be used to improve firm outcomes.

CHAPTER 1. INTRODUCTION

Performance evaluation and compensation decisions are often characterized by incomplete contracts, where individual contributions to firm value are not fully captured by formal contracts. In such decisions, managerial discretion allows managers to take into account information that is relevant to employee performance but is not reflected in objective performance measures (Baiman and Rajan 1995; Bol 2008; Maas et al. 2012). However, managerial discretion comes with costs. The flexibility in compensation decisions can provide managers with the opportunity to act on their own incentives and preferences, potentially in a way that would reduce the effectiveness of incomplete contracts (Fisher et al. 2005; Bol 2008). Eliminating discretion or setting up a fixed bonus pool, however, may not always be optimal (Banker and Datar 1989; Rajan and Reichelstein 2006). Thus, the problem remains with regard to how to motivate employee productivity when resource allocations are influenced by managers' discretion and self-interest. This study investigates potential mechanisms that could cause a loss of productivity under managerial discretion and provides insights on how managers can effectively implement discretion in incomplete contracts.

While theoretical work and empirical evidence suggests managerial discretion can reduce employee productivity when managers have conflicts of interests in rewarding employees (Milgrom and Roberts 1987; Milgrom 1988; Baiman and Rajan 1995; Fisher et al. 2005), prior work has not investigated the underlying mechanisms that cause the loss of productivity. Economic theory suggests that a loss of productivity can be driven by a belief that expected future rewards will not be sufficient to justify increased marginal effort.

Specifically, this could come in the form of compensation that is insufficiently sensitive to prior performance. However, research on gift exchange and reciprocity norms (e.g. Charness and Haruvy 2002; Falk 2007) suggests that the average level of compensation could play an important role in inducing production effort. In this study, I use an algorithm to remove managers' responsibility for compensation decisions to better understand whether employee productivity is driven by employees' trust in the rewarding system or the level of rewards paid to them.

Distinguishing the different mechanisms that drive a loss of productivity is important as it speaks to the different remedies managers can take to improve productivity in incomplete contracts. If productivity is driven by trust rather than the reward itself, then transparency in the rewarding process could be explored as a more effective way to improve productivity. On the other hand, if the primary concern is the level of rewards, then managers simply need to pay more to the employees for the inherent risk in incomplete contracts. I address this gap in the literature by examining how replacing managerial discretion with algorithm-generated bonus allocations that mimic managers' decision making affects employee productivity in incomplete contracts.[1]

To investigate this research question, I examine a setting where managers have discretion over the size of total employee compensation pool and are thus residual

[1] This study may also have implications for companies' use of algorithm-based automation (Schrage et al. 2019). There is an increasing trend that algorithm-based decision making replaces managerial subjective judgment in management control functions. For instance, IBM uses its Watson Analytics to predict employee attrition (Rosenbaum 2019) and determine the appropriate pay raises and promotions for employees (Hellard 2018). Companies such as Amazon and Jet Blue invest in complex algorithms to hire future employees and track current employees' performance (Dastin 2018; Logg et al. 2019). This study takes a first step to build an algorithm that simply mimics managers' decision making. Results can potentially shed light on the ways algorithm-based evaluations can be used to improve firm outcomes.

claimants of the firms' profits. In this setting, any additional rewards paid to the employees reduce managers' own wealth. I investigate this setting as prior research suggests that the incentive problem and concerns for managerial opportunism are more pronounced when managers have conflicts of interests in rewarding employees (Baiman and Rajan 1995; Fisher et al. 2005). I argue that changing the way in which discretion is implemented from human managers to algorithms can affect employees' interpretation of reward outcomes, which in turn affects their productivity, even though algorithms replicate the reward experience with human managers.

Two underlying mechanisms could potentially contribute to any different reactions to human versus algorithm-based bonus determinations. One mechanism comes from the social psychology theory on attribution (Kelley 1973; Crittenden 1983). Prior psychology work suggests that, when outcomes are perceived as attributable to actions of interested parties as opposed to natural occurrences, there is a stronger tendency for people to assign blame and a stronger preference for fair outcomes (Blount 1995). When managers determine the rewards, employees are more likely to attribute the reward outcomes to managers' intentions and self-interest, and as a result negatively (positively) reciprocate unfavorable (favorable) rewards; whereas when algorithms remove the discretion in decision making and managers only implement the rewards determined by algorithms, employees might be less likely to infer intentionality of the managers. Further, based on the negativity bias in attribution, people's tendency to infer causes of outcomes is stronger when rewards are unfavorable. Thus, I predict algorithms will alleviate employees' tendency to negatively reciprocate managers for unfavorable rewards, and as a result, induce higher productivity than managers.

Another mechanism that can potentially contribute to employees' different reactions is grounded in the credibility concerns in incomplete contracts. Theory on source credibility suggests that an action source that does not have vested interest in the outcomes is perceived as more trustworthy (Reinard 1988; Pornpitakpan 2004). Compared to managers' decision making, algorithms are more likely to be viewed as separated from a vested stake in the bonus outcomes, which can lead to perceived fairness and trustworthiness (McGarry and Hendrick 1974; Leventhal 1980). Further, recent research on algorithm appreciation has documented perceptions of algorithmic decisions as more consistent than human decision-makers (Dietvorst et al. 2015; Lee 2018). Thus, theories on source credibility and algorithm appreciation suggest that algorithm-generated bonus schemes can improve the credibility of the rewarding mechanism such that employees trust that their efforts will be rewarded fairly. This increased trust in the rewarding mechanism can lead to improved productivity. Hence, both the attribution and the credibility mechanisms suggest the same directional prediction that algorithms will improve productivity. The two mechanisms, however, differ in their prediction of employees' sensitivity to reward outcomes. Attribution theory indicates employees will respond more strongly to prior reward outcomes when they interact with managers than algorithms; whereas the higher credibility of algorithms suggests a stronger reaction to prior rewards under algorithms than managers. Therefore, I propose a directional prediction for employee productivity while using the pattern of observed results to help distinguish the underlying reasons for that behavior.

I use a computerized experiment with a 1 x 2 (manager responsibility: present versus absent) x 8 (period) design to investigate my research question. Participants interact in three-person groups with one manager and two employees in each condition. In each period, employees allocate the time to work on a real-effort task or to consume as leisure. Managers are granted discretion over the size of the employee total compensation pool, which is then split equally between the two employees. Consistent with an incomplete contract setting, the objective performance measure is subject to random noise and managers receive a noisy signal of employee group performance. Managers can use the discretion as desired in their compensation decisions and claim the residual profits. I manipulate between subjects the presence of managers' decision-making responsibility such that managers have full discretion in determining employees' rewards or they only implement the rewards determined by an algorithm designed to mimic a prior manager's rewarding decisions. The main outcomes of interest are employee productivity measured by group output, employee compensation, and the residual profits.

I build the algorithm based on the managers' decision making. Specifically, data from the managers is collected first. One manager is randomly selected and assigned to each group in the algorithm condition. The algorithm then mimics the selected manager's strategies, styles, and preferences, to the extent possible, to determine a bonus allocation that the selected manager would have decided in that situation. The managers in the algorithm condition are then forced to implement the predicted bonus allocation to reward the employees in each period. To determine the algorithm, I first classify the managers into two groups: those who reward employees contingent on the performance signal and those who do not. For those who reward employees based on the performance signal, I use regression models to predict their reward decisions; for those who do not reward based on the performance, their decisions are replicated in the algorithm condition. Overall, the algorithm is designed to extrapolate the managers' decision making to the new groups to capture, to the extent possible, what the managers would have decided in rewarding the employees in that situation. Thus, the experiment design aims to hold constant the reward strategy and outcomes, while varies the presence of managers' decision-making responsibility in rewarding employees.

The results from my experiment suggest an improvement in employee productivity when the bonus is determined by algorithms versus human managers. Importantly, the higher level of employee productivity is induced without sacrificing residual profits. These results suggest that employee productivity can be motivated by changing managers' responsibility in decision making, even though the reward strategy and outcomes are held constant. Consistent with the idea that it is hard for managers to establish credibility in rewarding employees in incomplete contracts, results show that the productivity-inducing effect from algorithms is stronger when employees' rewards are not contingent on the performance signal. Thus, when managers do not reward employees based on their performance and when trust is particularly needed, it could be helpful to remove responsibilities in setting the rewards from managers to improve productivity in incomplete contracts. Taken together, the results from my experiment suggest that employee productivity can be improved when managers' responsibility in compensation decisions is removed from the decision making setting. The loss of productivity in incomplete contracts documented in prior literature (Baiman and Rajan 1995; Fisher et al. 2005) is more likely to be driven by a lack of trust in the rewarding mechanism, rather than a lack of enough rewards to compensate employees.

This study contributes to the accounting literature on subjective performance evaluation. While the extant literature does not distinguish the underlying mechanisms for a less than optimal effort provision in incomplete contracts (Baiman and Rajan 1995; Fisher et al. 2005), this study advances our understanding that the credibility of the rewarding mechanisms is of more importance than the reward itself in influencing employee effort and productivity in incomplete contracts. I also extend the literature by demonstrating that removing managers' responsibility in deciding rewards is particularly helpful when managers do not implement pay-for-performance incentives and thus when trust in the rewarding mechanism is particularly needed.

The results of this study can have important practical implications for companies' design of incomplete contracts. While intuition suggests higher productivity only follows from higher pay, results of my study inform that productivity can be motivated by increasing the credibility in the compensation process, even when employees are not paid more. Thus, this study also speaks to a broader set of settings where managers are subject to credibility concerns in rewarding employees, even when they are not residual claimants. For instance, prior research suggests that managers use discretion for their own benefit when allocating a fixed bonus pool among employees (Bol 2008). Results from this study indicate that in those settings where resource allocations are influenced by conflicts of interests, procedures aimed at increasing the transparency of the decision process could be explored as a more effective way to improve the incentive effect of incomplete contracts.

This research should also be of interest to companies that are increasingly bringing artificial intelligence in their decision-making. Despite the rising importance of algorithmic decision-making in management control systems, not much is known about when and how the new technology can be helpful in inducing employee effort. As results from this study suggest that participants may expect algorithmic decision making to be more consistent and credible, transparency and consistency of the algorithms can be explored as more important attributes of algorithmic schemes.

CHAPTER 2. BACKGROUND AND HYPOTHESIS DEVELOPMENT

2.1 Background

Incentive compensation is typically characterized by incomplete contracts, where managerial discretion allows managers to incorporate private non-verifiable information into performance evaluations. Objective performance measures are often subject to uncontrollable random events, which reduces the incentive effect of compensation contracts (Bol 2008). The use of managerial discretion can filter out uncontrollable events, mitigate effort distortion, and thus allow better alignment of incentives (Baker et al. 1988; Banker and Datar 1989; Baker et al. 1994). The benefits of managerial discretion suggested in the literature are consistent with the wide use of discretion in performance evaluation decisions in practice (Gibbs et al. 2004; Hales and Williamson 2010).

Managerial discretion, however, comes with costs. While firms provide employees implicit incentives regarding how their efforts will be rewarded, employees must trust the managers to reward their efforts fairly (Fisher et al. 2005; Bol 2008; Hales and Williamson 2010). However, the unverifiable nature inherent in managerial discretion can provide opportunities for managers to act on their own incentives, potentially in a way that would reduce the effectiveness of incomplete contracts (Bol 2011). Prior research has documented managerial opportunism as one of the most pronounced problems with managerial discretion in incomplete contracts (Bol 2008). Potential solutions such as eliminating discretion or setting a fixed bonus pool based on contractible information may not always be optimal. Because first, objective contractible measures are often subject to uncertainties; second, a fixed bonus pool introduces interdependencies and sets up zero-sum games among employees, which imposes additional compensation risk and potentially hurts cooperation (Rajan and Reichelstein 2006). Thus, both solutions entail additional agency costs (Bol 2008). The problem remains with regard to how to motivate productivity when compensation decisions are subject to managerial discretion and conflicts of interests.

While a substantial body of research has investigated managerial discretion in incomplete contracts (e.g., Baiman and Rajan 1995; Fisher et al. 2005), the underlying mechanisms that can potentially cause a loss of productivity is not well understood. On one hand, economics theory indicates a lack of effort provision can be driven by a lack of employees' trust in the rewarding mechanism that their efforts will be rewarded fairly. That is, employees may not trust that managers would implement pay-for-performance incentives in discretionary bonus allocations. On the other hand, prior work in behavioral economics on gift exchange and reciprocity norms (e.g. Charness and Haruvy 2002; Falk 2007) suggests that employees will reciprocate managers for prior reward outcomes, so a loss of productivity can be driven by a lack of enough rewards. Thus, a lack of effort provision can be driven by a low compensation or a lack of credibility in the rewarding mechanism, regardless of the reward itself. Distinguishing the underlying mechanisms is important, as it speaks to the different remedies to improve employee productivity. If productivity is driven solely by rewards, then managers simply need to pay employees more. If productivity is more driven by employees' trust in the rewarding mechanism, then ways to improve the transparency of the rewarding process can be more effective to improve productivity in incomplete contracts. In this study, I use algorithmic decision-making to remove managers' decision-making responsibility in setting the rewards, while controlling the reward outcomes, to revisit the underlying mechanisms regulating employee effort provision in incomplete contracts.

2.2 Basic Setting

I examine a setting where employees engage in a production task and interact in groups with one manager and two employees for a finite number of periods.[2] The basic setting is similar to the one studied in Baiman and Rajan (1995), Fisher et al. (2005), and Maas et al. (2012). The aggregate performance measure, i.e., the summed total of each employee's output, is subject to an outside random noise, and thus is a noisy signal of employees' performance. The discretionary bonus pool is funded based on the performance signal. Managers are granted discretion in splitting the bonus pool between themselves and the two employees. The employees' total compensation pool is then split equally between the two employees. Managers are informed of the noisy performance signal, without knowing the random noise separately, and claim the residual profits after subtracting the compensation paid to the employees. Last, employees independently choose to allocate the time between work and leisure and earn payoffs from both activities.

The payoff function is designed such that employees have strong incentives to shirk, i.e., they earn a higher incremental payoff from leisure than work. Managers also have strong incentives to act opportunistically as they can use the discretion as desired in

[2] An alternative design choice is to use an infinitely repeated game with a random continuation rule. Whereas I expect the reputation effect in infinitely repeated games to curb opportunistic behavior in the human manager condition, I do not expect reputation to affect employee behavior to the same extent in the algorithm condition, as reputation would not be expected to affect algorithmic decision making. Thus, the design choice of a repeated game with a finite number of periods is suitable for the current research setting.

determining employees' payoff. Also, the noise component in the performance signal provides room for managers to engage in motivated reasoning and act on their own incentives. In this setting, backward induction in a finite number of periods will imply a Nash equilibrium of zero effort from employees and zero output for the firm. Deviation from Nash equilibrium will require cooperative norms between the employee, his co-worker, and the manager. Indeed, empirical results in behavioral economics suggest cases where social norms of trust and reciprocity are maintained such that people choose to cooperate rather than end up in equilibrium outcomes in prisoner's dilemma and public goods games (Fehr and Gachter 2000; Camerer 2003). Thus, prior research leads to the expectation of a greater than zero effort provision based on cooperative norms. Sustaining effort and output in this environment, however, requires employees to trust the manager to compensate them fairly for their collective efforts. Thus, this setting provides a clear equilibrium benchmark, while in the meantime, allows me to observe variation in employee productivity originating from changes in the presence of managers' responsibility in rewarding employees.

2.3 Hypotheses Development

In early rounds of the game, I expect cooperative norms to sustain a greater than zero output in the human manager condition, where employees trust their co-workers will not free ride and managers compensate employees appropriately. On the other hand, in the algorithm condition, employees might tentatively explore the rewarding rules of the algorithm at the beginning. Thus, I expect differences in employee productivity to be more likely to manifest in later rounds of the game.

In a repeated game, employees will update their expectations for future payoffs based on the outcomes in the current and previous rounds. Also, as managers' payoffs depend on employees' productivity, employees might reciprocate the managers by changing their effort provision for prior reward outcomes. Thus, employees' effort provision is driven by their updated beliefs about how they would be rewarded in the future periods and their intention to reciprocate the managers for prior reward outcomes. I argue that removing managers' responsibility in decision making can affect both mechanisms that influence productivity. First, moving from managers to algorithms, employees are less likely to attribute the reward outcomes to the managers, and thus less likely to reciprocate the managers for prior rewards. Second, employees might view algorithmic decisions as more consistent in rewarding their efforts, which influences their expectations of future reward prospects. In this section, I draw on related research in prior literature and discuss two mechanisms that could contribute to any different reactions to human versus algorithm-based bonus schemes.

2.3.1 Attribution of Reward Outcomes in Incomplete Contracts

Prior social psychology research on individual decision-making in social allocation settings indicates that people not only value fair outcomes, but also how the outcomes come to be (Ross and Fletcher 1985; Blount 1995). Relatedly, attribution theory describes how people make causal interpretations for events in their environment (Kelley 1967; Crittenden 1983). In a series of experiments, Blount (1995) shows that subjects are more likely to accept unfavorable offers if they come from a non-human random device than if they are chosen by a human participant. The underlying reasoning is that when outcomes are perceived as attributable to actions of interested parties, as compared to natural occurrences, there is a stronger tendency for people to assign blame and a stronger preference for fair outcomes (Camerer and Thaler 1995; Charness and Rabin 2002). On the contrary, when outcomes are attributed to non-human causes, people are less likely to perceive the causes as intentional and are less concerned with comparative payoffs (Blount 1995).

Moving from managers to algorithmic bonus determinations, the discretion is removed from managers' decision-making, which removes the social norms related to attribution and reciprocity. In bonus allocation settings where managers claim the residual profits, employees have a strong tendency to attribute reward outcomes to managers' intentions and self-interests. As a result, discretionary bonus implemented by managers can elicit a strong reaction from employees, where employees negatively (positively) reciprocate managers for the unfavorable (favorable) rewards. On the contrary, when algorithms are used to allocate the rewards, employees are less likely to infer intentionality of the managers, and thus less likely to reciprocate managers for the reward outcomes.

Further, attribution theory suggests that people exhibit negativity bias in the attribution of outcomes. People have a stronger tendency to infer causes when events are negative than positive or neutral events (Morewedge 2009). When employees receive favorable rewards, their motivation to infer causes is not strong. However, when a less than fair share of rewards is allocated, employees will react more negatively when the rewards come from a manager's decision than an algorithmic determination. Thus, I expect differences between managers and algorithms to be smaller when employees receive favorable rewards. This indicates the negative impact on productivity of low rewards from managers' decision making is likely to dominate the positive impact of high rewards.

Therefore, based on attribution theory and negativity bias, I propose that bonuses determined by algorithms will induce higher productivity than those determined by managers.

2.3.2 Credibility Concerns in Incomplete Contracts

Prior literature has recognized managerial opportunism as one of the most pronounced problems with managerial discretion in incomplete contracts (Milgrom and Roberts 1987; Baker et al. 1994; Baiman and Rajan 1995). In particular, when managers are residual claimants, they have strong incentives to under-evaluate employees' performance and reduce the compensation costs. As managers' self-interest reduces employees' trust that their efforts will be rewarded fairly, the credibility concerns can create additional agency costs in incomplete contracts.

While it is hard for managers to establish credibility in rewarding employees in incomplete contracts absent any enforcing mechanism, removing the responsibility in managers' decision making can potentially alleviate the credibility concerns. Theory on source credibility suggests that action sources that do not profit from the decision outcomes can be judged as more trustworthy (Hovland and Weiss 1951; Kelman and Hovland 1953). Compared to managers, algorithms that do not have discretion in decision making are more likely to be viewed as separated from the financial interests in the bonus outcomes, which can lead to perceptions of procedural fairness and trustworthiness (McGarry and Hendrick 1974; Leventhal 1980).

Further, recent research on algorithm appreciation has documented perceptions of algorithmic decision-making as more consistent, with fewer biases than human decision-makers (Dietvorst et al. 2015; Lee 2018). Algorithms are perceived as more capable in problems featuring external standards of accuracy (Dijkstra et al. 1998; Logg et al. 2019). This perception can lead to higher perceived credibility and consistency with algorithmic decision-making (Ryan and Ployhart 2000). Thus, when employees receive rewards that compensate for their performance, i.e., incentives that pay for performance, they are more likely to perceive consistency in the compensation promises in algorithms than human manager. So prior research on source credibility and algorithm perceptions suggests that employees are likely to perceive higher credibility and consistency with algorithms than human managers. This improved credibility for rewarding employees' performance can lead to higher productivity, even though algorithms replicate the rewarding decisions of the managers.

Overall, both mechanisms – one based on attribution theory and one based on credibility concerns – would suggest a directional prediction that algorithm-based bonus schemes will improve employee productivity compared to those determined by managers. Thus, I make the following directional prediction:

H1: *Algorithm-based bonus schemes will generate higher employee productivity than the bonus schemes determined by human managers.*

There may also be some countervailing forces that lead to the opposite direction of prediction. The algorithm-based decision-making in this study is very different from the operationalization of the non-human factors in prior social psychology literature. Prior studies operationalize the non-human action as a random number generator or a roulette wheel (Blount 1995; Camerer and Thaler 1995). The perception of randomness in earlier studies may not be present in the current setting. Algorithms built on historical data of human behavior can be perceived as a collection of human behavior rather than random non-human factors. Further, studies on algorithm aversion suggest that people distrust algorithmic output for subjective judgments (e.g., Yeomans et al. 2018). While algorithms can be perceived as more consistent than human managers (e.g., Lee 2018), people might also be averse to algorithms in subjective performance evaluations. These alternative perceptions will bias against finding the predicted results for overall productivity.

2.3.3 Sensitivity to Reward Outcomes

While both the two mechanisms suggest the same directional prediction, they differ in their predictions of employees' sensitivity to reward outcomes. Attribution theory and reciprocity norms indicate that individuals will react more strongly to reward outcomes in human than algorithm condition; whereas the enhanced credibility with algorithms suggest that employees might expect rewards to be more consistent and react more strongly to reward outcomes in algorithm than human condition. Thus, the two mechanisms predict different patterns of results in terms of the sensitivity of employee productivity to prior reward outcomes. The attribution mechanism indicates a steeper slope of productivity to prior reward outcomes in human than algorithm condition; whereas the credibility mechanism suggests a steeper slope in the algorithm condition. I do not make ex ante prediction for which mechanism will dominate, but distinguish the mechanisms through the test of the pattern of result. Overall, I investigate the following research question:

Research Question: *Will employee productivity be more sensitive to prior reward outcomes (high vs. low rewards paid to employees) in the human manager or the algorithm condition?*

CHAPTER 3. METHOD

3.1 Participants

I recruited 210 students from a large university in the southeastern United States to participate in one of the 13 sessions of a compensated online experiment. Participants received an average payment of $9 for approximately 45 minutes of participation in the study. Seventy-four percent of the participants are female. The average age of the participants is 20.8 years, and the average work experience is 1.76 years. Participants first joined a zoom meeting where the fundamental rules for the experiment were announced before they started the experiment. The online meeting before the experiment significantly reduced dropouts that are typical in online interactive experiments.[3] When the experiment started, all interactions took place via the computerized o-Tree program.

3.2 Experimental Task and Procedure

In both conditions, participants are randomly assigned to work in three-person groups composed of one manager and two employees (referred to as Manager, Worker A and Worker B in the experiment). Though my study focuses on the conflict of interests between managers and employees, I use two employees in this setting for several reasons. First, the group setting better reflects real-world scenarios where employees work in teams. Second, the free-rider problem associated with team work represents a more challenging incentive problem, where algorithms may be more beneficial to reinforce effort. Last, the two-employee setting allows me to draw comparison with results from earlier studies on

[3] One participant dropped out, causing that group of participants to not complete the experiment.

managerial discretion (e.g., Fisher et al. 2005). Further, it is worth noting that managers are present in both human manager and algorithm conditions. In doing so, my study removes managers' responsibility in the algorithm condition by forcing managers to use an algorithm to allocate the bonuses. This design choice effectively removes the decision-making responsibilities from the managers when allocating the bonuses.

Participants interact for 8 periods and remain in the same role and the same group throughout the experiment. Participants read instructions explaining the real-effort task, the rules in determining their payoff, and how the discretionary bonus pool will be allocated between managers and employees. To operationalize a real-effort task, participants work on a decoding task where they decode two-digit numbers into letters. In each period, employees have 2 minutes to earn compensation and are granted discretion to allocate the time to either work on the real-effort task or to consume as leisure. If they choose to consume as leisure, they will browse the internet as a leisure activity and earn compensation per second spent as leisure time. If they choose to work, their compensation will depend on the efforts of both workers and how the workers are rewarded for their work. To represent a situation where the objective performance measure is affected by uncontrollable random events, group output on the real-effort task is adjusted by a noise randomly chosen from a uniform distribution of -4 to +4. In each period, a discretionary bonus pool is funded based on the group output, i.e., the sum of the two workers' output and the random noise. Both employees' and managers' compensation depends on the size of the discretionary bonus pool. Employees' payoff is composed of their leisure payoff and their share of the discretionary bonus pool. Managers' payoff consists of their base pay and the remaining bonus pool after subtracting employees' compensation. Each player's compensation in each period takes the following form:

Employee's payoff = leisure payoff + ½ of total employee compensation

Manager's payoff = base pay + total discretionary bonus pool – total employee compensation

where:

Employee's leisure payoff = 0.5 tokens per second spent as leisure time;

*Total discretionary bonus pool = 5.4 * total group output on the real-effort task;*

Total group output = output from Worker A + output from Worker B + random noise;

Manager's base pay = 10 tokens per period;

Random noise = a randomly chosen integer from -4 to +4 using a uniform distribution.

In each period, a participant acting in the role of manager (human manager condition) or a computer algorithm (algorithm condition) splits the discretionary bonus pool between the manager and the employees in each group. The managers or the algorithm determines the percentage of the bonus pool allocated as total employee compensation, which is then split equally between the two employees. The managers claim the residual profits from the total discretionary bonus pool.

The experimental parameters are determined such that employees are provided incentives to shirk: the payoff per second from leisure exceeds the payoff one would earn if he/she spends one second on work and if the bonus pool is split equally among the three players. Based on a pilot test, participants on average decoded 0.21 items per second. Assuming the manager equally splits the discretionary bonus pool (i.e., manager receives 1/3 and each employee receives 1/3 of the bonus pool), each employee will earn an average compensation of 0.378 tokens per second spent on work (1/3 * 0.21 per second * 5.4 tokens per unit of output). Compared to the leisure payoff of 0.5 tokens per second, the marginal payoff from work represents a 0.75 marginal per capita return, which is commonly used in prior studies on public goods games (Ledyard 1995).

After each working period, employees learn their own output on the decoding task. Both managers and employees learn the amount of group output after noise (i.e., the noisy performance signal) and total discretionary bonus pool. However, the random noise is not disclosed to any of the players. This design choice allows room for both managers and employees to engage in self-serving reasoning in interpreting the bonus outcomes. After managers or the computer algorithm allocates the discretionary bonus pool, all players are informed of their own total payoff and the share of the bonus pool allocated to each player. Participants are also informed of their cumulative output and cumulative payoff at the end of each period.

Participants complete a short quiz after they read the instructions to ensure they understand the rules of the game and the mechanism in determining their payoff. Explanation for the instruction is provided if the participants answer the question incorrect, so that they need to answer all the questions correct before they can proceed. Next, all participants complete a hypothetical example where they play each of the three roles in turn to understand the decision and task of each player. All participants also complete a practice round of the decoding task that allows them to get familiar with the computer interface and alleviates any learning effect in the real game. Participants' roles are announced before they start the real game. After completing all 8 periods, participants complete a questionnaire that elicits their expectation of the distribution of payoff, their explanation for the bonus outcomes (open-ended), fairness perceptions, and attribution of their bonus outcomes.

3.3 Experimental Design

The experiment uses a 1 x 2 (manager responsibility: present versus absent) x 8 (period, within-subject) design. Data from the human manager condition is collected first, which serves as the database for the algorithm condition. In the human manager condition, in each three-person group, a participant acting in the role of manager determines the bonus allocation between himself/herself and the employees. In the algorithm condition, a human manager must follow the allocation indicated by the algorithm when splitting the bonus pool in each period. To determine the output of the algorithm, a manager in the historical database is randomly selected and assigned to each group.[4] The computer algorithm mimics the rewarding strategy of the assigned manager to capture what that manager would have decided for the bonus allocation for the group, and the manager in the group must use the predicted allocation to split the bonus pool in each period.

[4] The experiment adopts a semi-yoked design in the sense that the algorithm will ensure different managers assigned to each group in the algorithm condition such that there will not be an oversampling of any particular manager.

3.3.1 Determination of Algorithm-generated Bonus Schemes

To determine the algorithm used to predict each manager's rewarding decisions, I use a regression model for each individual manager that is based on their bonus allocation decision and the performance signal they observe in each period. The model includes performance signal (group output after random noise) and time effect (dummy variables for the last and first two periods) as explanatory variables to explain and predict each manager's rewarding behavior.[5] As percentage measure is used as the dependent variable, I use GLM with the logit link function as the estimation model. Regression models represent a commonly used and simplified method in machine learning for prediction of behavior patterns. This implementation is consistent with the methods used in prior algorithm aversion literature (e.g., Logg et al. 2019). Additionally, according to prior survey results, there is a high consensus among the surveyed participants about the definition of algorithms. Participants are familiar with algorithms as "a set of equations to find an answer" or "a procedure for computing a function" (Rogers 1987). Thus, the way I operationalize the algorithm-based bonus schemes is consistent with the general understanding of the concept of algorithm.

To ensure the algorithm captures individual manager's strategies, styles, and preferences, to the extent possible, managers are classified into two groups based on the regression results: those who reward employees contingent on the performance signal (with significant coefficient on the group output after controlling for time effects) and those who

[5] The performance signal (group output after random noise) is the only information that the managers observe before they make their decisions in each period, so I use it as the only observable variable to explain their decisions.

do not (with insignificant coefficient on the group output after controlling for time effects).[6] The algorithm is determined separately for the two groups of managers. First, for the managers who do not reward based on performance, the algorithm replicates the decisions of those managers, as observable data cannot provide a clear base to predict their rewarding behavior. Next, for the managers who reward employees contingent on the performance signal, the algorithm uses regression model predictions as the bonus allocations for the group. Further, to reflect the randomness in human decision making, the algorithm randomly draws one value within the 95% confidence interval of the predicted allocation, constrained to a boundary of zero and one. This design choice incorporates the randomness in human decision making into the algorithmic predictions, while ensuring the noise is within a reasonable bound. As the decisions are replicated for those managers who do not reward contingent on performance, it ensures that the algorithm does not add additional noise to the decisions of those managers who reward based on their own styles and preferences. Appendix A provides specific scheme of the algorithm to use for each category of managers; Appendix B provides example algorithmic determination of bonus outcomes for a typical manager from each of the two groups. Overall, the algorithm is designed to extrapolate managers' decision making to the new group to capture, to the extent possible, what the manager would have decided in rewarding the employees for that group.

[6] For those managers who reward employees contingent on performance but the contingency overlaps with a time effect in their behavior (coefficient on the group output turns insignificant after controlling for time effects), they are also classified as non-performance-contingent managers, as their behavior can be captured by a time trend.

Finally, the way the algorithm determines the bonus outcomes is transparently communicated to the participants in concise and understandable language. Their understanding of the algorithm is also tested in the quiz before the real-stage game to ensure they understand the attributes of the algorithm in allocating the bonus pool (see Appendix C for the language used in explaining the allocation principle in the human and algorithm conditions).[7]

[7] The experiment instructions provide more explanations for bonus allocations in the algorithm condition compared to the human manager condition. While it is a necessary design choice of my study to ensure that participants understand the algorithm, I acknowledge that it may potentially increase the credibility of the algorithm. Future research can further investigate algorithms with different degrees of transparency.

CHAPTER 4. RESULTS

I examine how changing the presence of managers' responsibility can affect employee productivity (measured by group output without random noise). I also investigate two other outcomes of interest: employee compensation and manager residual profit. Table 1 presents descriptives and ANOVA results with managers' responsibility (present versus absent) as the between-subject factor and period as the within-subject factor for each outcome measure. Also, Figure 1 displays the graphical results for each outcome measure across all eight periods. The data analyses are based on the participants that completed the study.[8]

4.1 Tests for Employee Productivity

Table 1A presents descriptive statistics and test results for employee productivity over all periods. Hypothesis 1 predicts that employee productivity will be higher when the rewards are implemented by algorithms than managers. Descriptive statistics and ANOVA results presented in Table 1A report a significant main effect of source of allocation on employee productivity (39.81 vs. 36.97, p = 0.038). As depicted in Panel A of Figure 1, difference in employee productivity between human and algorithm condition manifests in the later stage of the game (largest difference shows in period 5, t_{68} = 1.31, one-tailed p = 0.098, untabulated). While results in the human condition are generally consistent with results in Fisher et al. (2005) that employee productivity deteriorates over time, employee

[8] Only one participant did not complete the study. Untabulated tests indicate no statistically significant difference in gender, age, and work experience across the human and algorithm conditions, suggesting successful randomization of participants.

overall productivity is maintained especially in the later stages in the algorithm condition. Thus, algorithms can potentially mitigate the loss of productivity when responsibility in setting the rewards is removed from managers' decision making. As the algorithm simply mimics managers' rewarding behavior and employees are not compensated with a higher level of rewards, these results suggest that a less than optimal effort provision in incomplete contracts is more likely driven by a lack of trust in managers to pay employees for their performance rather than the level of rewards.

--- Insert Table 1 and Figure 1 about here ---

4.2 Tests for Employee Compensation and Manager Residual Profit

Besides employee productivity, I also examine the payoff of each players in the group. Table 1B and Panel B of Figure 1 present descriptives and graphical representation for employees' group work compensation (the amount of discretionary bonus pool received by employees). The results for employee compensation mirrors the results of employee productivity. Results in Table 1B indicate that a higher productivity translates into a higher work compensation for employees in the algorithm than human condition (131.73 vs. 116.62, $p = 0.015$). Further, as depicted in Panel B of Figure 1, employee compensation in the algorithm condition is maintained at a relatively high level in the later stage of the game (with significant difference between human and algorithm condition in period 5, $t_{68} = 2.04$, two-tailed $p = 0.045$ and period 8, $t_{68} = 1.96$, two-tailed $p = 0.054$, untabulated).

As managers claim the residual profits, any additional compensation paid to the employees reduces managers' own wealth. Thus, it is hard to predict *ex ante* whether manager's payoff will be improved by changing the presence of managers' responsibility in decision-making. Table 1C and Panel C of Figure 1 provide empirical results for managers' residual profits. As shown in Table 1C, managers' residual profits are not significantly different between the algorithm and human condition (84.10 vs. 83.63, p = 0.916). Further, results depicted in Panel C of Figure 1 do not suggest significant discrepancy in managers' compensation over time between the two conditions (all p values > 0.37 in each round). Thus, while employee productivity is improved with algorithmic allocations, managers' residual profits are kept at a similar level across the two conditions.

It is noteworthy that employees' higher payoff cannot be attributed to a more generous bonus scheme. Results in Table 1D suggest that employees' work compensation per unit of output is not significantly different between the two conditions (3.16 vs. 3.05, p = 0.210). Further, Panel D of Figure 1 depicts that employees' payoff per unit of output is maintained at a relatively stable level and does not differ between the two conditions over time (all p values > 0.40 in each round). Thus, after adjustment for the output, results on employee compensation suggest that employees are not paid a more generous reward scheme. However, their productivity and compensation outcomes are improved when the discretion in bonus allocation is delegated to the algorithms.

4.3 Tests for Percentage of Bonus Allocated to Employees

To identify the mechanism driving employees' effort provision in incomplete contracts, it is important that the difference in productivity is not driven by any difference in the reward outcomes. I compare the percentage of bonus allocated to the employees between the two conditions. Table 1E and Panel E of Figure 1 present the results for the average percentage of the bonus pool allocated to the employees across all periods and

Panel A of Figure 2 presents the distribution of allocation percentages in the human and algorithm conditions. Results suggest that the percentage of bonus allocated to the employees is not statistically different across the algorithm and human condition (0.59 vs. 0.57, p = 0.205). Further, the overall distributions of the allocation percentages do not suggest any systematic difference in the allocation outcomes between the two conditions (T-test of difference: t_{558} = 1.27, two-tailed p = 0.206; Kolmogorov–Smirnov test of distribution: p = 0.264). Thus, the productivity difference is less likely to be driven by differences in the distribution of reward outcomes between the two conditions.

--- Insert Figure 2 about here ---

Overall, results from my experiment suggest that employee productivity is improved once the decision making responsibilities is removed from the managers, even when the algorithm replicates the rewarding strategies, styles, and preferences of managers' decision making and when employees are not compensated with a more generous bonus scheme. Further, results from the distribution of allocation outcomes alleviate concerns for any differences in the bonus outcomes driving the results.

4.4 Effects of Performance-based Rewards Under Human vs. Algorithm

As managers are classified into those who reward employees contingent on the performance signal and those who do not, I investigate the effect of source of allocation and performance-based rewards on the outcome measures. Table 2 and Figure 3 summarize the descriptive statistics and test results for employee productivity based on whether the rewards are contingent on the performance signal. ANOVA results in Table 2 show a significant interaction effect between source of allocation and performance-contingent rewards on employee productivity ($p = 0.063$). The results indicate that employees respond positively in terms of a higher output when managers reward them based on their performance (41.86 vs. 36.64, $p = 0.000$). Follow-up simple effects tests suggest that algorithms are particularly helpful when rewards do not reflect employee performance ($p = 0.004$). That is, when managers reward employees based on their own strategies, preferences, or styles, algorithms can be particularly helpful in establishing the credibility in rewarding employees for their performance. Panel B of Figure 3 and untabulated ANOVA results on employee compensation also indicate that when rewards are not contingent on performance, employees can be better off when the same rewards are allocated by algorithms ($F_{1,528} = 6.21$, $p = 0.013$, untabulated). Panel C of Figure 3 and untabulated ANOVA results on managers' residual profit show a significant interaction effect between source of allocation and performance-contingent rewards ($F_{1,528} = 5.32$, $p = 0.022$, untabulated). Thus, managers can be better off when they allocate a performance-contingent reward; whereas when rewards do not reflect employees' performance, algorithms that mimic the managers' styles and preferences can improve managers' payoff.

--- Insert Table 2 and Figure 3 about here ---

4.5 Additional Analyses

4.5.1 Tests for Participants' Expected Bonus Allocations

Results from the main analyses are generally consistent with the idea that employees perceive the bonuses to be more consistent and based on performance when it is implemented by algorithms, even when the actual reward outcomes do not differ under algorithmic decisions. As a result, it would be interesting to investigate whether removing managers' responsibility also changes individual's expectations of bonus outcomes. In the ex post questionnaire, participants answered a series of questions eliciting their expectations of the percentage of bonus to be allocated to the employees given certain amount of group output, if they were to replay the game with the same group again.[9] Panel B of Figure 2 presents the distribution of the expected allocation percentage in the human and algorithm conditions. Results suggest that the overall distribution of the allocation percentages do not differ significantly between the two conditions (T-test of difference: $t_{1258} = 0.81$, two-tailed $p = 0.417$; Kolmogorov–Smirnov test of distribution: $p = 0.275$).

As participants indicated their expected allocations given certain group output, Panel C and Panel D of Figure 2 also plot the actual and expected allocations by group output in both human and algorithm conditions. Regression results of actual allocation percentage on group output indicates that while the reward outcomes are performance contingent in the algorithm condition ($\beta = 0.0039$, $t_{287} = 2.67$, two-tailed $p = 0.012$), rewards overall are not contingent on performance when managers allocate the bonus pool ($\beta = 0.0039$, $t_{287} = 2.67$, two-tailed $p = 0.012$). Accordingly, participants' expectations of bonus outcomes reflect this difference: participants expect rewards to be contingent on performance in the algorithm condition ($\beta = 0.0014$, $t_{647} = 2.67$, two-tailed $p = 0.011$), whereas such expectation does not appear in the human condition ($\beta = 0.0001$, $t_{611} = 0.31$, two-tailed $p = 0.758$). Interestingly, in the subgroups of participants that receive performance-contingent rewards, participants exhibit different expectations in human versus algorithm conditions.

[9] In particular, participants answer six questions regarding their expected bonus allocations, corresponding to six different ranges of group output, i.e., 1-10, 11-20, 21-30, 31-40, 41-50, 51-60 units of group output. This range is set up based on an earlier pilot test conducted on Prolific, though participants have higher output in the current experiment. Participants indicate their expected allocation percentage by dragging a bar to any point between 0% to 100%. Results in this section are based on the responses from all participants (including those playing the role of manager).

While participants' expectations reflect the performance-contingent rewards in the algorithm condition (β = 0.003, t_{215} = 5.18, p = 0.000), participants do not expect managers to be performance contingent even when they actually reward employees based on performance (β = -0.0002, t_{215} = -0.56, p = 0.586).

Taken together, while the overall distributions of expected allocations do not differ across the two conditions, participants seem to have different interpretations of performance-based pay when they interact with human versus algorithms. Participants' expectations of reward outcomes are in line with their actual reward experiences in the algorithm condition. However, they seem not to believe that managers will compensate them based on performance, even when managers actually do so. These results are consistent with the idea that it is hard for managers to establish credibility in rewarding employees for their performance in incomplete contracts, especially when managers have conflicts of interests in implementing their discretion.

4.5.2 Tests for Sensitivity to Prior Reward Outcomes

As two mechanisms can potentially contribute to the increased productivity under algorithmic bonus schemes, I distinguish the underlying mechanisms by examining employees' sensitivity to the reward outcomes. The mechanism based on attribution suggests a higher sensitivity to prior reward outcomes when employees interact with managers, whereas the mechanism based on credibility suggests a higher sensitivity under algorithmic decisions. The descriptive statistics and test results for employee productivity based on the bonus outcomes are summarized in Table 3 and depicted in Figure 4.[10] ANOVA results reveal a significant main effect of reward outcome (p = 0.000) and a significant interaction effect of source of allocation and reward outcome (p = 0.027). Follow-up simple effects tests suggest that the algorithm outperforms managers in eliciting a higher level of employee productivity when the overall rewards are high (43.47 vs. 38.44, p = 0.004); whereas the two rewarding mechanisms induce a similar level of productivity when the overall rewards are low (34.06 vs. 35.10, p = 0.619). Thus, while generous compensation can generally induce higher productivity, employees seem to respond more positively when the generous compensation is implemented by algorithms rather than managers. Overall, while the behavioral economics literature on gift exchange speaks to a psychological tendency to reciprocate, results from my experiment suggest that trust can play an incrementally positive role in influencing employees' productivity.

More importantly, results from another pair of simple effects indicate that employees are not that sensitive to reward outcomes when managers determine the bonus (p = 0.085), but react strongly to reward outcomes when the bonus is implemented by the algorithm (p = 0.000). These results generally support the idea that employees' productivity is more sensitive to reward outcomes when the bonus is allocated by algorithms compared to managers. Collectively, the results are consistent with the idea documented in prior literature that it is hard for managers to establish credibility in rewarding employees for their efforts in incomplete contracts (e.g., Bol 2008; Hales and Williamson 2010). The credibility of the rewarding mechanism can be improved when responsibility is removed

[10] Classification of overall reward outcomes as high versus low is based on whether the average percentage of the bonus pool allocated to the employees over all periods is above or below 60%.

from managers' decision making.

I also examine employees' compensation and managers' payoff based on the bonus outcomes. Panel B of Figure 4 suggests that algorithms induce higher employee compensation only when employees are paid high rewards ($F_{1,528}$ = 7.66, p = 0.006, untabulated), but not when an overall low rewards are allocated ($F_{1,528}$ = 0.001, p = 0.973, untabulated). Thus, the higher productivity under algorithmic determinations translates into better payoff outcomes for the employees when rewards are generous. Similarly, Panel C of Figure 4 indicates that generous reward reduces manager's payoff, and this effect dominates in both human and algorithm conditions ($F_{1,528}$ = 40.42, p = 0.000 in the human condition; $F_{1,528}$ = 19.94, p = 0.000 in the algorithm condition, untabulated).

Overall, the pattern of results of employees' productivity based on reward outcomes is in line with the credibility mechanism in which algorithmic decision-making is perceived as more credible in rewarding employees for their efforts and alleviates the credibility concerns with managerial discretion in resource allocations.[11]

However, results in this section should be taken with caveat that the pattern of results is sensitive to the classification of the overall reward outcomes as high versus low. Using alternative thresholds for classifying reward outcomes yields different pattern of results of observed behavior. For instance, using a fairness benchmark of 66% of bonus pool allocated to employees as the threshold, employees' productivity is sensitive to the reward

[11] While the credibility concern can originate from the fact that participants are interacting with unknown strangers in the experiment, it is worth noting that they also do not have prior experience with the algorithm used in the experiment. Nonetheless, participants seem to pick up the trust with the algorithms compared to the managers. This suggests that it is particularly hard for employees to trust the managers when they have conflicts of interests in their decision-making.

outcomes in both the human manager and the algorithm condition ($F_{1,528}$ = 33.22, p = 0.000 in the human condition; $F_{1,528}$ = 10.65, p = 0.001 in the algorithm condition, untabulated). Using a median split of 61.625% of bonus pool allocated to the employees as the threshold shows similar pattern of results: employee productivity is sensitive to the reward outcomes in both the human manager and the algorithm condition ($F_{1,528}$ = 13.84, p = 0.000 in the human condition; $F_{1,528}$ = 14.89, p = 0.000 in the algorithm condition, untabulated). Thus, results in this section should be taken with caution. More analysis would be needed in future work to distinguish the underlying mechanisms that contribute to the higher productivity in the algorithm condition.

--- Insert Table 3 and Figure 4 about here ---

4.5.3 Tests for Perceptions of Fairness

Recent research on algorithm appreciation suggests that algorithmic decisions are perceived as more consistent, with fewer biases than human decision-makers (Dietvorst et al. 2015; Lee 2018). This perception can lead to higher perceived procedural fairness (Ryan and Ployhart 2000), which in turn can explain employees' higher productivity under algorithm-based bonus schemes. In the post questionnaire, participants indicated the extent to which they were concerned with the relative payoffs between the workers and the managers and between the worker and their coworkers. Participants also indicated whether the two workers collectively or the manager received more than a fair share of the bonus pool. In addition, participants also indicated the extent to which they attribute the reward outcomes to the outside random noise, the coworker, or the manager. Overall, I do not find significant difference between the human and algorithm condition for participants' fairness perceptions or their attribution of outcomes (all $p \geq 0.129$).

While results do not show a significant main effect of source of allocation on participants' self-reported fairness perception, I explore whether the effect depends on the reward outcomes. Table 4 and Figure 5 present results of participants' fairness perceptions by reward outcomes.[12] The results for participants' concerns of relative payoffs mirror the results for employee productivity. ANOVA results suggest a significant interaction effect ($p = 0.023$) such that concerns for relative payoffs are stronger with managers than algorithms only when rewards are high ($p = 0.017$), but not when rewards are overall low ($p = 0.323$). Further, simple effects tests suggest that both measures show a similar pattern of results as in the main analyses. Participants' concerns for relative payoffs and their perceptions for a fair share of bonus are more sensitive to reward outcomes when rewards are allocated by algorithms compared to managers (for concern of relative payoff: $p = 0.000$ vs. $p = 0.845$ for algorithm and human conditions; for perception of fair allocation: $p = 0.004$ vs. $p = 0.070$ for algorithm and human conditions). Thus, results on participants' fairness perceptions corroborate the credibility concerns suggested in the main results. However, similar to the results in the previous section on employee productivity, results in this section should also be taken with caution as the pattern of results is sensitive to the classification of the overall reward outcomes as high versus low.

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

[12] Similar to earlier tests, classification of overall reward outcomes is based on whether the average bonus allocation over all periods is above or below 60%.

CHAPTER 5. CONCLUSIONS

Incentive compensation is often characterized by incomplete contracts. In a setting where managers have discretion over the size of employee compensation pool, prior work indicates that managerial discretion reduces employee productivity. While the agency problem with managerial opportunism is well documented in prior literature, it is not clear whether it is a lack of generous rewards or a lack of trust in managers that leads to a less than optimal level of effort provision in incomplete contracts. Recent technologies have driven management control systems to be more automated, which have increasingly replaced managers' subjective judgment in performance evaluation and compensation decisions. In this study, I experimentally investigate whether replacing managers' decision making with algorithmic bonus schemes that mimic managers' decision making improves employee productivity. I find that discretionary bonus pools allocated by algorithms generate higher employee productivity without sacrificing residual profits. Further, the productivity-inducing effect from algorithms is stronger when the rewards are not contingent on the performance. These results are generally consistent with the idea that it is hard for managers to establish credibility for rewarding employees for their efforts in incomplete contracts. Employee productivity is improved once responsibility is removed from managers' compensation decisions, even when employees are not paid more. This study advances our understanding of the behavioral factors regulating employee productivity in incomplete contracts.

The results of this study have important practical implications for companies' performance evaluations in incomplete contracts. There are many situations in practice where managers act as residual claimants over resources or have conflicts of interests in resource allocations. While my study investigates a particular role of managers, results can speak to a broader set of settings where managers are subject to credibility concerns in rewarding employees, even when they are not residual claimants. For instance, prior research suggests that managers use discretion for their own benefits and preferences when allocating a fixed bonus pool among employees, and employees have similar credibility concerns in such settings (Bol 2008). Results of my study inform that in those situations where managers' decision-making suffers from conflicts of interests, procedures aimed at increasing the transparency of the evaluation process could be explored as a more effective way to improve the incentive effect of incomplete contracts.

My study also have implications for firms that are increasingly bringing artificial intelligence in their decision-making. Despite the fast-increasing importance of algorithm-based decision-making in management control systems, not much is known about employees' reactions to the new technology and the situations under which it can be helpful. This research addresses this gap by studying algorithm-based decision-making in a setting where managerial opportunism is more pronounced. Results can inform potential desirable attributes of algorithms to facilitate resource allocations in settings where it is influenced by credibility concerns and conflicts of interests. While this study takes a first step to design an algorithm that mimics managers' compensation decisions, future research can explore other characteristics of algorithms. For instance, algorithms can implement certain optimization strategies such as learning from the good managers or removing the noise in human behavior. Further, the interaction of new technology with other control mechanisms would be promising avenues to pursue. For instance, how would employees react to algorithm-based performance evaluations when they are given the rights to appeal to the decisions from algorithms. Additionally, future research can explore alternative disclosure mechanisms that effectively "open" the black box of algorithms when employees' compensation is algorithmically determined. Last, the perception of algorithms can be subject to different social backgrounds and demographics of employees. Investigating the effect of algorithmic decision making in different populations would enhance our understanding of the effect of algorithms on employee productivity and firm outcomes. Overall, the intersection of information technology and management control systems represents fruitful avenue for future research.

F I G U R E 1

Figure 1. A - Average Employee Group Productivity by Period

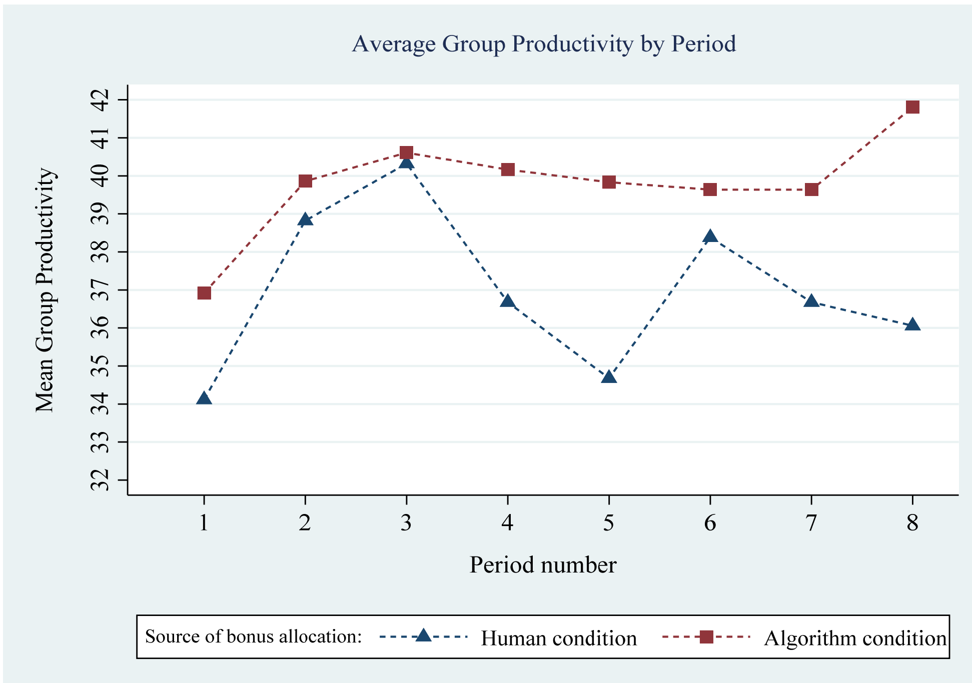

Figure 1. B - Average Employee Group Work Compensation by Period

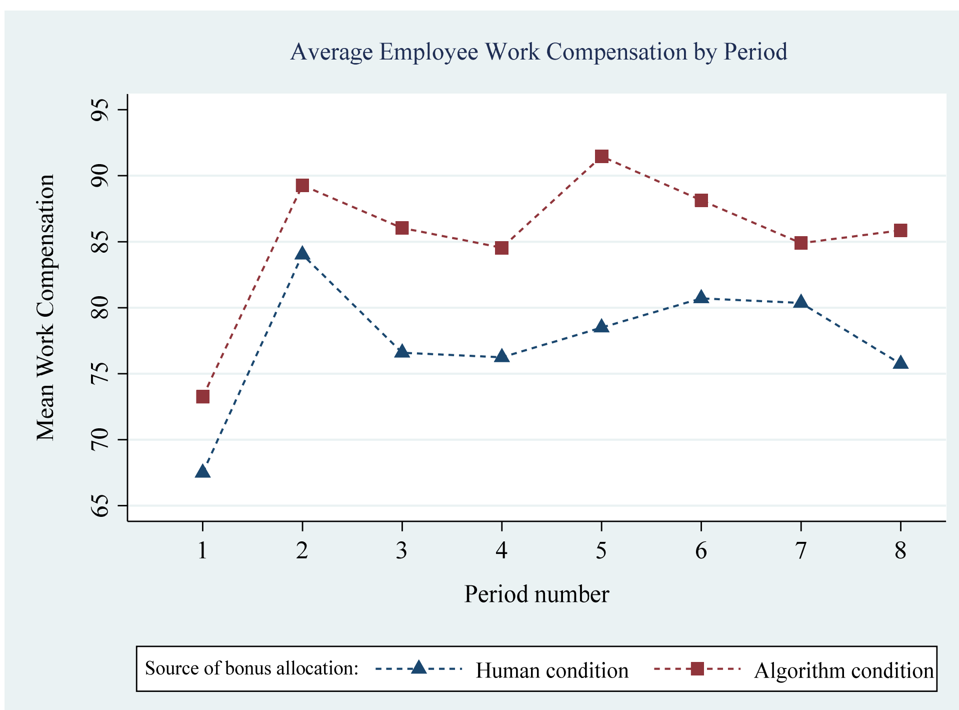

Figure 1. C - Average Manager Residual Profit by Period

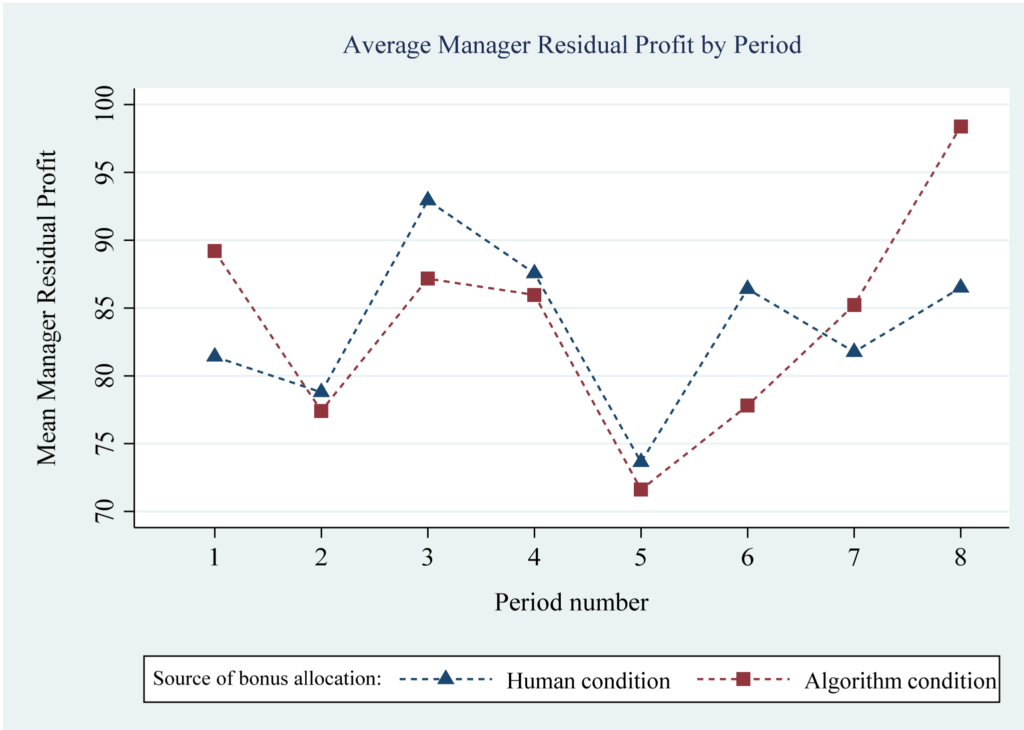

Figure 1. D - Average Employee Work Compensation Per Unit of Output by Period

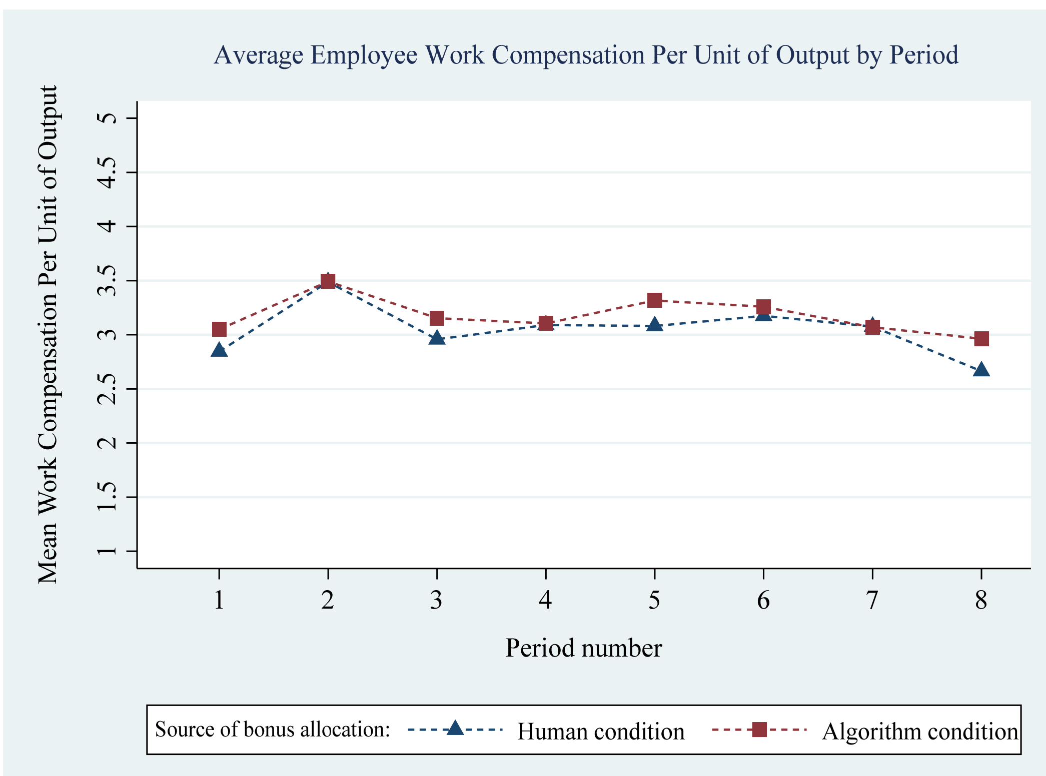

Figure 1. E - Average Allocation Percentage to Employees by Period

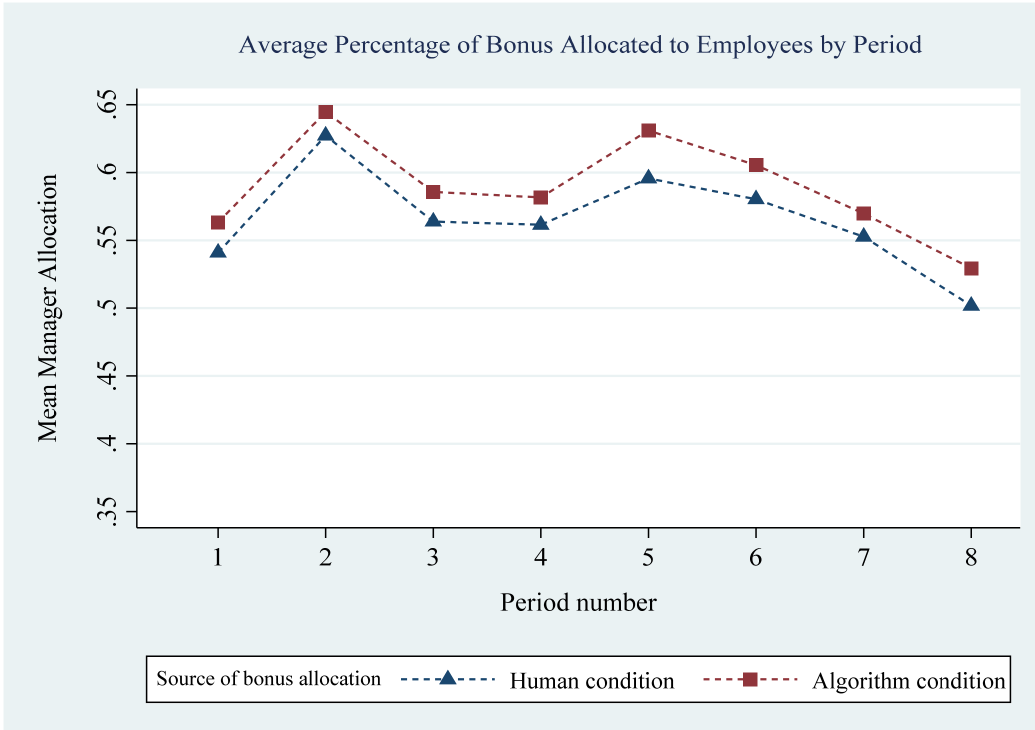

Figure 1 - Effect of Source of Allocation on Employee Productivity, Employee Work Compensation, Manager Residual Profit, and Employee Work Compensation Per Unit of Output by Period

This figure plots the average of employees' group productivity (Panel A), group compensation (Panel B), manager's residual profit (Panel C), employees' group work compensation per unit of output (Panel D), and percentage of bonus allocated to employees (Panel E) in each period. The source of allocation either comes from a human manager or from an algorithm that is designed to mimic the managers' decision making. Participants interact in three-person groups with two employees and one manager for 8 periods. Employee group productivity equals the sum of two employees' number of correctly decoded items in each group, without the random noise. Employee group work compensation equals the sum of two employees' compensation received from work in each group. Manager's residual profit equals the rest of the output bonus that has not been paid to the two employees in each group. Employee group work compensation per unit of output equals the sum of two employees' work compensation divided by the group productivity (output before noise) in each group. The percentage of bonus allocated to the two employees is determined either by the manager in the group or the computer algorithm that mimics the managers' decision making.

F I G U R E 2

Figure 2. A - Distribution of Actual Allocation Percentage in Human vs. Algorithm Condition[1]

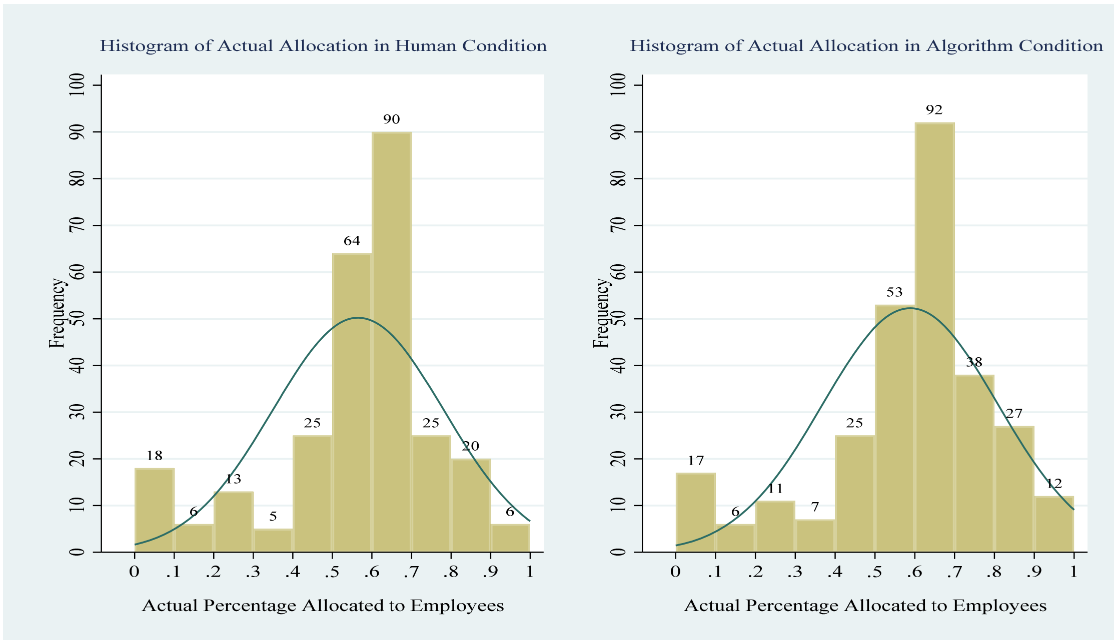

Figure 2. B - Distribution of Expected Allocation Percentage in Human vs. Algorithm Condition[2]

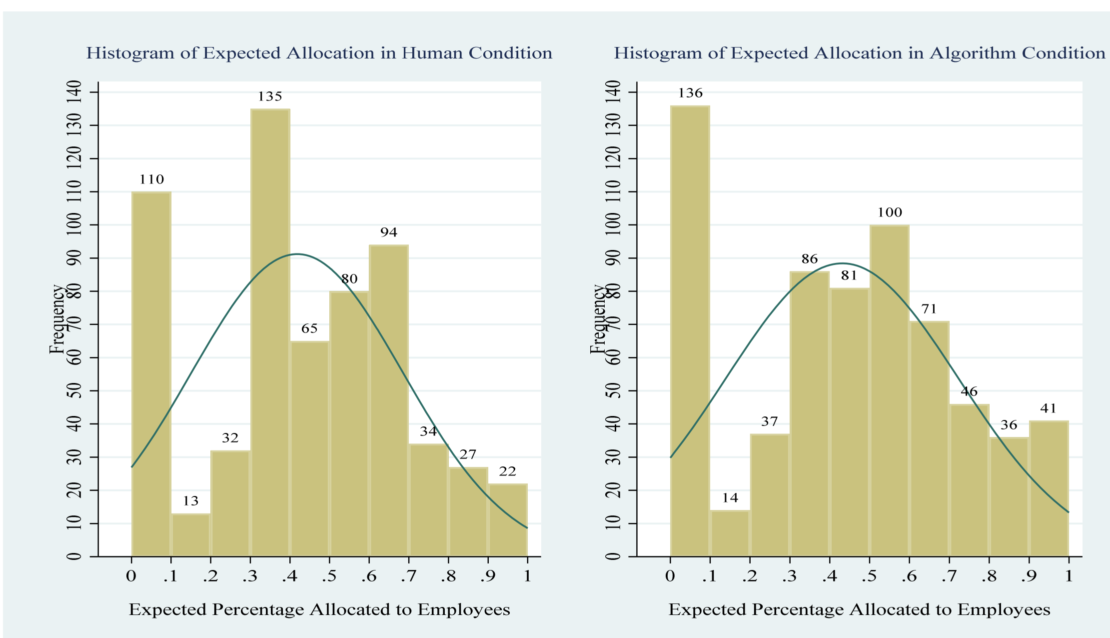

Figure 2. C - Actual Allocation Percentage by Group Output in Human vs. Algorithm Condition[3]

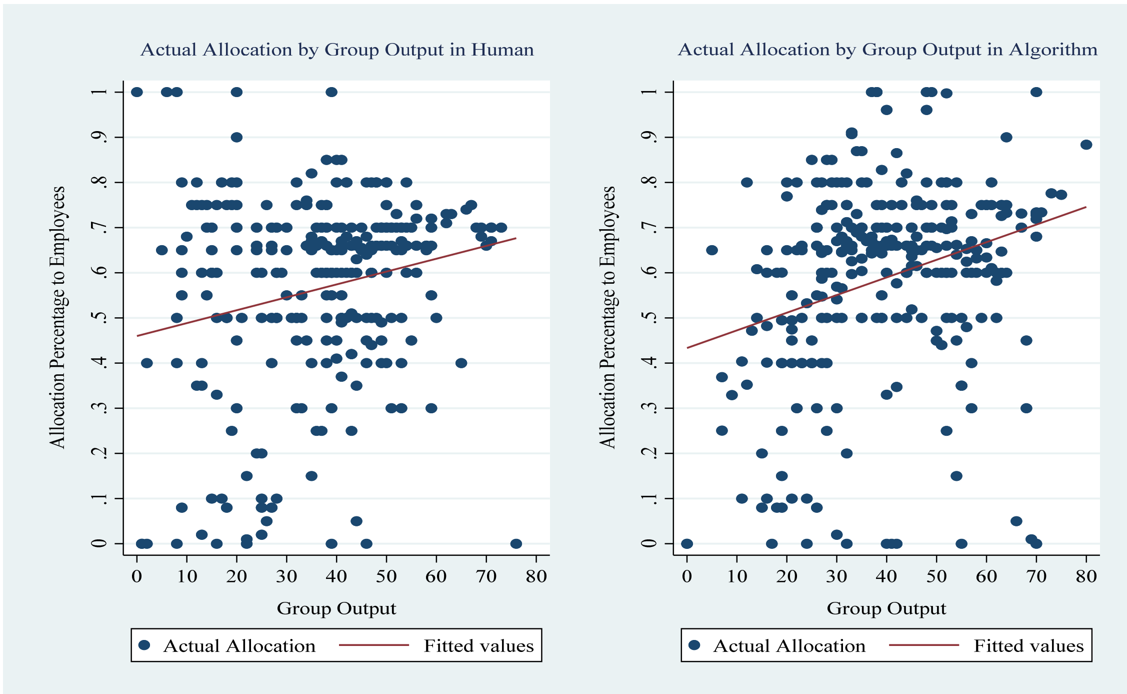

Figure 2. D - Expected Allocation Percentage by Group Output in Human vs. Algorithm Condition[4]

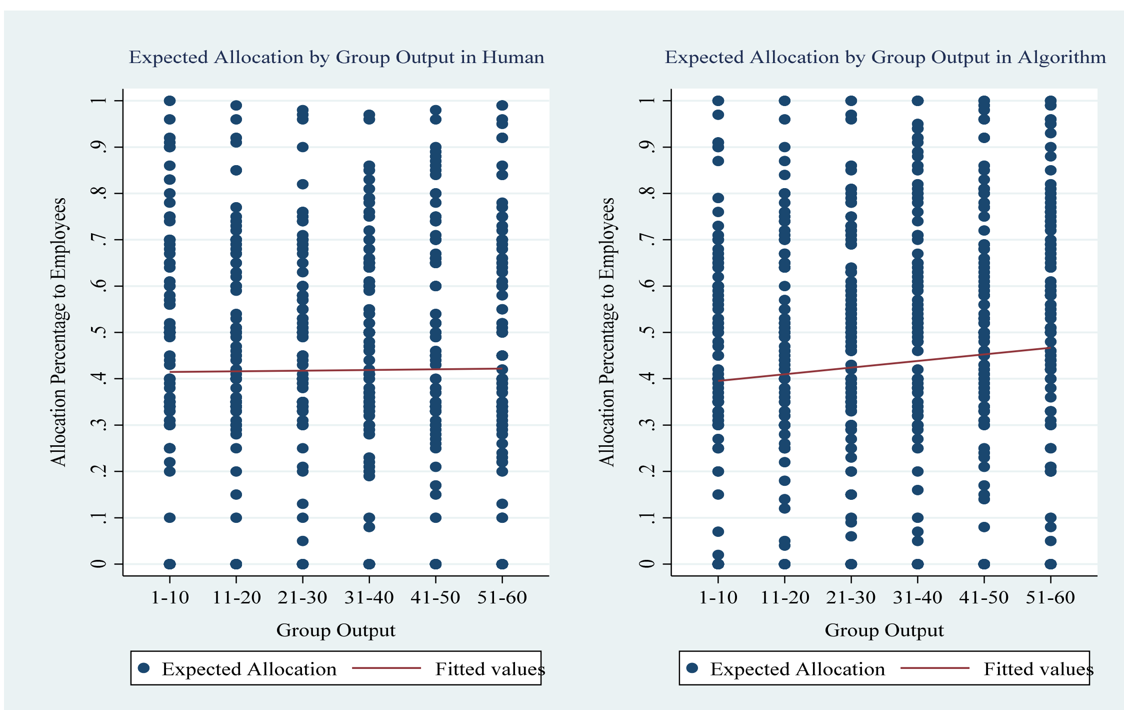

Figure 2 - Distribution of Percentage of Bonus Allocated to Employees

This figure plots the distribution of the actual percentage of bonus allocated to the employees in the human and algorithm condition (Panel A), the distribution of the expected percentage of bonus allocated to the employees in the human and algorithm condition (Panel B), the actual percentage of bonus allocated to the employees by group output in the human and algorithm condition (Panel C), and the expected percentage of bonus allocated to the employees by group output in the human and algorithm condition (Panel D). The source of allocation either comes from a human manager or from an algorithm that is designed to mimic the managers' decision making. Participants interact in three-person groups with two employees and one manager for 8 periods. The percentage of bonus allocated to the two employees is determined either by the manager in the group or the computer algorithm that mimics the managers' decision making. Group output equals the sum of two employees' number of correctly decoded items in each group, without the random noise. Expected percentage of bonus allocated to the employees is elicited in the ex post questionnaire. Participants responded to questions regarding the expected percentage of bonus to be allocated to the employees given certain group output, if they were to replay the game with the same group again. Participants answer six questions regarding their expected bonus allocation, corresponding to six levels of group output, i.e., 1-10, 11-20, 21-30, 31-40, 41-50, 51-60 units of group output. Participants indicate their expected allocation percentage by dragging a bar to any point between 0% to 100%.

[1] Test of distribution of actual allocation percentage: T-test of difference in allocation percentage between the two conditions: $p = 0.206$; Kolmogorov–Smirnov Test of distribution in allocation percentage between the two conditions: $p = 0.264$.

[2] Test of distribution of expected allocation percentage: T-test of difference in allocation percentage between the two conditions: $p = 0.417$; Kolmogorov–Smirnov Test of distribution in allocation percentage between the two conditions: $p = 0.275$.

[3] Regression results of actual allocation percentage on group output (clustered std. err. at group level): $\beta = 0.0029$, two-tailed $p = 0.106$ in the human condition; $\beta = 0.0039$, two-tailed $p = 0.012$ in the algorithm condition.

[4] Regression results of expected allocation percentage on group output (clustered std. err. at group level): $\beta = 0.0001$, two-tailed $p = 0.758$ in the human condition; $\beta = 0.0014$, two-tailed $p = 0.011$ in the algorithm condition.

F I G U R E 3

Figure 3. A - Average Employee Group Productivity over All Periods

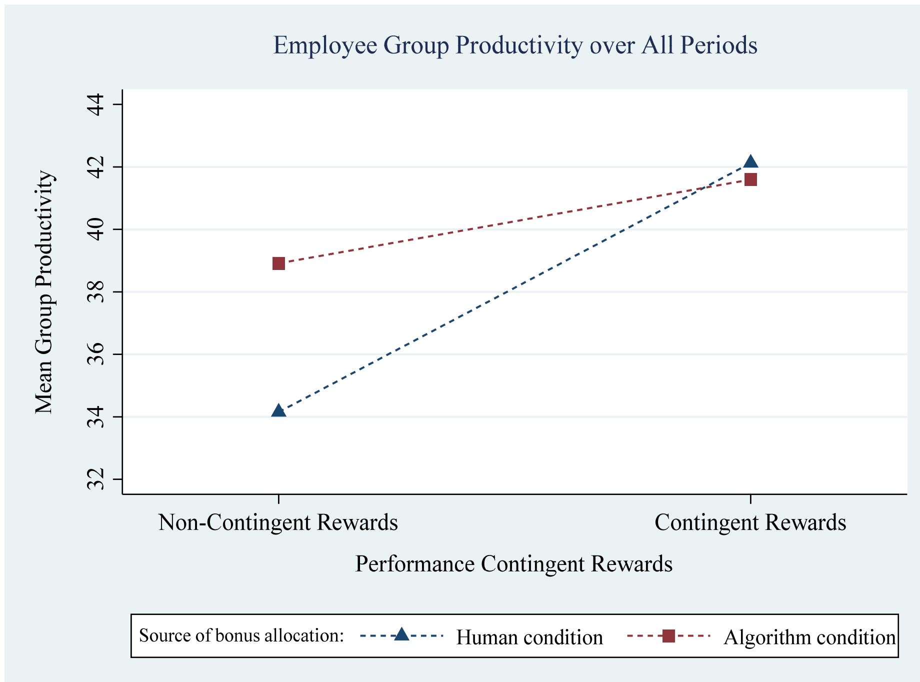

Figure 3. B - Average Employee Group Work Compensation over All Periods

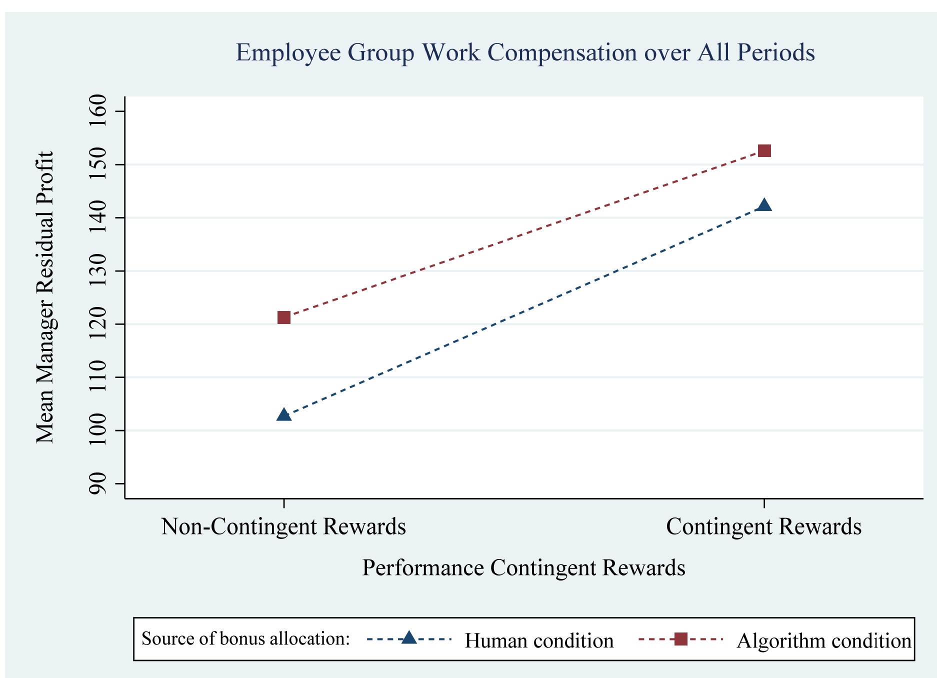

Figure 3. C - Average Manager Residual Profit over All Period

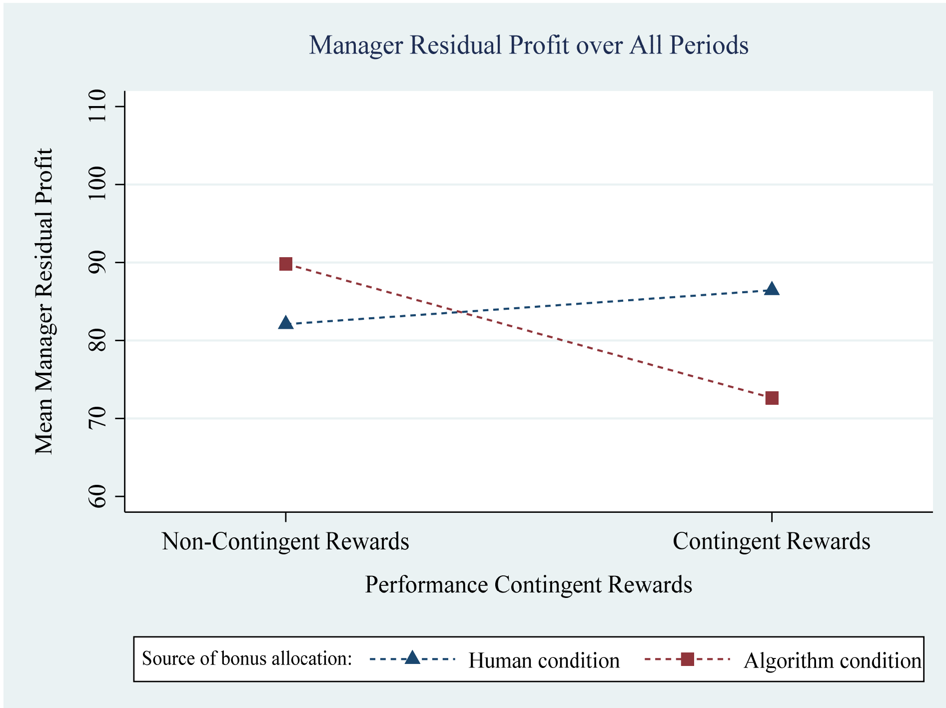

Figure 3 – Effect of Source of Allocation and Performance Contingent Reward on Employee Group Productivity, Employee Group Work Compensation, and Manager Residual Profit over All Periods

This figure plots the average of employees' group productivity (Panel A), group work compensation (Panel B), and manager's residual profit (Panel C) over all periods by whether employees receive rewards that are contingent on their performance signal. The source of allocation either comes from a human manager or from an algorithm that is designed to mimic the managers' decision making. Employee group productivity equals the sum of two employees' actual number of correctly decoded items in each group, without the random noise. Employee group work compensation equals the sum of two employees' compensation received from work in each group. Manager's residual profit equals the rest of the output bonus that has not been paid to the two employees in each group. Classification of performance contingent rewards is based on whether the performance signal significantly influences reward outcomes in the regression results.

F I G U R E 4

Figure 4. A - Average Employee Productivity over All Periods by Reward Outcome

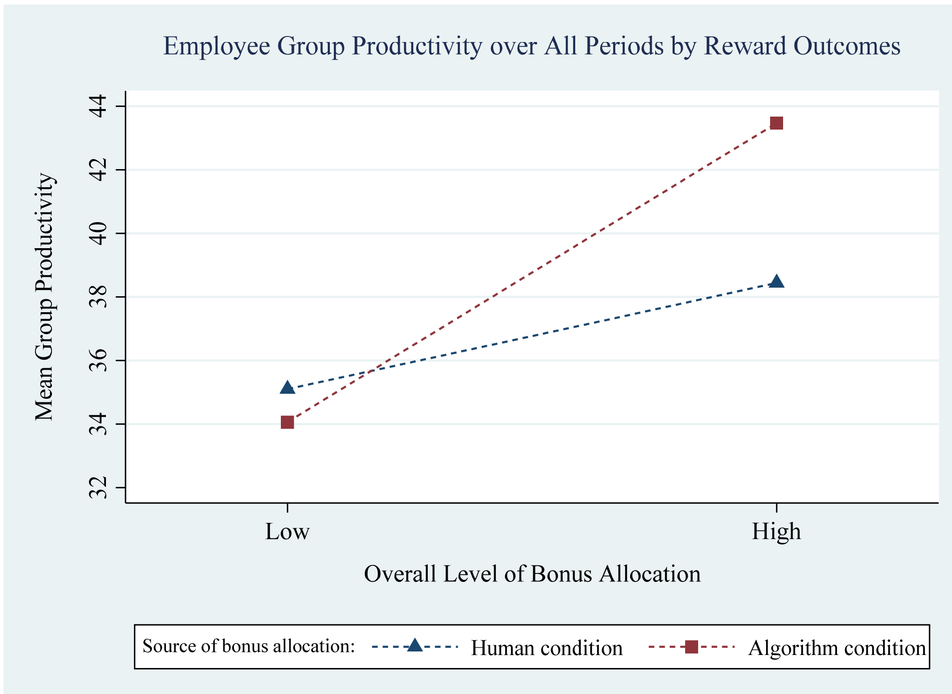

Figure 4. B - Average Employee Work Compensation by Reward Outcome

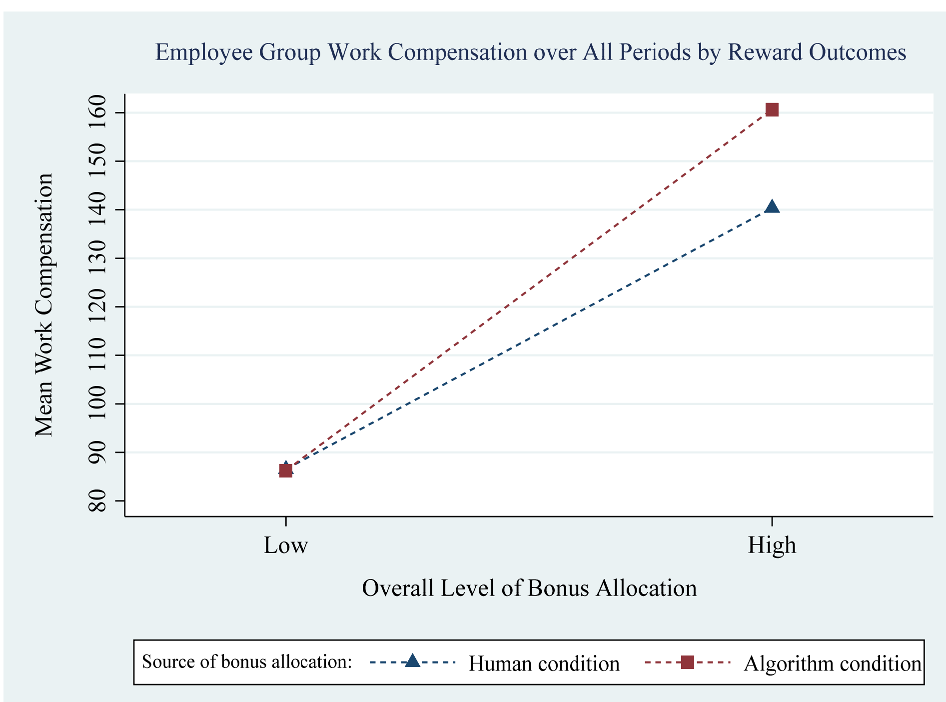

Figure 4. C - Average Manager Residual Profit by Reward Outcome

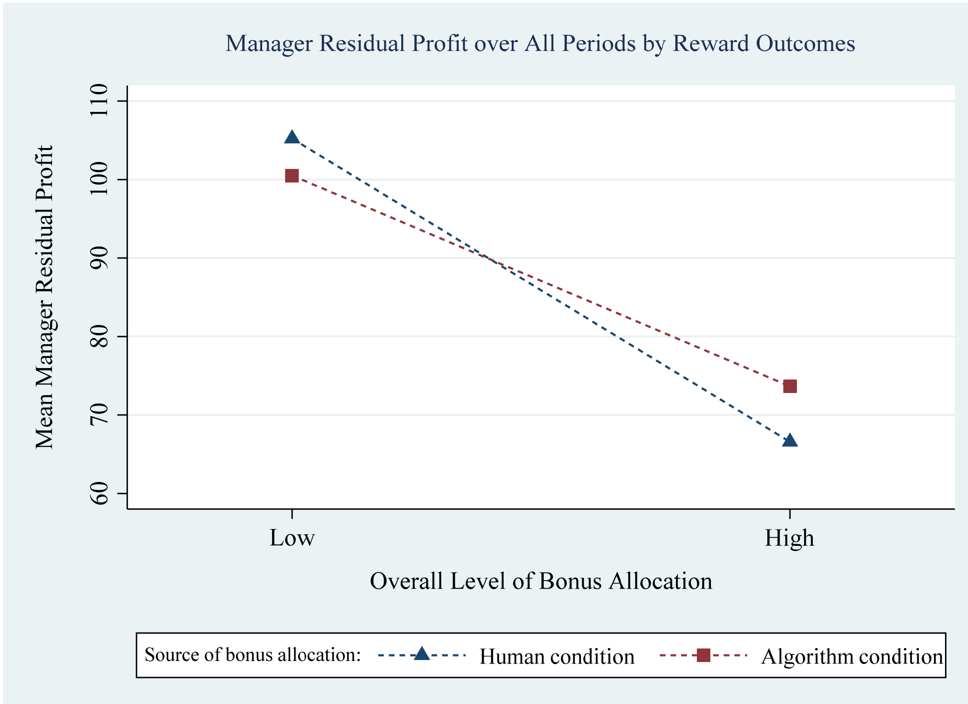

Figure 4. – Effect of Source of Allocation and Reward Outcome on Employee Group Productivity, Employee Group Work Compensation, and Manager Residual Profit over All Periods

This figure plots the average of employees' group productivity (Panel A), group work compensation (Panel B), and manager's residual profit (Panel C) over all periods by whether the overall rewards are low versus high. The source of allocation either comes from a human manager or from an algorithm that is designed to mimic the managers' decision making. Employee group productivity equals the sum of two employees' actual number of correctly decoded items in each group, without the random noise. Employee group work compensation equals the sum of two employees' compensation received from work in each group. Manager's residual profit equals the rest of the output bonus that has not been paid to the two employees in each group. Classification of high or low rewards in bonus allocations is based on whether the average bonus allocation over all periods is above or below 60%.

F I G U R E   5

Figure 5. A - Concern of Relative Payoff Between Workers and Managers

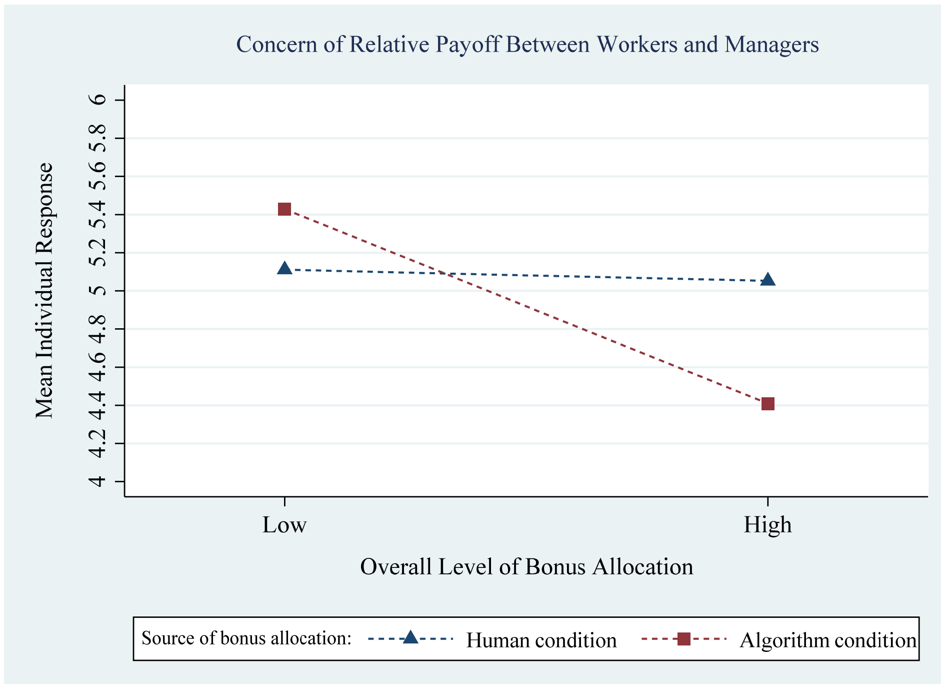

Figure 5. B - Perception of Fair Allocation Between Workers and Managers

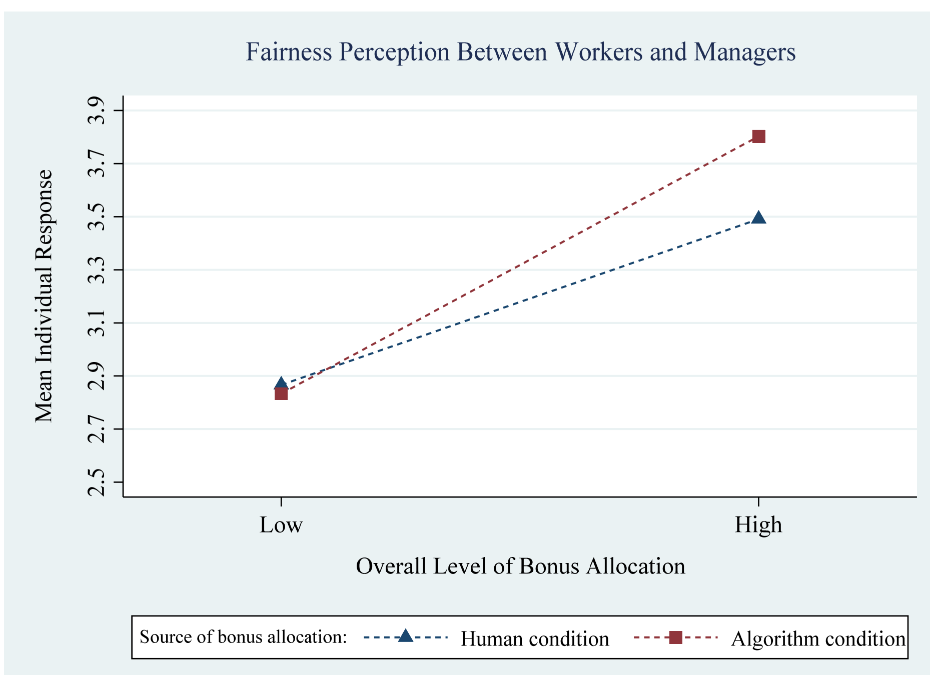

Figure 5. – Effect of Source of Allocation and Reward Outcome on Perception of Fairness Between Workers and Managers

This figure plots the average responses of participants' concern of the relative payoff between the workers and the managers (Panel A) and participants' perception of fair allocation of bonus between the workers and the managers (Panel B) by source of allocation and whether the rewards are overall low or high. The source of allocation either comes from a human manager or from an algorithm that is designed to replicate managers' decision making. Classification of high or low rewards in bonus allocations is based on whether the average bonus allocation over all periods is above or below 60%.

Participants were asked to indicate the extent to which they concerned with the relative payoff between the workers and the manager on a 7-point Likert scale with "1" labeled "Not at all" and "7" labeled "Very much." Participants were also asked to indicate whether the two workers collectively or the manager received more than a fair share of the output bonus on a 7-point Likert scale with "1" labeled "Definitely the manager," "3" labeled "Fair for both," and "7" labeled "Definitely the worker." Results in this figure include both manager and employee participants.

T A B L E 1

Table 1. A - Descriptive Statistics and ANOVA for Employee Group Productivity

Panel A: Descriptives for Employee Productivity over All Periods

	Source of Allocation	
	Human	Algorithm
Mean Group Productivity	36.97	39.81
	{12.99}	{13.20}
	n = 34	*n* = 36

Panel B: ANOVA for Employee Group Productivity

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	1129.94	4.33	0.038
Period	7	154.85	0.59	0.762
Source of Allocation*Period	7	66.26	0.25	0.971
Error	544	260.93		

Table 1. B - Descriptive Statistics and ANOVA for Employee Group Work Compensation

Panel A: Descriptives for Employee Group Compensation over All Periods

	Source of Allocation	
	Human	Algorithm
Mean Employee Compensation	116.62	131.73
	{58.57}	{61.04}
	n = 34	*n* = 36

Panel B: ANOVA for Employee Group Compensation

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	31928.48	5.93	0.015
Period	7	6477.17	1.20	0.299
Source of Allocation*Period	7	1238.65	0.23	0.978
Error	544	5380.01		

Table 1. C- Descriptive Statistics and ANOVA for Manager Residual Profit

Panel A: Descriptives for Manager Residual Profit over All Periods

	Source of Allocation	
	Human	Algorithm
Mean Manager Residual Profit	83.63	84.10
	{30.76}	{32.95}
	$n = 34$	$n = 36$

Panel B: ANOVA for Manager Residual Profit

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	30.71	0.01	0.916
Period	7	2848.00	1.04	0.402
Source of Allocation*Period	7	819.54	0.30	0.954
Error	544	2738.08		

Table 1.D - Descriptive Statistics and ANOVA for Employee Work Compensation Per Unit of Output

Panel A: Descriptives Percentage of Bonus Allocated to Employees over All Periods

	Source of Allocation	
	Human	Algorithm
Mean Employee Work Compensation Per Unit of Output	3.05	3.16
	{0.91}	{0.92}
	n = 34	n = 36

Panel B: ANOVA for Percentage of Bonus Allocated to Employees

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	2.31	1.57	0.210
Period	7	2.81	1.91	0.066
Source of Allocation*Period	7	0.25	0.17	0.991
Error	541	1.47		

Table 1 E - Descriptive Statistics and ANOVA for Percentage of Bonus Allocated to Employees

Panel A: Descriptives Percentage of Bonus Allocated to Employees over All Periods

	Source of Allocation	
	Human	Algorithm
Mean Allocation Percentage to Employees	0.54	0.56
	{0.22}	{0.22}
	n = 34	*n* = 36

Panel B: ANOVA for Percentage of Bonus Allocated to Employees

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	0.08	1.61	0.205
Period	7	0.10	2.05	0.048
Source of Allocation*Period	7	0.00	0.01	1.000
Error	544	0.05		

Table 1 – Effect of Source of Allocation on Employee Group Productivity, Employee Group Work Compensation, Manager Residual Profit, and Employee Work Compensation Per Unit of Output

This table presents the descriptives and ANOVA results for the effect of source of allocation (human vs. algorithm) and period on employee group productivity (Table 1A), employee group work compensation (Table 1B), manager's residual profit (Table 1C), employee group work compensation per unit of output (Table 1D), and percentage of bonus allocated to employees (Panel E). Panel A of each table presents the mean {standard deviation} of each outcome measure (n = number of groups in each condition). Panel B of each table reports the ANOVA results for the effects of source of allocation (human vs. algorithm) on each outcome measures. The source of allocation either comes from a human manager or from an algorithm that is designed to mimic the manager's decision making. Employee group productivity equals the sum of two employees' number of correctly decoded items in each group, without the random noise. Employee group work compensation equals the sum of two employees' compensation received from work in each group. Manager's residual profit equals the rest of the output bonus that has not been paid to the two employees in each group. Employee work compensation per unit of output equals the sum of two employees' work compensation divided by the group productivity (output before noise) in each group. The percentage of bonus allocated to the two employees is determined either by the manager in the group or the computer algorithm that mimics the managers' decision making. All reported p-values are two-tailed.

T A B L E 2

Table 2 - Effect of Source of Allocation and Performance Contingent Rewards on Employee Productivity

Panel A: Descriptives for Employee Output over All Periods

	Source of Allocation		
	Human	Algorithm	Mean {S.D.}
Performance-contingent rewards	42.13	41.59	41.86
	{11.47}	{15.18}	{13.16}
	n = 12	*n* = 12	*n* = 24
Rewards not contingent on performance	34.15	38.92	36.64
	{13.14}	{12.34}	{12.81}
	n = 22	*n* = 24	*n* = 46
Mean {S.D}	36.97	39.81	
	{12.99}	{13.20}	
	n = 34	*n* = 36	

Panel B: ANOVA for Employee Output

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Between-subjects				
Source of Allocation	1	564.56	2.22	0.137
Contingency	1	3574.46	14.07	0.000
Source of Allocation*Contingency	1	883.63	3.48	0.063
Within-subjects				
Period	7	220.58	0.87	0.532
Source of Allocation*Period	7	94.17	0.37	0.919
Contingency*Period	7	377.13	1.48	0.170
Source of Allocation*Contingency*Period	7	100.59	0.40	0.905
Error	528	254.12		

Panel C: Simple Effects

Effect of Source of Allocation	*df*	*F*-Statistic	*p*-Value
Effect of source of allocation under *non-performance-contingent rewards*	1	8.20	0.004
Effect of source of allocation under *performance-contingent rewards*	1	0.05	0.817

Effect of Performance Contingent Rewards	*df*	*F*-Statistic	*p*-Value
Effect of performance contingent rewards under *human*	1	15.53	0.000
Effect of performance contingent rewards under *algorithm*	1	1.80	0.179

Table 2 – Effect of Source of Allocation and Performance Contingent Rewards on Employee Productivity

This table reports the results of employees' productivity based on whether the rewards are performance contingent. Panel A contains the mean {standard deviation} of employees' group productivity in each of the conditions. Panel B reports the ANOVA results for the effects of source of allocation (human vs. algorithm), performance contingent rewards (contingent vs. non-contingent), and period on employees' group productivity. Panel C reports the results of simple effects tests. All reported p-values are two-tailed.

T A B L E 3

Table 3 - Effect of Source of Allocation and Overall Bonus Outcome on Employee Productivity

Panel A: Descriptives for Employee Productivity over All Periods

	Source of Allocation		
	Human	Algorithm	Mean {S.D.}
Overall high	38.44	43.47	41.14
	{15.06}	{11.53}	{13.35}
	n = 19	*n* = 22	*n* = 41
Overall low	35.10	34.06	34.60
	{9.96}	{14.00}	{11.87}
	n = 15	*n* = 14	*n* = 29
Mean {S.D}	36.97	39.81	
	{12.99}	{13.20}	
	n = 34	*n* = 36	

Panel B: ANOVA for Employee Productivity

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Between-subjects				
Source of Allocation	1	538.61	2.14	0.145
Reward Outcome	1	5501.35	21.81	0.000
Source of Allocation*Reward Outcome	1	1244.99	4.94	0.027
Within-subjects				
Period	7	124.30	0.49	0.840
Source of Allocation*Period	7	71.35	0.28	0.961
Reward Outcome*Period	7	123.71	0.49	0.842
Source of Allocation*Reward Outcome*Period	7	155.11	0.61	0.744
Error	528	252.26		

Panel C: Simple Effects

Effect of Source of Allocation	*df*	*F*-Statistic	*p*-Value
Effect of source of allocation when rewards are *high*	1	8.16	0.004
Effect of source of allocation when rewards are *low*	1	0.25	0.619

Effect of Reward Outcome	*df*	*F*-Statistic	*p*-Value
Effect of reward outcome under *human*	1	2.97	0.085
Effect of reward outcome under *algorithm*	1	23.99	0.000

Table 3 – Effect of Source of Allocation and Overall Bonus Outcome on Employee Productivity

This table reports the results of hypotheses tests of employees' productivity. Panel A contains the mean {standard deviation} of employees' group productivity in each of the conditions. Panel B reports the ANOVA results for the effects of source of allocation (human vs. algorithm), reward outcomes (generous vs. selfish), and period on employees' group productivity. Panel C reports the results of simple effects tests. Classification of overall reward outcome is based on whether the average bonus allocation over all periods is above or below 60%. All reported p-values are two-tailed.

T A B L E 4

Table 4 A - Effect of Source of Allocation on Concern of Relative Payoff Between Workers and Managers

Panel A: Descriptives for Concern of Relative Payoff Between Workers and Managers

	Source of Allocation		
	Human	Algorithm	Mean {S.D.}
Overall high	5.05	4.41	4.71
	{1.48}	{1.75}	{1.66}
	n = 57	*n* = 66	*n* = 123
Overall low	5.11	5.43	5.26
	{1.40}	{1.13}	{1.28}
	n = 45	*n* = 42	*n* = 87
Mean {S.D}	5.08	4.81	
	{1.44}	{1.61}	
	n = 102	*n* = 108	

Panel B: ANOVA for Reward Outcome and Concern of Relative Payoff

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	1.35	0.60	0.438
Reward Outcome	1	14.76	6.59	0.011
Source of Allocation*Reward Outcome	1	11.73	5.24	0.023
Error	206	2.24		

Panel C: Simple Effects

Effect of Source of Allocation	*df*	*F*-Statistic	*p*-Value
Effect of source of allocation when rewards are *high*	1	5.65	0.017
Effect of source of allocation when rewards are *low*	1	0.98	0.323

Effect of Reward Outcome	*df*	*F*-Statistic	*p*-Value
Effect of reward outcome under *human*	1	0.04	0.845
Effect of reward outcome under *algorithm*	1	11.91	0.000

Table 4 B - Effect of Source of Allocation on Perception of Fair Allocation of Bonus Between Workers and Managers

Panel A: Descriptives for Perception of Fair Allocation of Bonus

	Source of Allocation		
	Human	Algorithm	Mean {S.D.}
Overall high	3.49	3.80	3.66
	{1.57}	{1.90}	{1.75}
	n = 57	*n* = 66	*n* = 123
Overall low	2.87	2.83	2.85
	{1.67}	{1.70}	{1.67}
	n = 45	*n* = 42	*n* = 87
Mean {S.D}	3.22	3.43	
	{1.64}	{1.88}	
	n = 102	*n* = 108	

Panel B: ANOVA for Reward Outcome and Perception of Fair Allocation of Bonus

Source	*df*	Mean Square	*F*-Statistic	*p*-Value
Source of Allocation	1	0.98	0.33	0.566
Reward Outcome	1	32.28	10.84	0.001
Source of Allocation*Reward Outcome	1	1.51	0.51	0.477
Error	206	2.98		

Panel C: Simple Effects

Effect of Source of Allocation	*df*	*F*-Statistic	*p*-Value
Effect of source of allocation when rewards are *high*	1	0.01	0.928
Effect of source of allocation when rewards are *low*	1	0.10	0.318

Effect of Reward Outcome	*df*	*F*-Statistic	*p*-Value
Effect of reward outcome under *human*	1	3.29	0.070
Effect of reward outcome under *algorithm*	1	8.10	0.004

Table 4 – Tests of Post Questionnaire - Perception of Fairness Between Workers and Managers

This table presents the descriptives and ANOVA results for the effect of source of allocation (human vs. algorithm) and reward outcome on participants' concern of the relative payoff (Table 4A) and participants' perception of fair allocation of bonus (Table 4B) between the workers and the managers. Panel A of each table contains the mean {standard deviation} of participants' responses in each of the conditions. Panel B of each table reports the ANOVA results for the effects of source of allocation (human vs. algorithm) and reward outcomes (high vs. low) on participants' responses. Panel C of each table reports the results of simple effects tests of participants' responses. Classification of overall reward outcome is based on whether the average bonus allocation over all periods is above or below 60%.

Participants were asked to indicate the extent to which they concerned with the relative payoff between the workers and the manager on a 7-point Likert scale with "1" labeled "Not at all" and "7" labeled "Very much." Participants were also asked to indicate whether the two workers collectively or the manager received more than a fair share of the output bonus on a 7-point Likert scale with "1" labeled "Definitely the manager," "3" labeled "Fair for both," and "7" labeled "Definitely the worker." Results in this table include both manager and employee participants. All reported p-values are two-tailed.

APPENDIX A. SCHEME OF ALGORITHM TO USE FOR EACH INDIVIDUAL MANAGER

Table A. 1 - Scheme of Algorithm To Use for Each Individual Manager

Start from 34 managers

Those who reward **not contingent** on team output	Count	Algorithm to use
No time effects:		
zero variation (std. dev. = 0)	1	
Significant intercept:		
Stick around fairness benchmark (0.5, 0.6, 0.66)	3	
Stick around generous benchmark (0.8)	2	
Mix of generous and selfish	5	
Insignificant intercept:		
highly variant	3	Replicate the allocation
Stick around selfish benchmark	2	
Significant time effects:		
Beginning of the game effect (significant coefficient for dummy of first two periods)	2	
End of the game effect (significant coefficient for dummy of last two periods)	2	
Time effects overlap with performance effect (after control for time effects, coefficient on performance turns insignificant)	2	
Total	22	

Those who reward **contingent** on team output	Count	Algorithm to use
Performance effect & time effects (after control for time effects, still have significant coefficient on performance):		
Round effect	2	Regression prediction and control for round number, random pick within confidence interval
Beginning of the game effect	2	Regression prediction and control for dummy of first two periods, random pick within confidence interval
Performance effect & no time effects	8	Regression prediction, random pick within confidence interval
Total	12	

Table A. 1 – This table lists the scheme of algorithm to use for each category of individual managers.

Managers are classified into two major groups based on the regression model predictions: those who reward based on the performance signal and those who do not reward based on the performance signal. Algorithm is built up for each individual category of the managers.

APPENDIX B. EXAMPLES OF ALGORITHMIC DETERMINATION OF BONUS OUTCOMES

Figure B.1 - Managers Reward Contingent on the Performance Signal

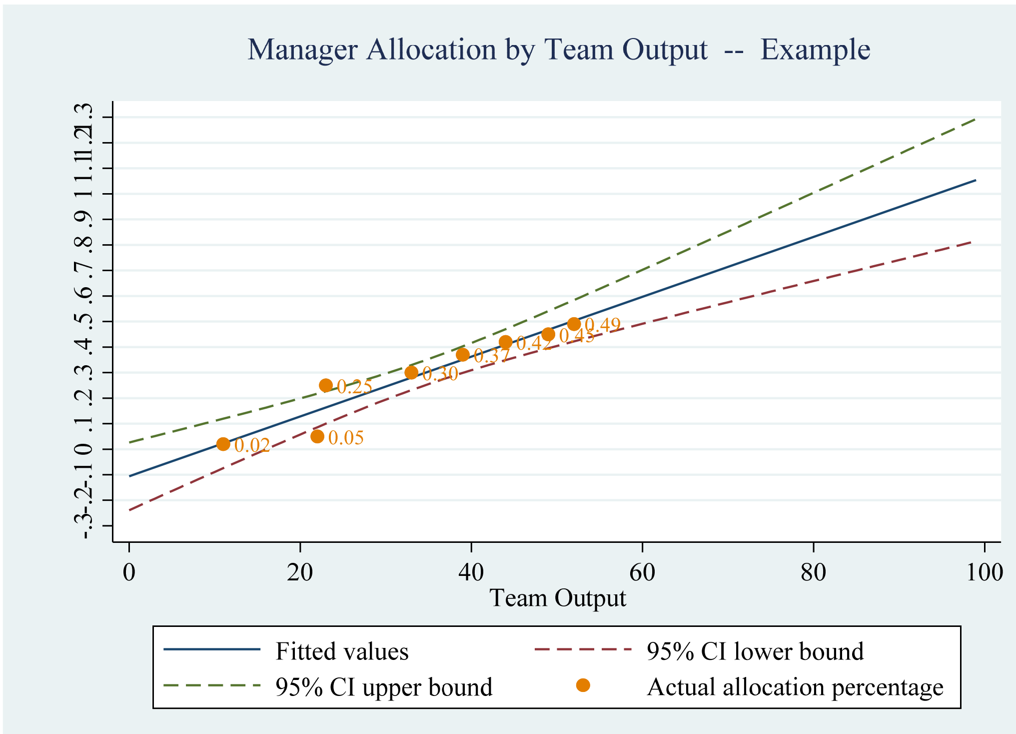

Figure B. 1 - Example of manager's bonus allocation that is contingent on the performance signal.

For instance, the manager's bonus allocation plotted in the figure is significantly positively correlated with the performance signal that he/she observes (i.e., the team output) ($\beta = 0.066$, two-tailed $p < 0.001$).

For those managers, the algorithm uses model predictions to extrapolate to the new situation. To reflect the randomness in human decision making, the algorithm randomly draws one value within the 95% confidence interval of the predicted allocation, constrained to a boundary of zero and one (i.e., between the lower bound and upper bound in the figure). The estimation model uses GLM with logit link function and robust standard error.

Figure B.2 - Managers Reward Not Contingent on the Performance Signal

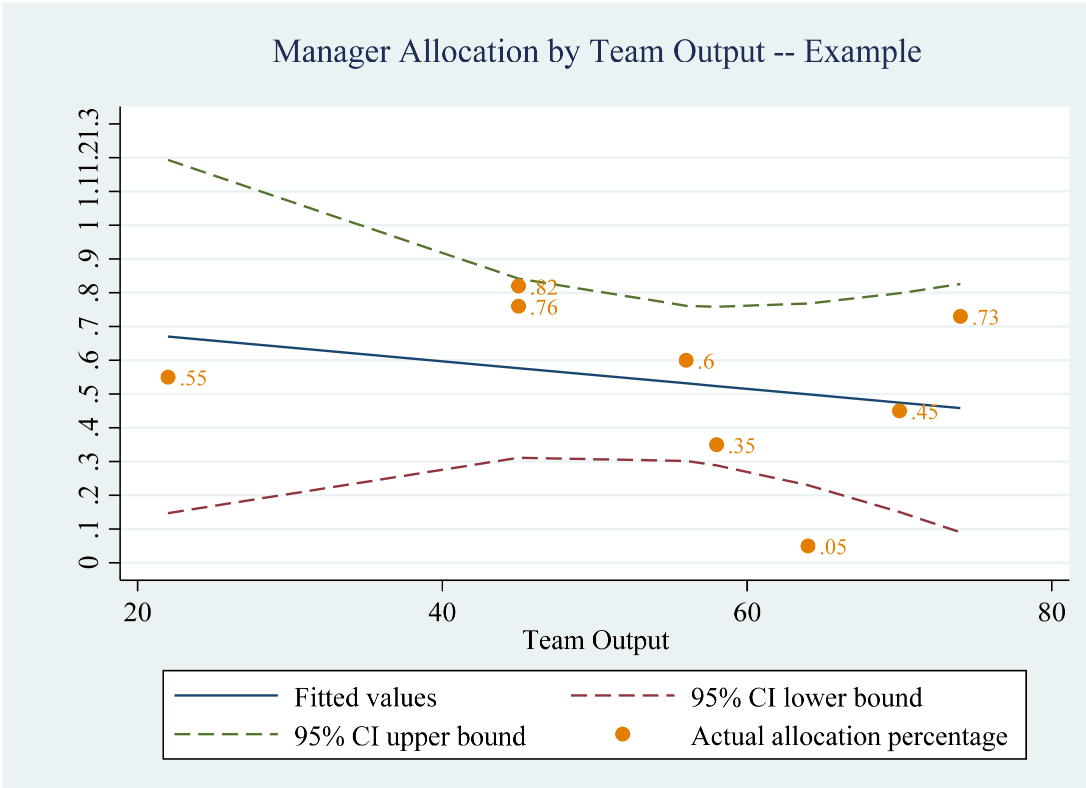

Figure B.2 - Example of manager's bonus allocation that is not contingent on the performance signal.

For instance, the manager's bonus allocation plotted in the figure is *not* significantly correlated with the performance signal that he/she observes (i.e., the team output) (β = -0.017, two-tailed p = 0.386).

For those managers, the algorithm replicates the decisions of those managers, as observable data cannot provide a clear base to predict their rewarding behavior.

APPENDIX C. EXPERIMENTAL INSTRUCTIONS FOR BONUS ALLOCATION

Allocation of Output Bonus

[Human condition]

In each round, **the manager** will decide how to split the output bonus between himself/herself and the two workers. The amount the manager allocates to the workers can range from zero up to 100% of the output bonus. That is, the manager can give nothing to the workers, can give the entire output bonus to the workers, or anywhere in between.

Specifically, the manager will determine, out of the output bonus, how much to allocate to the workers and how much to retain as his/her own bonus. The workers' bonus will be split **equally** between the two workers. The manager **cannot** differentially reward the two workers.

[Algorithm condition]

In each round, **a computer algorithm** will decide how to split the output bonus between the manager and the two workers. The amount the algorithm allocates to the workers can range from zero up to 100% of the output bonus. That is, the algorithm can give nothing to the workers, can give the entire output bonus to the workers, or anywhere in between.

Specifically, the computer algorithm will determine, out of the output bonus, how much to allocate to the workers and how much to allocate to the manager. The workers' bonus will be split **equally** between the two workers. The algorithm **cannot** differentially reward the two workers.

[Only for the algorithm conditions]

Algorithm to Allocate Output Bonus

The algorithm has been developed to predict managers' behavior in splitting the output bonus, based on a large sample of real managers' allocation decisions. These managers played the same game as you are playing today and with the same compensation structure.

These managers exhibited different patterns of behavior, reflecting different strategies, styles, and preferences. The algorithm is designed to try to **capture** those strategies, styles, and preferences.

Specifically, the algorithm will **randomly select one manager** from the large sample and use data about that manager's allocation behavior to determine all of the allocations for your group. In other words, the algorithm uses the observable data and factors that appeared to influence that manager's strategies, styles, and preferences when he or she made allocation decisions. Then, the algorithm uses data from your group's situation and **extrapolates** that manager's allocation behavior to your group's situation to capture, **to the extent possible**, what that manager would have decided for your group's bonus allocation for that round.

The algorithm only selects **one manager** per group for all rounds and will not use data from other managers. Please be aware that the manager in your group cannot influence the algorithm’s allocation decision. He or she will be paid based on whatever allocation comes from the algorithm.

REFERENCES

Bailey, W. J., Hecht, G., & Towry, K. L. (2011). Dividing the pie: The influence of managerial discretion extent on bonus pool allocation. *Contemporary Accounting Research*, 28(5), 1562-1584.

Baiman, S., & Rajan, M. V. (1995). The informational advantages of discretionary bonus schemes. *The Accounting Review*, 557-579.

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

Baker, G. P., Jensen, M. C., & Murphy, K. J. (1988). Compensation and incentives: Practice vs. theory. *The Journal of Finance*, 43(3), 593-616.

Banker, R. D., & Datar, S. M. (1989). Sensitivity, precision, and linear aggregation of signals for performance evaluation. *Journal of Accounting Research*, 27(1), 21-39.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*.

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

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

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

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: Ultimatums, dictators and manners. *Journal of Economic Perspectives*, 9(2), 209-219.

Camerer, C. (2003). Behavioral Game Theory: Experiments in Strategic Interaction. *Princeton University Press*. Princeton, NJ.

Charness, G., & Haruvy, E. (2002). Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games and Economic Behavior*, 40(2), 203-231.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3), 817-869.

Crittenden, K. (1983). Sociological aspects of attribution. *Annual Review of Sociology* 9: 425–46.

Dastin, J. (2018). "Amazon scraps secret AI recruiting tool that showed bias against women." See https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155-1170.

Dijkstra, A., De Vries, H., & Roijackers, J. (1998). Long-term effectiveness of computer-generated tailored feedback in smoking cessation. *Health Education Research*, 13(2), 207-214.

Du, F., Tang, G., & Young, S. M. (2012). Influence activities and favoritism in subjective performance evaluation: Evidence from Chinese state-owned enterprises. *The Accounting Review*, 87(5), 1555-1588.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, 75(5), 1501-1511.

Fehr, E., & Gächter, S. (2000). Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives,* 14(3), 159-181.

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

Fisher, A. (2019). “An Algorithm May Decide Your Next Pay Raise.” See https://fortune.com/2019/07/14/artificial-intelligence-workplace-ibm-annual-review/

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

Hales, J., & Williamson, M. G. (2010). Implicit employment contracts: The limits of management reputation for promoting firm productivity. *Journal of Accounting Research*, 48(1), 51-80.

Hellard, B. (2018). “IBM Watson can predict just how productive you are.” See https://www.itpro.co.uk/technology/31493/ibm-watson-can-predict-just-how-productive-you-are

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635-650.

Kelley, H. H. (1967). Attribution theory in social psychology. In Nebraska symposium on motivation. *University of Nebraska Press*.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, 28(2), 107.

Kelman, H. C., & Hovland, C. I. (1953). " Reinstatement" of the communicator in delayed measurement of opinion change. *The Journal of Abnormal and Social Psychology*, 48(3), 327.

Ledyard, J. O. (1995). Public goods: A survey of experimental research. In Handbook of Experimental Economics, edited by H. H. Kagel, and A. E. Roth, 111-194. Princeton, NJ: Princeton Uni-versity Press.

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society, 5(1), 2053951718756684.

Leventhal, G. S.(1980): What should be done with equity theory? New approaches to the study of fairness in social relationships. *Gergen et al.(Hg.): Social exchange: Advances in theory and research*. New York, 27-55.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90-103.

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

McGarry, J., & Hendrick, C. (1974). Communicator credibility and persuasian. *Memory & Cognition*, 2(1), 82-86.

Milgrom, P. R. (1988). Employment contracts, influence activities, and efficient organization design. *Journal of Political Economy*, 96(1), 42-60.

Milgrom, P., & Roberts, J. (1987). Informational asymmetries, strategic behavior, and industrial organization. *The American Economic Review*, 77(2), 184-193.

Morewedge, C. K. (2009). Negativity bias in attribution of external agency. *Journal of Experimental Psychology*: *General*, 138(4), 535.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243-281.

Rajan, M. V., & Reichelstein, S. (2006). Subjective performance indicators and discretionary bonus pools. *Journal of Accounting Research*, 44(3), 585-618.

Reinard, J. C. (1988). The empirical study of the persuasive effects of evidence the status after fifty years of research. *Human Communication Research*, 15(1), 3-59.

Rogers, H. (1987). Theory of recursive functions and effective computability, Vol. 5. New York: McGraw-Hill.

Rosenbaum, E. (2019). “IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs”. See https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html

Ross, M., & Fletcher, G. J. (1985). Attribution and social perception. *The Handbook of Social Psychology*, 2, 73-114.

Ryan, A. M., & Ployhart, R. E. (2000). Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of Management*, 26(3), 565-606.

Schrage, M., Kiron, D., Hancock, B., & Breschi, R. (2019). "Performance Management's Digital Shift." *MIT Sloan Management Review,* 0, 1-23.

Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology*, 2(4), 325-342.

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making*, 32(4), 403-414.