Behavioral Decision-Making in Operations Management

by

Jiawei Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Business Administration)
in the University of Michigan
2022

Doctoral Committee:

Professor Damian Beil, Co-Chair
Professor Stephen Leider, Co-Chair
Professor Yan Chen
Professor Izak Duenyas
Associate Professor Yesim Orhun
Professor Tanya Rosenblat

Jiawei Li

jiawli@umich.edu

ORCID iD: 0000-0002-7945-8537

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF APPENDICES

ABSTRACT

In this dissertation, I explore the role of human decision-making in operations management. In the first part of the dissertation, I compare the performance of teams and individuals in two canonical operational settings and find that teams do not always outperform individuals. For Newsvendor inventory decision-making, teams exhibit a similar degree of the "pull-to-center" bias and order too close to the mean of demand. In a strategic information sharing game, teams outperform individuals in terms of earning profits only as the retailer who has a clear strategy to benefit from misreporting demand signals. I then implement a framework that studies team decision mechanisms by analyzing text chats between team members. I find that the logic leading to "optimal/rational" decision-making may or may not be frequently mentioned or found compelling in team discussions. In the second part of the dissertation, I explore how humans make dynamic resource allocation decisions. The decision maker - a product development manager - is given a limited budget for design improvement opportunities that arrive sequentially; the opportunities are beneficial but implementing them is costly. I find that most subjects perform well in a simple setting where the design change cost is constant throughout the decision horizon, but subjects perform heterogeneously in the more complex and realistic increasing cost setting. In the latter setting, some top performers effectively decompose the problem into two simpler sub-problems, each of which they handle nearly optimally as in the simpler experimental condition. I test managerial interventions based on these insights and show they can improve subject performance. In the third part of the dissertation, I study a procurement process where the manufacturer first holds an auction with suppliers for a prototype product; after the auction, renegotiations will happen as design changes are needed. I use game theory to analyze a practical mechanism that constrains such renegotiation by pre-limiting design change expenditures. I find that such a mechanism may not benefit the manufacturer depending on the supply base condition. A follow-up experiment is proposed to study humans' ability to implement this mechanism.

CHAPTER 1

Introduction

This dissertation explores the role of human decision-making in operations management (OM). Many operational decisions, such as inventory decisions, are cognitively challenging as decision-makers need to jointly consider different aspects of the problem and account for substantial trade-offs. Yet, due to practical challenges, not all of these decisions can be automated. When people make these decisions, they may be subject to various behavioral forces such as limitations to cognitive abilities and therefore perform differently from standard (rational) model predictions. This difference between predictions and reality motivates me to use behavioral operations management (BOM) research methods, primarily controlled laboratory experiments, to understand human decision-making in OM contexts. My research covers various topics in OM, including inventory management, product development, and procurement management.

The field of BOM is still young and developing with many unexplored yet important questions. My research in BOM focuses on these unexplored avenues and aims to answer the following two research questions: (1) when will people make good/bad operational and OM decisions and what is the decision mechanism behind this; (2) how can we devise managerial interventions to bring out the best performance in human decision-making. The key research methodology is laboratory experiments (both in-person and online), which allows me to collect data that captures human behavior in a controlled environment. In addition, I also develop novel analytical models to serve as rational decision-making benchmarks. Overall, my goal of the research is to identify major behavioral mechanisms in order to generate insights that speak to applied settings and improve managerial practices.

Below, I present a summary of my three research papers in the dissertation. The three papers are then presented in Chapter 2, Chapter 3, and Chapter 4, respectively. I conclude in Chapter 5.

1. *How do teams make operational decisions?* Teams are prevalent in real-world operational decision-making. Past studies in behavioral economics have shown that teams are more cognitively sophisticated and more self-interested, but these are based on abstract and simple decision contexts. In Chapter 2, I seek to understand how teams make complex inventory

and strategic information sharing decisions and how they differ from individuals.

2. *How do people make dynamic decisions?* Dynamic decisions are challenging as human decision-makers need to understand the inter-connection among decision periods and plan accordingly. In Chapter 3, I study how people make dynamic resource allocation decisions with a budget constraint in the context of product development management. When they fall short of making optimal decisions, I propose managerial treatments to improve their performance.

3. *How well do people understand the linkage between decision stages?* Change orders post contract award in procurement auctions have been frequently observed in practice. Anticipating this, suppliers will alter the way they compete during the procurement auction. To manage this, the original equipment manufacturer (OEM) needs to jointly consider the auction stage and the post-auction stage. In Chapter 4, I analyze a new practical mechanism to manage the OEM's expenditure from both the procurement auction and change orders and propose ways to assess people's ability to use this new mechanism.

CHAPTER 2

Team Decision-Making in Operations Management

2.1 Introduction

Business decisions are often made in a team setting. For example, in a large consumer goods firm I interact with, managers reported that at one large factory they have a dedicated scheduling team managing production, a different team responsible for demand forecasting, and a management team coordinating overall operations. Yet, in the operations management community, the implication of *team decision-making* has not been well understood. Specifically, when should I expect that teams would make better, or worse, decisions than individuals? Current research in behavioral operations management (BOM) has largely used *individuals* to conduct experiments and develop managerial insights. The goal of this chapter is to begin studying team decision-making in operational contexts by identifying if and when teams make better operational decisions compared to individuals.

To study team decision-making in operations, I focus on settings that have very well-established benchmarks for individual decision-making. This leads us to use Newsvendor decisions and information sharing in a Newsvendor context. Both are important in practice and have been studied in BOM with individuals as subjects [Schweitzer & Cachon (2000), Özer et al. (2011)]. Moreover, collectively they cover the two main streams of operational decisions: tactical decisions where managers engage in pure problem-solving tasks, and strategic decisions where managers or companies interact with each other to determine operational outcomes. I consider a small team setup (two team members) with aligned information and aligned preferences; I employ the consensus-based decision-making scheme where the two team members are required to form a mutual agreement to make a team decision. This team setup is simple to employ in the lab and captures many real-world team decision-making settings. By comparing the performance of teams and individuals in these decision settings, I provide a first step in evaluating the potential benefits of implementing team decision-making in operations. In this chapter, I also pay special attention to understanding the team decision mechanism - the way decisions are formulated through intra-team discussions and interactions. This is important for two reasons. First, the team is an important decision-unit in many core operational settings, so understanding team decision mechanisms has direct managerial importance. Second, understanding team decision mechanisms allows us to better understand why teams do, or do not, outperform individuals. Past research (presented below) has suggested a general mechanism for how team discussions connect with team decision outcomes, but in abstract and general economic/psychology decision-making contexts. It is therefore important to study whether and how the general mechanism extends to operational contexts.

There is reason to suspect that teams might reach different decisions from individuals. Research in psychology and behavioral economics has found that teams are better at making many tactical and strategic decisions. For tactical decisions, teams are found to be more cognitively sophisticated than individuals and better at understanding the probabilistic nature of tasks [Charness et al. (2010), Charness et al. (2007)]. For strategic decisions, teams are found to have stronger self-interested preferences (are more selfish), are more skeptical towards the opponent (are more untrusting), and are better able to reason from the opponent's perspective. As a result, teams' actions are more consistently aligned with game-theoretic predictions [Cox (2002), Cooper & Kagel (2005)]. The primary mechanism for team differences is "effective persuasion" happening within the team, via the team's decision-making discussions. In tactical decisions, particularly when a "correct answer" exists (such as the solution to a word puzzle), the team member who derives the answer can effectively persuade other team members to follow his/her answer by explaining the logic behind it, so that the whole team can converge to choosing the "correct answer". In strategic settings, the "correct answer" may not exist; however, arguments related to "doing best for the team", along with the logic behind it, stand out to be compelling arguments and strongly drive team decision outcomes [Charness & Sutter (2012)]. In both scenarios, the key behind team-individual decision differences is the emergence of a *compelling argument* in team discussions.

In Newsvendor decision-making, past BOM studies have shown that individuals fail to make optimal Newsvendor decisions: They often order too close to the mean - the "pull-to-center" bias [Schweitzer & Cachon (2000), Thonemann & Becker-Peth (2018)]. If teams are capable of more sophisticated thinking (enabled by effective persuasion within the team), they should be better at overcoming the pull-to-center bias by deriving classical Newsvendor logic (tradeoff behind overage and underage costs) and use it as the compelling argument in team discussions. If so, teams will act closer to the optimum compared to individuals. For information sharing in the Newsvendor context, I employ the setting introduced by Özer et al. (2011): A retailer (he) privately observes the demand forecast and strategically sends a demand signal to the supplier (she); the supplier updates her belief regarding the demand and makes her Newsvendor decision. Game-theoretic analysis suggests that the signal sent by the retailer will be uninformative, and the supplier will never trust this signal. Interestingly, Özer et al. (2011) and Özer et al. (2014) find that individual retailers and suppliers exhibit a certain degree of trust and trustworthiness. To study the impact of teams in this setting, I consider a two-by-two factorial design: team/individual retailer vs. team/individual supplier. Given results from behavioral economics, it seems reasonable to expect that arguments related to being more untrustworthy (untrusting) will become the compelling argument for retailers (suppliers) in team discussions; as a result, compared to the baseline pure-individual configuration (individual retailer vs. individual supplier, the replication of Özer et al. (2014)), team retailers will be more untrustworthy and team suppliers will be more untrusting. The above conjectures form the baseline hypotheses of our study. However, I note that these conjectures are based on results from psychology and behavioral economic studies, where abstract games/decision tasks are used; in concrete operational settings, context-related elements may trigger different kinds of decision heuristics and/or different norms of decision-making. Therefore, careful behavioral studies (such as ours) are needed to determine whether these conjectures hold in operational settings.

In the standalone Newsvendor task, surprisingly I find that teams *fail to* outperform individuals and exhibit a similar degree of "pull-to-center" bias. By analyzing the text chats, I find that teams are able to agree on an argument through team discussions. However, the Newsvendor solution, in particular the logic behind it, is neither prevalent nor particularly persuasive in decision-making. Instead, team members find other arguments compelling, with the top two being general mental dispositions of being aggressive or conservative. As a result, teams as a whole fail to systematically outperform individuals. In a robustness check, I conduct a follow-up experiment where I give teams extensive opportunities to learn and explore; I find a similar outcome.

For our information sharing context, against the pure individual benchmark, I find that team retailers are indeed more untrustworthy compared to individual retailers. On the supplier side, team suppliers are more untrusting compared to individual suppliers when paired with an individual retailer (with a similar but statistically insignificant difference when paired with a team retailer). By analyzing the team's text chats, I find that the discussion between being trustworthy or not as retailer, and trusting or not as supplier is the key factor driving team decisions. However, the discussion dynamics are different between retailers and suppliers. The retailer's strategic problem is more straightforward, with team retailers quickly converging on being untrustworthy. On the other hand, suppliers appear to find their situation more ambiguous, with many teams switching between being trusting and untrusting, and teams continuing to find arguments to be trusting and arguments to be untrusting both compelling, even at the end of the experiment. The difference in discussion dynamics is paralleled by a difference in decision effectiveness (measured by expected profits) between retailers and suppliers. Team retailers' untrustworthiness enables them to increase their expected profits compared to individual retailers. By contrast, team suppliers' untrusting behavior does *not* lead to systematic profit advantage: The expected profits are highly variable, like the variability in their chats, such that the expected profits are not significantly different between team and individual suppliers; as a result, suppliers do not benefit from team decision-making.

Moving forward, our results suggest that teams will perform differently from individuals, to the extent that a compelling argument emerges in team discussions. When companies consider whether to implement team decision-making, it is therefore important to think carefully about whether and which compelling argument(s) will emerge during team discussions, and whether the compelling argument will promote the outcome the company wants. For tactical inventory decisions, I show that the "textbook" solution may not be prevalent or persuasive in team discussions, so that teams may fail to outperform individuals. Hence, our results suggest that team decision-making is not a silver bullet tool to improve tactical decisions. For our strategic decision setting, I observe differences in the team decision mechanism between the two sides of the supply chains: Retailers have a clear and compelling argument to be untrustworthy leading to superior performance for teams, while team suppliers are more conflicted and exhibit little performance difference.

This chapter offers four major contributions. First, this chapter is among the first examinations of team decisions in operations, a prevalent feature of industry. Second, I confirm that the general team decision mechanism identified in abstract contexts extends to operational contexts; that is, the team decision mechanism can be studied by identifying the compelling arguments in team chats that drive teams' final decision outcomes. Third, I identify two key challenges in the study of team decision-making in operational contexts: (1) Because of the complexity in operational decision-making, there could be multiple (potentially conflicting) compelling arguments, pulling teams in different directions; (2) It could be difficult to predict *ex-ante* what the compelling arguments will be. Fourth, I identify cases in an operationally relevant setting where team decision-making does improve outcomes (retailers) and does not improve outcomes (suppliers and standalone Newsvendors), as well as using the tools described above to demonstrate the underlying mechanisms for those outcomes. In general, our study confirms the need to carefully study team decision-making in operational contexts, and I find that teams do *not* always outperform individuals. Behavioral studies, in particular text chat analysis, prove to be useful tools to explore the team decision mechanism and explain team decision outcomes, and can be applied to future studies on team decision-making in other operational contexts.

2.2 Literature Review

Research in behavioral economics suggests that teams behave differently from individuals in systematic ways. Researchers have studied team decision-making mainly in two contexts: tactical decision-making [see for example Kocher & Sutter (2005), Charness et al. (2007), Charness et al. (2010), and Sheremeta & Zhang (2010)] and self-interested strategic decision-making [see Cox (2002), Cooper & Kagel (2005), and Feri et al. (2010)]. For a comprehensive review, see Charness

& Sutter (2012). The team setup I use in our experiments has been the primary format for past tactical/strategic experiments: small teams, aligned information, aligned preferences, and consensus-based decision-making scheme; hence, I expect that the insights from these studies should extend to operational contexts. For tactical decision-making, teams are found to be more *rational* in the sense that they are more cognitively sophisticated and make fewer errors. This is particularly true for "Eureka" problems, where the correct answer is obviously true once identified. Charness et al. (2007) consider a decision task where subjects learn the state of the world by drawing balls. They find that teams are better at performing Bayesian updating and making decisions based on the principle of first-order stochastic dominance. Charness et al. (2010) study teams' performance in the "Linda paradox" proposed by Tversky & Kahneman (1983). They find that teams are much more likely to choose the correct option that is statistically more probable, and are much less affected by the conjunction fallacy. In all these contexts, a clear "correct answer" exists, and team members are able to utilize the logic behind it in team discussions in order to drive and improve team decision outcomes.

For strategic decisions, several studies have shown that, compared to individuals, teams are found to be more *selfish* and *strategic*; teams pay more attention to monetary payoffs, and act much more consistently with what game theory predicts [Charness & Sutter (2012)]. In the market entrant signaling game where the weak incumbent can pretend to be the strong incumbent by setting a high production quantity, Cooper & Kagel (2005) find that teams are much more likely to play the equilibrium strategy, and this is driven by their ability to reason from the opponent's perspective. In the beauty-contest game, where subjects are asked to pick a number that is p times the average of all other subjects' numbers ($p < 1$), game theory predicts that everyone should pick the smallest possible number. Individuals typically pick a number that is far above the minimum, while teams tend to pick a significantly lower number [Kocher & Sutter (2005)]. In the trust game where the sender sends money to the receiver, followed by money returned by the receiver, a rational sender will send zero to the receiver because a rational receiver will not return any positive amount back to the sender. Hence, any deviation from the rational behavior must be due to the fact that the sender trusts the receiver and the receiver is being trustworthy. Individuals have been found to send and return significantly positive amounts of money, while teams tend to send or return lower amounts, showing a lower degree of trust or trustworthiness [Cox (2002), Kugler et al. (2007)].

Given its importance, the topic of team decision-making is relatively under-studied in BOM. To our knowledge there are just three existing papers that examine team decision-making in operations, in a Newsvendor setting, albeit under various experimental interventions: providing the subjects with multiple decision proposals [Gavirneni & Xia (2009)], endowing one team member with superior information [Laya & Pavlov (2015)], or placing subjects in a multi-round tournament setting where contestants (teams or individuals) are in competition and learn the winning decision made each round [Wu & Seidmann (2015)]. Our focus is different — I study a standard Newsvendor setting, along with the information sharing game. To align with the information sharing game, our Newsvendor setting is distinct in that the mean of demand changes in each round. I also design our study to uncover the decision mechanisms underlying the observed decisions using the team chats. Interestingly, our finding that the classical Newsvendor logic is not uniquely persuasive per se helps to organize some of the observations of team performance in these three papers, which find that team Newsvendors largely perform similarly to individuals, save one (of four) treatments in Wu & Seidmann (2015) where teams perform slightly better.

2.3 Decision Contexts and Hypotheses

I now introduce our study's decision context, a modified version of the information sharing game proposed by Özer et al. (2011). The information sharing game, when appropriately adjusted for experimental purposes, allows us to study both the standalone Newsvendor decision and the information sharing in the Newsvendor context.

The supply chain consists of one supplier (she) and one retailer (he). The supplier produces for the retailer who sells to the end customers. The end customer demand is uncertain. Relative to the supplier, the retailer has better demand information due to his proximity to the market. The retailer determines how truthfully he wants to share his demand information with the supplier. The supplier then interprets the information and makes her production decision. This is a one-shot game, so reputation does not play a role. To represent this in our experiments, retailers and suppliers are anonymous and they are randomly re-matched from round to round.

Formally, the end customer demand D equals $X + \xi$, where X and ξ are random variables. X is known as the *demand forecast*, and it follows the cumulative distribution function (c.d.f.) $F(\cdot)$ over support $[X_l, X_u]$. ξ is the market uncertainty and is distributed on $[\xi_l, \xi_u]$ with mean 0 and c.d.f. $G(\cdot)$. I assume $X_l + \xi_l > 0$ to ensure a positive demand. In our experiments, X is uniformly distributed between 100 and 400; ξ is uniformly distributed between -75 and +75.

The event sequence of the game is as follows: (1) The retailer privately observes the realized value of X, and he delivers a signal $\tilde{X}$ to the supplier. The demand uncertainty ξ remains uncertain to both the retailer and supplier. (2) The supplier receives the demand signal $\tilde{X}$, updates her belief about end demand, and makes her production decision Q (at cost $c \cdot Q$). (3) The demand uncertainty ξ is realized. The retailer orders D from the supplier, receives min(D, Q) from the supplier, pays the supplier $w \cdot \min(D, Q)$, and sells these units to the end customers to obtain revenue $p \cdot \min(D, Q)$. To keep the context simple, all the price and cost parameters are exogenous. The specific parameters I use in the experiment are: $p = 140, w = 100, c = 80$. The (prior) distributions of X and ξ, the event sequence, and the price/cost parameters are all common knowledge. The only piece of asymmetric information is the realized value of X privately observed by the retailer.

In our experiments, all the subjects play the above information sharing game *in two different settings*. In the first setting, the "computerized retailer setting", the computer plays the retailer role and all the subjects (teams and individuals) play as the supplier. The computer always reports the signal *truthfully* to the supplier, and the supplier knows this. Hence, the supplier knows the exact demand distribution when making the production decision Q. This setting corresponds to the standalone Newsvendor setting, and it allows us to study how teams and individuals make the Newsvendor decision. In the second setting, the "human retailer setting", both the retailer and the supplier are played by human subjects. This setting corresponds to the Newsvendor under demand information sharing, and it allows us to study how teams and individuals make strategic decisions in the context of inventory planning. In fact, for each team/individual, I need to combine the data from *both* settings to determine how strategic the team/individual is. This point will be made clear in Section 2.3.2. I now present the details of each setting.

2.3.1 The Computerized Retailer Setting

In the computerized retailer setting, because the computerized retailer always reports the signal truthfully and the supplier knows this, it is straightforward for the supplier to update her belief regarding the customer demand: The mean of demand is just the signal she receives. Let Q_{CR} be the supplier's production decision in the computerized retailer setting. Notice that X changes from round to round while the interval of demand uncertainty ξ does not. Namely, the optimal decision can be expressed as a function of X. With the parameters introduced above, the optimal production decision is: $Q^*_{CR} = X - 45$. Hence, I can use $(X - Q_{CR})$, the production adjustment relative to the mean X, to measure the supplier's performance in making the standalone Newsvendor decision. In the experiment, subjects make the standalone Newsvendor decision multiple times; so, the average value $\overline{(X - Q_{CR})}$ measures the supplier's overall ability in making the Newsvendor decision.

2.3.2 The Human Retailer Setting

In the human retailer setting, both the retailer and the supplier are played by human subjects. Let $\tilde{X}$ be the signal sent by the retailer, and Q_{HR} be the production decision made by the supplier in the human retailer setting after receiving $\tilde{X}$.

On the retailer's side, note that since he does *not* bear any inventory risk in this game, it is always beneficial for the retailer to report the highest amount possible, X_u, regardless of the true value of X, in order to induce the supplier to produce more for him. However, past behavioral research suggests that people are not totally untrustworthy, even in non-repeated game settings among strangers or anonymous participants [Fehr et al. (1998), Charness et al. (2004)]. Therefore, I may expect a spectrum of trustworthy behavior. A fully trustworthy retailer will simply set $\tilde{X} = X$ to tell the truth to the supplier. An *untrustworthy* retailer will distort the signal to act in his own benefit: The more untrustworthy he is, the more he will inflate the signal towards X_u. I use the degree of inflation $(\tilde{X} - X)$ to measure his degree of (un)trustworthiness.

On the supplier's side, she needs to consider two questions to determine Q_{HR}: (1) to what extent she trusts the signal sent by the retailer, i.e., how to update her belief regarding the end customer demand; and (2) how to make the production decision with the updated demand distribution. Her response to the first question depends on how trusting she is. If she is fully *trusting* towards the retailer, she will believe the signal to be true and set it to be the updated mean of demand. If she is fully *untrusting*, the signal is useless and she does not update her belief. Her response to the second question depends on how sophisticated she is in making the standalone Newsvendor decision.

In the human retailer setting, I only observe the supplier's final production decision Q_{HR}, which reflects her answer to both of the questions jointly. To disentangle the supplier's degree of *trusting*, I need to contrast her production adjustment $(\tilde{X} - Q_{HR})$ in the human retailer setting with her average production adjustment in the computerized retailer setting $(\overline{X - Q_{CR}})$, similar to Özer et al. (2011). Then, the difference $(\tilde{X} - Q_{HR}) - (\overline{X - Q_{CR}})$ measures the supplier's trusting behavior. If the supplier is very trusting, her behavior in the two decision contexts should be quite similar, indicating that $(\tilde{X} - Q_{HR}) - (\overline{X - Q_{CR}})$ should be close to 0. On the other hand, the more *untrusting* she is, the more she thinks the signal is inflated and adjusts further away from it by setting a lower Q_{HR}, which translates into a higher value of $(\tilde{X} - Q_{HR}) - (\overline{X - Q_{CR}})$.

2.3.3 Hypotheses in the Computerized Retailer Setting

I now hypothesize teams' behavior for the Newsvendor decision. Based on teams' generally superior performance in solving tactical problems (see Section 2.2), I conjecture that teams will outperform individuals in Newsvendor decision-making:

Hypothesis 1. *With the computerized retailer, teams make better Newsvendor decisions than individuals, i.e., teams' production decision Q_{CR} is closer to the optimum Q^*_{CR}.*

Given past research, the superior team performance should be driven by a corresponding decision process: A compelling argument should emerge both within team discussions and across different teams [Hill (1982), Cooper & Kagel (2005)]. Specifically, in a typical team, each team member may have different ideas about how decisions should be made. During team discussions, if a team member can propose an argument that the teammate finds sufficiently compelling, then the team will converge to making a decision driven by that compelling argument. When I consider this dynamic across different teams, it is possible that different teams may find different arguments compelling; however, if there exists an argument that is widely recognized by many teams to be compelling, then teams as a whole will perform differently from individuals. If this compelling argument points teams towards better decisions, then their overall performance will improve.

In the computerized retailer setting, the Newsvendor logic (tradeoff between overage and underage costs) is the key to making the optimal Newsvendor decision. I anticipate that teams are better able to utilize this logic to overcome the pull-to-center bias and make decisions closer to the optimum. From the above discussion, I can also summarize the three features for an argument to be compelling: (1) it should be frequently mentioned (by many teams); (2) when mentioned in team discussions, the other team members find it to be persuasive; (3) it drives the team's final decision outcomes. Therefore, I anticipate the arguments related to the classical Newsvendor logic should have all three features.

Hypothesis 2. *With the computerized retailer, in team discussions, the argument related to the classical Newsvendor logic is: (1) frequently mentioned (by many teams); (2) found persuasive in team discussions; (3) strongly driving team decision outcomes.*

2.3.4 Hypotheses in the Human Retailer Setting

Özer et al. (2011) and Özer et al. (2014) have shown that individuals exhibit a certain degree of trust and trustworthiness in the information sharing game. Similar observations have been made for teams in trust games [Cox (2002), Kugler et al. (2007)]. Therefore, regardless of whether the roles (retailer/suppliers) are played by teams or individuals, I should observe that as retailers the signals they send will be positively associated with the actual mean of demand X, and as suppliers their production decisions will be positively associated with the signals they receive from the retailers. This captures the idea that some useful information will be transmitted in the information sharing game.

Hypothesis 3. *In the human retailer setting, regardless of whether the roles (retailer/suppliers) are played by teams or individuals,*

(a) The signal $\tilde{X}$ sent by the retailer is positively correlated with the private demand forecast X.

(b) The supplier's production decision Q_{HR} is positively correlated with the signal $\tilde{X}$ she receives.

To compare team and individual performance in the context of trust in supply chains, a natural starting point is to consider the two "pure" supply chain configurations: pure-team (team retailer vs. team supplier); and pure-individual (individual retailer vs. individual supplier). The pure-individual configuration is a replication of the game in Özer et al. (2011) and Özer et al. (2014), and it serves as the baseline for our analysis. For the pure-team configuration, if I suppose teams are indeed more untrustworthy and untrusting as suggested by the behavioral economics literature (see Section 2.2), then I should observe a higher degree of untrusting and untrustworthiness compared with the pure-individual configuration. In addition, I also note that in strategic settings, the decision-maker's actions are driven by his/her (1) own decision tendency, and (2) strategic reaction to opponent behavior. Therefore, for any observed team-individual difference in this comparison, it is not entirely clear which of the two elements is the primary factor. To help disentangle this, I introduce two additional "mixed" configurations: team retailer vs. individual supplier, individual retailer vs. team supplier. This helps us to further explore whether and how teams' degree of untrustworthiness/untrusting changes, *as a function of the opponent identity*. Given that ours is among the first studies on this topic, I take the conservative approach to conjecture that the opponent identity does *not* affect teams' degree of (un)trusting and (un)trustworthiness. (Using the classical trust game, Kugler et al. (2007) also find that the opponent identity does *not* influence teams' degree of (un)trusting and (un)trustworthiness as the decision-maker.) I summarize the above discussions in the following hypothesis.

Hypothesis 4. *In the human retailer setting, compared with the pure-individual baseline:*

(a) as the retailer, regardless of the opponent identity, teams are more untrustworthy compared with individuals in the pure-individual configuration. That is, retailer inflation is higher in the pure-team and team retailer-individual supplier configurations.

(b) as the supplier, regardless of the opponent identity, teams are more untrusting compared with individuals in the pure-individual configuration. That is, supplier reductions are larger in the pure-team and individual retailer-team supplier configurations.

As in the tactical setting, I also anticipate that a team decision process driven by compelling strategic arguments will underlie the differences in behavior. In strategic settings, past studies have identified that the argument associated with being more selfish and caring less about members outside the group (opponent) usually stands out to be the compelling argument both within team discussions and across different teams. For example, in the classical trust game, teams are found to send and return less money because the arguments of being untrusting or untrustworthy stand out to be compelling in this setting [Cox (2002), Kugler et al. (2007), Charness & Sutter (2012)]. This has direct implications on our human retailer setting. On the retailer side, between the discussion of trustworthy vs. untrustworthy, I expect the argument of being untrustworthy should stand out to be the more compelling argument. Thus in the text chat analysis, the argument to be untrustworthy should be (1) frequently mentioned; (2) persuasive; (3) strongly driving decisions, while the argument to be trustworthy should fail to have at least one of these three features. On the supplier side, between the arguments of trusting vs. untrusting, similarly I should expect that only the argument of being untrusting has all three features and therefore is the compelling argument in team discussions. Thus I have the following hypothesis.

Hypothesis 5. *In human retailer team discussions, regardless of opponent identity, I expect:*

(a) for retailers, the argument to be untrustworthy is (1) frequently mentioned (by many teams); (2) found persuasive in team discussions; (3) strongly driving team decision outcomes; on the other hand, the argument to be trustworthy should lack at least one of the three features.

(b) for suppliers, the argument to be untrusting is (1) frequently mentioned (by many teams); (2) found persuasive in team discussions; (3) strongly driving team decision outcomes; on the other hand, the argument to be trusting should lack at least one of the three features.

2.4 Experimental Design

Our experiment proceeds in three stages, summarized in Figure 2.1. All subjects play both the computerized and human retailer games, with the majority of their decisions as either an individual or a team (based on their treatment). Specifically, all subjects first play the computerized retailer game individually for three periods. This serves as a training stage within our experiment, familiarizing subjects with the decision context of the experiment before potentially joining a team in the next stage. However, subjects do receive payments based on their decision outcomes in this stage. At the end of the first stage, depending on their treatment subjects are either paired into a team for the remainder of the session or remain as an individual decision-maker.

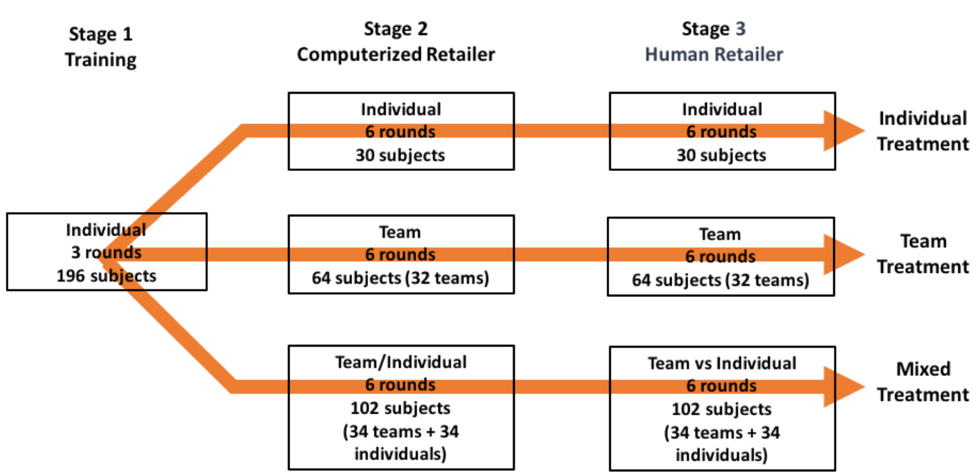

Figure 2.1: Experimental Design

In the second stage, all subjects play 6 rounds of the computerized retailer game. I can address our first research question of team versus individual performance in a tactical setting by comparing performance between team and individual decision-makers. Additionally, having all subjects play the computerized retailer game before the human retailer game allows subjects in teams to get comfortable with the team decision-making format in the simpler game, and it gives us a measure of their baseline inventory decisions with perfect information. This allows us to separate strategic versus tactical considerations in analyzing the human retailer game.

Finally, in the third stage, all subjects play six rounds of the human retailer game. The roles of supplier and retailer are randomly assigned at the beginning of each round. I conduct three between-subjects treatments: subjects play either the pure individual, the pure team, or the mixed treatment where teams play against individuals. To eliminate reputation effects retailers and suppliers are randomly and anonymously matched each period (with no consecutive pairings allowed). The 3 treatments I developed are given below:

1. **The individual treatment:** Subjects make decisions individually throughout the experiment. This can be considered a replication of the experiment in Özer et al. (2011).

2. **The team treatment:** Subjects are formed into two-person teams at the beginning of stage 2, and they keep making decisions in that team for the remainder of the experiment. So, in this treatment in the human retailer setting, I will have teams playing against teams.

3. **The mixed treatment:** At the beginning of stage 2, two-thirds of the subjects form two-person teams. The remaining one-third continue decisions as individuals. This assignment stays unchanged for the remainder of the experiment. In this treatment in the human retailer setting, I always have either team retailers playing against individual suppliers or the other way around.

I test Hypotheses 1-2 by comparing team and individual performance in the computerized retailer game, and test Hypotheses 3-5 by comparing trust and trustworthiness of teams to individuals in the human retailer setting.

The experiment timeline and treatment are public information for all the subjects. That is, subjects know whether they will play as a team/individual and whether they are playing against a team/individual. At the end of each round, subjects receive the following feedback information in a dashboard: the decision they have made in that round, the realized end-customer demand, the profits they have earned in that round and their accumulated profit over the previous rounds. To assist comparisons across treatments, in the experiment, I pre-generate a list of X and ξ, and apply them to all three treatments. The experiment is conducted in an economics laboratory in a large public university in the United States. The participants are undergraduate and graduate students from the university. The participation time is around 75 minutes. Subjects gain 5 dollars show-up fee upon joining the experiment and earn additional payments based on their performance in the experiment. The average payoff is 20 dollars. I use Z-tree to program the experiment [Fischbacher (2007)] and ORSEE for subject recruitment [Greiner (2004)].

2.4.1 The Setup of Teams

In in the team treatment and mixed treatment, at the beginning of stage 2 (computerized retailer setting), the subjects are randomly paired to form two-person teams. They keep making decisions in the same team for the remainder of the experiment. In our experiments, team members do not sit physically together; instead, they interact with each other using a computer program, where a text chat function is provided to assist the communication process. (There is no cross-team chat communication.) This is a way to help us externalize their thought processes. In Section 2.5 and 2.6, I will make use of these chats to study teams' decision mechanisms.

In order to eliminate potential confounding factors and to keep the team decision context simple, in this chapter I consider a small team setup (two members in each team) with aligned information and aligned interests. Specifically, inside a team the information is fully transparent: The two team members observe the same information and get the same feedback. The two team members receive the same payoff as the team's payoff, i.e., the payoff is not split between the two team members; this is to align team members' payoffs and to keep the monetary payoff comparable across treatments. This setup has the following advantages. (1) A small team setup is simple to work with, and past research has shown that increasing the team size from team of one to two is enough to generate significant behavioral differences [Cooper & Kagel (2005), Charness et al. (2007)]. (2) This setup allows us to focus on the question of whether or not teams, when fully aligned, make better decisions than individuals, and removes the effect that teams may perform worse due to misaligned incentives or untruthful sharing of information within the team.

In our experiments, I employ a consensus-based decision-making scheme: The two team members need to jointly agree on a value (Q if supplier, $\tilde{X}$ if retailer) in order to form a team decision. The detailed rule is as follows: The team decision procedure consists of two types of actions — making a proposal and agreeing to a proposal. When a round starts, the two team members chat with each other to discuss, and they can make numerical proposals to reflect their thoughts. The latest proposals from both of them will be shown on the screen. The final team decision is made when one team member picks the latest proposal from the other team member and confirms it. If no confirmation is made within the given time limit, the computer will randomly pick one member from the team and use his/her latest proposal as the team's decision. If the round ends before either team member makes a proposal, the computer will randomly generate a number from the full possible demand range (i.e., between 25 and 475). Both teams and individuals are given two minutes in each round. Subjects are given one additional minute in the *first round* in stage 2 and stage 3. Each round ends once all the decisions have been made.

Subjects in teams have two means of reaching a decision: sending a numerical proposal and sending a text message. Because text chat analysis will be a major source of insight into the decision-making process, it is important to note that it is the primary means by which teams reach consensus. In our experiments, on average teams make 1.57 proposals in each round, i.e., most teams make 1 or 2 numerical proposals in each round. Recall how teams make decisions in our experiment: A team decision is reached only when one member's numerical proposal is accepted by the other team member. Hence, the *minimal* number of numerical proposals to reach a team decision in each round is 1. The fact that I see the majority of proposals being 1 or 2 shows that there is little back-and-forth when making numerical proposals. From a complementary angle, I find that most of the discussions regarding how decisions should be made happen via text chats. In particular, chats that contain the quantity of production decisions are among the most frequently seen chats (such chats constitute $\frac{1}{8}$ of the total chats in the standalone Newsvendor task). Therefore, even when trying to communicate the production quantities they desire, team members rely on text chats rather than numerical proposals. Hence, text chats are a good representation of how teams make decisions in our experiment.

Finally, for the decision time I find that more than 97$\%$ of teams are able to reach an agreement within the given time limit; in addition, in both the standalone Newsvendor and the information sharing in the Newsvendor contexts, on average teams spend less than 35$\%$-45$\%$ of the time given. Hence, most team decisions come from collaboration and relatively quick agreement by members, rather than extensive internal conflict. In other words, I find evidence that most teams are able to form a consensus in their decision-making processes.

2.5 Analysis in the Computerized Retailer Setting

I now discuss the results in the computerized retailer setting (standalone Newsvendor task).

2.5.1 Hypothesis 1: Decision Outcomes in the Computerized Retailer Setting

The optimal ordering decision in the experiment is $Q^*_{CR} = X - 45$; thus, the pull-to-center bias discussed before Hypothesis 1 means that subjects order close to the mean X and do not make a large enough reduction from the mean. Hereafter I simply use "reduction" to refer to the supplier's production adjustment relative to the signal $\tilde{X}$ she receives. In the computerized retailer setting, the signal is the same as the true demand mean X; therefore, the reduction is just the reduction from X. If Hypothesis 1 is true, teams should make a *larger* reduction from the mean when ordering, making their orders *closer* to the optimum.

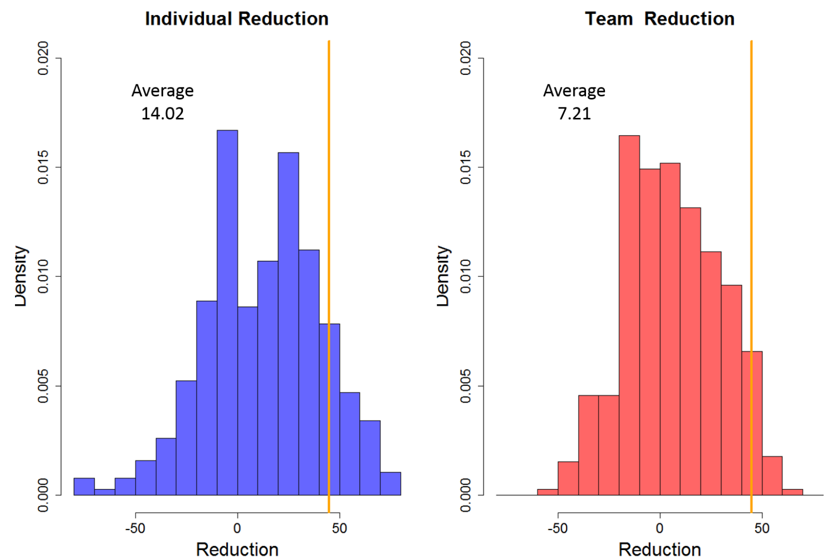

Figure 2.2: Reduction in the Computerized Retailer Setting

Figure 2.2 depicts the observed stage 2 reduction relative to the demand mean X for individuals and teams. The optimal level of reduction (45 units) is depicted using the orange lines. On average, individuals reduce 14.02 units from the mean, which is smaller than the optimum and is too close to the mean, consistent with the pull-to-center bias. Quite surprisingly, teams do not reduce more than individuals. Instead, there is more pull-to-center bias exhibited, with teams reducing on average by 7.21 units. The Wilcoxon rank-sum test on average reduction confirms the difference to be significant ($p = 0.02$). To study this more formally, I conduct a random-effects regression analysis. I consider the function:

$$Q_{CR,it} = \text{Intercept} + \beta_X \cdot X_t + \beta_{team} \cdot Team_i + \beta_t \cdot t + v_i + \epsilon_{it}. \tag{2.1}$$

In this and all the subsequent regressions, t is the round and i is the "decision-making unit" (either the individual or the team, depending on the treatment). Given that two members inside a team share the same information to reach a mutual agreement to produce a team decision and receive the same payoff, in the regression I only have *one* data point for each team in a round. The variable $Q_{CR,it}$ represents the decision-making unit i's production decision made in round t in the computerized retailer setting. X_t represents the demand forecast, which is also the mean of customer demand subjects receive in round t. Notice that in our design X_t is identical for all teams/individuals in the same round. $Team_i$ is the dummy variable for the decision-making unit i being a team or not. The round variable t controls for time trends. v_i is the decision-unit specific error term, and ϵ_{it} is the independent error term.

Table 2.1: Computerized Retailer Setting: Newsvendor Decision Regression Analysis

Variable	Estimate of the Coefficient
X_t	$1.00(0.01)^{***}$
$Team_i$	$6.81(2.96)^{**}$
t	$-0.20(0.56)$
Intercept	$-13.62(3.88)^{***}$

Notes: Random-effects GLS regression with balanced panel data, clustering on the decision unit level. 778 observations over 6 rounds, from 66 teams and 64 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1,^{**} p < 0.05,\ ^{***}p < 0.01$.

The results of Regression (2.1) are summarized in Table 2.1. I see that the intercept is significantly negative, meaning that on average, subjects do reduce from the mean to set their production decisions. However, the estimate for β_{team} is significantly positive, indicating that teams on average reduce less from the mean, hence end up being further away from the optimum. Hence I reject Hypothesis 1 — teams do *not* make better Newsvendor decisions than individuals. (I refrain from concluding that teams perform worse than individuals for three reasons: (1) as will be discussed below, teams and individuals end up earning similar profits; (2) in our follow-up experiment where subjects make many more Newsvendor decisions, teams end up performing similarly as individuals; see Section 2.7 for details; (3) Hypothesis 1 mainly concerns whether teams will make *better* decisions; therefore, concluding that teams make no better decisions than individuals is consistent with rejecting the null hypothesis.)

Finally, I further analyze the profit implication of team decision-making in standalone Newsvendor tasks. I focus on the expected profit of subjects' production decisions to remove the impact of randomness. Furthermore, because the mean of demand X changes from round to round, making the scaling of profit change along with it, I use normalization to make the profit comparable across rounds to be suitable for running panel regression. Specifically, I divide subjects' expected profit in each round by the *optimal expected profit* they could have earned in that round (derived from setting the reduction to be 45 units). The result is a ratio smaller or equal to 1 where 1 represents earning the optimal profit; I call this the "normalized profit ratio".

I find that, on average, teams have a similar normalized profit ratio as individuals (0.65 versus 0.66). The following random-effects GLS regression (see (2.2)) confirms that the difference between teams and individuals is insignificant (estimate for $\beta_{team} = -0.01$, p-value > 0.70; clustering is conducted on the decision unit level). Hence, the profit analysis further confirms that teams indeed make no better standalone Newsvendor decisions compared to individuals.

$$ProfitRatio_{CR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + \beta_t \cdot t + v_i + \epsilon_{it}. \quad (2.2)$$

2.5.2 Hypothesis 2: Team Decision Mechanism in the Computerized Retailer Setting

Studying the decision mechanism has been the central topic in many behavioral studies. Yet, because individual subjects' thought processes cannot be directly observed, the current methods researchers have tried are either prohibitively expensive (e.g., the neuroscience approach) or inaccurate (for example, the "think aloud" approach that requires people to write what they think when making decisions). In this section, with the help of team chats I am able to directly study the decision mechanism in a Behavioral OM problem. Recall that in our design I require team members to reach a *mutual agreement* to form a team decision with the help of textual communication (chats). This design makes the team chats an important part of teams' reasoning processes. Hence, by studying the team chats and relating them to the team's decision outcome, I will be able to analyze the team decision mechanism.

I employ two main usages of the text chat analysis. First, ex-ante, I expect Newsvendor logic to be a compelling argument in team discussions by having all three features: (1) frequently mentioned (by many teams); (2) found persuasive in team discussions; (3) strongly driving team decision outcomes. Chat analysis helps us to directly address our conjectures. Importantly, I can then connect the results with teams' behavior, as a way to understand the mechanism behind differences/indifferences between teams and individuals. Second, text chat analysis also helps to explore whether other arguments stand out as compelling in team discussions.

5.2.1. General Method Introduction

Following the method in Cooper & Kagel (2005), I use a three-step approach to conduct the chat analysis. This general approach is applied to both the standalone Newsvendor decision-making and information sharing in a Newsvendor context (discussed in Section 2.6).

In the first step, I develop a coding scheme (classification of the text chats) based on an initial review of the chats and various Newsvendor decision theories. The full coding scheme can be found in Appendices A.1 and A.2. (The complete chat data is also available from the authors upon request.) I have two coding schemes: one for the standalone Newsvendor task (computerized retailer setting), and one for the information sharing game (human retailer setting). Each code in the coding scheme captures a type of argument in team discussions. Table 2.2 provides examples of the codes in the standalone Newsvendor coding scheme.

Table 2.2: Examples of Codes and their Identification

NV Decision Formulation Codes	Identification of the Code
Newsvendor Economic Reasoning	Discussing balancing of overage and underage costs.
Durable Strategy	Formulating a concrete, long-lasting strategy in making production decisions.
Aggressive	Expressing the desire be aggressive by making a large production decision.
Conservative	Expressing the desire to be conservative by making a low production decision.
Waste Aversion	Expressing aversion for potential waste due to over-production.
Stock-out Aversion	Expressing willingness to produce more in order to avoid stockout.
Risk Seeking	Explicitly mentioning that they want to take more risk.
Risk Averse	Explicitly mentioning that they want to be risk-averse or be safe.
Loss Aversion	Expressing aversion for loss in profits.
Mean Anchoring	Using mean (X) of demand as the decision benchmark and making production adjustments.
Forecast-dependent Strategy	The value of the forecast (mean X) affects teams' production decisions.
Demand Chasing Low	Using low demand realization in the past to predict the demand to be low again.
Demand Reversal Low	Using mean low demand realization in the past to predict the demand to be high.

Second, I recruit five coders; three using the Newsvendor coding scheme to code the computerized retailer setting in the experiment, and two using the information sharing game coding scheme to code the human retailer setting. I take the average of the three coders' results in the Newsvendor coding work, and the average of the other two coders' results in the information sharing game coding work, as the coding results for our further analysis. The five coders are trained with the experiment background and the coding scheme, and they conduct the coding work *independently*. I use two metrics to evaluate coder consistency — simple correlation and Cohen's (joint) Kappa test — and find good consistency among the coders. Details of the tests are provided in Appendix A.3. To better understand the coding work, consider the following dialogue from a team in the computerized retailer setting. A and B refer to the two members of the team.

> A: "I think its always better to be a bit conservative on the numbers."
>
> B: "Okay, that's fine with me."
>
> A: "Because you don't lose as much when you under produce as you do when you over produce."

The chat "always better to be a bit conservative on the numbers" expresses the desire to be conservative, and also uses it as a decision strategy that should be carried out in future rounds (notice the word "always"); therefore, it is coded as *Conservative* and *Durable Strategy*. In the chat "you don't lose as much when you under produce as you do when you over produce", team member A reasons using the Newsvendor logic; therefore, it is coded as *Newsvendor Economic Reasoning*.

In the third step, I explore the coding results and relate them with teams' decision outcomes to determine whether and how the arguments will influence teams' final decisions. For each argument (as represented by a code), I will analyze whether it has any of the three features for it to be compelling in team discussions: (1) frequently mentioned (by many teams); (2) persuasive; (3) driving decisions. This concludes the general method of conducting chat analysis in our study.

Below, I employ this general method to the standalone Newsvendor context to understand our surprising finding: Teams fail to outperform individuals. Section 5.2.2 addresses features (1) and (2), and Section 5.2.3 addresses feature (3).

5.2.2. Newsvendor Argument Frequency & Persuasiveness Analysis

Table 2.3: Computerized Retailer Setting (Stage 2): Chat Analysis Coding Results in the Experiment

Code	Experiment Frequency	Number of Teams Having Mentioned the Code	Intra-Team Acceptance Rate
Conservative	55	39	75%
Aggressive	28	29	70%
Risk Averse	26	23	71%
Mean Anchoring	24.33	26	74%
Loss Aversion	14.33	18	67%
Durable Strategy	8.67	15	62%
Risk Seeking	6.67	8	70%
Mean-Dependent Strategy	6.67	9	85%
Newsvendor Economic Reasoning	5	7	60%
Demand Reversal Low	4	5	83%
Demand Reversal High	2.67	7	87.5%
Demand Chasing Low	0	0	NA
Demand Chasing High	0	0	NA
Stock-out Aversion	0	0	NA
Waste Aversion	0	0	NA

Note: There are in total 66 teams in the experiment. The non-integer frequency for some codes is due to the fact that I am averaging across three independent coders. The experiment frequency could be smaller than the number of teams having mentioned the code when not *all* three coders attach the code to the chat. For example, if only 1 of the 3 coders attach a code to a piece of chat, the resulting frequency (from this piece of chat) is 0.33, but I will consider to be mentioned by one (more) team.

When a code applies to a discussion, I say the discussion "mentions" that code. For a code to be frequently mentioned by many teams (feature (1)), it should both have a high coding frequency among all the codes and be a popular discussion topic for many different teams (codes that frequently appear but are relegated to only a few teams would not satisfy feature (1)). Per columns 2 and 3 in Table 2.3, I find alignment between these two measures: codes that are frequently mentioned overall are also mentioned by many teams. Our results make clear that the classical Newsvendor logic fails to be a frequent argument: the Newsvendor Economic Reasoning code has a low coding frequency (only 3% of all codes) and is only mentioned by a few teams. Instead, I observe that the discussions are mainly related to five arguments: two from general mental dispositions (*Aggressive/Conservative*), two from utility preferences (risk preferences, loss aversion), and one from mean anchoring (*Mean Anchoring*).

To operationalize the analysis for persuasiveness (feature (2)), I ask a coder to track each piece of coded chat and consider the response from the other team member. There are three possible results: (i) accept, meaning that the other team member agrees with what is proposed; (ii) reject; (iii) no direct response. I call the rate of acceptance the "intra-team acceptance rate" for each code. The fourth column of Table 2.3 presents the results. Overall I find that most arguments are at least somewhat persuasive: Intra-team acceptance rates are higher than 50% for all codes with positive frequency. However, Newsvendor Economic Reasoning has the *lowest* intra-team acceptance rate among all codes. Therefore the Newsvendor logic, with its low frequency and relatively weak persuasiveness, could end up having very little impact on teams' decision outcomes, despite its importance in enabling teams to overcome the pull-to-center bias and derive the optimal Newsvendor solution. I will formally analyze this using regression analysis below.

5.2.3. Newsvendor Decision Driver Regression Analysis

To conduct a formal analysis regarding how different arguments affect teams' final decision outcomes (feature (3)), I consider the following regression model(s): I incorporate coding frequency as additional independent variable(s) into Regression (2.1). I run it with only the data from teams (therefore, the team dummy variable in Regression (2.1) is not included). The results are almost unchanged if I include a team dummy variable and run it with the full data set (including individuals). I consider the impact of the codes in the round that it is mentioned, i.e., the coding frequency variables are round-specific (with subscript it). Significant coefficients therefore indicate that team decisions move in a systematic direction when a particular topic is discussed. Hence this analysis shows which arguments influence team decision-making outcomes.

In the regression, I consider two types of specifications: (a) include only a single chat code at a time; (b) include all codes jointly. For example, with only the code *Aggressive*, I have:

$$Q_{CR,it} = \text{Intercept} + \beta_X \cdot X_t + \beta_t \cdot t + \beta_{Agg} \cdot Aggressive_{it} + v_i + \epsilon_{it}. \tag{2.3}$$

This specification is simply the original Regression (2.1) with the additional $Aggressive$ variable, which refers to the coding frequency of the code $Aggressive$ for team i in round t. Table 2.4 presents the regression results for the top five (by frequency) chat codes, along with *Newsvendor Economic Reasoning*. The remaining regression results are omitted for brevity but are available upon request.

The results are presented in Table 2.4. I confirm that conversing about *Newsvendor Economic Reasoning* does *not* have a significant impact on decision outcomes - its estimate is not significant

Table 2.4: Computerized Retailer Setting (Stage 2): Newsvendor Chat Regression Analysis

Code	(1)	(2)	(3)	(4)	(5)	(6)	Full
Nvdr. Econ. Reas.	-0.93(10.00)						6.47(10.88)
Loss Aversion		-2.01(6.19)					5.68(6.87)
Mean Anchoring			3.36(2.99)				0.37(3.77)
Aggressive				24.78(4.96)***			20.30(5.96)***
Conservative					-16.98 (3.38)***		-15.46(3.42)***
Risk Averse						1.95(3.78)	4.66(3.21)

Notes: Random-effects GLS regression with balanced panel data, clustering on the decision unit level. 395 observations over 6 rounds, from 66 teams. Columns 2-7 report the estimates for the specification with a single code included. Column 8 (Full) reports the results with all codes included. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

and the magnitude of the estimate is also much smaller than other major codes. On the other hand, I find that the codes *Aggressive* and *Conservative*, which capture subjects' general mental dispositions to make a large or low production decision, are driving teams' decisions away from the mean of demand X. Each time a team talks about being conservative, on average, its decision goes down by 16.98 units, which brings the team closer to the optimum. Each time a team talks about being aggressive, its decision increases by 24.78 units, bringing it further away from the optimum. Discussion of *Mean Anchoring* does not drive decisions to deviate from the mean, which is expected given its purpose. Discussion of *Loss Aversion* or *Risk Averse* does not have a significant impact on teams' decision outcomes.

In summary, I reject Hypothesis 2: I find that among the three features necessary for an argument to be compelling, classical Newsvendor logic lacks two features and performs weakly on the third. Instead, I find that the arguments related to general mental dispositions (*Aggressive* and *Conservative*), utility preferences (*Loss Aversion*, *Risk Averse*), and anchoring on the mean (*Mean Anchoring*) are most frequently mentioned and found persuasive by teams. Among them, the arguments of general mental dispositions strongly influence teams' final decision outcomes, but only the argument of *Conservative* drives teams' decisions to be closer to the optimum.

2.6 Analysis for the Human Retailer Setting

Recall that for the human retailer setting I use a two-by-two experimental design (team/individual retailer versus team/individual supplier), resulting in four supply chain configurations. Figure 2.3 visually overviews our experimental results, using box plots to depict the decision outcomes for retailers and suppliers, for the four configurations. (Abbreviations are used to denote different configurations; for example, TR-IS stands for the team retailer-individual supplier configuration. Blue boxes are for individuals as the decision-maker, and red boxes are for teams are the decision-maker.) Besides the percentiles (as shown in the box plots), I also include the mean inflation/reduction.

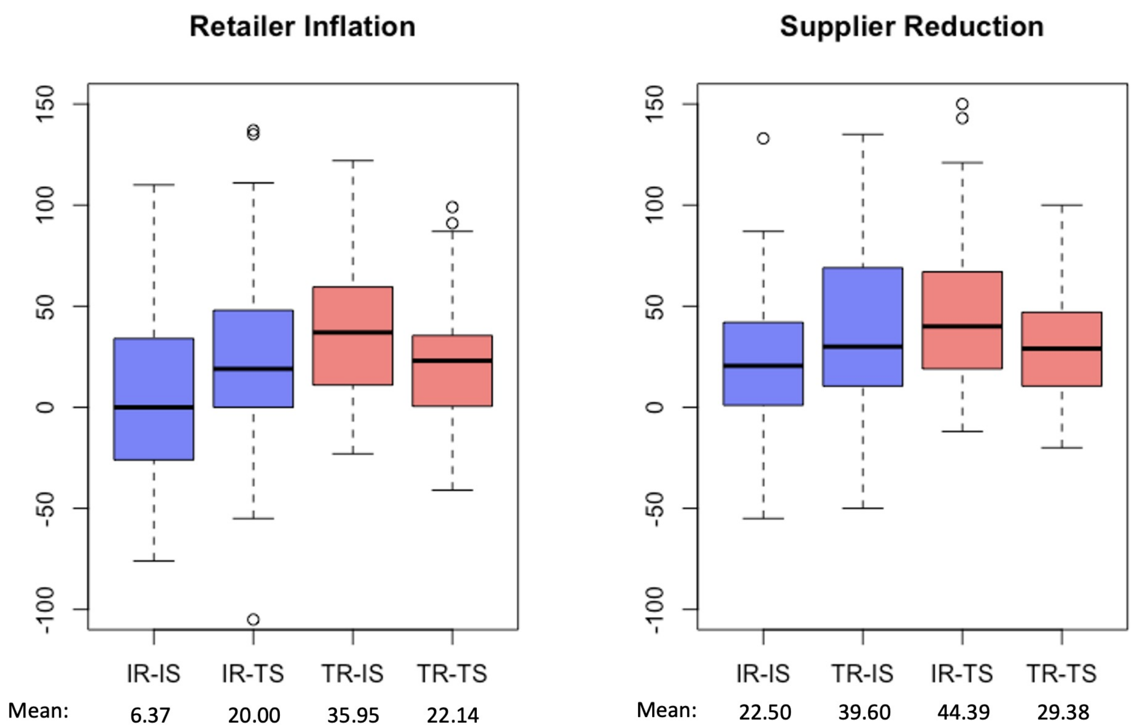

Figure 2.3: Boxplots for Decision Outcome Comparisons in the Human Retailer Setting.

2.6.1 Hypothesis 3: Information Transmission in Supply Chains

Hypothesis 3 covers whether useful information is transmitted in supply chains. To test Hypothesis 3a, I examine "retailer inflation", $\tilde{X} - X$, the amount that retailers inflate their demand signals over the true demand forecast. A value of 0 means that retailers are fully trustworthy. I observe that retailers in all four supply chain configurations on average tend to inflate (are untrustworthy) when reporting their signals $\tilde{X}$. However, I also note that retailers rarely inflate dramatically relative to the true demand signal X, i.e., their reported signals are largely based on the true demand signals; this is further validated by the high correlations between $\tilde{X}$ and X in all of the four supply chain configurations (ranging from 0.70 to 0.90). I thus confirm Hypothesis 3a. To test Hypothesis 3b, I examine "supplier reduction", $\tilde{X} - Q$, the amount of reduction suppliers made in their production decisions relative to the signals they receive from the retailer. From Figure 2.3 I also observe that few suppliers make extremely large/small reductions, and moreover I also find a high correlation between $\tilde{X}$ and Q in all four supply chain configurations (ranging from 0.82 to 0.92). I therefore confirm Hypothesis 3b. To summarize, although retailers inflate their forecasts, their signals are nonetheless informative and suppliers interpret them as such.

2.6.2 Hypotheses 4 and 5: Trust/Trustworthiness of Teams and Individuals

I now turn to Hypothesis 4 and 5 that cover retailers and suppliers' trust/trustworthiness behavior. For retailers, I note from Figure 2.3 that while the median retailer in the pure-individual configuration makes zero inflation, the median retailer played by teams makes positive inflation. That is, compared against the baseline pure-individual configuration, team retailers seem to be more untrustworthy. For suppliers, I find that team suppliers make larger reductions (are more untrusting) compared to suppliers in the pure-individual configuration, although I need to control for their standalone Newsvendor decision-making behavior and the retailer's signal to be able to draw meaningful conclusions. Below, I first use regression to study Hypothesis 4 that relates to decision outcome analysis. I then explore the decision mechanism, i.e., test Hypothesis 5.

For expositional clarity, I present the analysis separately for the retailer and supplier. I also explore the *profit implication* of subjects' decision-making. Doing so requires us to put both sides of the supply chain together; therefore, I delay the related results to an independent Section 2.6.3.

2.6.2.1 Hypothesis 4a: Decision Outcome for Retailers

I consider Regression (2.4). The variable $\tilde{X}_{it}$ refers to the signal that the retailer i sends in round t, and the other variables are as defined in Regression (2.1). In our primary specification I follow Özer et al. (2014) in regressing the signal sent on the true mean X, which allows for a more flexible estimation. The estimation results remain mostly unchanged if I directly consider the amount of inflation ($\tilde{X} - X$) as the dependent variable.

$$\begin{aligned}\tilde{X}_{it} =& \text{Intercept} + \beta_X \cdot X_t + \beta_{team} \cdot Team_i + \beta_{VT} \cdot VersusTeam_i \\ &+ \beta_{TvT} \cdot Team_i \cdot VersusTeam_i + t \cdot \beta_t + \epsilon_{it}.\end{aligned} \tag{2.4}$$

With this regression, the linear combination of $\beta_{team} + \beta_{VT} + \beta_{TvT}$ measures the difference in retailer signal inflation between the pure-team configuration and the pure-individual configuration; the estimate of β_{team} measures the difference between the Team Retailer-Individual Supplier configuration and the pure-individual configuration.

The results summarized in Table 2.5 formalize our earlier observations about the left panel of Figure 2.3: Compared with the baseline pure-individual configuration, the retailer signal inflation

Table 2.5: Human Retailer Setting: Retailer Signal Decision Regression Analysis

Variable	Estimate of the Coefficient
Intercept	$36.45(9.26)^{***}$
X_t	$0.88(0.02)^{***}$
t	$2.58(0.80)^{***}$
$Team_i$	$25.93(8.19)^{***}$
$VersusTeam_i$	$9.09(8.19)$
$Team_i \cdot VersusTeam_i$	$-21.47(11.42)^{*}$
$Team_i + VersusTeam_i + Team_i \cdot VersusTeam_i$	13.55(8.54)

Notes: Random-effects GLS regression with unbalanced panel data. 390 observations over 6 rounds, from 66 teams and 64 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$,$^{**}p < 0.05$, $^{***}p < 0.01$. For completeness, I report all the pair-wise comparison results in Appendix A.4.1.

is (directionally) larger in the pure-team configuration (estimate=13.55, p-value=0.11) and significantly larger in the team retailer-individual supplier configuration (estimate=25.93, p-value$<$0.01). A joint hypothesis test supports the conclusion that the two pair-wise comparisons are jointly significant (the null hypothesis that the difference is insignificant in both pair-wise comparisons is rejected at p-value$<$0.01). Hypothesis 4a's general conjecture that team retailers are less trustworthy holds, although the opponent identity does play a role in the magnitude of the effect. It is also worth noting that Regression (2.4) reveals a strong learning effect: the estimate for the round variable $t = 2.58$, p-value$<$0.01. This suggests that, with experience, retailers understand the benefit of inflating signals; they thus keep pressing harder in this direction over time.

2.6.2.2 Hypothesis 5a: Decision Mechanism for Team Retailers

I now study how team retailers make their decisions. I use the same chat analysis approach as I did in Section 2.5.2, except that here I apply the human retailer setting coding scheme (available in Appendix A.2.).

Coding results are presented in Table 2.6. I observe that arguments related to strategic reasoning (*Own Strategy*, *Opponent*, *Feedback*) and trustworthiness (*Untrustworthy*, *Trustworthy*) are indeed frequently mentioned both in the sense of having a high coding frequency and being mentioned by many different teams, and I note that the top 5 codes and their ordering are identical between the two opponent identities. This suggests that strategic reasoning and issues related to trust/trustworthiness are indeed central to retailer decision-making. In addition, I observe that both the arguments of *Untrustworthy* and *Trustworthy* have high intra-team acceptance rates, with *Untrustworthy* being slightly more persuasive.

Table 2.6: Retailer Side Coding Frequency and Persuasiveness Analysis

Retailer Side	Individual as the Opponent	Team as the Opponent	Total Frequency	Number of Teams Having Mentioned this Code	Intra-Team Acceptance rate
Own Strategy	55	67	122	51	75.0 %
Opponent	45	57	102	39	65.0%
Untrustworthy	30	35.5	65.5	41	79.2 %
Feedback	18	28.5	46.5	33	76.3%
Trustworthy	11	13	24	21	70.8%
Future Strategy (Same Role)	0.5	7.5	8	10	93.3%
Regret Aversion	2.5	2	4.5	5	66.7%
Future Strategy (Opposite Role)	4	3.5	7.5	5	60.0%
Autocorrelation	0	0	0	0	NA

Note: There are 66 teams in total.

Hypothesis 5a states that the argument to be *Untrustworthy* should (1) frequently appear in chats, (2) be found persuasive, and (3) affect retailers' decisions. From Table 2.6, I confirm (1) and (2) for *Untrustworthy*. Meanwhile, I note that for the counter-argument *Trustworthy*, its coding frequency is indeed smaller (about half of *Untrustworthy*) and subjects generally found it to be (directionally) less persuasive. However, it is certainly *not* the case that arguments related to trustworthiness are never mentioned, nor do subjects find them to be totally unpersuasive. This is in fact consistent with our results related to Hypothesis 3, where I confirm that effective information transmission is still taking place in supply chains.

Nonetheless, after further investigation, I find that these two arguments are mostly mentioned when it is early in the game: 53$\%$ (58$\%$) of the mentions of *Untrustworthy* (*Trustworthy*) happen when teams play the role of retailers *for the first time*. As teams gain more experience, they gradually stop mentioning *Trustworthy* while they continue to talk about *Untrustworthy*: Only 19$\%$ of *Trustworthy* happen in the second half (rounds 4-6) of the information sharing game, while 34$\%$ of *Untrustworthy* happen in the second half. Hence, between the two arguments, team retailers seem to gradually converge to the argument of *Untrustworthy*. Along with the fact that *Untrustworthy* has a much higher coding frequency compared to *Trustworthy*, our results here suggest that subjects seem to find *Untrustworthy* to be the single compelling argument between the two, especially when they gain experience in the game.

Building on this, I conduct a regression analysis to formally consider the decision drivers in the information sharing game. To test Hypothesis 5a part (3), I focus on analyzing the trustworthiness dynamics in the information sharing game by including the related codes.

$$\begin{aligned}\tilde{X}_{it} =& \text{Intercept} + \beta_X \cdot X_t + \beta_t \cdot t + \beta_{UTR} \cdot Untrustworthy_i \\ &+ \beta_{VT} \cdot VersusTeam_i + v_i + \beta_{UTRVT} \cdot Untrustworthy_i \cdot VersusTeam_i + \epsilon_{it}.\end{aligned} \tag{2.5}$$

$$\begin{aligned}\tilde{X}_{it} =& \text{Intercept} + \beta_X \cdot X_t + \beta_t \cdot t + \beta_{TR} \cdot Trustworthy_i \\ & + \beta_{VT} \cdot VersusTeam_i + v_i + \beta_{TRVT} \cdot Trustworthy_i \cdot VersusTeam_i + \epsilon_{it}.\end{aligned} \tag{2.6}$$

As before, I use only the team data for our regressions. For retailers, the effect of *Untrustworthy* is estimated with Regression (2.5). The estimate for β_{UTR} captures the effect of *Untrustworthy* when the opponent is an individual supplier; the linear combination for $\beta_{UTR} + \beta_{UTRVT}$ captures the effect of *Untrustworthy* when the opponent is a team supplier. The effect of *Trustworthy* is similarly considered in Regression (2.6). I note that the coding frequency of *Untrustworthy* and *Trustworthy* are considered as the *summation* across all rounds for team i in the information sharing game. This summation configuration is useful when subjects' discussion of codes tend to happen only in a certain part of the game, which fits the situation on the retailer side. Recall our earlier results that more than half of the mentions of trustworthy/untrustworthy happen when teams play the role of retailers for the first time. The regression result with the summation configuration will demonstrate whether mentioning this argument will have a *long-lasting impact* on retailers' decision-making.

The results are summarized in Table 2.7. Directionally, talking about being untrustworthy increases the amount of inflation, while talking about being trustworthy decreases it. However, for both opponent identities, only the argument of being untrustworthy is significant (p-value $<$ 0.05). Therefore, between the arguments of *Trustworthy* and *Untrustworthy*, *Untrustworthy* is the only argument that can drive team decision-making outcomes in the information sharing game, consistent with Hypothesis 5a. (I further note that if I instead use the round-specific configuration of these two codes, the qualitative nature of our results still hold, but the estimate for *Untrustworthy* is no longer significant (p-value $>$ 0.20). The reason is a mismatch between the time trend of chats and inflation behavior: Retailers tend to inflate more over time, but the coding frequency of *Untrustworthy* is predominantly in the early part of the game. Meanwhile, the fact that its estimate is significant in the summation configuration demonstrates that teams tend to internalize the understanding of the benefit of *Untrustworthy* such that: (1) teams who have *ever* mentioned *Untrustworthy* have an overall higher level of inflation; (2) in later rounds, teams can continue to inflate (more) and do so without having to mention *Untrustworthy*.)

Finally, I further study why *Trustworthy* does not have a significant impact on inflation. As stated above, *Trustworthy* is infrequently mentioned when subjects gain experience in the information sharing game. This suggests that although teams may initially find the argument to be trustworthy compelling, they "abandon" this argument as they gain more experience, leaving *Untrustworthy* the only compelling argument. To further validate this, I compare the group of retailers who have talked about *both* being *Trustworthy* and *Untrustworthy* with the group of retailers who have only talked about being *Untrustworthy* throughout the information sharing game. I find that

Table 2.7: Human Retailer Setting (Stage 3): Team Retailer Strategic Behavior Opponent Effect Analysis

Untrustworthy	Individual Supplier as the Opponent	Team Supplier as the Opponent
Total Coding Frequency	30	35.5
Corresponding Estimate	9.11(4.77)**	7.20(3.46)**
Trustworthy	Indivdiual Supplier as the Opponent	Team Supplier as the Opponent
Total Coding Frequency	11	13
Corresponding Estimate	-7.50(8.68)	-4.23(5.60)

Notes: Random-effects GLS regression with unbalanced panel data. 199 observations over 6 rounds for each regression, from 66 teams. Robust standard errors in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the degree of inflation between these two groups is very similar (p-value > 0.50 in panel regression for the dummy variable distinguishing between these two groups; this conclusion is valid either when I condition on the opponent identity or I pool all the data together). That is, mentioning *Trustworthy* does *not* have a significant impact on the retailer's strategy when *Untrustworthy* is also present; the group that mentions both arguments quickly converge to recognizing *Untrustworthy* as the single compelling argument and therefore end up with a similar degree of signal inflation. Overall, I conclude that *Untrustworthy* is indeed the single compelling argument between the two in determining the degree of inflation. The evidence supports Hypothesis 5a.

2.6.2.3 Hypothesis 4b: Decision Outcome for Suppliers

Recall that supplier reduction, $\tilde{X} - Q$, is the amount of reduction suppliers made in their production decisions relative to the signals they receive from the retailer. Compared to the baseline pure-individual configuration, Hypothesis 4b predicts that supplier reductions will be greater in supply chain configurations where the supplier is a team. By revisiting Figure 2.3, I can see that this is at least directionally true when I compare the pure-team configuration with the pure-individual configuration (mean comparison 29.38 vs. 22.50) and when I compare the individual retailer-team supplier configuration with the pure-individual configuration (44.39 vs. 22.50), although I also note that the difference seems quite weak for the former comparison.

To formalize this, I utilize Regression (2.7) where variable $Q_{HR,it}$ refers to supplier i's production decision in period t in the human retailer setting, and $\tilde{X}_{it}$ refers to the signal that supplier i receives from the human retailer. I use $(\overline{X - Q_{CR}})_i$, supplier i's average Newsvendor reduction in the computerized retailer setting, to control for heterogeneity in Newsvendor behavior. I include both $\tilde{X}_{it}$ and $(\overline{X - Q_{CR}})_i$ as independent variables to keep the estimation flexible. Directly using the amount of reduction $(Q - \tilde{X})$ or the "net difference" $((\tilde{X} - Q_{HR})_{it} - (\overline{X - Q_{CR}})_i)$ as the independent variable will not change our conclusions.

$$Q_{HR,it} = \text{Intercept} + \beta_{signal} \cdot \tilde{X}_{it} + \beta_{team} \cdot Team_i + \beta_{VT} \cdot VersusTeam_i + \beta_{TvT} \cdot Team_i \cdot VersusTeam_i + \beta_{NV} \cdot (\overline{X - Q_{CR}})_i + t \cdot \beta_t + \epsilon_{it}. \tag{2.7}$$

The linear combination $\beta_{team} + \beta_{VT} + \beta_{TvT}$ measures the difference in reduction between the pure-team and the pure-individual configurations; the estimate of β_{team} measures the difference between the individual retailer-team supplier configuration and the pure-individual configuration.

Table 2.8: Human Retailer Setting: Supplier Production Adjustment Decision Regression Analysis

Variable	Estimate of the Coefficient
Intercept	$17.94(7.99)^{**}$
$\tilde{X}_{it}$	$0.90(0.02)^{***}$
t	$0.21(0.73)$
$(\overline{X - Q_{CR}})_i$	$-0.78(0.14)^{***}$
$Team_i$	$-23.15(6.75)^{***}$
$VersusTeam_i$	$-17.30(6.68)^{***}$
$Team_i \cdot VersusTeam_i$	$33.64(9.33)^{***}$
$Team_i + VersusTeam_i + Team_i \cdot VersusTeam_i$	-6.81(7.02)

Notes: Negative estimates means a lower production quantity, i.e., a larger reduction adjustment from the signal. Random-effects GLS regression with unbalanced panel data. 390 observations over 6 rounds, from 66 teams and 64 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$. For completeness, I also report all the pair-wise comparison results in Appendix A.4.1.

The regression results reported in Table 2.8 confirm our previous observations from Figure 2.3. Compared with the baseline pure-individual configuration, team suppliers indeed make larger production reductions in both the pure-team configuration and the individual retailer-team supplier configuration; however, the difference in the former comparison is directional and highly insignificant: estimate= -6.81, p-value>0.30. Therefore, I find support for Hypothesis 4b when the opponent is an individual retailer, but *not* when the opponent is a team retailer. I also note the lack of learning effect on the supplier side (estimate for the time variable t =0.21, p-value>0.70). That is, suppliers do not become increasingly untrusting even after they gain experience.

2.6.2.4 Hypothesis 5b: Decision Mechanism for Team Suppliers

I now analyze the decision mechanism on the supplier side through chat analysis. The general method and analysis sequence here is the same as for the retailer side. I note that I also coded for the pure Newsvendor behavior on the supplier side, but find the coding frequency to be quite small compared to the strategic and trust-related coding results ($< 10\%$). This suggests that in the human retailer setting, the discussions related to trust/strategic issues become first-order; the discussions related to standalone Newsvendor decision-making generally falls away. Therefore, in the analysis below I do not include the Newsvendor-related coding results, and instead focus on presenting the coding results from the information sharing game coding scheme.

In Table 2.9 I find that, similar to the retailer side, the arguments related to strategic reasoning (*Own Strategy*, *Opponent*, *Feedback*) and trust (*Trusting*/*Untrusting*) are indeed frequently mentioned both in the sense of having a high coding frequency and being mentioned by many different teams, consistent with what I have conjectured. I also note that the top five codes are identical on both sides of the supply chains (three related to strategic issues, plus two trust/trustworthiness related codes); these 5 codes together constitute more than 85% of the total frequencies.

Table 2.9: Supplier Side Coding Frequency and Persuasiveness Analysis

Supplier Side	Individual as the Opponent	Team as the Opponent	Total Frequency	Number of Teams Having Mentioned this Code	Intra-Team Acceptance rate
Own Strategy	47.5	43.5	91	44	67.0%
Opponent	39	43.5	82.5	41	60.6%
Feedback	31	23	54	32	77.8%
Untrusting	23	27.5	50.5	29	69.3%
Trusting	4.5	18	22.5	21	57.8%
Future Strategy (Same Role)	0	2	2	2	25.0%
Regret Aversion	1.5	4	5.5	5	54.6%
Future Strategy (Opposite Role)	1	6.5	7.5	9	66.7%
Autocorrelation	0	0	0	0	NA

Notes: There are 66 teams in total.

Hypothesis 5b conjectures that arguments to be *Untrusting* will be frequently mentioned and found persuasive by suppliers. This is confirmed in Table 2.9. However, I note that the counter-argument of *Trusting* also has a high coding frequency and is also persuasive. Therefore, it is certainly not the case that subjects completely reject the idea of being trusting, and both *Trusting* and *Untrusting* are potentially influential on teams' final decision outcomes. When I consider the timing of when these two codes are mentioned during the information sharing game, interestingly, I find the dynamics on the supplier side to be quite different from those on the retailer side. Specifically, for suppliers I observe that: (1) they do not tend to talk about *Trusting* and *Untrusting* in the same round; (2) the discussion between trusting vs. untrusting are less restricted to the first round teams act as the supplier and, instead, tend to continue in the second half of the game (rounds 4-6): 38% of *Untrusting* and 36% of *Trusting* happen in the second half. Hence, even towards the later rounds of the information sharing game, team suppliers do not seem to settle on either one of the two arguments and instead find *both* arguments compelling.

I then conduct a regression analysis to formally consider the supplier's decision drivers in the information sharing game. Akin to our analysis for the retailer, I conduct an analysis for the codes

Untrusting and *Trusting* with Regressions (2.8) and (2.9), respectively. Here, however, their coding frequencies are *round-specific*. This is motivated by the observation that related discussions happen throughout the information sharing game, and I try to capture such dynamics.

$$\begin{aligned} Q_{HR,it} =& \text{Intercept} + \beta_{signal} \cdot \tilde{X}_{it} + +\beta_{NV} \cdot (\overline{X - Q_{CR}})_i + \beta_{UTS} \cdot Untrusting_{it} \\ &+ \beta_{VT} \cdot VersusTeam_i + \beta_{UTSVT} \cdot Untrusting_{it} \cdot VersusTeam_i + t \cdot \beta_t + v_i + \epsilon_{it}. \end{aligned} \tag{2.8}$$

$$\begin{aligned} Q_{HR,it} =& \text{Intercept} + \beta_{signal} \cdot \tilde{X}_{it} + +\beta_{NV} \cdot (\overline{X - Q_{CR}})_i + \beta_{TS} \cdot Trusting_{it} \\ &+ \beta_{VT} \cdot VersusTeam_i + \beta_{TSVT} \cdot Trusting_{it} \cdot VersusTeam_i + t \cdot \beta_t + v_i + \epsilon_{it}. \end{aligned} \tag{2.9}$$

Table 2.10: Human Retailer Setting (Stage 3): Team Supplier Strategic Behavior Opponent Effect Analysis

Untrusting	Individual Retailer as the Opponent	Team Retailer as the Opponent
Total Coding Frequency	23	27.5
Corresponding Estimate	-15.53(4.42)***	-9.14(3.96)**
Trusting	**Individual Retailer as the Opponent**	**Team Retailer as the Opponent**
Total Coding Frequency	4.5	18
Corresponding Estimate	32.86(12.25)***	15.85(4.49)***

Notes: Random-effects GLS regression with unbalanced panel data. 197 observations over 6 rounds for each regression, from 66 teams. Robust standard errors in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. Note that a negative estimate means a larger reduction from the signal $\tilde{X}$, and vice-versa.

The regression results are summarized in Table 2.10, and they reveal a qualitatively different story than what I saw for the retailer (Table 2.7). I now have *both* the arguments of *Trusting* and *Untrusting* strongly driving decision outcomes; this is true for both opponent identity conditions. Recall that teams do not tend to talk about both *Trusting* and *Untrusting* in the same round. In other words, it is *not* the case that one argument is "objecting to" the other argument; instead, different teams find both of them persuasive in different rounds, and they are thus both influential on final decision-making. As a result, both arguments are compelling during team discussions, contrary to what I expected, and the results do *not* support Hypothesis 5b. In addition, I note that *Trusting* is much more likely to be mentioned when the opponent is a team (frequency 18 vs. 4.5), along with a larger overall impact on decision outcomes ($18 \cdot 15.85 > 4.5 \cdot 32.86$). This explains the lack of (extra) reductions from team suppliers when the opponent is a team.

2.6.3 Expected Profit Analysis

In this subsection, I present here a formal analysis of subjects' expected *profits* in the information sharing game. In doing so, I recognize that it is vital to put both sides of the supply chains together so that I can consider the joint effect of their decision-making. To see this, consider the retailer side as an example. The retailer's profit is dependent on the supplier's response to the signal she receives and her subsequent production decisions. If the team retailer's untrustworthy behavior provokes an untrusting production adjustment from the supplier, then the team retailer may not end up benefiting from this in terms of profits. In analyzing the effectiveness of subjects' decision-making, I will also see that results from analyzing the team decision mechanism play a key role in helping us understand whether retailers and suppliers can benefit from team decision-making.

I consider the expected values of the profits by taking the expectation over the possible demand realizations. The supply chain profit is the sum of retailer and supplier profits. I normalize subjects' expected profits, given the changing mean of demand and the changing roles, to make performance more comparable across rounds. For the expected profit of each party (retailer, supplier, integrated supply chain), I respectively normalize (divide) them by the optimal profit each party can earn in the corresponding round. The results are summarized in Table 2.11 for each party and separately for the four supply chain configurations. (As a special note, I see that team suppliers do *not* earn high profits in the individual retailer-team supplier configuration despite their very untrusting behavior there. This is driven by the prevalence of *overly* large reductions from the true demand mean X, and the fact that the supplier's profit function is unimodal and peaks at 45 units; 12% of the reductions from the true demand mean X in that configuration are larger than 90 units, resulting in a significant profit loss.)

Table 2.11: Two-by-Two Decision Effectiveness Analysis

Retailer Average Relative-to-Optimum Ratio	Team Retailer	Individual Retailer
Team Supplier	0.90	0.86
Individual Supplier	0.90	0.88
Supplier Average Relative-to-Optimum Ratio	**Team Retailer**	**Individual Retailer**
Team Supplier	0.75	0.75
Individual Supplier	0.67	0.76
Supply Chain Average Efficiency Ratio	**Team Retailer**	**Individual Retailer**
Team Supplier	0.95	0.91
Individual Supplier	0.92	0.93

Notes: Cells report normalized profit relative to optimum possible profit. For retailers, optimal profit is if supplier produces such that demand is always satisfied. For suppliers, optimal profit is when the reduction from *true demand mean* X is 45 units. For the supply chain, optimal profit is a reduction from the *true demand mean* of $\frac{75}{7}$ units.

Managers evaluating whether or not to implement team decision-making need to consider both (1) whether the company itself should choose to use a team or individual decision-maker, and (2) how the opponent's identity (team/individual) affects this decision. To evaluate this, I continue to use regressions to conduct pair-wise comparisons among the four supply chain configurations, separately for retailers and suppliers. The results and the corresponding regression methods are provided in Appendix A.4.2; below I refer to the comparison results when needed.

For retailers, I observe that they have strong incentives to choose to be a team, and this is robust to the identity of the opponent. Specifically, I note that the two configurations with team retailers have the highest average expected profit ratios (0.90 for both of them). When compared with the two configurations with individual retailers, the difference is directional when compared with the pure-individual configuration (0.90 vs. 0.88, p-value>0.20 in both comparisons) and significant compared with the Individual Retailer-Team Supplier Configuration (0.90 vs. 0.86, p-value<0.03 in both comparisons). These two configurations themselves are very similar (0.90 vs. 0.90, p-value>0.50). From a more general perspective, these results suggest that implementing team decision-making "statistically dominates" individual decision-making - the two configurations with team retailers are better or no worse than the two configurations with individual retailers. Therefore, these results suggest a simple strategy for retailers: They can implement team decision-making without worrying too much about their supply chain partner's configuration/reaction.

On the supplier side, I find that suppliers may also prefer to be a team. Both configurations with team suppliers have higher expected profit ratios compared to the Team Retailer-Individual Supplier configuration (0.75 vs. 0.67 in both comparisons) while being similar to the pure-individual configuration (0.75 vs. 0.76 in both comparisons). However, regression analysis reveals that none of these comparisons are significant (p-value>0.20). This may seem a bit surprising given that the difference in the former comparisons (0.75 vs. 0.67) seems quite large. However, I find that the supplier expected profit ratios are highly variable, which ultimately drives the comparisons to be insignificant. (In fact, the variation in all four configurations is so strong that *none of* the pair-wise comparisons is significant.) Specifically, I find that the standard deviations of expected profit ratios range from 0.38 to 0.47 on the supplier side; on the retailer side, they only range from 0.08 to 0.13. In general, the ambiguous nature of suppliers' strategic environment creates large variation in suppliers' decisions and effectiveness of their decisions (expected profits), leading to the result that suppliers fail to gain a systematic advantage from implementing team decision-making. Nonetheless, given that retailers have strong incentives to choose to be a team, directionally it is still better for suppliers to be a team as well (0.75 vs. 0.67). Hence, suppliers may still want to implement team decision-making.

The above results stand in parallel with the results in chat analysis. Recall that, from chat analysis, I find the retailer's strategic problem is more straightforward, with team retailers quickly converging on being untrustworthy (i.e. inflating more). On the other hand, suppliers find their situations to be more ambiguous, with many teams switching between being trusting and untrusting, and teams continuing to find both being trusting and untrusting compelling even at the end of the experiment. These results help us understand why team retailers are able to consistently benefit from team decision-making, while team suppliers suffer from the large variation in their performance (expected profits) which curbs an advantage from team decision-making.

Finally, some managers may be interested in total rather than individual profits. This is evaluated by the supply chain efficiency ratio results in Table 2.11. Would such managers draw different conclusions regarding whether to implement team decision-making? In our experiments I find that, interestingly, the two goals turn out to be aligned. To see this, I note that the pure-team configuration has the highest efficiency among all the configurations; the difference is significant against the Individual Retailer-Team Supplier configuration (0.95 vs. 0.91, p-value=0.04) and directional in the other two configurations (0.95 vs. 0.92, 0.95 vs. 0.93; p-value>0.25 in both comparisons). The intuition is as follows. Note that the supply chain optimal effective reduction is $\frac{75}{7}$, which is much smaller than the optimal effective reduction for suppliers (45 units); therefore, achieving supply chain efficiency requires the retailer's role in inducing the supplier to have a lower effective reduction, but not so low that it gets drastically smaller than $\frac{75}{7}$. Recall our results from decision outcome analysis: Teams are more untrustworthy and untrusting compared to individuals, but the effect is moderated in the pure-team configuration, especially for suppliers. Hence, the pure-team configuration achieves such balance, pushing the effective reduction to the level that is closest to the supply chain optimum in our experiments.

2.6.4 Summary and Further Discussions

To summarize the results in the information sharing game, I find that teams will perform differently from individuals *to the extent that a compelling argument emerges in team discussions*. On the retailer side, as I expected, team retailers are indeed more untrustworthy (i.e., inflate more) compared to individual retailers. From chat analysis, I find that the problem is more straightforward for retailers: *Untrustworthy* emerges as the compelling argument over *Trustworthy* when teams gain more experience. On the supplier side, however, team suppliers find the situation to be more ambiguous: Many teams switch between being trusting and untrusting, and they find both of *Trusting* and *Untrusting* compelling even at the end of the experiment. The high frequency and impact of *Trusting* in the pure-team configuration explains the result that team suppliers' decisions are *not* statistically different from individuals. The result on the supplier side echoes the results in the standalone Newsvendor task. There, I find that teams perform no better than individuals, and the *Newsvendor Economic Reasoning* is neither prevalent nor particularly persuasive in team discussions; meanwhile, teams find many other arguments compelling, such as mental dispositions of *Aggressive* and *Conservative*. In general, when teams find multiple, potentially conflicting arguments to be compelling, they may not have a performance advantage over individuals.

Building on this, I also identify a close connection between the results from chat analysis and the results from expected profit analysis. I find that team decision-making benefits the company (in terms of improving profit) when the compelling arguments *unanimously* point to the direction of performance-improvement. For suppliers in the standalone Newsvendor and information sharing game, as I have stated above, the compelling arguments point to different directions of decision-making. Teams therefore are unable to gain a systematic profit advantage over individuals. For team retailers in the information sharing game, the argument of *Untrustworthy* is consistent with the direction of performance-improvement; the fact that it wins over the argument of *Trustworthy* leads teams to earn more profits. For managers who wish to implement team decision-making, it is therefore important to carefully consider what the compelling argument(s) would be, and whether they will promote the outcome the company wants.

2.7 Robustness Check: The Extended Newsvendor Experiment

In this section I further explore our previous results for the standalone Newsvendor setting, where Hypotheses 1 and 2 were fully rejected (in contrast to the human retailer setting where Hypotheses 3, 4, and 5 had at least partial support). The finding that teams fail to outperform individuals in the standalone Newsvendor task is surprising, especially with the finding that the classical Newsvendor logic does *not* emerge as the compelling argument as I expected. Yet, it is valuable to consider the robustness of this finding from two other angles. First, it is possible that when more decision-making opportunities are presented, teams will be able to eventually learn and find this logic to be persuasive and therefore outperform individuals. Second, so far I have only examined the low critical ratio condition. It is possible that Newsvendor logic is a more compelling argument in the high critical ratio case. This motivates us to conduct a second wave of experiments - the extended Newsvendor experiment - as a robustness check for our results in the standalone Newsvendor task.

2.7.1 Design and Results of the Extended Newsvendor Experiment

In the extended Newsvendor experiment, I drop the human retailer setting (stage 3) of the experiment and focus solely on the computerized retailer setting (stage 2). The experiment is conducted under two parameter conditions: (1) the low critical ratio condition (LCR) condition, where the parameters are the same as the main experiment ($w = 100, c = 80$), and (2) the high critical ratio condition (HCR), where I set $w = 40, c = 8$ for a critical ratio of $\frac{4}{5}$ and the optimal decision being $X + 45$. Within each parameter condition, I establish two treatments: the individual treatment and the team treatment. Both treatments are identical to the corresponding treatment in the main experiment, except that stage 2 is extended from 6 rounds to 20 rounds, and I use a new sequence of X and demand realizations; this sequence is kept the same across sessions for different treatments and parameter conditions (LCR or HCR) for experimental control purposes.

Table 2.12: Extended Newsvendor Regression Analysis

Variable	LCR Condition	HCR Condition
X_t	$0.99(0.01)^{***}$	$1.02(0.01)^{***}$
$Team_i$	$-5.08(5.95)$	$-4.56(5.84)$
t	$-0.37(0.15)^{**}$	$0.24(0.13)^{*}$
Intercept	$-1.00(4.79)$	$7.36(4.35)^{*}$

Notes: Random-effects GLS regression with balanced panel data, clustering on the decision unit level. In LCR, I have 617 observations over 20 rounds, from 15 teams and 16 individuals. In HCR, I have 639 observations 16 teams and 16 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1,^{**} p < 0.05$, $^{***}p < 0.01$.

Next, similar to Section 2.5, I consider subjects' expected profits and normalize them by dividing the optimal expected profit in the corresponding round. I then apply Regression (2.2) to LCR and HCR separately. Again I find that teams do *not* earn more profits compared to individuals, in either LCR or HCR (p-value for the team dummy variable $>$ 0.10 in both conditions). Hence, I confirm that teams indeed make no better Newsvendor decisions compared to individuals even with extensive opportunities to learn and explore.

Because the extended experiment is structurally the same as the computerized retailer setting in the main experiment, I can apply the same analysis method, i.e., apply Regression (1) to rounds 4-23 in the extended Newsvendor experiment. The results are summarized in Table 2.12. I find that the estimate for the $Team$ variable is insignificant and small in magnitude in both LCR and HCR, suggesting that teams perform similarly compared to individuals. Hence, our conclusion in the main experiment extends to the extended Newsvendor experiment: Teams perform *no better* than individuals in standalone Newsvendor decision-making.

I also conduct text chat analysis similar to the main experiment, and confirm that our conclusions from the main text hold in the extended Newsvendor experiment. The results are summarized in Appendix A.5. As a summary of the key results, I find that: (1) the argument of *Newsvendor Economic Reasoning* has no significant impact on decision outcomes, in either HCR or LCR; (2) *Aggressive* in HCR and *Conservative* in LCR strongly drive team decision outcomes. Hence, I continue to conclude that *Newsvendor Economic Reasoning* is *not* a compelling argument in team discussions, even when enough decision opportunities are present. On the other hand, general mental dispositions are still found compelling in team discussions.

2.8 Conclusion

In this chapter, I study how teams make decisions in two canonical operational contexts: standalone Newsvendor and Newsvendor with information sharing. I find that whether teams will perform differently from individuals depends critically on the team decision mechanism, as reflected in what team members find compelling in intra-team chats. When playing the role of the supplier, both in the standalone Newsvendor setting and the information sharing game, teams find multiple (potentially conflicting) arguments compelling, which pull team decisions in different directions, and overall the teams do not outperform individuals. By contrast, as the retailer in the information sharing game, the compelling argument points to the direction of self-interest, and team retailers outperform individual retailers in terms of earning profits. Our study offers a general research methodology for studying team decision-making in operations, both in terms of how team-based experiments can be designed and how systematic analysis can be conducted to analyze both decision outcomes and the team decision mechanism. Future studies can continue to explore team decision-making in other operational contexts. In this process, I hope that our methodology will provide a useful point of reference.

In our experiments, I require the two members of a team to reach a *mutual agreement* to form a team decision. Of course, there exist other team decision rules, for example, the majority rule. This rule is not quite applicable to our experimental setting, given that each team only has two members. It will be interesting to study how our results and insights can be extended to the context of large group decision-making. Another important decision rule is the "authority rule with discussion", where team members discuss and report their opinions to the manager, who is the *only person* responsible for making the decision. This decision rule is effectively an *individual decision-making* rule, while the focus of this chapter is *team decision-making*.

Future research can extend our work in many directions. One possibility is to consider operational contexts where members of the same team have direct conflicts of interests and preferences, which is not the focus of this chapter. De-biasing tools for teams could be investigated. Incorporating de-biasing treatments and subsequently analyzing team chats can help us identify the tools that are more efficient in facilitating logical reasoning. Being among the first studies in BOM to explore team decision-making, and given the prevalence of teams in real-world business and operations decision-making, I believe there are plenty of opportunities for future research building on our work.

CHAPTER 3

Human Decision-Making in Dynamic Resource Allocation

3.1 Introduction

Many operational contexts involve dynamic decision-making. Examples include inventory planning, resource allocation, and dynamic pricing decisions. In the authors' recent interaction with a major automotive OEM in the United States, their product (vehicle program) development process involves making sequential decisions on how to allocate limited financial resources to improve the product design over the course of the development process. This problem belongs to an important class of operational problems - *dynamic resource allocation under a budget constraint*. For the OEM, due to practical challenges in implementing decision automation, all such decisions are currently — and for the foreseeable future will continue to be — handled by human managers instead of algorithms. Therefore, the product development manager's ability to conduct careful planning is important to the success of the product; yet, given that this is a dynamic process rife with uncertainty, managers could be affected by decision biases or could employ simplifying heuristics when making decisions. Understanding their behavior and developing managerial interventions to improve managers' performance is therefore important. In this chapter, I focus on how human decision-makers actually make such dynamic decisions and the design of effective managerial treatments for performance improvement. I address the following three research questions. (1) How will subjects perform (relative to the optimal policy[1]) in such dynamic resource allocation problems? (2) If subjects deviate from the optimal policy, what is the underlying mechanism/drivers? (3) How should I design managerial treatments to achieve performance improvement? Answering these three research questions will generate direct managerial benefits for the OEM and also offer insights for future research on human dynamic decision-making in operations.

[1]I obtain the optimal policy by solving a standard formulation of the problem where the decision-maker seeks to maximize expected payoff.

Specifically, in this chapter I consider the following stylized dynamic resource allocation problem, motivated by our interactions with the OEM. The decision-maker in our experiment is the product development manager. During the multi-year process of product development, design improvement opportunities (or "opportunities" for brevity) arise with varying benefits. The manager is given a limited budget to implement these opportunities; therefore, naturally not all the opportunities can be pursued. Due to the long engineering lead-time of implementing the opportunities and the fast-progressing engineering process, the manager needs to decide right away whether to implement the opportunity when presented to him/her, and the opportunities are considered lost if not immediately implemented. In other words, it is not possible to "wait and bundle" these opportunities. For the manager, the goal is to allocate the limited budget to maximize the total benefit collected from these opportunities. This setup has the advantage of capturing the key challenge the OEM faces in reality while being a well-defined dynamic programming (DP) problem, which allows us to establish a clear benchmark to evaluate human subject performance. I also note that this model extends beyond the context of product development management, and has direct application in other important managerial topics as well, such as in project management, revenue management, and financial investment (see Section 3.3).

I consider two conditions of this dynamic resource allocation problem representing different degrees of complexity. The first condition, which I call the "Constant Cost" condition, is the simplest condition I study. In this condition, the cost to implement each opportunity is constant throughout the project. Studying this condition helps us to understand how subjects will behave when given the best chance to perform well. Since the cost is constant, subjects can focus on capturing opportunities with decent benefit values, i.e., being "selective" in the opportunities to implement. In the second condition, which I call the "Increasing Cost" condition, I incorporate another practical consideration from the OEM's setting: for many of their products, the cost to implement opportunities increases later in the project. In the Increasing Cost condition of our study, the cost to implement opportunities doubles when entering the second half of the project. This seemingly minor change in the cost structure dramatically changes the optimal policy. It introduces an additional tradeoff between capturing benefits and spending budget when the cost is still low in the first half of the project, making it more challenging for subjects to perform well. The direct consequence of this is that subjects can no longer afford to be so "selective" early in the project and should instead spend the majority of their budget during the first half of the project.

Using laboratory experiments, I find that subjects' behavior in the Constant Cost condition aligns with the general structure of the optimal policy; they are sufficiently patient and selective in implementing opportunities. In the Increasing Cost condition, however, subjects perform substantially worse and exhibit a high degree of individual heterogeneity in performance. One major source behind the poor performance is that a sizable portion of subjects is too selective (relative to the optimal policy) in the first half of the project. They end up saving too much budget before entering the second half of the project, where the cost is substantially higher; their overall performance suffers as a result. After considering a wide range of behavioral mechanisms, I find that many bottom performers either significantly overestimate the continuation value in solving the dynamic problem, or they simplify the problem by following the "Selective Heuristic": They focus on implementing the opportunity with a high benefit value and are unresponsive to the cost-increase information.

On the other hand, to get insight into improving subject performance in the Increasing Cost condition, I look at the decision rules reported by top performers. I find that the very best performers also conduct simplification. Interestingly, they derive another decision heuristic which effectively decomposes the problem into simpler sub-problems. In particular, they divide the project into two halves based on whether or not the cost has increased. Then, within each half of the project, the cost is constant, making this (sub-problem) effectively a Constant Cost condition problem which they can solve quite optimally. Meanwhile, subjects utilizing the decision decomposition approach need to decide how much budget to allocate to the two halves of the project; I find those top performers are able to allocate the budget in a way that reflects the cost-increase information.

Motivated by this observation, in designing managerial treatments, I explore how I can share the decision decomposition idea to future subjects with improve their performance. I operationalize this by guiding subjects to conduct budget planning: Subjects are asked to consider how much budget to allocate before and after the cost increase. Such budget planning clearly decomposes the problem into two sub-problems based on the cost-increase information. Subjects with a clear budget-spending goal in each half of the project should then be able to act as if they were solving a Constant Cost condition problem.[2] In the two managerial treatments I propose, I vary the extent to which I guide subjects to conduct such budget planning. In the first treatment, the "decision prompt" treatment, I simply introduce the decision decomposition idea by asking subjects to conduct budget planning at the beginning of every project. I also remind them of the cost increase information, but I do not share any specific budget allocation plans. Interestingly, subjects land on many different budget allocation plans, some of which are unresponsive to the cost differences between the two sub-problems (such as allocating an equal amount of budget to the two halves of the project despite their cost differences in implementing opportunities); subjects in this treatment therefore fail to achieve a systematic performance improvement. In the second treatment, the "sharing best practices" treatment, I share the specific way top performers allocate their budget to the two halves of the project; I also provide the reasoning behind their budget planning as

[2] I note that conditioning on subjects following a specific budget allocation plan, there are multiple ways to implement the plan in every step of decision-making which will then lead to varying performance outcomes. In other words, our treatment is far from dictating every move of subjects' decision-making.

offered by the top performers. Subjects are free to decide whether or not to follow the suggested budget plan and, if so, how to carry out the plan in every step of their decision-making. I find that this treatment very effectively improves subjects' overall performance and results in a substantial portion of subjects acting close to the optimal policy.

To summarize, our results suggest that subjects can perform reasonably well in dynamic resource allocation problems with simple structures (such as the Constant Cost condition in our setting). This suggests that humans understand the key "now vs. future" tradeoff in dynamic resource allocation problems, and they understand the importance of saving. For complex dynamic resource allocation problems, subjects perform worse overall and they exhibit a large degree of heterogeneity in the way they approach the problem. Interestingly, some subjects are able to simplify the problem in reasonable ways and achieve good (or even close-to-optimum) performance. Specific to our context, a small portion of subjects effectively decompose the complex Increasing Cost condition problem into simpler, manageable sub-problems (Constant Cost condition problems). More interestingly, many subjects can benefit from being advised to follow such a heuristic and achieve close-to-optimum performance. Our results shed light on the conditions where humans can perform well in dynamic resource allocation problems and provide guidance on possible directions in designing managerial treatments for performance improvement.

3.2 Literature Review

The type of decision I consider (dynamic resource allocation under a financial constraint) belongs to the general category of dynamic decision-making problems. Dynamic decision problems have been studied in behavioral and experimental economics, although with very different contexts and models. Specifically, two types of problems have been extensively studied: dynamic consumption [Hey & Dardanoni (1988), Brown et al. (2009), Duffy (2016)] and dynamic search [Hey (1982), Shin & Ariely (2004), Schunk & Winter (2009)]. The first type is related to a central model in macroeconomics: How a representative individual makes his/her consumption and saving decisions over time. The subject makes decisions in T periods. In every period, s/he receives an income stream (whose value is random ex-ante), and decides how much to save versus consume. The goal is to maximize the total discounted utility from all T periods. Given that the individual's utility function is concave in the amount of consumption in every period, it is optimal to save some income when the realized income is high so that s/he can spend a decent amount of money when the income turns out to be low in other periods, i.e., the "consumption smoothing" behavior. A very robust finding has been that human subjects spend too much when the income is high and therefore do not save enough for the future. This phenomenon can be explained by the limited-forward looking bias: Subjects are not able to correctly account for the potential benefits from the future and therefore act myopically [Duffy (2016)]. Our model is structurally different. I consider how subjects consume one given income (budget) toward various beneficial opportunities with ex-ante random benefits (utilities). With these differences, I find that behavior in our setting is qualitatively different than in the consumption literature (see the end of the next paragraph).

In the second type of problem studied in behavioral and experimental economics - dynamic search - the individual engages in a sequential search process for a list of options with ex-ante unknown utilities; the actual utility of each option is revealed after search. The subject searches these options sequentially, pays a cost for each additional step of searching, and decides when s/he should stop searching and pursue the best among the searched options. The goal is to maximize utility, taking into account the cost of search [Hey (1982)]. In this type of decision task, the robust finding has been that subjects do not search enough [Seale & Rapoport (1997), Seale & Rapoport (2000)]. In a survey paper, Schunk & Winter (2009) compare five types of decision rules, with the purpose of identifying the decision policies/heuristics subjects are employing. Many of the decision rules are specific to the search context, such as the heuristic to continue searching until finding an option with a certain utility (threshold heuristic), or the heuristic to stop searching when the latest searched option is less valuable than the previous option. Among them, limited forward-looking types of decision rules again stand out as the decision rules that very well describe a decent proportion of subjects' actions.[3] Our decision context is structurally different. Subjects in our context will necessarily go through *all* the decision periods and will eventually observe the expected benefits from all the opportunities. Therefore, the key tradeoff is not when to stop searching; instead, subjects focus on allocating the budget to implement (a subset of) opportunities. With such differences, I indeed find that subject behavior in our settings is *not* well explained by the limited forward-looking bias. This result further underscores the importance of conducting careful behavioral studies in operational contexts as the bias found in other contexts may not extend to operational contexts.

Product development and project management have emerged as important topics in operations management. They have recently attracted significant amounts of interest in behavioral operations management (BOM) with the goal of understanding how human decision-making affects the process of generating and executing innovative ideas. The decision context of this chapter is related to product development, although our core model has wider applications to other operations management (OM) problems as well. In a survey chapter, Grushka-Cockayne et al. (2018) conclude that research in product development in BOM has mainly focused on three topics corresponding to different stages of the product development process: (1) the new idea creation process, especially

[3] In the context of dynamic search, the limited forward-looking bias predicts that subjects will undervalue the potential benefit from continuing the search process, and therefore will tend to be satisfied with the value from the searched options. This very well explains the subjects' tendency to end searching too soon (insufficient search).

on how different mechanisms (brainstorming, innovation contests, open innovation) affect the generation of innovative ideas; (2) the planning process intended for turning creative ideas into reality, and the biases people exhibit in conducting project planning; (3) the project execution process in terms of meeting goals and deadlines. Our study focuses on a new aspect of product development - how to optimally allocate financial resources to improve the product over the course of its development. This is an understudied topic in the current product development/project management literature in BOM, yet it is a crucial topic in practice based on our interactions with the OEM.

In BOM, two streams of literature consider decision problems where subjects make dynamic (sequential) decisions: supply chain coordination decision-making (the beer game), and behavioral revenue management. The field of BOM has a long history of conducting experiments to study subjects' behavior in the beer game [Sterman (1989), Croson & Donohue (2006)]. The beer game simulates a decentralized supply chain with several decision-makers that have different incentives. It is a multi-period game; in every period, each party decides the production or ordering quantity from the upstream supply chain partner and tries to fulfill the demand from the downstream supply chain partner. Systematic inefficiency and sub-optimal decision-making have been widely observed in experiments of the beer game, due to the issues of the parties optimizing locally, not sharing information, etc. Our decision context is an individual decision-making problem and, therefore, does not involve interacting and coordinating with other parties in supply chains. Hence, our setting focuses on a different kind of complexity from the beer game. There, the challenge comes from the system-driven complexity of the coordination among different parties and responding to the feedback that comes from other parts of the system in a continuous manner. In our setting, the challenge is the cognitive complexity in trying to optimize for a single decision-maker DP problem.

In the behavioral revenue management literature, Bearden et al. (2008) consider how the retailer can optimally sell to customers who arrive sequentially to maximize revenue. The retailer has multiple units to sell over a fixed selling season. In each period, one customer arrives with a certain probability; given that a customer arrives, the retailer receives an offer (value random ex-ante) from the customer and decides whether to sell one unit of product. They find that most subjects understand the general structure of the optimal policy. However, subjects' tendency to (incorrectly) reject valuable offers when the remaining inventory is high and (incorrectly) accept less valuable offers when the remaining inventory is low leads to revenue loss from the optimum. Two follow-up studies directly utilize the setup in Bearden et al. (2008), and add the consideration of multitasking (managing several independent projects simultaneously) [Bendoly (2011)] and providing real-time feedback with revenue management key performance indicators (KPIs) [Bendoly (2013)]. The model studied in these papers is structurally similar to the Constant Cost condition of our study. Our study differs from this stream of literature in three key aspects. First, I introduce the more challenging and practical situation - the Increasing Cost condition - to compare with the Constant Cost condition. This allows us to take a significant step forward in understanding subjects' dynamic resource allocation decisions. I find that although subjects act quite consistently with the general structure of the optimal policy in the Constant Cost condition, confirming the results in Bearden et al. (2008), subjects' performance falls dramatically in the (seemingly) slightly more complicated condition - the Increasing Cost condition. Second, the contrast between the two conditions also allows us to further derive the key behavior behind subjects' poor performance in the Increasing Cost condition - some subjects are acting as if the cost were constant throughout. Third, the treatments I develop are based on the more challenging condition - Increasing Cost condition - instead of the Constant Cost condition. Our managerial treatments are also more extensive as I cover a wide range of treatments with different formats (decision prompt, sharing best practices, and extra monetary incentives), and are different from what have been considered above in revenue management (providing real-time KPIs). I contribute to the literature by demonstrating how managerial treatments with these general formats can be incorporated with context-specific insights (Decision Decomposition Heuristic) to be effective.

More generally, as has been pointed out in a recent survey paper by Donohue et al. (2020), research evaluating managerial treatments to improve subjects' performance in operational decision-making is still in its infancy. So far, most of the work has been done in the context of Newsvendor decision-making, where subjects are found to consistently exhibit the pull-to-center bias [Schweitzer & Cachon (2000), Becker-Peth & Thonemann (2018)]. Treatments that have been considered include, but are not limited to, individual learning that lets subjects make many rounds of independent Newsvendor decisions [Bolton & Katok (2008), Bostian et al. (2008)], giving subjects an hour-long training session prior to decision-making or even explicitly showing the expected profit function to subjects [Bolton et al. (2012)]. The general result is that these (very strong) treatments help to improve performance but do not completely de-bias subjects. In the majority of these treatments, the information provided in the treatments come from the experimenter. In behavioral economics, Brown et al. (2009) consider the social-learning treatment in dynamic consumption problems, where they find that making use of a basket of past decision-making examples as decision suggestions can effectively improve subject performance. Nonetheless, Carbone & Duffy (2014) find that sharing the average consumption amount in each period, which is not directly related to the driver behind subjects' (poor) performance (i.e., not targeted at addressing the limited forward-looking bias), could lead to *worse* performance in dynamic consumption problems. In general, designing effective treatments is far from trivial and requires careful behavioral studies to identify potential drivers behind poor performance and subsequently tailor treatments. I contribute to the study of managerial treatments in BOM in several aspects. First, I consider a wide range of treatments in different formats - decision prompt, sharing best practices, and providing extra monetary incentives. All of these treatments have direct practical uses; yet, there have been relatively few studies regarding how to effectively implement these treatments in OM, particularly in the context of dynamic operational decision-making. Second, I am able to explain why the treatments work/do not work thanks to the experimental design I employ (see Section 3.6). Finally, our results suggest that I need *not* go so far as to dictate/demonstrate every step of decision-making (as in Brown et al. (2009)) to achieve performance improvement. Instead, guiding subjects with effective decision heuristics can already lead to substantial performance improvement in dynamic resource allocation problems, and can do so in an economical way for the firm.

An important finding of this chapter is that sharing best practices of a particular ilk - decision decomposition – can be effective in improving subject performance. While there are many types of best practice advice offered by top performers in our setting, not all of which are likely to be effective (see Appendix B.2 for a sampling of best performer advice, which can be quite convoluted), I zero in on decomposition-based advice as promising advice to share, and demonstrate its effectiveness at improving performance. Somewhat relatedly, Song et al. (2018) showed that sharing best practices directly was not effective, and suggest changing the process by which best practices are shared, with a "pull" approach and action-item outcomes from best practice discussions. Our approach is complementary to Song et al. (2018), in that I do not focus on process (how to share) but rather the type of advice (what to share). Our results suggest that in our dynamic decision context, sharing decomposition-based advice can be effective.

3.3 Decision Context and Formulation

Our leading example captures a product development manager's decision-making in improving product design, and is directly motivated by our interaction with a major US automotive OEM. During the R&D process, the product development manager faces a handful of beneficial design change opportunities. Typically, these opportunities are driven by technology advancements and marketing needs, such as the availability of a new technology or in response to a marketing survey that captures updated customer preferences. These design change, once implemented, will bring benefits to the product in terms of increasing the expected sales. Meanwhile, these opportunities are also costly to implement. To summarize, there are three key properties of these opportunities:

1. These opportunities are *optional* to implement.[4]

2. These opportunities arrive in a *sequential* order. This is because the events driving those opportunities (such as technology advancements) typically don't happen at the same time.

[4]Another class of design changes is the compulsory design changes, which are typically engineering or safety-related. For those design changes, they have to be made regardless of the cost.

3. The manager cannot “wait and bundle” the decision-making for these opportunities. As I have mentioned in the Introduction, the OEM’s project progresses very fast. Combined with the fact that there are usually time gaps between these opportunities, in practice, the manager does *not* wait and bundle the implementation decision-making for several opportunities.

For the manager, s/he is given a fixed amount of budget to be spent on these opportunities at the beginning of the project. Because the budget is limited and the opportunities are costly, not all of the opportunities can be implemented, and the manager needs to allocate the budget wisely in a dynamic manner. Modeling this process leads to the core model of this chapter.

In addition, to make the problem manageable, I introduce the following three assumptions.

(1) The total number of opportunities that will arrive is known to the manager at the beginning of the project. In practice, by reviewing past projects, the manager will have a good idea about the total number of opportunities to arrive. I am imposing a somewhat stronger assumption by assuming that it is a fixed and known value.

(2) The benefit/cost to these opportunities follows a known distribution to the manager. In practice, the manager can estimate the benefit/cost distribution from past projects.

(3) The opportunities are independent of each other. This is the strongest assumption I impose. Having this assumption greatly simplifies the decision-making context.

In Appendix B.1, I offer a mathematical formulation of our core model, along with explaining the structure of its optimal policy.

3.4 Experimental Design and Hypotheses

In our experiments, I further consider two conditions of the decision context that differ in their level of complexity. In the first decision context, which I call the “Constant Cost” condition, the cost to implement each opportunity is fixed (and known) throughout the project. The benefit of each opportunity is random ex-ante and follows a distribution known to the manager. This is the simplest context I can consider for this class of problems while still capturing the key aspects of the problem in practice - dynamically allocating limited financial resources to beneficial opportunities with uncertainty. In the second condition of the decision context, which I call the “Increasing Cost” condition, the cost doubles in the second half of the project (the fact that cost will double is known to decision-makers). This seemingly simple change dramatically changes the optimal policy from the Constant Cost condition as it introduces an additional key tradeoff: Subjects in the Increasing Cost condition should balance between capturing high-value opportunities and spending the budget when the cost is still low. Due to this additional tradeoff, it becomes more challenging for subjects to perform well in the Increasing Cost condition.

In this study, I conduct two waves of experiments. The first wave establishes subjects' performance in the dynamic resource allocation problem in two conditions as stated above: the Constant Cost condition and the Increasing Cost condition. This is a between-subject study where subjects play *either* one of the two conditions. From the first wave of experiments, I find that subjects perform poorly in the Increasing Cost condition (shown in Section 3.5). I therefore conduct the second wave of experiments (with a new group of subjects) where I develop managerial treatments to improve subject performance in the Increasing Cost condition. In this section, I introduce the design of the first wave of experiments. The design of the second wave of experiments will be introduced in Section 3.6 after I discuss subjects' performance in the first wave of experiments in Section 3.5.

3.4.1 Design of the Two Decision Conditions

I first give the parameter details for the Constant Cost condition. The experiment session consists of managing 5 independent projects in a row, each of which consists of making decisions for 10 design improvement opportunities. At the beginning of each project, the manager (subject) is presented with a total budget of 5000 experimental currency units (ECUs) that can be allocated to implement the opportunities. For each opportunity, its benefit follows a three-point distribution with equal probability: high benefit of 6000 ECU, medium benefit of 4000 ECU, and low benefit of 2000 ECU. Each opportunity, if implemented, will incur a cost of 1000 ECU. In other words, out of the 10 opportunities in each project, the manager can implement at most 5 of them. The 10 opportunities are independent of each other. At the end of each project (10th decision), the subject is presented with a summary table regarding his/her performance in the current project, including the total benefits collected and the total budget spent. If the subject still has an unspent budget on hand, it will be added to the total payoff from the current project. A history table recording all of the subject's past decisions and performance is always presented to facilitate learning and decision-making. Within each experimental session, I control for the sample path to be identical for all subjects; across different experimental sessions, I use different sets of sample paths. At the end of the experiment, 1 out of the 5 projects is randomly selected to determine the subject's actual monetary payoff, with an exchange rate of 2000 ECU to 1 dollar.

I now proceed to the Increasing Cost condition. In the Increasing Cost condition, I address the practical situation where cost increases in the second half of the project; in our experiment, I double the cost to be 2000 ECU in the second half of the project. Subjects get a notification of this when they enter the second half of the project to ensure their awareness of this piece of information. In designing the Increasing Cost condition, instead of simply holding all the parameters except the cost to be identical, I argue that it will be more important to keep the more general aspects of decision-making environment comparable between the two conditions: (1) subjects should have enough resources to implement a comparable number of opportunities; (2) the benefit values should always be strictly higher than the cost to avoid trivial or boundary decisions; (3) the overall expected profits between the two conditions should be comparable. Driven by the above considerations, I subsequently make several changes to other experiment parameters in the Increasing Cost condition: (1) the total budget increases from 5000 to 6000 so that subjects can make 4 or 5 decisions in every project; (2) the benefit values increase by 1000 for every realization, i.e., the benefits in the Increasing Cost condition are now 7000-5000-3000, with equal probability. The effect from these two changes, along with the increased cost in the second half, leads to a comparable expected profit between the two conditions. Finally, in the Increasing Cost condition of the experiment, these settings are also presented to subjects at the beginning of the experiment so that there is no "surprise" when they enter the second half of each project.

In the experiment, in order to elicit each subject's full strategy for each decision, I employ the so-called "strategy method". Specifically, in every decision period, *before* I show the subject the realized benefit, I ask the subject to think carefully about his/her *minimum willingness-to-accept* given his/her remaining budget and decision periods.[5] To make the decision more intuitive, I restrict the selection set to be one of the three benefit values. After the subject reports his/her minimum willingness-to-accept, I then reveal the realized benefit in this decision period, and the decision of whether to implement the opportunity in this period is *automatically carried out* based on his/her minimum willingness-to-accept.[6]

At the end of both conditions of the experiment, subjects perform three additional diagnostic tasks.[7] The three tasks are: (1) risk preference task, based on Holt & Laury (2002); (2) the cognitive reflection task Frederick (2005); (3) the Hit-15 task Carpenter et al. (2013). In addition, I also

[5]As stated in Appendix B.1, subjects should implement a threshold policy, i.e., if a subject wants to accept a certain benefit value, then all the benefit values higher than this should also be accepted. I therefore ask subjects to choose one out of the three benefit values as their minimum benefit-to-accept to help subjects avoid making trivial mistakes. In a related study (a different dynamic decision-making problem), Kagan et al. (2020) find that more than 90% of subjects effectively implement such a threshold policy although they are not restricted to do so.

[6]Regarding the strategy method, Brandts & Charness (2011) conduct a survey that summarizes 29 papers which directly compare the impact of using the strategy method vs. the "direct response" method on experiment outcomes (subject decisions). They conclude that very few papers are able to find a significant difference between these two methods: only 4 out of 29. Hence, this gives us confidence that utilizing the strategy method in our context would not significantly impact subjects' decision outcomes. Moreover, Brandts & Charness (2011) point out that the difference is more likely when there are only two possible contingencies for subjects to consider, because "[O]ne will be better able to imagine being in each alternative when there are fewer possibilities". In our setting, in every decision period, subjects face three possible contingencies (high-medium-low benefit realizations).

[7]Results related to these tasks are not included in this chapter since they are not central to our discussions here. The related results are in an experimental methodology paper that is available upon request.

distribute a questionnaire to collect subjects' additional demographic information and ask them to share with us any decision rule(s) they have used in managing the 5 product development projects. The experiment lasts around 1.5 hours, and each subject earns between 15 and 25 dollars. The subjects are undergraduate and graduate students recruited through ORSEE [Greiner (2004)] from a large public university in the Midwestern United States. The experiments are conducted online using a newly-developed platform ZTREE Unleashed that allows us to stream ZTREE [Fischbacher (2007)] to subjects.[8] The instructions are available in the Appendix B.3.

3.4.2 Performance Metrics and Hypotheses

It is useful to clearly define how I evaluate subjects' performance in both conditions of the decision task. As noted in Section 3.3, the optimal policy is to have an acceptance threshold, which is a function of both the remaining budget on hand and the remaining number of periods. With the experiment parameters, I am able to numerically calculate the optimal thresholds for all possible states (t, b_t). Given our experimental setup, subjects should then choose the benefit value that is *just above* (or equal to) the optimal threshold to make an optimal decision.

In examining the data, I want to measure the quality of subjects' decisions and, in particular, identify the direction and magnitude of their mistakes. However, I note that the optimal threshold will change across states quite a bit even if the implied action does not. Therefore, the magnitude of a mistake will differ: Choosing to accept 4000 is a bigger mistake when the optimal threshold is 5500 than when it is 4200. To account for this intuition, I define the decision gap as:

$$\text{Decision Gap} = \begin{cases} 0, & \text{if Observed Decision is Optimal} \\ \text{Observed Decision - Optimal Threshold}, & \text{Otherwise.} \end{cases}$$

A positive decision gap means the subject is being more *selective* than optimal (i.e., choosing thresholds higher than the optimum and therefore accepting too few opportunities), while a negative decision gap means the subject is being too accepting. With this, I can examine the following for each subject: (1) summarize the proportion of decisions with a decision gap of 0, which corresponds to the proportion of decisions that are optimal for the subject; (2) examine the magnitude and direction of mistakes for the subject.

Next, it is equally important to understand the outcome of the subjects' decisions - measured in terms of the profit they collect in each project and then the entire session. For the profit evalu-

[8] Before the shutdown caused by COVID-19, I were able to run several pilot sessions for the first wave of experiments with the same subject pool and using the physical lab in the large public university. The pilot sessions cover around 30 subjects in each of the two conditions (Constant Cost and the Increasing Cost); I find the results from "offline" sessions to be comparable with what I derive in online sessions. The related results are discussed in an experimental methodology paper by the authors, available upon request.

ation, a natural benchmark is the optimal profit - what the subject would earn if s/he followed the optimal policy; this establishes the general upper bound for the profit performance.[9] Meanwhile, I note that the lower bound profit performance is not 0 as subjects are guaranteed a positive return from implementing opportunities. Therefore, I consider the following performance lower bound - conservative profit - that is derived from implementing the first 5 opportunities in every project regardless of their benefit values.[10] With these two profit benchmarks, I am able to establish the following profit performance metric - the normalized profit ratio.

$$\text{Normalized Profit Ratio} = \frac{\text{Actual Profit} - \text{Conservative Profit}}{\text{Optimal Profit} - \text{Conservative Profit}}. \tag{3.1}$$

For each subject, the normalized profit ratio measures how much his/her earned profit is an improvement from being totally "conservative" compared to following the optimal policy. A ratio of 1 means that the subject is earning as much as following the optimal policy; a ratio of 0 means that the subject is earning as much as being totally "conservative"; a negative ratio means that the subject is doing worse than even being totally "conservative." This ratio is again calculated at the "session level"; for example, the actual profit is the sum of the profits in the 5 projects.

I am now ready to establish our hypotheses with respect to subjects' performance in these two conditions. First, I expect that subjects will not be able to follow the optimal policy in either condition of the problem. For the decision gap, when I calculate the proportion of decisions that are optimal (with a decision gap of 0) for each subject, it should rarely be 1. For the normalized profit ratio, I expect that subjects can rarely hit the ratio of 1 (achieving optimal profit).

Hypothesis 6. *Subjects perform significantly worse than the optimal policy in both conditions of the design change decision task, measured by (a) decision gap; (b) normalized profit ratio.*

Next, I would like to compare the subjects' performance between the two conditions of the decision task. In the Constant Cost condition, since the cost stays unchanged throughout, subjects can mainly focus on capturing high-value opportunities as the way to maximize their earnings in the project. In the Increasing Cost condition, on the other hand, it is necessary to also incorporate the cost aspect into consideration. In particular, in the first half of the project, there will be a tradeoff between earning benefits versus spending out the budget when the cost is still low. Given

[9] Note that it is possible that the subject can beat this upper bound by luck since the optimal policy maximizes the *expected* benefit-to-go in every decision period.

[10] I also consider an alternative performance lower bound - the "random strategy" benchmark - where subjects are assumed to implement any opportunity with a probability of 50%, regardless of the benefit realization, until the budget is insufficient. I find that all of our results below remain robust: (1) subjects perform significantly worse in the Increasing Cost condition compared to the Constant Cost condition, and (2) the results in Section 6 regarding the effectiveness of different managerial treatments in improving subject performance. I choose to utilize the conservative profit benchmark in the main text because it does not require extensive simulation to derive the expected payoff and because it is a strategy actually being utilized by several subjects (as seen in their self-reported decision rules).

this extra layer of complexity, I conjecture that subjects will perform worse in the Increasing Cost condition compared to the Constant Cost condition; see Appendix B.1 for a further discussion of the optimal policies in this two conditions. This motivates our second hypothesis.

Hypothesis 7. *Between the two conditions of the design change decision task:*

(a) Subjects will make more mistakes in the Increasing Cost condition compared to the Constant Cost condition. That is, subjects in the Increasing Cost condition will have lower proportions of decisions that are optimal and/or larger values of the decision gaps.

(b) Subjects will have lower normalized profit ratios in the Increasing Cost condition compared to the Constant Cost condition.

3.5 Experiment Results - Comparison Between the Two Conditions

This subsection analyzes and compares subjects' performance in the two conditions of the decision context: the Constant Cost condition and the Increasing Cost condition. I have in total 22 subjects in the Constant Cost condition and 48 subjects in the Increasing Cost condition.[11]

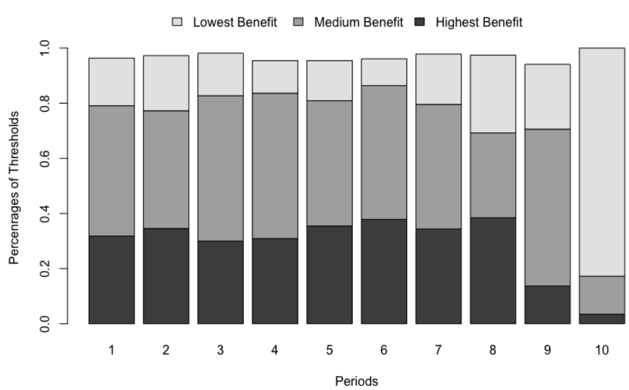

Figure 3.1: Constant Cost Condition Threshold by Round

Firstly, I illustrate subjects' actual decision-making pattern in these two conditions by presenting subjects' threshold decisions. For better visual illustration, I present their decisions in terms of

[11]The reason for having a larger subject size in the Increasing Cost condition is due to the high variability in subjects' performance in this condition, and I seek to utilize a larger subject size for more power in the conclusions. Our results here (both qualitative and quantitative) are further validated with earlier offline (lab) sessions (not included here for the consistency of presentation) as well as with different online experiment protocols. The related results are in a separate experimental methodology paper, available upon request.

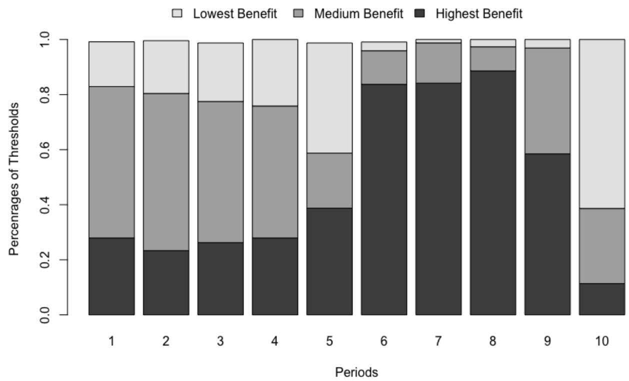

Figure 3.2: Increasing Cost Condition Threshold by Round

the period and compress the budget dimension.[12] The results are presented in Figures 3.1 and 3.2. The horizontal axis lays out the 10 decision periods for a project. The bar for each period illustrates the composition of different threshold decisions made from all subjects in the corresponding condition.[13] In the Constant Cost condition, I find that subjects are general quite "selective": Most of them choose either the medium benefit or the highest benefit as the threshold for the majority of the project. Such a pattern only changes until they have reached the very end of the project, where they lower the threshold to spend out the remaining budget. Such a behavior is in general consistent with the optimal policy: Given that the budget is limited (they can only implement 5 out of 10 opportunities), it is rational for subjects to focus on capturing valuable opportunities, until it is at the end of the project where it is optimal to spend out all the budget.

On the other hand, in the Increasing Cost condition, interestingly I find that subjects are also generally very selective in the first 4 periods in choosing either the medium benefit or the highest benefit as the threshold. In period 5, however, their tendency to choose the lowest benefit as the threshold dramatically increases in response to the fact that a cost increase is imminent (staring in period 6). Then, from period 6 and onward, most subjects start off being very selective (selecting the highest benefit as the threshold) and gradually lower their threshold until the end of the project. The behavior in the second half of the project is rational since the cost there is much higher, making them only have enough budget to implement 1 or 2 opportunities out of 5 periods. The behavior in the first half, however, is not-so rational. In fact, with our current parameters, it is *never* optimal for subjects to select the highest benefit as the threshold in the first 4 periods in the Increasing Cost

[12]The full presentation will involve presenting a matrix with two dimensions: decision period and remaining budget. However, some of those states are rarely visited, so presenting subjects' threshold decisions in that form will appear messy. Here, I choose to compress the budget dimension and focus on the time (decision period) dimension in our illustration.

[13]As an illustration, in Figure 3.1, I see that around 30 % of the threshold decisions made in Period 1 are choosing the highest benefit as the threshold.

condition. Nonetheless, a decent portion of subjects make such decisions. This will prevent them from utilizing the low cost opportunities in the first half of the project.

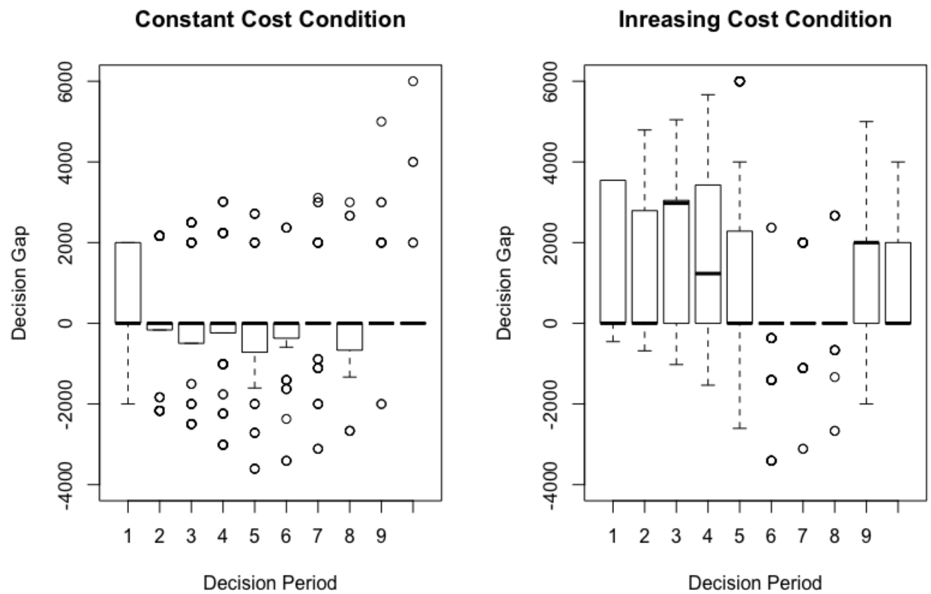

Figure 3.3: Decision Gap Demonstration and Comparison

Next, in Figure 3.3, I present the box plots of decision gaps in each period for both conditions.[14] I confirm that the way people make mistakes is structurally different between the two conditions. In the Constant Cost condition, I find that the median decision gap is 0 for all 10 periods. On the other hand, subjects in the Increasing Cost condition are mostly making higher-than-optimum decisions (positive decision gaps) in the first 5 periods; the bias is systematic in periods 3 and 4 where the median subject is making higher-than-optimum decisions. That is, subjects do *not* seem to understand the characteristic of the optimal policy that it is preferable to choose lower thresholds in the first half of the project. Moreover, the magnitudes of the mistakes are also visually larger in the Increasing Cost condition compared to the Constant Cost condition, particularly in the first half of the project (periods 1-5). To formally show this, for each subject, I calculate the average value of the *absolute* decision gap per decision and find this value to be (1) significantly larger than 0 in both conditions (one-sample t-test, p-value<0.01), confirming Hypothesis 6a;[15] (2) significantly larger for subjects in the Increasing Cost condition compared to subjects in the Constant Cost condition

[14]As I have noted above, the decision gap is a function of both the decision period and remaining budget. Here, again, I compress the budget dimension and only present the decision gap distributions with respect to the time dimension (decision period), for a more visual presentation.

[15]Alternatively, for each subject, I also calculate the simple proportion of decisions that are consistent with the optimal policy. I find that the distribution of this proportion is significantly smaller than 1 in either condition (one-sample t-test, p-value<0.01).

(rank-sum test, p-value$<$0.01). I therefore confirm Hypothesis 7a.

3.5.1 Impact from Decision Outcomes

From the above analysis, I find that subjects are generally consistent with the optimal decision-making in the Constant Cost condition, but they choose threshold substantially higher than the optimum early in the project in the Increasing Cost condition. In the Increasing Cost condition, such a pattern is particularly problematic, because subjects fail to spend their budget quickly enough and are then "forced" to spend their budget in the second half of the project when the costs are higher. This subsection further unpacks the impact of subjects' decisions, particularly in the Increasing Cost condition.

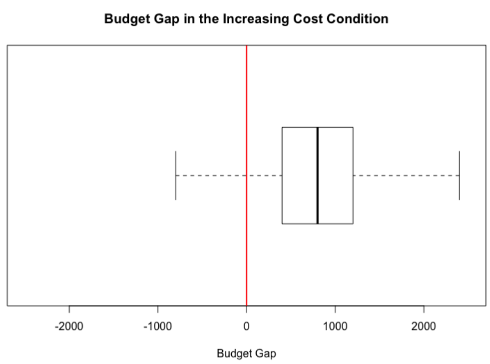

Figure 3.4: Budget Gap

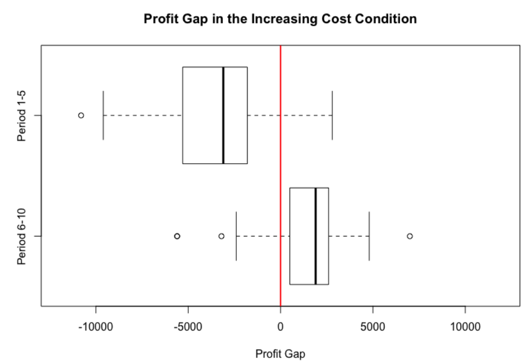

Figure 3.5: Profit Gap

In the Increasing Cost condition, the direct result of subjects' threshold decisions is that they could end up saving too much budget entering the second half of the project. To demonstrate this, I compare the subjects' remaining budget when entering the second half (period 6) to the situation where subjects are assumed to always follow the optimal policy. I call the difference to be the "budget gap";[16] a positive budget gap means that subjects are saving too much compared to the optimum (had they always followed the optimal policy), and vice-versa. This is demonstrated in Figure 3.4. I see that the majority of subjects end up having a higher-than-optimum remaining budget when entering the second half.

For such budget allocation behavior, I should expect that subjects will be earning more than the optimum in the second half - because they have more budget. However, this comes at the cost of potentially earning less in the first half since fewer financial resources are allocated there. Ultimately, subjects' performance compared to the optimum depends on a tradeoff - whether the

[16]The budget gap is calculated at the "session level" for each subject. That is, each data point in Figure 3.4 is a subject's average budget gap over 5 projects in the session. The situation is similar for the profit gap (discussed later).

extra earnings from the second half can outweigh the losses in the first half. To illustrate this, I calculate how much profit each subject earns in the two halves of each project; I then compare this to how much the subject would have earned in the two halves of the same project, *had they always followed the optimal policy throughout*. The comparison between the actual and optimal profit in the corresponding parts of the project (first or second half) will be called the "profit gap"; a positive profit gap means that subjects are earning more than the optimum (in that specific part of the project), and vice-versa. The profit gaps for the first half (period 1-5) and the second half (period 6-10) of the project in the Increasing Cost condition are demonstrated in Figure 3.5.

I find that by deviating from the optimal path and saving "too much" for the second half, subjects indeed end up having an overall positive profit gap in the second half of the project - the average gap is around 1500. However, the profit decrease in the first half is much worse - the average is around -3500. As a result, subjects end up earning less than optimal profit in the whole project. In summary, I observe that subjects' tendency to be overly selective in implementing opportunities creates systematic deviations from the way the budget should be allocated to the two halves of the project, and ultimately negatively impacts their profits.

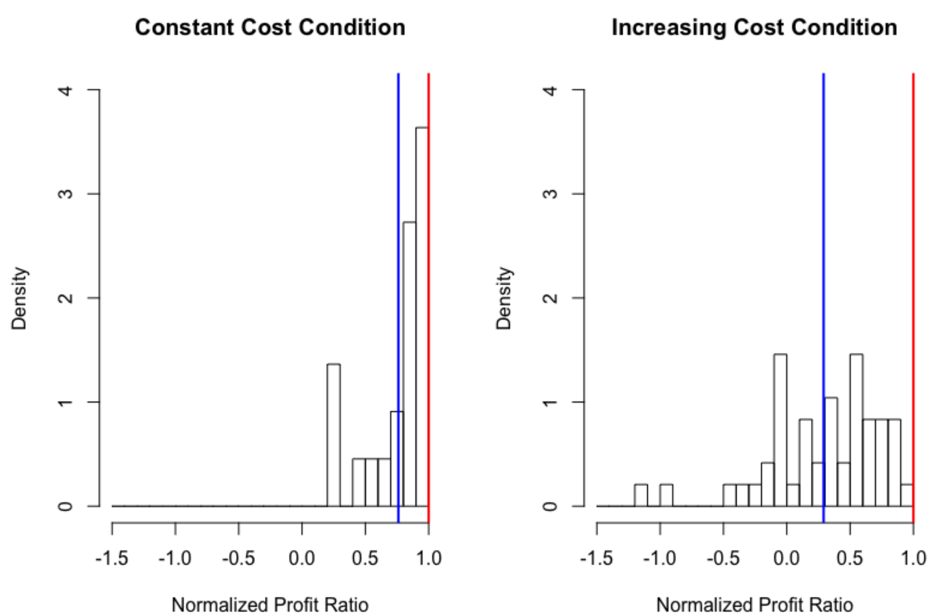

Figure 3.6: Normalized Profit Ratio Comparison

Finally, I demonstrate subjects' normalized profit ratios in the two conditions in Figure 3.6. I find that both distributions are significantly smaller than 1 (one-sample t-test, p-value$<$0.01), confirming Hypothesis 6b. When comparing the two distributions, I find that the ratios are significantly higher in the Constant Cost condition compared to the Increasing Cost condition: mean value comparison 0.76>0.29; rank-sum test returns a p-value<0.01. I also note that the degree of individual heterogeneity is much larger in the Increasing Cost condition, and a decent proportion of subjects (14 out of 48) fail to outperform the simple benchmark of "implementing the first 5 opportunities". Therefore, I confirm Hypothesis 7b. Hence, indeed, I find that subjects' perform much more in the more complex Increasing Cost condition. More importantly, from the above analysis, I find that this is driven by a systematic deviation from the optimal policy: Choosing thresholds that are higher than the optimum early in the project, and saving too much budget to the second half of the project where the cost is substantially higher. In the next subsection, I will further unpack the behavioral mechanisms driving subjects' performance in the Increasing Cost condition.

3.5.2 Behavioral Mechanism in the Increasing Cost Condition

In this subsection, I perform a comprehensive decision mechanism analysis, with a focus on the Increasing Cost condition. Our goal is two-fold: (1) forming a better understanding of the potential decision drivers in the Increasing Cost condition, which I discuss here, and (2) getting inspiration for the design of managerial interventions to improve subject performance, which I discuss in the next section.

As an overview, our analysis starts with a wide range of mechanisms that are ex-ante plausible in the Increasing Cost condition. I then fit each of the proposed mechanisms to subjects' actual decision outcomes. This exercise allows us to take a comparative perspective and utilize a data driven approach to determine if a mechanism well describes a particular subject's actions (or "well describes a subject" for short).

The mechanisms I consider here fall into two general categories, depending on whether or not subjects have the sophistication to solve the full DP problem. If subjects are able to do so, then their deviation from the optimal policy will come from their mis-specification of the parameters or utility terms. Classes 1-3 below cover three classes of mechanisms that fall into these categories, with different ways of parameter/utility mis-specifications. Alternatively, if subjects do *not* have full sophistication to solve the DP problem, then they could be simplifying certain aspects of the problem to derive their decision policies. Classes 4-6 below layout three classes of mechanisms that differ in the aspects subjects simplify. In each class of mechanisms, I note that there will usually be an inherent parameter. Hence, it is also important for us to specify the range of parameters I consider, which will be discussed below along with the discussion of the mechanisms.

1. **Overestimation of Continuation Value Mechanism**, which assumes that subjects overestimate the continuation value in solving the DP problem. In particular, I assume that the continuation value $V_t(b_t)$ is always multiplied by a markup parameter $\beta > 1$. I consider four

β values: 1.05, 1.10, ..., 1.20. A $\beta > 1.20$ will be too large and will not bring any further changes to the (predicted) decision policy compared to $\beta = 1.20$.

2. **Cost Underestimation Mechanism**, which assumes that subjects in the first half of the project systematically underestimates the cost in the second half of the project.[17] I consider 5 perceived cost values in the second half: $c = 1000, 1200, ..., 1800$. Note that a perceived cost of 1000 means that subjects would be acting as if the cost were constant throughout when they are in the first half of the project.

3. **Anticipated Regret Mechanism**, which assumes that subjects will experience a negative utility r if they encounter a design change opportunity that has a higher benefit value than an opportunity they have previously implemented, but they *do not* have the budget to implement this more valuable opportunity. The negative utility is in accordance with the perceived loss in opportunity.[18] I consider two negative utility values - $r = 1000$ or 2000 - to capture whether or not subjects factor into the cost difference between early and late opportunities.

4. **Limited Forward-Looking Mechanism**, which assumes that subjects are acting as if there were only a few decision periods left instead of correctly perceiving all the remaining periods. I consider three "lengths" for people's tendency to be (limited) forward-looking: periods = 1, 2, or 3. Past research in dynamic consumption problems found that people who are limited forward-looking rarely plan beyond 3 periods [Brown et al. (2009)].

5. **"Selective Heuristic"**, which assumes that subjects only select the highest benefit value as the threshold from the beginning of the project, until the number of remaining budget matches with the number of remaining periods times cost-to-implement; then, they will select the lowest benefit as the threshold. Such a heuristic is frequently observed in subjects' self-reported decision rules in the Increasing Cost condition.

6. **"Decision Decomposition Heuristic"** , which assumes that subjects have a pre-defined budget spending goal on each half of the project; within each half, the decisions are made optimally given the budget allocation plans; I consider 5 ways to allocate the budget: 5000 to

[17] When they are in the second half of the project, I assume they make the decisions optimally. In general, the idea is that subjects can correctly perceive the cost when they are in the corresponding phase of the project (first half or second half). The mistake comes from perceiving the cost for future phases which, in our model, happens when they perceive the cost for Phase 2 (second half of the project) from Phase 1 (first half of the project).

[18] If difference in the more valuable opportunity and the previously implemented opportunity with the lowest benefit value is adjacent (medium vs. low, high vs. medium), then the negative utility is r. If the difference is between high and low opportunity values, then the negative utility is $r + 2000$. The most straightforward r value is 2000, the pure benefit difference between opportunities. I also consider $r = 1000$ to capture the fact that those more valuable opportunities happen in the second half of the project (when the budget runs out), and the cost there from implementing opportunities also increases by 1000. This cost difference between early and late opportunities could mitigate the negative utility of regret.

first half and 1000 to the second half, 4000 to the first half and 2000 to the second half, ..., 1000 to the first half and 5000 to the second half. This exhausts all the budget allocation plans. Such heuristics are also frequently observed in subjects' self-reported decision rules.

From this, I derive a total of 6 classes, 20 mechanisms (when I consider different parameters). I also include the optimal policy into our mechanism analysis, making it a total of 21 mechanisms. For these 21 mechanisms, I then match their predictions with every subjects' actual decision outcomes. From this, I derive a large number of decision fitting rates: $21 * 48 = 1008$ decision fitting rates. Such "large data" of 1008 fitting rates is very useful as it allows us to form a holistic view and determine the decision fitting rate that is "high enough". Specifically, these 1008 fitting rates form an empirical distribution, and I take the 85th percentile of the empirical distribution - which gives 65.85% - as a threshold.[19] I say that a mechanism "well describes" a subject if the corresponding matching rate is higher than or equal to this threshold. With this, I can then calculate the number of subjects each mechanism well describes and use this as the key performance metric to evaluate each mechanism.

Below in Figure 3.7, I present the number of subjects each class of mechanisms can well describe, with the best fitting parameter chosen. I use red (blue) to denote the classes of mechanism from the first (second) category discussed above. Overall, I find that a slight degree of value overestimation well describes the largest number of subjects: With a parameter of $\beta = 1.05$, it is able to well describe 20 out of 48 subjects. The first three classes of mechanisms (with the corresponding parameters) all outperform the optimal policy in the number of subjects they can well describe. This is consistent with our earlier decision gap result (shown in Figure 3.3) that many subjects in the Increasing Cost condition deviate from the optimal policy by being too selective early on. I also note an important result - the Limited Forward-looking mechanisms perform very poorly here in capturing subjects' actual decision-making. The limited forward-looking behavior is frequently proposed as one of the plausible behavioral explanations in many other dynamic decision problems. As I have noted in the Literature Review Section, the dynamic resource allocation problem I consider here is structurally different from many other dynamic problems; our results here showcase the importance of conducting careful behavioral research in dynamic resource allocation problems to identify subject behavior.

The above exercise also allows for an important extension: The identification of a best *combination* of mechanisms to jointly well describe as many subjects as possible. After all, there is a high degree of individual heterogeneity in the Increasing Cost condition. I may therefore require different mechanisms to separately capture good performing or bad performing subjects. For ex-

[19]For such an empirical distribution, the range is 0 to 9615%, with a median value of 47.37%. I also note that the qualitative nature of our results below remain unchanged if I choose a lower percentile of 75th, which will result in a threshold of 60%.

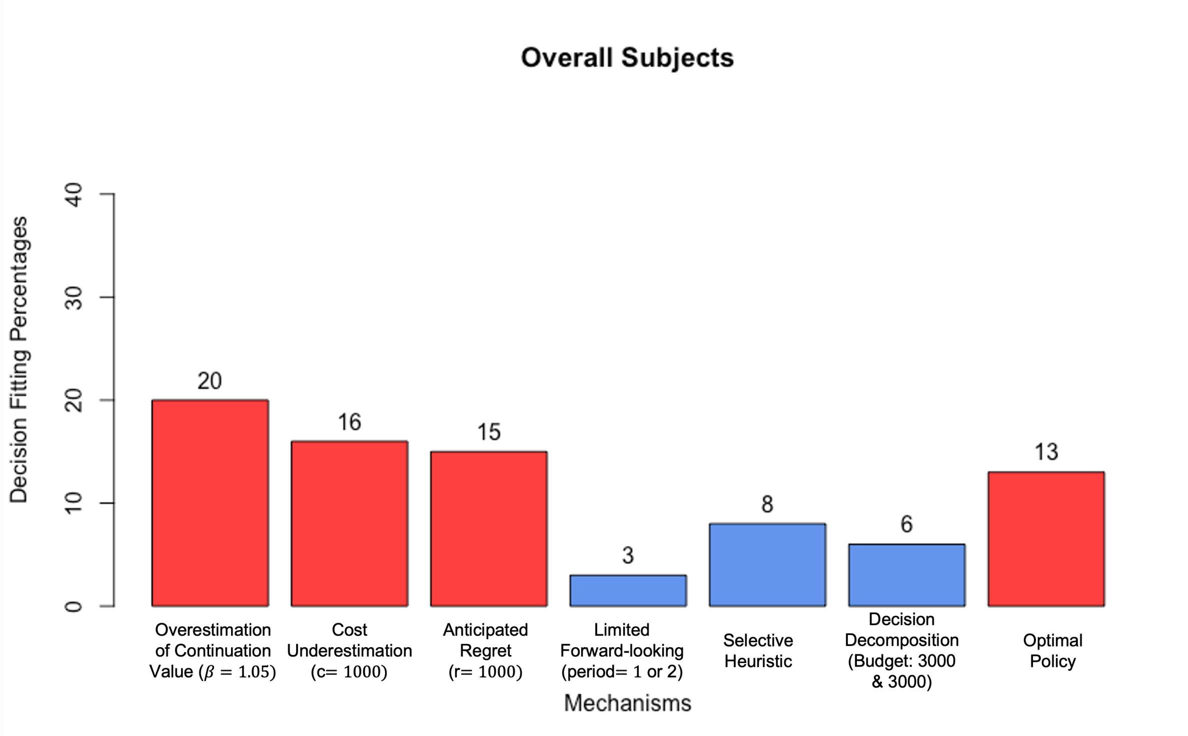

Figure 3.7: Decision Mechanism Analysis Fitting Overall Subjects

ample, the Selective Heuristic captures bad performing subjects very well: Out of the 8 subjects it can well describe, 7 are bottom performers (bottom 1/3 in terms of the normalized profit ratio). On the other hand, for the people who are well described by the optimal policy, naturally most of them will be good performing subjects. To study such best combination, for a given number of mechanisms, I simply iterate all the possible mechanism combinations. For example, when I consider a combination of 2 mechanisms, I will have a total of ${}_{21}C_2 = 210$ combinations. For each combination, I simply identify the total number of subjects that can be well described by at least one mechanism in the combination.

The results are summarized in Table 3.1. In addition to presenting the best single mechanism to capture as many subjects as possible, I also present the best combination of 2, 3, or 4 mechanisms that can jointly well describe as many subjects as possible.[20] I have two leading observations: (1) the minor Value Overestimation mechanism ($\beta = 1.05$) is present regardless of the number of mechanisms in the combination. This mechanism captures people making ”small mistakes”; (2) the Selective Heuristic is also always present. This mechanism, as explained above, captures people making “big mistakes”, i.e., bottom performing subjects. Overall, I find that our approach

[20]With a combination of 3 mechanisms (4 mechanisms), I derive “ties”, i.e., several different combinations can well describe 31 (34) subjects. Interestingly, I find that, among those combinations, they only differ in the “additional mechanism”.

here is useful as I am able to capture the majority of subjects with only a few (2 or 3) mechanisms.

Table 3.1: Increasing Cost Condition Behavioral Mechanism Combination Analysis

Number of Mechanisms in the Combination	1 Mechanism	2 Mechanisms	3 Mechanisms	4 Mechanisms
Best Combination	Value Overestimation ($\beta = 1.05$)	Value Overestimation ($\beta = 1.05$) + Selective Heuristic	Value Overestimation ($\beta = 1.05$) + Selective Heuristic +1 Additional Mechanism	Value Overestimation ($\beta = 1.05$) + Selective Heuristic +Cost Underestimation (c=1.0) +1 Additional Mechanism
Number of Subjects Jointly Well Describes	20 [42%]	28 [58%]	31 [65%]	34 [71%]

Notes: There are 48 subjects in total in the Increasing Cost condition. Percentages in brackets are showing the proportions relative to the total number of subjects

3.6 Improving Decision-Making Through Managerial Treatments in the Increasing Cost Condition

In Section 3.5, I observed that subjects' performance falls dramatically in the more complicated situation - the Increasing Cost condition. I also conducted decision mechanism analysis to understand the drivers behind subjects' performance. These findings offer the basis for us to design managerial treatments for subject performance improvement.

In designing the treatments, I refrain from directly telling subjects what the optimal decision/policy is. The reason is that, in reality, the situation could be more complicated than the simple settings in our experiments, such that it is impractical to derive or implement the optimal policy[21]: This, after all, is a key reason why human decision-makers are prevalent in practice, while of course insights based on stylized but tractable models are still useful. In this study, the tractability of the model and the ability to easily compute the optimal policy brings two key benefits: (1) it enables us to quantify the profit loss from subject behavior and the improvement from managerial treatments; (2) the structure of the optimal policy can inform the design of managerial treatments (see the discussions below). Meanwhile, our stylized context still captures the core of the dynamic resource allocation problem - how to allocate limited resources to collect benefits from opportunities that arrive sequentially. As a result, the behavior I have identified in this study and the insights regarding why the managerial treatments work/do not work could well extend into those more complicated situations.

[21] For example, the distribution of benefits could be unknown or highly uncertain, or the time when costs increase could be uncertain. In fact, the OEM I interact with has observed that the degree of cost increase as well as when costs will increase vary significantly from project to project, making it much more difficult to develop an "overall simple optimal policy."

3.6.1 Design of the Managerial Treatments for the Increasing Cost Condition - Motivation

I first discuss the motivation behind our managerial treatment design, with implementation details explained in the next subsection. In general, I recognize that there are (at least) three ways that can motivate our design of effective managerial treatments to improve subject performance.

Firstly, I can study the behavior of poorly performing subjects. From the previous section, I learn that many bad performing subjects are are well described by the Selective Heuristic; that is, they are overly selective early in the project and are unresponsive to the cost-increase information. Therefore, I can design a managerial treatment to address such behavioral pattern.

Secondly, I can study the behavior of subjects who perform very well. Also from the previous section, I learn that many subjects are slightly overestimating the continuation value, and quite a few of them (13 out of 48) are actually well-described by the optimal policy. However, this is not that useful as a motivation - obviously I will want subjects to perform closer to the optimal policy, but I have decided to refrain from directly telling subjects what the optimal policy is. Interestingly, when I look at the *top performing subjects* (top 2 out of 48), I find that they seem to be simplifying the problem in a particular way, which could serve as our motivation for treatment design. In particular, these top performing subjects decompose the problem into two manageable sub-problems based on whether or not the cost has increased. They then allocate a 4000 (ECU) budget to the first half and a 2000 (ECU) budget to the second half.[22] In fact, this corresponds to (one of) the Decision Decomposition Heuristics I included in the above Class 6. Such a decomposition idea is potentially useful to guide future subjects for performance improvement.

Lastly, I may consider ways to transform or re-frame the problem into another problem that is more manageable to human subjects. There are different ways to design this, depending on the problem structure. Interestingly, I note that the decision decomposition idea mentioned above is such a method to re-frame the problem in the Increasing Cost condition. To see this, after decomposing the problem based on whether or not the cost has increased, subjects effectively derive two Constant Cost condition problems. This could be useful because, from earlier results, I know that subjects in general can handle the Constant Cost condition problem quite optimally. I also note that, by prompting subjects to act on the cost-increase information, such a re-framing is useful to guide subjects to move away from following the Selective Heuristic in the entire project.

Overall, the decision decomposition idea emerges as a natural candidate to improve future subject performance because (1) it draws from top performers' decision-making insights, (2) it addresses bad performers behavioral tendency, and (3) it is a logical method to re-frame/decompose

[22] For these top 2 subjects, this heuristic matches with at least 80% of their actual decisions, and is directly observed in their self-reported decision rules (see Appendix B.2).

the complex problem into something manageable to general subjects. To further validate the usefulness of this approach, I also compare it with the optimal policy in the Increasing Cost condition. I find that, with appropriately allocated budget to the two sub-problems (4000 to the first half and 2000 to the second half, what the top performing subjects are doing), such a decision decomposition method closely resembles the optimal policy;[23] This adds to our confidence that the approach of decision decomposition through budget planning, when conducting appropriately, can lead to a performance improvement for subjects. In the rest of the section, I discuss how to design managerial treatments around effectively sharing the decision decomposition idea to future subjects.

3.6.2 Design of the Managerial Treatments for the Increasing Cost Condition - Implementation

In this subsection, I present the details of how I operationalize the sharing of Decision Decomposition Heuristic with future subjects. Specifically, at the beginning of every project, I will guide subjects to conduct *budget planning* based on whether or not the cost has increased. I develop two managerial treatments that both center around the idea of conducting decision decomposition through budget planning, but differ in the extent to which they offer guidance to subjects.

In the first treatment, the decision prompt treatment, the goal is to introduce the general idea of decision decomposition through budget planning. In the process, I offer no guidance on how they should be allocating the budget, and I make no reference to the fact that the decision decomposition through budget planning idea is coming from past best performing subjects. Specifically, at the beginning of every project in this treatment, I give subjects 1 minute to make a budget allocation plan - how much budget to be spent before and after the cost increase of the incoming project. I also remind subjects (once more) that the cost will double in the second half of the project, but I provide no more further guidance on how the budget allocation should be made. Subjects proceed to the project after finishing the budget planning. I note that the budget allocation plan is *not* binding in any way, and I also allow subjects to change their budget allocation plans from project to project.[24]

If simply prompting subjects to focus on the idea of decision decomposition is not enough, then an alternative idea is to directly share how the top performers are conducting decision decomposition through budget planning. This leads to our second treatment, the "sharing best practices treatment". In this treatment, I directly share their proposed budget allocation plan (4000 in the

[23]To see this, I compare the decision policy from this specific decision decomposition method with the optimal policy. I find that the two have a perfect alignment *when subjects have perfect execution*, i.e., when they *always* follow the suggested decision policy from either the specific decision decomposition method or the optimal policy.

[24]Collecting their budget plans in the experiment allows us to see: (1) whether their understanding of the problem changes (as indicated by the change in reported budget plans) as they go through 5 independent projects; (2) whether and how different budget plans correlate with their actual performance.

first half and 2000 in the second half), along with the reasoning behind it (the cost increase information), as the sharing best practices information provided to subjects in this treatment.[25] Subjects are free to decide whether to follow such a budget allocation plan and, if so, how to carry it out in making the ten implementation decisions of the project. The complete information I share can be found in the Appendix B.3.

So far, the two managerial treatments I propose are based on detailed knowledge of subject behavior in the dynamic resource allocation problems. Their designs are therefore context-dependent. In this study, I also include a "general purpose" treatment that does *not* require pre-existing knowledge of subject behavior or biases in the decision context. This third treatment, the "payoff scaling" treatment, is based on the idea that I may use extra monetary incentives to induce subjects to try harder in their decision-making; their performance may improve as a result. This treatment has the advantages of being easy to design and straightforward to implement, and using extra monetary incentives to induce performance improvement is quite common in practice. Moreover, this treatment helps to address the concerns in experiments where the payoff may not be enough to incentivize subjects when solving complex problems [Enke et al. (2020)]. Specifically, in this treatment, I double all the parameters of the Increasing Cost condition (for example, the lowest benefit increases from 3000 ECU to 6000 ECU) while keeping the ECU-to-dollar exchange rate unchanged. To control for the effect where subjects' risk-taking behavior may greatly change after receiving a large number of monetary payoffs [Thaler & Johnson (1990)], I normalize the final payoff in each project by subtracting the initial budget (12000 ECU) from it. After the normalization, subjects still earn about 50% more monetary payoffs.

In designing and implementing the above three treatments in the second wave of experiments, I use the same set of sample paths as the Increasing Cost condition in the first wave of experiments for a better control across different treatments. I hereafter refer to the data from the Increasing Cost condition as the "baseline" as it serves as the benchmark to evaluate whether the treatments have achieved performance improvement for subjects. The second wave of experiments is also a between-subject study. I recruit around 50 (new) subjects into each of the three managerial treatments to be comparable with the baseline Increasing Cost condition. The instructions of the

[25] I also note that our "sharing best practices" treatment is in fact a "curated" sharing best practices treatment as I am *not* simply sharing the best practices as-is. As I can see from the reported decision rules in Appendix B.2, those decision rules from top performers tend to be lengthy and complicated; different top performers also have different "detailed plans" in every decision period, despite the fact they share the same budget allocation plans. I also further note that our "curated" sharing best practices approach is complementary to the findings by Song et al. (2018). In their paper, the authors first concluded that simply sharing the best practices as is (which they call the "push" approach) is not so effective in improving the emergency department efficiency. They then moved to a "pull" approach by clearly identifying the top performing workers, which then allows for (1) engaging discussions with top performing workers for actionable items, and (2) direct observation of top performing works' actions on a daily basis. In our approach here, I am also moving away from simply "pushing" the best practices out as is, and I instead summarize the best practices into actionable items - the specific decision decomposition method - for future subjects.

decision prompt treatment and the sharing best practices treatment are available in the Appendix B.3. The instructions for the payoff scaling treatment are the same as that in the Increasing Cost baseline, except for the necessary changes related to doubling the parameters in this treatment.

3.6.3 Managerial Treatments: Experiment Results

This subsection presents the experimental results from the three managerial treatments, along with the Increasing Cost condition baseline. I focus on discussing the normalized profit ratio here since it summarizes the quality of the subject's overall decision-making throughout the experiment.[26] The results are illustrated in Figure 3.8.

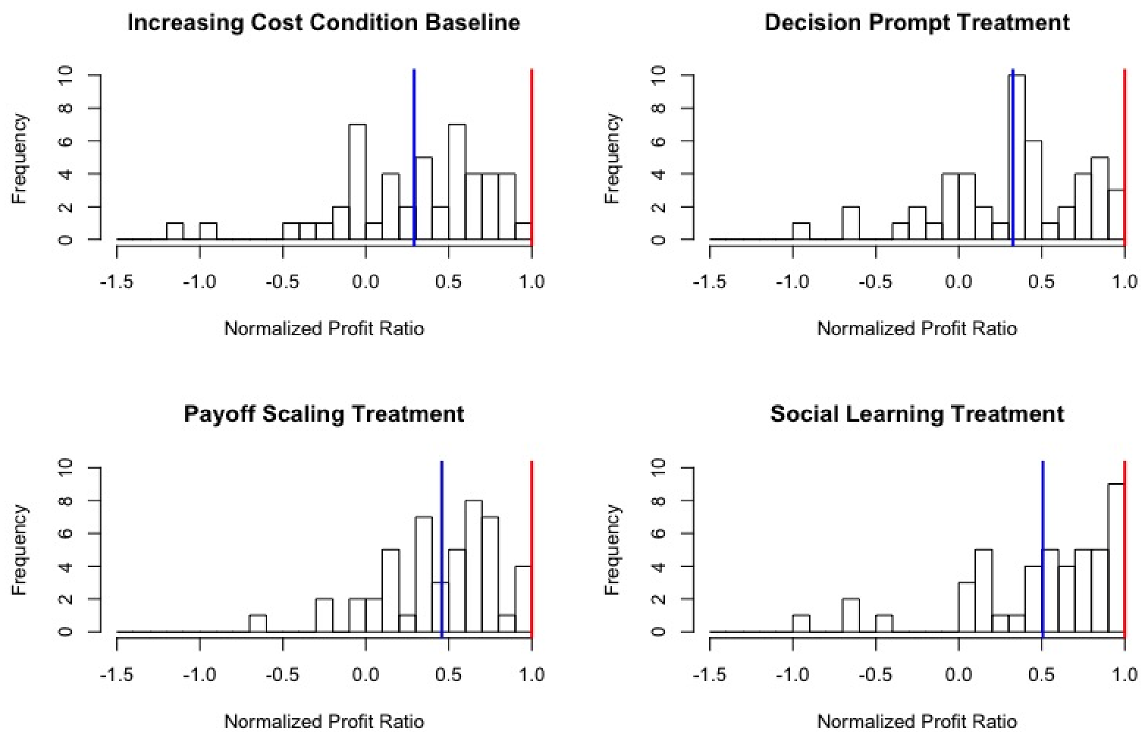

Figure 3.8: Normalized Profit Ratio Comparison

I find that all the three managerial treatments are able to increase the average performance: the average normalized profit ratio is 0.29 for the baseline, 0.33 for the decision prompt treatment, 0.46 for the payoff scaling treatment, and 0.51 for the sharing best practices treatment. In terms

[26]If I instead focus on analyzing the decision gap, the qualitative nature of the comparison results between the three treatments and the Increasing Cost baseline will hold, with the slight modification that the improvement from the payoff scaling treatment will be less significant using that metric. The sharing best practices treatment, when compared with the Increasing Cost baseline, remains the most effective treatment as it significantly improves the proportion of decisions that are optimal and significantly decreases the decision gap.

of the magnitude of improvement, only the payoff scaling treatment and the sharing best practices treatment lead to significant profit improvement (rank-sum test p-value$= 0.09$ and <0.01, respectively). From this result, I can conclude that addressing limited attention alone (with the decision prompt treatment) is insufficient to improve subject performance.

Our aim is to understand why each of the treatments works/does not work. The analysis will not only help us better understand our existing results, but also shed light on the design of effective treatments in future studies. In conducting further analysis, I will make use of both more detailed performance metrics (such as subjects' decision time) as well as treatment-specific data I collect. Table 3.2 summarizes various performance metrics in different treatments, which I will refer to in our discussions below. I first discuss the payoff scaling treatment as it belongs to a different category (not utilizing the decision decomposition idea) compared to the other two treatments.

Table 3.2: Performance Metrics in the Baseline Increasing Cost Condition and Three Managerial Treatments

	Condition/Treatment			
Metric	Increasing Cost Baseline	Payoff Scaling	Decision Prompt	sharing best practices
Average Normalized Profit Ratio	0.29 (0.07)	0.46* (0.05)	0.33 (0.06)	0.51*** (0.07)
Average Decision Time Per Round (in Seconds)	11.67 (0.69)	13.31** (0.72)	10.95 (0.53)	12.10 (0.64)
Proportion of Decisions as the Highest Benefit in Round 1-5	0.29 (0.04)	0.22 (0.03)	0.30 (0.03)	0.16*** (0.03)
Budget Gap	845.83 (102.49)	1325.00 (172.65) [662.50*]	787.76 (101.42)	317.39*** (78.53)
Profit Gap from Round 1-5	-3445.83(412.19)	-5058.33 (661.34) [-2529.17**]	-3179.59 (416.59)	-1430.44***(313.74)
Profit Gap from Round 6-10	1433.33 (344.68)	1975.00 (610.01) [987.50]	1297.96 (369.73)	31.88*** (236.67)

Notes: Standard Errors are in parentheses. All the metrics are calculated at the session level. Brackets in the budget gap and profit gap numbers in the payoff scaling treatment shows the normalized number (divided by 2) to be comparable with the other three conditions. Stars denote the comparison results against the corresponding metric in the Increasing Cost Baseline condition, using the rank-sum test for all the metrics except the metric in the third row, where the proportion test is used. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. I have 48, 48, 49, and 46 subjects in these four treatments (from left to right), respectively.

3.6.4 Payoff Scaling Treatment Further Analysis

In the payoff scaling treatment, I double the numerical parameters in the Increasing Cost condition while holding the ECU-to-dollar exchange rate fixed. By doing this, I dramatically increase the payoff differences between good and bad decision-making, with the intention of inducing subjects to put more effort into their decision-making. To evaluate the success of this treatment in inducing more effort, I consider subjects' average time spent in decision-making; this measurement method has been used in behavioral economics studies [Enke et al. (2020)]. Also, longer decision time has been associated with better profit performance in the Increasing Cost baseline: I regress the normalized profit ratio on average decision time (per decision, in seconds) and subjects' performance in the three diagnostic tasks, and find the estimate for the average decision time to be significantly positive (estimate=0.03; p-value=0.04).

The results are summarized in the second row of Table 3.2. I find that the time subjects spend is the longest in the payoff scaling treatment and is substantially higher than that in the baseline Increasing Cost condition (rank-sum test p-value=0.03); this result goes hand-in-hand with the performance improvement from the Increasing Cost baseline. Therefore, our result suggests that it is possible to achieve performance improvement in the Increasing Cost decision context by inducing subjects to try harder. However, this may require substantial economic incentives.

3.6.5 Decision Prompt Treatment Further Analysis

In the decision prompt treatment, I first note an interesting result that subjects on average spend the least amount of time per decision out of the four treatments. This suggests that our decision prompt of budget planning does help the subjects develop plans for the upcoming 10 decisions in the project. However, the fact that the decision prompt treatment fails to systematically improve subjects' performance also suggests that performing budget planning alone is insufficient. To further explore this, I review the subjects' budget allocation plans entered in the program for all 5 projects. I find that subjects can be put into one of the following three categories:

(1) Subjects who change their budget plans throughout the 5 projects: 12 subjects with an average normalized profit ratio of 0.21;

(2) Subjects who stick with or converge to a plan *different from* the "good budget allocation plan" of spending 4000 ECU in the first half and 2000 ECU in the second half: 12 subjects with an average normalized profit ratio of 0.08.[27]

(3) Subjects who stick with or converge to the "good budget allocation plan": 25 subjects with an average normalized profit ratio of 0.50.

Subjects' behavior in category (1) and (2) above offers further insights into why subjects, even with full awareness of the cost-increase information and with the decision prompt, would perform badly. The subjects who keep changing their budget plans seem to be having difficulties in making

[27]Regarding the specifics of these alternative plans: (1) 8 subjects converge to a plan of 2000 in the first half and 4000 in the second half, which is equivalent to implementing 2 opportunities in both halves of the project; (2) 1 subject with the plan of 3000 in both halves; (3) 3 subjects with a plan of 5000 in the first half and 1000 in the second half, which is the same as achieving the "conservative benchmark" in profit. People who have the first two kinds of plans are exhibiting a tendency to achieve "equality/balance" between the two halves of the project, either in terms of the number of opportunities to implement or the amount of budget spent.

use of the decision prompt. Subjects who converge to plans different from the good budget plan are making use of the decision prompt but generating “wrong conclusions”, which leads to even worse performance. The managerial implication is that companies should be well aware of the possibility that their employees are either not understanding the decision prompt or, even worse, using the decision prompts but reaching undesirable conclusions; the latter could be especially problematic since these employees may be confident of their undesirable conclusions since they are the result of implementing the decision prompts.

Subjects’ performance in category (3) demonstrates that deriving the “good budget allocation plan” is indeed associated with good performance, confirming its usefulness in the Increasing Cost condition. Nonetheless, I note that only half of the subjects are able to derive/converge to this plan. In fact, the performance of the other subjects who cannot derive this plan is so bad that the decision prompt treatment overall results in an insignificant improvement from the Increasing Cost baseline. This demonstrates that simply prompting subjects to the decision decomposition idea is insufficient to achieve a systematic performance. There is the need to provide more guidance on how to correctly implement this idea in this context, which is exactly what I offer in the sharing best practices treatment.

3.6.6 Sharing Best Practices Treatment Further Analysis

Out of the three managerial treatments, the sharing best practices treatment leads to the most significant performance improvement. This result highlights that subjects can effectively absorb the decision decomposition idea to achieve a significant performance improvement, provided that they are also offered with the guidance on how to implement this idea. I also note that this treatment is most successful in de-biasing subjects away from being overly selective early on and achieve close-to-optimal performance.

In the third row of Table 3.2, I present the proportion of decisions in the first 5 periods that choose the highest benefit as the minimum benefit-to-accept. I find that, compared to the Increasing Cost baseline, only the subjects in the sharing best practices treatment significantly decrease such preference (rank-sum test p-value$<$0.01). As a result, subjects in the sharing best practices treatment have the closest-to-zero budget gap and profit gaps among the four conditions, i.e., being the closest to the optimal policy. On the other hand, neither the payoff scaling treatment nor the decision prompt treatment achieves a similar degree of success in preventing subjects from being too selective. As a result, the improvement in budget gaps and profits gaps is not as significant as that in the sharing best practices treatment.

From a complementary angle, the success of the sharing best practices treatment is further demonstrated in the shape of the profit distributions. From Figure 3.8, I see that the shape of the distribution in the sharing best practices treatment centers around the optimum. The payoff scaling treatment, while achieving an overall performance improvement relative to the baseline, does not dramatically change the shape of the distribution.[28] Finally, similar to the decision mechanism analysis in the previous section, here I also try matching the Selective Heuristic and the optimal policy to subjects' actual decision outcomes from the three treatments as well as the Increasing Cost baseline condition. The results are presented in Table 3.3. I find that the sharing best practices treatment has the lowest fitting rate for the Selective Heuristic and the highest fitting rate for the optimal policy.

Table 3.3: Number of Subjects Well Described by the Mechanisms in Different Treatments

Treatment	Mechanisms: Selective Heuristic	Mechanisms: Optimal Policy
Increasing Cost Baseline	8 [17%]	13 [27%]
Decision Prompt	7 [14%]	11 [25%]
Payoff Scaling	4 [8%]]	10 [21%]
Sharing Best Practices	2 [4%]	24 [52%]

Notes: A uniform threshold of 65.83% is applied to determine the number of subjects well described in all four treatments. Percentages in brackets are showing the proportions relative to the total number of subjects in the corresponding treatment.

Overall, the success from this treatment shows that sharing the decision decomposition idea with guidance on how to implement it can significantly improve subject performance. Our result has two direct managerial implications. First, for complex DP problems that need to be handled by humans, I do *not* need to go so far as dictating every step of decision-making to achieve a performance improvement; this is important because the approach of sharing the optimal policy could be undesirable when the optimal policy is hard to derive or is unintuitive to human subjects. Instead, I find that identifying effective simplifying heuristics - such as the Decision Decomposition Heuristic - can be a useful approach to guide subjects and improve their performance. Second, specifically to dynamic resource allocation problems, I have identified a class of such problems that is "solvable" to humans. These are the problems that are either "stationary" with unchanged parameters/distributions over time, or can be decomposed into such "stationary" sub-problems.

[28] Specifically, I find that only the sharing best practices treatment has a significantly higher proportion of subjects with a normalized profit ratio≥ 0.75 compared to the baseline (0.30 vs. 0.15; proportion test p-value=0.07) while the other two treatments do not (p-value>0.40).

3.7 Conclusion

In this study, I consider a dynamic resource allocation problem where subjects act as the product development manager and sequentially allocate financial resources to improve the design of the product. I set up and compare subjects' performance in two conditions: the simple Constant Cost condition and the more practical yet challenging Increasing Cost condition. I find that subjects perform reasonably well in the Constant Cost condition but substantially worse in the Increasing Cost condition and exhibit a large degree of individual heterogeneity in performance. In particular, many subjects in the Increasing Cost condition slightly overestimate the continuation value when solving the DP problem; meanwhile, many bad performing subjects are overly selective early on in the project and end up saving too much budget to the later part of the project where the cost is substantially higher. To improve subject performance in the Increasing Cost condition, I share the decision decomposition idea to future subjects: It prompts subjects to decompose the complex Increasing Cost condition problem into manageable Constant Cost condition sub-problems. I find that such a decision decomposition idea, along with some guidance on how to implement it, can effectively improve future subject performance.

Our results help to establish the conditions and boundaries of good decision-making in dynamic resource allocation problems. When facing a stationary environment with fixed parameter/distribution of parameter over time, subjects are capable of performing well. When the parameter is changing over time, not all the subjects can respond to this structural change effectively. However, when the parameters are changing in a way that allows for effective decision decomposition, subjects will be able to perform well, because they can either derive how to decompose the problem by themselves or absorb the advice on the decision decomposition idea. Related to this, our results from managerial treatments not only confirm that sharing best best practices is a useful approach to achieve performance improvement, more important, the results also illustrate "which best practice" to convey - the insight of cleverly decomposing the complex DP problem into manageable sub-problems.

It is also worth pointing out that the goal of the sharing best practices treatment is to propose a design that has the highest possibility to achieve a performance improvement, following the idea of sharing the decision decomposition idea. One may imagine a different treatment that focuses on further studying the behavioral drivers behind the success of this treatment. For example, one potential driver behind the performance improvement comes from the fact that, in this treatment, subjects are made aware of the existence of "best performing subjects". This is a form of social comparison that has been utilized to improve subject performance; see Cadsby et al. (2019) for a review. With this, one may consider setting up a pure "social comparison" treatment and study its effectiveness. In this treatment, subjects will be made aware of the existence of top performing subjects; they may also be explicitly benchmarked against those top performing subjects. If such a treatment is effective to improve performance, then I can confirm that "social comparison" alone can also achieve performance improvement in complex operational settings.

Future research can continue to explore human subject behavior and performance in other dynamic decision problems. Our work points out that researchers may want to focus on the kind of complex dynamic problems that can be decomposed into manageable sub-problems. With this, it is necessary to understand: (1) the specific way to properly decompose the problem, which could vary based on the specific problem context and may require analysis by theoretical researchers, and (2) whether the sub-problem is indeed "manageable" for human subjects, which may require conducting experiments by behavioral researchers. Collaboration between theoretical and behavioral researchers will help us identify other complex dynamic problems that are "solvable" for human subjects.

Finally, in some other DP problems, such decision decomposition heuristic may or may not approximate the optimal policies. Nonetheless, our two key managerial messages could still hold: (1) Top performers may be able to identify ways to effectively simplify the problem and form decision heuristics, and (2) subjects can effectively learn from such decision heuristics and achieve a good performance. Future research that considers human behavior in DP problems should utilize the experimental design to collect decision policies from good/top performers. Once effective heuristics are identified, they will become natural candidates for performance improvement with human subjects. I believe that there are plenty of opportunities in the study of human decision-making in dynamic problems.

CHAPTER 4

Procurement with Change Order Renegotiation

4.1 Introduction

The authors were recently approached by a US-based original equipment manufacturer (OEM) to analyze their current procurement and cost management process. The OEM observes that, from rich historical data, the parts they wanted at the procurement stage were usually *not* the parts that they actually used for final production. The reason is that there is usually a two-to-three year engineering testing period between the procurement stage and the final mass production. During that period, engineers work to put together the parts acquired from suppliers, test the overall functionality and stability, and raise concerns to fix or change the design of the parts based on engineering needs. Therefore, the final price the OEM pays could be much higher than the price at the initial procurement stage.

How to manage these design changes and the corresponding price increase has become a key managerial question for the OEM. The OEM I work with has a long history of outsourcing the production of the parts. The way the OEM currently sources and manages such parts is very simple. At the supplier selection stage, the OEM specifies his initial needs for the product.[1] The suppliers compete by submitting their design prototypes, along with their bids for the prototype. The OEM tests the functionality of the prototypes, and picks the supplier who can produce a qualified prototype with the lowest bid. During the engineering testing period, the OEM negotiates with the (winning) supplier to determine the price to handle the design change. This mechanism has the advantage that it is straightforward to implement; I thereafter refer to it as the "simple mechanism". In addition, it does not require the OEM to write sophisticated state-contingent contracts at the supplier selection stage. This is particularly important in the industrial context, given that the OEM usually has limited knowledge of the design, and is therefore unable to specify ex-ante what the possible design changes will be.

However, the simple mechanism has the clear disadvantage that it gives the winning supplier

[1] Throughout this chapter, I use "he" to refer to the OEM, and "she" to refer to the supplier.

significant bargaining power in handling the design change. For the OEM, it is usually not possible to ask a different supplier to handle the design change: The winning supplier holds the intellectual property of the prototype she suppliers, and she is also the only one that understands the design and production details. Therefore, the OEM is "held-up" by the winning supplier in the engineering testing period. This can also be seen in how the pricing of design change is currently managed. In the engineering testing period, the buyer will receive requests from engineers and the marketing department regarding proposed design changes and the associated (estimated) benefits. The buyer will then contact the winning supplier to get a quote for handling the design change. The design change will be implemented if the quote is below the benefits, and will otherwise not be implemented. Therefore, in principle, the winning supplier can price in a way that fully squeezes the OEM's benefits from the design change.

The OEM is currently under pressure to achieve cost reduction, and there are two potential approaches. The first approach is to improve their sourcing and supplier management process. In particular, the OEM is evaluating an alternative mechanism that can potentially constrain the supplier's pricing power for the design change. This mechanism is a direct extension of the above simple mechanism: At the beginning of the whole process (procurement auction), the OEM announces a budget cap for the maximal amount he is willing to spend on the post-auction design change; I thereafter refer to this as the "budget mechanism". All the other elements of the mechanism remain unchanged, and the supplier who submits a feasible design with the lowest bid wins the auction. Supporters of the budget mechanism believe that it directly constrains the supplier's ability to gain from the design change and cuts the corresponding expenditures. On the other hand, some other people raise the concern that a more holistic view should be taken, especially regarding how the budget will affect the dynamics at the supplier selection stage.

The second approach the OEM is considering is to expand their supply base. For a long time, the OEM has primarily been sourcing from US-based suppliers (which I call the "incumbent supplier" hereafter). Recently, the OEM has started to qualify overseas suppliers (which I call "new entrant suppliers" hereafter) into the supply base. Compared to the incumbent supplier, new entrant suppliers in general have a much lower production cost for the prototype (initial design). However, it is expected their ability to handle major design changes is much weaker. When a design change request comes in, engineers from the incumbent supplier can more efficiently produce a solution due to their extensive collaboration experience with the OEM. For new entrant suppliers, this is much more difficult given that this is likely their first collaboration with the OEM. It may take them several iterations to produce a solution that can meet the OEM's design change request, leading to higher overall engineering costs. In general, it is expected that new entrant suppliers have a more efficient per unit variable cost for the simpler, more standardized product such as the prototype (initial design) but less efficient in handling the more tailored design change requests if the design needs to be changed prior to mass production.

In this chapter, I intend to study the unexpected ways that initiatives can affect total costs, especially when I consider both approaches together. Our main model focuses on the stylized case where I have one incumbent supplier and one new entrant supplier. The key insight of this chapter is that the performance comparison between the simple mechanism and the budget mechanism depends critically on the relative competitiveness comparison between the incumbent and the new entrant supplier. Interestingly, I find that the simple mechanism outperforms the budget mechanism when the new entrant supplier is sufficiently efficient in producing the prototype. That is, moving to the seemingly more sophisticated budget mechanism can hurt the OEM in the very setting that is likely to occur in practice.

The key insights are as follows. The budget mechanism has two impacts: (1) it benefits the OEM in handling the design change because a budget cap is imposed; (2) it hurts the OEM at the supplier selection stage because the suppliers will bid higher, as their expected windfall profits from handling the design change reduce. The overall effect of the budget mechanism depends on the comparison between these two forces. Meanwhile, note that imposing a budget will have an asymmetric impact on the two suppliers. The incumbent supplier's windfall profit reduces dramatically due to the budget, so she will bid a lot higher in the procurement auction. The new entrant supplier's profit is less dependent on the windfall profit because of her higher costs to handle the design change. However, she may become unable to handle the design change for the given budget: This happens when her change cost is higher than the budget imposed. Now, suppose the incumbent supplier wins the auction, then the overall effect is acceptable for the OEM: He pays more at the procurement stage, but he can benefit from the design change since the business goes to the incumbent supplier (with a properly chosen budget). However, if the new entrant supplier wins, the OEM is hurt on both stages. The ending price at the procurement stage is higher (note that it is determined by the incumbent supplier's dropout bid), *and* the business goes to the new entrant supplier, who may not be able to handle the design change, so there is no way for the OEM to benefit from it. Therefore, the OEM is hurt by the budget mechanism. In this chapter, I also propose an experimental design that considers how humans will actually implement the budget mechanism, standing from the viewpoint of the OEM. The key question I aim to address in the experiment is that, given the complex tradeoffs involved in implementing the budget mechanism, human managers of the OEM may or may not be able to set the budget wisely, and the experiment aims at gaining insights into their systematic patterns in budget-setting behavior.

Finally, I note that our model has broader applications in many other contexts, thanks to the flexible structure of our model. In general, our model applies to any context where there is an upfront auction stage and a post-auction stage, during which the winning party and the auction organizer will interact once more. I note that such interactions need *not* be contract renegotiation.

Examples include any supply activities that involve after-sale services, such as airplane engine supply & maintenance. Another line of examples includes government contracting in infrastructure building; see the Literature Review section for more examples and related studies in economics.

This chapter is organized as follows. Section 4.2 reviews the literature. Section 4.3 presents the setup of the model. Section 4.4 analyzes the simple mechanism and the budget mechanism. Section 4.5 compares these two mechanisms. Section 4.6 conducts numerical studies. Section 4.7 proposes an experimental design based on the theoretical models. Section 4.8 concludes this chapter. All proofs are relegated to the Appendix C.

4.2 Literature Review

The study on auctions with post-auction activities in OM is limited. The Economics literature has a long history of studying renegotiation. Tirole (1986) and Hart & Tirole (1988) are classical references for renegotiation in contracting. Bajari & Tadelis (2001) compare the performance of a fixed-price contract with a cost-plus contract in an incomplete contracting setting. For auction with renegotiation, several papers consider the possibility of renegotiation as an integrated part of the buying mechanism. In these papers, the buyer has the option to make use of renegotiation to extract more surplus right after seeing the outcome of the auction. Waehrer (1995) analyzes the situation where the procurer and the winner can renegotiate a new contract. Wang (2000) and Shachat & Tan (2015) consider a setup where the procurer can reject all the bids and negotiate with the supplier who placed the lowest bid. In Herweg & Schwarz (2018), the buyer is aware of both the baseline design and a fancier design. The buyer runs a simple price-only auction, but needs to choose which design to put forward for auction. After the auction concludes, the buyer and the winning supplier can renegotiate to deviate from the design put forward by the buyer in the auction. Hence, renegotiation is a way to further improve the surplus of both parties. Our model is different as the design change is a separate event which happens at a later time point away from the auction. Meanwhile, I do share a similar incentive structure as the previous papers, where the post-auction activities can affect bidders' (suppliers) auction behavior as suppliers expect to gain from the post-auction renegotiation.

This chapter also relates to the literature on auctions with asymmetric bidders. Maskin & Riley (2000) show that the revenue equivalence theorem fails to hold when bidders are asymmetric in their types. In particular, different practical mechanisms (such as the sealed bid auction vs. the English auction) will generate different expected revenue for the seller, and the performance comparison between different mechanisms depends on the relative strength between the two suppliers. In general, the practical mechanism(s) will *not* implement the optimal mechanism, even with simple extensions such as adding the reserve price. The literature on asymmetric auction has continued to focus on evaluating different practical mechanisms. See Hafalir & Krishna (2008) for a comprehensive review. This chapter also considers an environment with (two) asymmetric bidders. Different from the previous literature that typically draws a clear distinction between a "strong" and a "weak" supplier, in this chapter, the two suppliers are both "strong" but in *different dimensions*: Compared to the new entrant supplier, the incumbent supplier is more efficient (stronger) in handling the design change, but less efficient (weaker) in producing the baseline design. Given the difficulty of implementing the optimal mechanism, our focus is on comparing two practical mechanisms - the simple mechanism vs. the budget mechanism. In Section 4.6, I am able to derive the optimal mechanism (under extra assumptions about the knowledge structure), which serves as a lower bound for performance evaluation of the practical mechanisms.

Finally, there is also a growing body of empirical literature that considers the impact of renegotiation on project management. Bajari et al. (2014) identify and estimate the economic impact of post-auction renegotiation in California highway construction. Ryan (2020) uses a structural model to evaluate the impact of strategic contract renegotiation on overall cost spending in an Indian energy project.

4.3 The Model

The stylized model I consider has two suppliers: i and j. Their private types are denoted as θ_i and θ_j, both of which are i.i.d. over over $[\underline{\theta}, \overline{\theta}]$ according to $H(\theta)$. The procurement problem I consider has two stages: (1) the supplier selection (procurement) stage, where the suppliers submit prototypes and their bids for their own prototype; (2) the post-auction stage, where the winning supplier may be involved to handle the design change request. Each supplier's cost has two parts: the cost to produce the prototype, denoted as $C(\theta)$, and the cost to handle the design change once it is requested, denoted as $\Delta C(\theta)$. In this chapter, I focus on the situation of linear costs, i.e., I have: $C(\theta) = C \cdot \theta; \Delta C(\theta) = \Delta C \cdot \theta$. The linear cost assumption is standard in procurement auction literature, for example Chen (2007) and Duenyas et al. (2013).

To be consistent with the practical situation I am interested in, in this chapter, I focus on the case of the asymmetric supply base. Without loss of generality, denote supplier i as the incumbent supplier and supplier j as the new entrant supplier. Both suppliers have been qualified to compete for the business at the supplier selection stage. However, as has been discussed in the Introduction, compared to the incumbent supplier, the new entrant supplier is more efficient in producing the prototype but less efficient in handling the design change request. Therefore, the *parameters* in their cost functions are different, and follow the relationship of $C_i > C_j, \Delta C_i < \Delta C_j$.[2]

[2]One could certainly impose stronger assumptions where the new entrant supplier's design change cost is *always* stronger than the incumbent supplier's change cost: $\Delta C_j \cdot \theta_j > \Delta C_i \cdot \theta_i,\ \forall \theta_i, \theta_j$.

4.3.1 The Design Change

To consider the impact of the design change, I analyze the problem backward by starting at the post-auction stage. When the design change is needed at the post-auction stage, it will bring a benefit M to the OEM. The realized value of M is seen by both parties.[3] Due to the winning supplier's strong position in the post-auction stage, she will be able to price the service of handling the design change in such a way that completely squeezes the OEM's benefit from the design change. In practice, the winning supplier makes a take-it-or-leave-it quote for the design change to the OEM. When the winning supplier's change cost is lower than the realized benefit m, her quote can just be the realized value of m, and the design change request is resolved. When the winning supplier's change cost is higher than the realized benefit m, she will necessarily charge a price higher than the realized benefit to at least break even, and the OEM will choose to turn down the quote since it is not worth it.

Moving backward, at the procurement stage, the design change is expected to happen with probability q, which I normalize to 1 in this chapter; all our results hold without this normalization. The benefit of M is also random ex-ante (at the procurement stage), and is assumed to be drawn from a distribution $F(m)$ over $[\underline{m}, \overline{m}]$. For supplier i with type θ_i, under the simple mechanism, the expected benefit from the design change is:

$$B_i^S(\theta_i) = \int_{\Delta C_i \cdot \theta_i}^{\overline{m}} m - \Delta C_i \cdot \theta_i dF(m). \tag{4.1}$$

In the next section, I will discuss how this expected value changes when a budget b is imposed. Finally, following common procurement management literature, I assume that all the above setup is common knowledge. That is, apart from the private types θ, all the other setup is known among all the parties (the OEM and the two suppliers).

4.4 Analysis of the Two Mechanisms

This section presents the analyses and equilibrium predictions of the two mechanisms. I use supplier i for illustration when appropriate; the case for supplier j follows the same analysis method.

[3] In practice, the supplier can usually infer the usage from the design change request. An alternative model setup is that the supplier only knows the distribution of M but not the realization value of it. In that case, the supplier will always submit a fixed quote.

4.4.1 The Simple Mechanism

In the procurement stage of the simple mechanism, supplier i's expected benefit from design change as being stated in (1):

$$B_i^S(\theta_i) = \int_{\Delta C_i \cdot \theta_i}^{\overline{m}} m - \Delta C_i \cdot \theta_i dF(m).$$

Therefore, in the auction stage, supplier i will include the expected benefit $B_i(\theta_i)$ into her bidding behavior. Therefore, her bid is in fact *lower* than the supplier's production cost for the prototype. Formally:

$$DB_i^S(\theta_i) = C_i \cdot \theta_i - B_i^S(\theta_i). \tag{4.2}$$

I impose the regularity assumption that such drop put bid is *increasing* in the type. In the simple mechanism, the supplier competes by lowering their bids at the auction stage. It is their dominant strategy to stay in the auction so long as the outstanding price is higher than their dropout bids. Formally, I have the following proposition.

Proposition 1. *In the simple mechanism, supplier i and j have the dominant strategy to decrease their bids to stay in the auction, until the outstanding price is below their dropout bids $DB_i^S(\theta_i)$, $DB_j^S(\theta_j)$. The auction ends at $max\{DB_i^S(\theta_i),\ DB_j^S(\theta_j)\}$, and the supplier who stays in the auction wins.*

4.4.2 The Budget Mechanism

This mechanism is identical to the simple mechanism, except that a *budget* b is imposed on the expense of handling the design change. When analyzing and presenting the results in the budget mechanism, I will consistently refer back to the simple mechanism to better understand the impact of budget b.

The value of b is chosen by the OEM, announced at the supplier selection stage, and *cannot* be renegotiated in the future. From the supplier's viewpoint, the budget b has two potential impacts: (1) it decreases the supplier's expected gain from handling the design change; (2) it makes it unprofitable for some types of suppliers to handle the design change. The actual impact will depend on the value of b.

4.4.2.1 Case 1: $b \geq \Delta C_i \theta_i$.

First, if the budget b is higher than the supplier's cost to handle the design change, then the probability that this supplier will be able to handle the design change remains unchanged. She will want to handle the design change so long as the realized benefit m is higher than her cost $\Delta C_i \cdot \theta_i$. However, her ability to squeeze the OEM's benefit is constrained by the budget b: When the realized benefit is higher than b, supplier i is only able to charge the OEM b. Formally, her expected benefit from the design change now becomes:

$$\begin{aligned} B_i^b(\theta_i) &= \int_{\Delta C_i \cdot \theta_i}^{b} m - \Delta C_i \cdot \theta_i dF(m) + \int_b^{\overline{m}} b - \Delta C_i \cdot \theta_i dF(m) \\ &= B_i^S(\theta_i) - \int_b^{\overline{m}} (m-b)dF(m) \end{aligned}$$

I observe from this that her expected benefit decreases by $\int_b^{\overline{m}}(m-b)dF(m)$. Interestingly, this term is independent of her type or cost parameters. This will play an important role when I consider the impact of having a budget b. Her dropout bid then becomes:

$$\begin{aligned} DB_i^b(\theta_i) &= C_i \cdot \theta_i - B_i^b(\theta_i) = C_i \cdot \theta_i - [B_i^S(\theta_i) - \int_b^{\overline{m}} (m-b)dF(m)] \\ &= DB_i^S(\theta_i) + \int_b^{\overline{m}} (m-b)dF(m) \end{aligned} \tag{4.3}$$

That is, her dropout bid under the budget mechanism is higher than that in the simple mechanism, by an amount independent of her cost or type.

4.4.2.2 Case 2: $b < \Delta C_i \theta_i$.

In this case, supplier i cannot handle any design change, regardless of the realized benefit m. This is because her cost to handle the design change is higher than the largest amount she can earn from the design change, which is b. In this case, her expected benefit from the design change is simply 0, and her dropout bid is just her production cost for the prototype.

$$DB_i^b(\theta_i) = C_i \cdot \theta_i = DB_i^S(\theta_i) + \int_{\Delta C_i \cdot \theta_i}^{\overline{m}} m - \Delta C_i \cdot \theta_i dF(m) \tag{4.4}$$

From (4.4), I can see that the supplier's dropout bid also increases in this case, compared to the simple mechanism. However, the increased amount is now dependent on her type and cost parameters.

Moreover, notice that since $b < \Delta C_i \theta_i$, I have: $\int_{\Delta C_i \cdot \theta_i}^{\overline{m}} m - \Delta C_i \cdot \theta_i dF(m) < \int_b^{\overline{m}}(m-b)dF(m)$. That is, the increased amount is smaller than what it would have been if she were able to do the design change under budget b. I will explain the managerial implication of this when I compare the two mechanism in the next section.

In the supplier selection stage of the budget mechanism, it is still in dominant strategy for both suppliers to bid towards her dropout bids. Formally, I have the following proposition.

Proposition 2. *In the budget mechanism, supplier i and j have the dominant strategy to decrease their bids to stay in the auction, until the outstanding price is below their dropout bids $DB_i^b(\theta_i)$, $DB_j^S(\theta_j)$. The auction ends at $max\{DB_i^b(\theta_i),\ DB_j^b(\theta_j)\}$, and the supplier who stays in the auction wins.*

4.5 Comparing the Two Mechanisms

I am now ready to formally compare these two mechanisms. I take the perspective of the OEM and consider the expected *total* transfer he pays under each mechanism.[4] The total transfer covers *both stages*: the supplier selection stage and the post-auction design change stage.

The transfer during the post-auction design change stage requires further discussion. There, the OEM incurs two potential cash flows. First, suppose the OEM is able to resolve the design change with the help of the winning supplier, then he receives the realized benefit m; in practice, this could come from marketing considerations of attracting more customers. Second, the OEM pays the winning supplier for being able to offer a solution for the design change.

I use the simple mechanism as the performance benchmark for our analysis. In this mechanism, the winning supplier has all the bargaining power in handling the design change and will squeeze all the OEM's benefit from the design change. Therefore, the OEM incurs a transfer of exactly 0 in resolving the design change. That is, the OEM's expected total transfer in the simple mechanism is, as the benchmark:

$$\int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} \max\{DB_i^S(\theta_i), DB_j^S(\theta_j)\} dH(\theta_j) dH(\theta_i) \tag{4.5}$$

Imposing a budget cap b will have an impact on both stages. As I have seen in Section 4.2, imposing a budget b will increase the dropout bids from both suppliers because the expected benefit from the design change decreases; therefore, it hurts the OEM at the procurement stage. Meanwhile, imposing a budget may benefit the OEM in the design change stage: The budget puts an upper bound on how much the OEM will pay the winning supplier. Therefore, which mechanism is better for the OEM depends on the strength of these two opposite forces.

[4] Note that I calculate this value here for the purpose of policy evaluation, and it does *not* mean that the OEM knows the details of the design change.

4.5.1 Impact of Budget b on Mechanism Comparison

Naturally, the size of these two forces depends on the selection of b value. Therefore, the comparison between the two mechanisms will also depend on the selection of b. Below, I first analyze the performance comparison between the two mechanisms as a function of b.

4.5.1.1 A Large b Value

I first begin by considering the simplest case: b is so high that *any* supplier i and j can still handle the design change: $b > \Delta C_i\overline{\theta}, b > \Delta C_j\overline{\theta}$. That is, the impact of b is only on the payment made for the design change, but not whether the supplier can handle it or not. Because of this, intuitively, it should be very attractive to the OEM: Doing so reduces the OEM's payment for the design change while does not reduce the supplier's ability to handle the design change. Interestingly, I find that this actually has zero *total impact* on the OEM's payment once I take into account the higher auction bids.

Specifically, in this case at the procurement stage, both suppliers dropout bids follow Case 1 in Section 4.2, and increase by $\int_b^{\overline{m}}(m-b)dF(m)$. Therefore, the winner in the two mechanisms *remain unchanged*, and the ending price increases by exactly $\int_b^{\overline{m}}(m-b)dF(m)$, which is an amount independent of the types or parameters of the two suppliers.

In the post-auction design change stage, the OEM receives the realized benefit m. However, his payment to the winning supplier is now at most b. Therefore, the OEM incurs a positive cash flow whenever the realized benefit is larger than b. Hence, his expected gain from the design change is no longer 0, but a positive value $\int_b^{\overline{m}}(m-b)dF(m)$.

Now, when I put the two stages together, I have an interesting observation. For the OEM, the ending price at the procurement stage increases by $\int_b^{\overline{m}}(m-b)dF(m)$, but he earns exactly $\int_b^{\overline{m}}(m-b)dF(m)$ from the design change stage; therefore, the total transfer in the budget mechanism is exactly the same as that in the simple mechanism. This is formalized in the following proposition.

Proposition 3. *If b is set at a high value such that $b > \Delta C_i\overline{\theta}, b > \Delta C_j\overline{\theta}$, then the simple mechanism and the budget mechanism give the same expected total transfer for the OEM.*

4.5.1.2 A Small b Value

I now consider the other extreme case. Suppose b value is so small that *no types* of supplier i or j can handle the design change: $b < \Delta C_i\underline{\theta},\ b < \Delta C_j\underline{\theta}$. This seems to make intuitive sense when one only considers the *expense* for handling the design change. However, there are two additional impacts: (1) the OEM can never receive the benefit from the design change; (2) the ending price in the supplier selection stage is higher. Therefore, overall the OEM will be strictly worse off. Formally, I have the following result:

Proposition 4. *If the b is low enough such that neither supplier can handle the design change, i.e., $b < \Delta c_i \cdot \underline{\theta}, b < \Delta c_j \cdot \underline{\theta}$, then the budget mechanism is worse than the simple mechanism for the OEM in total transfer, both from an ex-ante point of view (expected value) and an ex-post point of view (any realization of types).*

The intuition is as follows. Consider the post-auction design change stage first. In the simple mechanism, although the OEM does need to spend a lot of money in involving the winning supplier to handle the design change, such spending is actually "justified" by the benefit of the design change. Therefore, the OEM incurs no "net spending" from the design change. The budget mechanism also incurs 0 "net spending" for the OEM in handling the design change, but this is simply because neither supplier can handle it. Meanwhile, at the supplier selection stage, the OEM will enjoy a low ending price in the simple mechanism because the suppliers bid in a way anticipating the windfall profits from the design change stage. On the other hand, the suppliers in the budget mechanism will not be able to bid such a low price because they anticipate no future gain after earning the business. Putting the two stages together, I can see that the simple mechanism is a better choice. In fact, its advantage is so strong that it is better than the budget mechanism *for any realization of supplier types*.

4.5.1.3 An Intermediary b Value

I now consider the most complicated case: an intermediary b value: $\Delta C_i \underline{\theta} \leq b \leq \Delta C_j \overline{\theta}$. In this case, for suppliers i and j, it may be that some types of them will be able to handle the design change, while some other types cannot. In general, such a b value will divide the type space into different regions depending on whether the suppliers can handle the design change; these regions need to be analyzed separately.

Region 1. Both Suppliers Can Handle the Design Change

Region 1 covers the situation where *both* suppliers can handle the design change: $\Delta C_i \cdot \theta_i \leq b, \ \Delta C_j \cdot \theta_j \leq b$. In this region, our result from Section 4.5.1.1 applies: The simple mechanism and the budget mechanism performance the same for the OEM.

Region 2. Neither Suppliers Can Handle the Design Change

Region 2 covers the situation where *neither* suppliers can handle the design change: $\Delta C_i \cdot \theta_i \geq b, \ \Delta C_j \cdot \theta_j \geq b$. In this region, our result from Section 4.5.1.2 applies: The simple mechanism performs strictly better than the budget mechanism for the OEM.

Region 3. Supplier i Can Handle the Design Change While Supplier j Cannot Not

This is the most interesting region. Because supplier i can still handle the design change under the budget b, the OEM will be able to benefit in the post-auction design change stage *if* eventually supplier i wins the auction. When supplier i wins, the dropout bid will be determined by supplier

j. However, as I have shown in Section 4.2.2, the increase in supplier j's dropout bid is *smaller* than the OEM's benefit in the post-auction stage. Therefore, so long as supplier i is selected, the OEM will be able to benefit from the budget mechanism.

However, I should note that supplier i in this region does not always win the auction. In particular, if she loses the auction, then the OEM will be hurt by the budget mechanism. The business goes to supplier j, who cannot handle the design change (so there is no way for the OEM to benefit in the post-auction stage); meanwhile, the ending price in the supplier selection stage is determined by supplier i, which increases by $\int_b^{\overline{m}}(m-b)dF(m)$. Therefore, the OEM is worse off in the budget mechanism.

Therefore, the overall effect in this region is unclear; I can find different numerical examples where the budget mechanism performs better or worse in this region, depending on the cost parameters and type distributions.

4.5.1.4 Summary and Discussions

As I have seen above, the ex-ante comparison between the two mechanisms depends critically on the value of b chosen:

- When b is large, the two mechanisms perform the same for the OEM.
- When b is small, the budget mechanism performs worse.
- When b is in an intermediary value, the exact ex-ante comparison depends on the parameters and type distributions. Note that if the budget mechanism is a worse choice in this region, I can then directly conclude that *no b value* can make the budget mechanism a better choice (from an ex-ante viewpoint). Otherwise, the b value should be chosen in this region for the budget mechanism to outperform the simple mechanism.

As an obvious next question, I would like to know: What is the supply base condition that will make it possible/impossible for the budget mechanism to outperform the simple mechanism? Understanding this will be useful in guiding the OEM's future practice. For the supply base condition where the budget mechanism cannot outperform the simple mechanism, surely the OEM should not implement it.

Even for the supply base condition where the budget mechanism can stand out, I should understand two key questions. First, how easy it is to find a b value that can lead to performance improvement, and how "robust" it is to ex-post type realizations? Ideally, I would like to find the b value that can achieve performance improvement for *any realization of types.* This has important managerial implications in the current OEM's situation because knowing the *distribution* of supplier types and suppliers' beliefs for the type space is extremely challenging in practice. Second, how large is the benefit? If the improvement is not significant and sensitive to type realizations, then it may be better for the OEM to simply implement the original simple mechanism. The second question is considered in Section 4.6 where I conduct extensive numerical studies. For the first question, I am now ready to present our results.

4.5.2 Impact of Supply Base on Mechanism Comparison

Recall that supplier i is assumed to be less efficient in the base cost parameter: $C_i > C_j$, but less efficient in the change cost parameter: $\Delta C_i < \Delta C_j$. For the OEM I work with, supplier i is their current (incumbent) supplier, and supplier j is the new (new entrant) supplier they admit to their supply base. In the language of our model, their managerial questions can be stated as follows: How to select the mechanism in a way that matches the situation of supplier i and j? In particular, will the budget mechanism be a good choice over the simple mechanism if the supplier j is strong (in being efficient in producing the prototype)?

Our results from the previous subsection help us address this question. From Section 4.5.1, I conclude that only an intermediary value of budget b (region 3) will be possible for the budget mechanism to perform better; in particular, the budget mechanism is better *only when* supplier i (who can handle the design change) ends up winning the auction. Therefore, intuitively, if supplier j is very strong, then supplier i is less likely to win, making the budget mechanism less likely to benefit the OEM.

This result is formalized in the following proposition. Our formal comparison of the strength of the two suppliers is captured by their dropout bids in the simple mechanism. A higher dropout bid means a weaker supplier (in terms of costs).

Proposition 5. *Suppose the most efficient supplier i (type $\underline{\theta}$) has a strictly larger dropout bid compared to the least efficient supplier j (type $\overline{\theta}$) in the simple mechanism, then the budget mechanism performs no better than the simple mechanism for the OEM both in terms of ex-ante and ex-post total transfer.*

Therefore, if supplier j is so strong relative to supplier i, then the budget mechanism is *guaranteed* to perform worse relative to the simple mechanism. I note that this is a fairly strong result because the budget mechanism is better for any realization of types. Therefore, I can even extend the results to any distribution function over the type space, and I can forgo the assumption that the distributions are common knowledge among the players. This is extremely important in practice because the common knowledge assumption is very difficult to test or verify in practice.

On the other hand, if supplier j is weak relative to supplier i, then supplier i wins more often, and it is also more likely for the budget mechanism to perform better. This intuition is formalized in the following proposition.

Proposition 6. *If the least efficient supplier j (type $\overline{\theta}$) has a strictly larger dropout bid compared to the least efficient supplier i (type $\overline{\theta}$) in the simple mechanism, then there exists a range of budget values b such that the budget mechanism is a better choice for the OEM compared to the simple mechanism in terms of total transfer, both in the sense of ex-ante (expected total transfer) and ex-post (for any realization of types).*

Putting these two results together, I can have a holistic view regarding the performance comparison between the two mechanisms as a function of the relative strength of the two suppliers. When supplier j is sufficiently weak (in the base cost parameter), the budget mechanism *always* leads to better performance for the OEM compared to the simple mechanism. When I increase supplier j's strength, it becomes less and less likely (in terms of type realizations) that the budget mechanism can benefit the OEM. When supplier j is sufficiently strong, the budget mechanism *never* leads to better outcomes for the OEM.

With this, I can provide an answer to the OEM. If they want to admit a very strong supplier j to the supply base, the original simple mechanism is a better choice. On the other hand, if they want to admit a weak supplier j, then the budget mechanism, with a properly chosen budget b, may lead to a better outcome.

4.6 Numerical Studies

The purpose of this section is to establish *how much* the OEM can benefit from implementing the budget mechanism. Based on Proposition 6, the budget mechanism performs the best when supplier j is relatively weak. However, it is important to understand how large this advantage is. I should note that the implementation of the budget mechanism is a careful exercise. As I have pointed out in Section 4.5.1, the mis-selection of budget value b can have severe negative consequences on the total transfer, and careful evaluation needs to be done to determine the optimal b. It will be necessary for the budget mechanism to perform sufficiently better compared to the simple mechanism in order to justify the effort. Otherwise, the OEM may as well just use the simple mechanism.

In addition, it is also important to establish a lower bound for the expected total transfer. In our context, such a lower bound can be derived by solving the "optimal mechanism". So far, I have been assuming that the OEM cannot write a contingent contract regarding the design change details and associated payments. Now, suppose the OEM is able to do this, then I will be able to apply the classical techniques from Myerson (1981) to establish the optimal mechanism. Note again that the optimal mechanism is hard to implement in practice since it requires the OEM to have significantly more knowledge about the decision environment; nonetheless, it provides the lower bound for how good any practical mechanisms can possibly achieve. Below, I first formally establish the "optimal mechanism" in Section 4.6.1, which I then use as the lower bound for the expected total transfer when I compare the two practical mechanisms in Section 4.6.2.

4.6.1 A Lower Bound of the Expected Total Transfer

Suppose the OEM has knowledge of the design change details and each supplier's change cost *parameter* ΔC, she can write the contingent contract regarding the design change details and associated payments; suppose further that the OEM can commit to not to renegotiate the contract, then I am able to establish the optimal mechanism by relying on the Revelation principle and the classical techniques from Myerson (1981). Following the standard notation, let $T_i(\tilde{\theta}_i)$ denote the expected transfer supplier i receives by reporting her type to be $\tilde{\theta}_i$ at the procurement auction stage; this does *not* include the payment she will receive from handling the design change.[5] Let $W(\tilde{\theta}_i)$ be the associated expected winning probability. Finally, I define $G_i(\tilde{\theta}_i) = 1 - F(\Delta C_i \tilde{\theta}_i)$ be the supplier's probability of being allowed to handle the design change from reporting $\tilde{\theta}_i$. At the procurement stage, supplier i's expected total utility from reporting $\tilde{\theta}_i$, while her true type being θ_i, is:

$$U_i(\tilde{\theta}_i; \theta_i) = T(\tilde{\theta}_i) + W(\tilde{\theta}_i)[C_i \cdot \tilde{\theta}_i - C_i \cdot \theta_i + (\Delta C \cdot \tilde{\theta}_i - \Delta C_i \cdot \theta_i) \cdot G_i(\tilde{\theta}_i)]. \tag{4.6}$$

This expression is due to the fact that the OEM can now contract on the design change cost $C_i \tilde{\theta}_i$ with supplier i. In other words, the only way for her to benefit from the design change is by misreporting $\tilde{\theta}_i$. The notation for supplier j is symmetric.

By the revelation principle, it suffices to analyze the truth-telling equilibrium, i.e., making sure that the expected transfer and winning probability are selected such that each supplier is incentive compatible (to truthfully reveal her type). The results are similar to the monotonicity requirement I have on the expected winning probability in the classical mechanism design literature, except that they need to be slightly adjusted because there are now two associated costs: cost for the baseline design and the design change cost. I have the following result:

Proposition 7. *The direct mechanism is incentive compatible if and only if:*

1. $W_i(\theta_i)(C_i + \Delta C_i \cdot G_i(\theta_i))$ *and* $W_j(\theta_i)(C_j + \Delta C_j \cdot G_j(\theta_j))$ *are both non-increasing.*
2. $U_i(\theta_i; \theta_i) = U_i(\overline{\theta}; \overline{\theta}) + \int_{\theta_i}^{\overline{\theta}} W_i(x)(C_i + \Delta C_i \cdot G_i(x))dx;$
   $U_j(\theta_j; \theta_j) = U_j(\overline{\theta}; \overline{\theta}) + \int_{\theta_j}^{\overline{\theta}} W_j(x)(C_j + \Delta C_j \cdot G_j(x))dx.$

[5] This requirement is inconsequential; I can easily rewrite the model such that the $T_i(\tilde{\theta}_i)$ also includes the expected transfer from handling the design change

Part 1 of this proposition is the familiar monotonicity requirement for the allocation rule (winning probabilities), now adjusted to account for costs in both stages. The expressions in part 2 can be derived by applying the envelope theorem. With this, I am ready to establish the optimal allocation rule in the optimal mechanism in the following proposition. Essentially, I am selecting suppliers based on a version of the "virtue value".

Denote the virtual values for the two suppliers: $v_i(\theta_i) = \frac{H(\theta_i)}{h(\theta_i)}(C_i + \Delta C_i \cdot G_i(\theta_i)) + C_i \cdot \theta_i - B_i^S(\theta_i)$; $v_j(\theta_j) = \frac{H(\theta_j)}{h(\theta_j)}(C_j + \Delta C_j \cdot G_j(\theta_j)) + C_j \cdot \theta_j - B_j^S(\theta_j)$. I have the following result:

Proposition 8. *If $v_i(\theta_i)$ is increasing in θ_i and $v_j(\theta_j)$ is increasing in θ_j, then in the optimal mechanism (in the sense of achieving the lowest expected total transfer for the OEM), the supplier with the lower virtual value wins the auction.*

The terms in the virtual value deserves further discussion. I use supplier i for illustration. The terms $C_i + \Delta C_i \cdot G(\theta_i)$ represents the OEM's expected total transfer from selecting supplier i if she has full information of supplier i's type. The first term $\frac{H(\theta_i)}{h(\theta_i)}(C_i + \Delta C_i \cdot G_i(\theta_i))$ is the information rent the OEM needs to pay to induce truth-telling from supplier i.

4.6.2 Numerical Comparison of the Two Practical Mechanisms

For numerical studies, I consider the following set of parameters as it reflects the cost structures of one part: $C_i = 3, \Delta C_i = 2, C_j = 3, \Delta C_j = 5$. The types are uniformly distributed over $[1, 2]$. The b in the budget mechanism is selected optimally. Moreover, I further consider the impact of decreasing the c_j, which captures the situation that the OEM is able to identify an even more efficient new entrant supplier.

As I have indicated in Proposition 5 and 6, the budget mechanism is most beneficial when the new entrant supplier is not so efficient (high C_j); when the new entrant supplier is very efficient, the two mechanisms perform the same because there is no value of budget b that could make the budget mechanism outperform the simple mechanism.[6]

I also have a very important observation: The two practical mechanisms, in particular the simple mechanism, perform the closest to the optimal mechanism in the middle range of C_j. Note that I am in an asymmetric supplier base, and ΔC_j is a lot larger than ΔC_i. Therefore, when C_j is close to C_i, the new entrant supplier is in a complete cost disadvantage to the incumbent supplier. When I decrease C_j, the new entrant supplier becomes more and more competitive. In the middle of the range, the two suppliers are actually similarly competitive in the simple mechanism, and competition with a "even" supply base leads to an outcome close to the "optimal" bound.[7]

[6]In this case, it is optimal to choose a very high b, so that the two mechanisms perform the same.

[7]In some area, the optimal bound is slightly higher than the budget mechanism performance due to randomness in (estimated) performance from simulation.

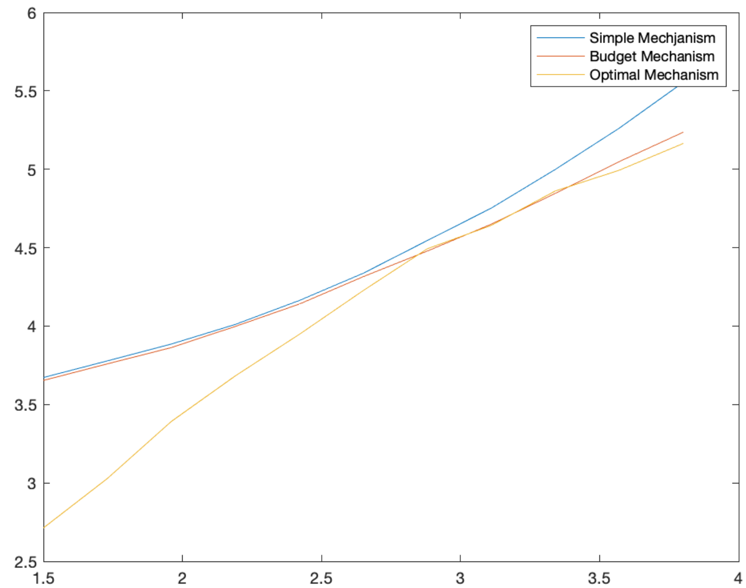

Figure 4.1: Impact of C_j

However, as I further decrease C_j, neither mechanism gets a significant performance improvement, while the "optimal" bound drops sharply. This is because with a low C_j, the auction (in either practical mechanism) is almost always won by the new entrant supplier. In other words, the ending price is almost always determined by the incumbent supplier, whose strength does not change in this numerical study. Therefore, further decreasing C_j brings no benefit to the OEM. The "optimal" mechanism, however, has a steady performance improvement because it induces truth-telling from *both* suppliers, and the OEM will be able to benefit from a more and more efficient new entrant supplier as C_j drops.

4.7 Proposed Experimental Design

In this section, I offer an overview of the experiments. Overall, the goal of the experiment is to consider human subjects' ability to utilize the budget mechanism. In the experiments, subjects' payoff will be proportional to the OEM's expected total transfer. I note that the OEM seeks to minimize the expected total net transfer, i.e., the smaller the value, the better for the OEM. Hence, to motivate subjects to minimize the number, I will set up their payoff to the extent that they can achieve a "cost saving". That is, I will pre-announce a fixed "cost target" C_{payoff}. When subjects make the budget decision and the OEM's total transfer is realized, I will calculate how much the net transfer is below the cost target of C_{payoff}, and use this difference to determine the subjects' actual payoff. In other words, the higher the saving is, the more beneficial it is for the subjects.

As I have discussed in the previous section, there are conditions where the budget mechanism can benefit the OEM (with a properly chosen budget value) and conditions where it is never possible to benefit the OEM. I note that I can vary the supply base condition to determine which scenario subjects will be in, and this can be achieved by varying one out of the four cost parameters, such as by varying *only* C_i.

With this, I consider the following two treatments.

1. The "useless budget" treatment, where the supply base condition is such that it is never possible to choose a budget value to benefit the OEM. In other words, it is optimal to select a very high, effectively non-binding budget value b.

2. The "useful budget" treatment, where the supply base condition is such that it is possible to choose a budget value to benefit the OEM. In particular, it can make the OEM better off both ex-ante and ex-post.

These two treatments will help to establish the two extreme conditions, which will then help us understand subjects' ability to utilize the budget mechanism. Regarding their behavior, I first note that it may be challenging for subjects to act optimally in the "useless budget treatment" because it is rather counter-intuitive: They are presented with a tool, but the optimal decision is effectively not to use it (not to set a binding budget). I therefore propose the following hypothesis.

Hypothesis 8. *In the "useless budget" treatment, a decent portion of subjects will choose a binding budget and therefore achieve sub-optimal performance.*

Regarding the "useful budget" treatment, subjects' performance lies in their ability to discover the existence of an appropriate budget value. Because I do not have any ex-ante theory against their ability to do so, I propose a neutral hypothesis.

Hypothesis 9. *In the "useful budget" treatment, subjects can converge to the budget values that can benefit the OEM, both ex-ante and ex-post, as they gain experience in the experiments.*

I will implement a between-subject experiment, meaning that subjects will be in either one of the two treatments. Within each treatment, subjects will make the budget decisions repetitively. Supplier types and the benefit values are independent from round to round. At the end of the experiment, I will randomly select one round and use the subjects' performance in that round to determine his/her payoff in the experiment.

4.8 Conclusion

In this chapter, I model and analyze the OEM's benefit and cost in implementing two practical mechanisms - the simple mechanism and the budget mechanism. I find that, quite interestingly, implementing the seemingly more sophisticated mechanism, the budget mechanism, can bring unfavorable outcomes to the OEM in terms of the total transfer (the money spent on procurement and handling the design change). Motivated by practical situations, in this chapter, I consider an asymmetric supply base: a incumbent supplier and an new entrant supplier. Different from traditional literature in asymmetric auctions, both suppliers in this chapter have their own strengths, but in *different dimensions*: Compared to the new entrant supplier, the incumbent supplier is more efficient in handling the design change but less efficient in producing the baseline design (prototype).

I find that when the new entrant supplier is relatively "weak" in the sense of having a high baseline design production cost (although still lower than that of the incumbent supplier), the budget mechanism can benefit the OEM with a properly chosen b. On the other hand, when the new entrant supplier is very "strong" in the baseline design production cost, the budget mechanism can never outperform the simple mechanism. In practice, the OEM clearly wants to admit a very efficient new entrant supplier to the supply base, provided that she is qualified. Therefore, our key managerial insight is that, in this situation, sticking with the original simple mechanism is better than trying to implement the seemingly sophisticated budget mechanism.

By comparing the two practical mechanisms with the theoretical lower-bound - the "optimal" mechanism - I find that the simple mechanism performs quite close to the optimum when the supply base is "balanced", i.e., when the *overall* cost (that considers both the baseline cost and the change cost) is similar between the two suppliers. This result offers important managerial insights from a slightly different angle: Suppose the OEM wants to stick with the simple mechanism, then the best way for him to get a performance guarantee is to form a "balanced" supply base. I also propose an experimental design that can help evaluate human subjects' actual ability to make use of the budget mechanism. In particular, I propose varying the supply base conditions and test if human subjects can correctly set the budget value, with the goal of achieving savings for the OEM.

Future research can extend our analyses in various dimensions. For example, one can consider the implication of other practical mechanisms in this context. Another possible extension is to relax the linear cost assumption. Doing so will impose great challenges in deriving the "optimal" mechanism, but the analysis with the practical mechanisms should still be feasible.

CHAPTER 5

Conclusion and Future Work

In this dissertation, I consider how humans make complex operational decisions in a wide range of settings. I cover various practical contexts, including inventory management, supply chain management, product management, and procurement management. I consider both tactical and strategic decision-making, as well as single-period and multi-period/multi-stage decision-making. Two messages emerge. First, humans are, in general, unable to make optimal decisions; here, the "optimality" is defined by models that assume fully rational, profit-maximizing decision-making. Humans deviate from such a benchmark because they are affected by decision biases or influenced by social preferences. Second, even in complex operational contexts, it turns out that the deviation from the optimal benchmark is *not* random. Through careful experimental design and data analysis, one can capture such systematic deviation. The resulting knowledge is useful as it lays down the foundation for future work to better capture real-world human decision-making. Meanwhile, there is also a cautionary note that there could be more than one way subjects deviate from the optimal benchmark. What seems to be messy in the data may actually be a combination of several different ways of deviation. This is seen most clearly in the third chapter of the dissertation, where I find that humans deviate from the optimal policy in the Increasing Cost condition of the dynamic resource allocation problem. Through our behavioral mechanism analysis, I am able to identify several distinct mechanisms that can capture different sub-groups of subjects in the Increasing Cost condition. Hence, allowing for a flexible approach is important for future researchers when studying behavioral mechanisms in complex operational settings.

Moving forward, as a behavioral researcher, I have often heard criticisms of the relevance of conducting behavioral research in a world where humans are increasingly being replaced by machines and algorithms in the decision-making processes. From my personal interaction with some of the major US manufacturers and retailing companies, my take on this matter is "yes and no". Yes, algorithms are taking up an increasingly larger portion of the decision-making processes. However, the human element must always exist in the decision-making process because humans will possess the knowledge and insights that are unknown to the algorithms, and humans can foresee the potential mistakes through reasoning instead of having to wait for the mistakes to happen and then “learn” from the data. To this end, I believe that the future of behavioral research in operations management lies in the understanding of how humans and algorithms can better work together for the purpose of improving efficiency and achieving social goods.

APPENDIX A

Appendix for "Team Decision-Making in Operations Management"

A.1 Standalone Newsvendor Task Coding Scheme

Appendices A.1 and A.2 cover the coding schemes we provide to the coders to code the team chats.

1. Decision Formulation

Newsvendor Economic Reasoning

The team/one team member reasons using the Newsvendor decision-making logic. Specifically, the team discusses the cost for *both* over-producing and under-producing, and use the reasoning to support making production decisions.

Example: "We lose 80 if we over-produce but we only lose 20 if we under-produce, so let's play safe."

Durable Strategy

The team formulates a concrete strategy to act in making production decisions. Tentative proposals should NOT be included. Directional statements (go higher/lower) should be included *only when* it has been formulated as a decision rule that should be carried out for a certain period of time.

Example: "I think we should *always* go below the mean X."

Aggressive: Being Aggressive in Making the Production Decision

The team expresses the desire be aggressive by making a large production decision. Directional statements (go high) can be included.

Example: "Let's make a large production decision in this round."

Conservative: Being Conservative in Making the Production Decision

The team expresses the desire to be conservative by making a small/low production decision. Directional statements (go low) can be included.

Example: "I want to go below the mean in this round."

Risk Seeking: Being Risk-Seeking in Making the Production Decision

The team explicitly mentions that they want to take more *risk*.

Example: "High risk high reward."

Risk Averse: Being Risk-Averse in Making the Production Decision

The team explicitly mentions that they want to be *risk-averse* or be safe.

Example: " If we overshoot we will lose a lot of money and I don't want to take too much risk, so let's play it safe."

Loss Aversion

The team expresses the aversion for potential loss in profits.

Example: "If we do so it's guaranteed that our profit won't go negative."

Waste Aversion

The team expresses the aversion for potential waste due to *over-production*.

Example: "I don't want to have leftovers in production, so let's go low."

Stock-out Aversion

The team expresses the desire to produce more in order to avoid *stock-out*.

Example:"I think we should go high. If we don't produce enough we may not be able to fulfill all the demand."

Mean Anchoring

The team expresses using the mean (X) as the decision benchmark and making decisions relative to the mean. In particular, the adjustment is independent of the value of X.

Example: "I think we can just pick a number around the forecast information in each round, say from -10 to +10."

Mean-Dependent Strategy

The team's adjustment relative to the mean is affected by the value of mean.

Example: "Let's produce X+10 when demand mean X is high, and produce X-10 when X is low."

2. Demand Reversal

Demand Reversal High: Demand Reversal Decision-Making When Demand was High

Subjects want to make a low production decision because the demand was high in the previous round.

Demand Reversal Low: Demand Reversal Decision-Making When Demand was Low

Subjects want to make a high production decision because the demand was low in the previous round.

3. Demand Chasing

Demand Chasing High: Demand Chasing When Demand was High

Subjects want to make a high production decision because the demand was high in the previous round.

Demand Chasing Low: Demand Chasing When Demand was Low

Subjects want to make a low production decision because the demand was low in the previous round.

A.2 Information Sharing in the Newsvendor Context Coding Scheme

I have divided the coding scheme for the supply chain information sharing game into the following three sub-categories:

1. Strategic Behavior

Own Strategy: Talking about Own Strategy

The team discusses the general objectives of their actions/strategies. The directional statements should be included (go higher/lower). However, non-tactical number proposing statements should not be coded.

Example: "As the supplier, I think we can do -50 this round to be conservative."

Opponent: Thinking from the Opponent's Perspective

The team discusses how the opponent will set the strategy, or how the opponent will respond to the team's action/strategy.

Example 1: "I think the retailer will always go high by a large amount."

Example 2: " They will produce more when we pick a higher value, so let's go high. "

2. Trust Behavior

Trustworthy: Expressing Willingness to be Trustworthy as the Retailer

Expressing willingness to be truth-telling.

Example: "At least we are not fooling them."

Untrustworthy: Expressing Willingness to be Untrustworthy as the Retailer

Expressing willingness to inflate, potentially by a large amount.

Example: "We want them to produce as many as possible, so its best for us to give a higher prediction."

Trusting: Expressing Willingness to be Trusting as the Supplier

Expressing confidence for their opponent or the message they receive.

Example: "I think we can trust the retailer."

Untrusting: Expressing Willingness to be Untrusting as the Supplier

Expressing skepticism for the message they receive.

Example: "I'm not trusting them, definitely lying."

3. Team Dynamics

Regret: Expressing Regret

The team expresses regret for what they had done in the last round/past few rounds. In particular, the team discusses **counterfactual** situations.

Example: "We should have gone even higher as the retailer last round."

Feedback: Using Feedback from Last Round

Referring to realized outcomes in previous rounds **to facilitate decision making for the current round**.

Example: "So again, same strategy, go high when we report since that worked last time"

Future Strategy (Same Role): Forward Looking and Formulating Strategies in Advance for Being in the Same Role

Discussing what the team should do if they play **in the same role** in future rounds.

Example: "Let's go even lower if we are **still** the supplier next round."

Future Strategy (Opposite Role): Formulating Strategies in Advance for Being in the Different Role

The team discusses strategies for being in the **different** role in future rounds.

Example: "We inflate a lot as the retailer this round. So in the future if we are the supplier, we should be super conservative."

Autocorrelation: Expressing Doubt about the Randomness in Demand Realization

The team expresses doubt about the random nature of demand realization.

Example: "Demand in the last few rounds suggests that there's a pattern."

A.3 Coder Consistency Rate Analysis

In the standalone Newsvendor task, I recruit three coders to apply the Newsvendor coding scheme to the Newsvendor settings: main experiment, extended Newsvendor experiment LCR and HCR condition. To measure their coding consistency, in particular for the round level coding scheme, I consider two metrics: simple correlation, and Cohen's joint Kappa. (The correlation is calculated as the average of the three pair-wise correlation for the three coders.) The results are summarized in Table A.1. I observe great consistency among all coders, with correlation coefficients above 0.50 and Cohen's joint Kappa above 0.40 for all codes. Therefore, I have good confidence in our chat analysis based on these coding results. For the two coders in the information sharing in the Newsvendor context, I also measure their consistency rate by considering the above two metrics. The results are summarized in Table A.2 and A.3. Here I also observe good coder consistency using all three metrics. In particular, I have high consistency rates (Kappa>0.40) for the trust/trustworthiness codes I consider as the main decision drivers.

Table A.1: Coder Consistency Analysis in the Standalone Newsvendor Task

Code	Correlation	Cohen's Kappa	Pass Kappa? (> 0.40)
Newsvendor Economic Reasoning	0.55	0.48	Yes
Mean Anchoring	0.54	0.51	Yes
Aggressive	0.66	0.63	Yes
Conservative	0.67	0.65	Yes
Risk Seeking	0.68	0.67	Yes
Risk Averse	0.72	0.71	Yes
Durable Strategy	0.50	0.48	Yes
Loss Aversion	0.55	0.55	Yes
Stock-out Aversion	/	/	/
Waste Aversion	/	/	/
Forecast-dependent Strategy	0.58	0.53	Yes
Demand Chasing High	/	/	/
Demand Chasing Low	/	/	/
Demand Reverse High	/	/	/
Demand Reverse Low	/	/	/

Notes: I do not show the statistics for six codes because their coding frequencies are mostly 0 in all Newsvendor settings. A simple correlation above 0.50 is considered as excellent consistency according to Cooper & Kagel (2005). For the Cohen's Kappa test, a Kappa larger than 0.20 is considered as acceptable; a Kappa between 0.40 and 0.60 is considered as a moderate level of agreement, and a Kappa large than 0.60 is considered as a substantial level of agreement [Landis & Koch (1977)].

Table A.2: Coder Consistency Analysis in the Information Sharing in the Newsvendor Context - Retailer Side

Code	Correlation	Cohen's Kappa	Pass Kappa? (> 0.20)
Untrustworthy	0.60	0.42	Yes
Trustworthy	0.78	0.37	Yes
Question	0.64	0.52	Yes
Own Strategy	0.56	0.20	Yes
Opponent	0.84	0.58	Yes
Feedback	0.52	0.40	Yes
Regret	0.50	0.66	Yes
Future Strategy (Same Role)	0.50	0.20	Yes
Future Strategy (Opposite Role)	0.96	0.91	Yes
Autocorrelation	/	/	/

Table A.3: Coder Consistency Analysis in the Information Sharing in the Newsvendor Context - Supplier Side

Code	Correlation	Cohen's Kappa	Pass Kappa? (> 0.20)
Untrusting	0.62	0.48	Yes
Trusting	0.76	0.58	Yes
Question	0.71	0.54	Yes
Own Strategy	0.48	0.19	No
Opponent	0.76	0.58	Yes
Feedback	0.42	0.30	Yes
Regret	0.68	0.44	Yes
Future Strategy (Same Role)	/	/	/
Future Strategy (Opposite Role)	0.55	0.66	Yes
Autocorrelation	/	/	/

A.4 Human Retailer Setting Regression Result Details

This section reports all the pair-wise comparison results in the information sharing game (human retailer setting). Appendix A.4.1 covers the results in decision outcome analysis. Appendix A.4.2 covers detailed regression configurations and corresponding results in expected profit analysis.

A.4.1 Decision Outcome Analysis

The results are presented in Table A.4.

Table A.4: Configuration Trust and Trustworthiness Pair-wise Comparison Analysis

Retailer Decision Pair-wise Comparison	Average Inflation	Coefficient	Estimation
A1. $R_T S_T$-$R_I S_I$	22.14 vs. 6.37	$\beta_{team} + \beta_{VT} + \beta_{TvT}$	13.55(8.54)
A2. $R_T S_T$-$R_T S_I$	22.14 vs. 35.95	$\beta_{VT} + \beta_{TvT}$	-12.37(7.97)
A3. $R_T S_T$-$R_I S_T$	22.14 vs. 20.00	$\beta_{team} + \beta_{TvT}$	4.46(7.97)
A4. $R_T S_I$-$R_I S_I$	35.95 vs. 6.37	β_{team}	25.93(8.19)***
A5. $R_I S_T$-$R_I S_I$	20.00 vs. 6.37	β_{Vt}	9.09(8.19)
Supplier Decision Pair-wise Comparison	**Average Reduction**	**Coefficient**	**Estimation**
B1. $R_T S_T$-$R_I S_I$	29.38 vs. 22.50	$\beta_{team} + \beta_{VT} + \beta_{TvT}$	-6.80(7.02)
B2. $R_T S_T$-$R_T S_I$	29.38 vs. 39.60	$\beta_{team} + \beta_{TvT}$	10.49(6.49)
B3. $R_T S_T$-$R_I S_T$	29.38 vs. 44.39	$\beta_{VT} + \beta_{TvT}$	16.34(6.49)
B4. $R_T S_I$-$R_I S_I$	39.60 vs. 22.50	β_{Vt}	-17.30(6.68)***
B5. $R_I S_T$-$R_I S_I$	44.39 vs. 22.50	β_{team}	-23.15(6.75)***

Notes: The average decision level (second column) are for illustration only; the comparisons are all conducted using Regressions (4) and (7). On the supplier side, a negative estimation represents a *larger* reduction in the regression configurations. The method is random-effects GLS regression with unbalanced panel data - 780 observations over 6 rounds, from 66 teams and 64 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.4.2 Expected Profit Analysis

On the retailer side, I consider Regression (A.1). With appropriate linear combinations, Regression (A.1) allows us to conduct all the pair-wise comparisons among the four supply chain configurations, except for the comparison between the two mixed configurations; for this comparison, I can simply compare teams and individuals within the mixed treatment. The analyses for suppliers and for the supply chain efficiency are conducted in the same way, except that the corresponding metric is substituted in as the dependent variable.

$$\begin{aligned} RetRatio_{HR,it} = \text{Intercept} &+ \beta_{team} \cdot Team_i + \beta_{VT} \cdot VersusTeam_i \\ &+ \beta_{TvT} \cdot Team_i \cdot VersusTeam_i + t \cdot \beta_t + v_i + \epsilon_{it}. \end{aligned} \tag{A.1}$$

$$SuppRatio_{HR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + \beta_{VT} \cdot VersusTeam_i + \beta_{TvT} \cdot Team_i \cdot VersusTeam_i + t \cdot \beta_t + v_i + \epsilon_{it}. \tag{A.2}$$

$$Efficiency_{HR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + \beta_{VT} \cdot VersusTeam_i + \beta_{TvT} \cdot Team_i \cdot VersusTeam_i + t \cdot \beta_t + v_i + \epsilon_{it}. \tag{A.3}$$

Table A.5 summarizes the pair-wise comparison results for the expected profit ratios for the three

Table A.5: Configuration Decision Expected Profit Ratio Pair-wise Comparison Analysis

Retailer Ratio Pair-wise Comparison	Average Ratio	Coefficient	Estimation
C1. R_TS_T-R_IS_I	0.90 vs. 0.88	$\beta_{team} + \beta_{VT} + \beta_{TvT}$	0.02(0.02)
C2. R_TS_T-R_TS_I	0.90 vs. 0.90	$\beta_{VT} + \beta_{TvT}$	0.00(0.02)
C3. R_TS_T-R_IS_T	0.90 vs. 0.86	$\beta_{team} + \beta_{TvT}$	0.04(0.02)**
C4. R_TS_I-R_IS_I	0.90 vs. 0.88	β_{team}	0.02(0.02)
C5. R_IS_T-R_IS_I	0.86 vs. 0.88	β_{Vt}	-0.02(0.02)
Supplier Ratio Pair-wise Comparison	**Average Ratio**	**Coefficient**	**Estimation**
D1. R_TS_T-R_IS_I	0.75 vs. 0.76	$\beta_{team} + \beta_{VT} + \beta_{TvT}$	0.01(0.09)
D2. R_TS_T-R_TS_I	0.75 vs. 0.67	$\beta_{team} + \beta_{TvT}$	0.08(0.08)
D3. R_TS_T-R_IS_T	0.75 vs. 0.75	$\beta_{VT} + \beta_{TvT}$	0.00(0.08)
D4. R_TS_I-R_IS_I	0.67 vs. 0.76	β_{Vt}	-0.07(0.09)
D5. R_IS_T-R_IS_I	0.75 vs. 0.76	β_{team}	0.01(0.08)
SC Efficiency Ratio Pair-wise Comparison	**Average Ratio**	**Coefficient**	**Estimation**
E1. R_TS_T-R_IS_I	0.95 vs. 0.93	$\beta_{team} + \beta_{VT} + \beta_{TvT}$	0.02(0.02)
E2. R_TS_T-R_TS_I	0.95 vs. 0.92	$\beta_{VT} + \beta_{TvT}$	0.02(0.02)
E3. R_TS_T-R_IS_T	0.95 vs. 0.91	$\beta_{team} + \beta_{TvT}$	0.04(0.02)**
E4. R_TS_I-R_IS_I	0.92 vs. 0.93	β_{team}	-0.00(0.02)
E5. R_IS_T-R_IS_I	0.91 vs. 0.93	β_{Vt}	-0.02(0.02)

Notes: The average decision level (second column) are for illustration only; the comparisons are all conducted using Regressions (A.1), (A.2), and (A.3). The method is random-effects GLS regression with unbalanced panel data - 390 observations over 6 rounds, from 66 teams and 64 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

parties (retailer, supplier, integrated supply chain). The results cover all the pair-wise comparisons among the four supply chain configurations, except for the comparison between the two mixed configurations. To address this, I consider the following regression functions for these three parties. The regression results are summarized in Table A.6.

$$RetRatio_{HR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + t \cdot \beta_t + v_i + \epsilon_{it}. \tag{A.4}$$

$$SuppRatio_{HR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + t \cdot \beta_t + v_i + \epsilon_{it}. \tag{A.5}$$

$$Efficiency_{HR,it} = \text{Intercept} + \beta_{team} \cdot Team_i + t \cdot \beta_t + v_i + \epsilon_{it}. \tag{A.6}$$

Table A.6: Decision Effectiveness Pair-wise Comparison Analysis - Between the Two Mixed Configurations

Retailer Ratio Pair-wise Comparison	Average Ratio	Coefficient	Estimation
C6. $R_T S_I$-$R_I S_T$	0.90 vs. 0.86	β_{team}	0.04(0.02)**
Supplier Ratio Pair-wise Comparison	**Average Ratio**	**Coefficient**	**Estimation**
D6. $R_I S_T$-$R_T S_I$	0.75 vs. 0.67	β_{team}	0.06(0.07)
SC Efficiency Ratio Pair-wise Comparison	**Average Ratio**	**Coefficient**	**Estimation**
E6. $R_T S_I$-$R_I S_T$	0.92 vs. 0.91	β_{team}	0.02(0.02)

Notes: The average decision level (second column) are for illustration only; the comparisons are all conducted using Regressions (A.4)-(A.6). The method is random-effects GLS regression with unbalanced panel data - 204 observations over 6 rounds, from 34 teams and 34 individuals. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

A.5 Extended Newsvendor Experiment - Chat Analysis

This section provides further results for subject performance (especially for teams) in the extended Newsvendor experiment.

Similar to the text chat analysis conducted in the main experiment, here I focus on the three features for the codes to determine whether they become compelling arguments in team discussions. The coding frequency summary of both parameter conditions and the corresponding intra-team acceptance rates are presented in Table A.7. In addition, to address the potential concern of large variation due to having too small a coding frequency in one setting, here I also pool the data between the two Newsvendor settings in the extended experiment (LCR and HCR) and consider the pooled agreement rate for each code.

For the *Newsvendor Economic Reasoning*, I find that, compared to the main experiment, subjects indeed find it to be more persuasive when more opportunities are presented. However, the total coding frequency of it remains low, making it unclear how much an impact it will have on teams' final decision outcomes. Meanwhile, I continue to find that general mental dispositions (*Aggressive* in HCR, *Conservative* in LCR) continue to be both frequently mentioned and found persuasive in team discussions.

Regression analysis is then conducted by applying the same regression model in the main experiment. The results are summarized in Table A.8 and Table A.9. I confirm that: (1) *Newsvendor Economic Reasoning* has no significant impact on decision outcomes, in either HCR or LCR; (2) *Aggressive* in HCR and *Conservative* in LCR strongly drive team decision outcomes. Therefore, I

Table A.7: Extended Newsvendor Experiment Coding Frequency

Code	LCR Frequency	Intra-Team Acceptance Rate	HCR Frequency	Intra-Team Acceptance Rate	Pooled Intra-Team Acceptance Rate
Newsvendor Economic Reasoning	4.67	100%	1.33	71%	90.0%
Mean Anchoring	14.67	61%	5	81%	66.1%
Aggressive	9.67	64 %	30.33	87%	81.7%
Conservative	19.67	64%	13.33	85%	73.0%
Risk Seeking	4	25%	7.33	68%	52.9%
Risk Averse	7.33	36%	5.33	63%	47.3%
Durable Strategy	8.67	38 %	8	91%	63.3%
Loss Aversion	8	52%	1	/	53.6%
Stock-out Aversion	0.33	/	1	/	/
Waste Aversion	0.33	/	0.33	/	/
Forecast-dependent Strategy	2	100%	2.33	43%	71.4%
Demand Chasing High	0	/	0.33	/	/
Demand Chasing Low	0	/	0	/	/
Demand Reverse High	0.67	/	2.33	71%	66.7%
Demand Reverse Low	2.33	83%	2	100%	91.7%

Notes: The agreement rate for the codes with a frequency less than 2 is not considered. The pooled intra-team
Notes: acceptance rate also incorporates the data from the main experiement

conclude that *Newsvendor Economic Reasoning* is *not* a compelling argument in team discussions, even when enough decision opportunities are present. On the other hand, general mental dispositions are still found compelling in team discussions. Therefore, our conclusion from chat analysis in the main experiment also holds in the extended Newsvendor experiment. In a longer version of this chapter (available upon request), I further explore these findings and identify the Newsvendor decision theories that are consistent with our results here.

Table A.8: Extended Newsvendor Experiment LCR Chat Regression Analysis

Code	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Full
Nvdr. Econ. Reas.	-10.05(7.12)							-8.85(8.47)
Durable Strategy		-9.69(3.56)***						-9.06(5.87)
Mean Anchoring			-2.05(5.96)					2.88(6.83)
Aggressive				16.44(7.65)**				17.76(8.31)**
Conservative					-4.86 (2.90)*			-2.37(3.43)
Risk Seeking						7.17(7.22)		9.72(9.53)
Risk Averse							-7.97(4.00)**	-8.55(4.20)**

Notes: Random-effects GLS regression with balanced panel data, clustering on the decision unit level. 319 observations over 20 rounds, from 16 teams. Columns 2-8 report the estimates for the specification with a single code included. Column 9 (Full) reports the results with all codes included. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Finally, I also note that I have two additional designs of the extended Newsvendor experiment. First, at 4 randomly selected rounds of stage 2, I elicit each team members' individual inclination in making Newsvendor decisions prior to making their team decisions. This allows us to directly study how preferences are integrated inside teams *after* teams are formed. The results are discussed

Table A.9: Extended Newsvendor Experiment HCR Chat Regression Analysis

Code	(1)	(2)	(3)	(4)	(5)	(6)	(7)	Full
Nvdr. Econ. Reas.	-16.99(32.42)							-7.90(20.07)
Durable Strategy		6.22(1.93)***						-6.16(3.19)*
Mean Anchoring			-7.15(8.82)					-17.81(11.18)
Aggressive				10.80(2.17)***				11.84(1.82)***
Conservative					-17.70 (3.55)***			-13.20(3.01)***
Risk Seeking						11.88(4.37)***		2.22(5.20)
Risk Averse							19.57(9.03)**	24.98(8.49)***

Notes: Random-effects GLS regression with balanced panel data, clustering on the decision unit level. 319 observations over 20 rounds, from 16 teams. Columns 2-8 report the estimates for the specification with a single code included. Column 9 (Full) reports the results with all codes included. Robust standard errors are in parentheses. Significance is denoted: $^{*}p < 0.1$, $^{**}p < 0.05$,$^{***}p < 0.01$.

formally in the longer version of this chapter available upon request. The key observation is that the more capable team member influences, but does not fully determine, the team's final Newsvendor decisions. Second, at the end of the whole extended Newsvendor experiment, I add a separate stage to elicit each individual's risk preference and the corresponding team's risk preference with the standard Holt and Laury risk measurement table [Holt & Laury (2002)]. The decision mechanism is structurally identical to what I use in Newsvendor preference elicitation. I find that including the risk measures from individuals and/or teams does *not* help us explain team decision-making outcomes. Therefore, to keep our narration concise I do not include the corresponding analysis in this chapter.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

APPENDIX B

Appendix for "Human Decision-Making in Dynamic Resource Allocation"

B.1 Mathematical Formulation of the Model

In this appendix, I present the mathematical formulation of the our core model, discussed in Section 3.3. I also cover the general structure of the optimal policy, including in the two conditions I consider in this chapter: the Constant Cost condition and the Increasing Cost condition.

B.1.1 Model Formulation

To formalize this, the manager has an initial budget of B to be allocated to T opportunities to arrive. Each arrival of an opportunity is a decision period denoted by t, $t = 1, 2, ..., T$; the budget available at the beginning of period t is denoted by $b_t \leq B$. The state of the system is, therefore, (t, b_t). In every decision period, the manager makes a binary decision: whether or not to implement the opportunity. Each opportunity brings in a benefit of u_t and, if implemented, will lead to a cost of c_t subtracted from the budget ($b_{t+1} = b_t - c_t$). Ex-ante, u_t is a bounded random variable with a density/probability mass of $f(u_t)$. c_t is assumed to be non-random to simplify the decision problem; this also makes our model consistent with the application in revenue management (see Bearden et al. (2008)).[1] I further assume the smallest value of u_t would still be larger than c_t to reflect the fact that I am considering optional, beneficial opportunities. Also, the total budget is scarce: $B < \sum_{t=1}^{T} c_t$, such that not all the opportunities can be implemented. If the opportunity is not implemented, it is gone and will not be revisited again in the future. The project is finished

[1]In engineering practices, each opportunity is presented with a specific benefit and cost (u_t and c_t here), and the decision making from the manager is based on evaluating these two numbers as well as considering future design change opportunities. The value of u_t can be considered as the "best guess" of the benefit at the time it is being evaluated. From our interaction with our industrial partner, having specific numbers of u_t and c_t simplifies the discussion when a implementation decision is needed upon the arrival of each opportunity.

either when it reaches the end of $t = T$ or when there is insufficient remaining budget for one opportunity to be implemented. Unspent funds at the end will be added to the total benefit collected. The manager's job is to maximize the total amount of benefit collected.

I also note that our core model can be directly applied to formulate problems in many other managerial contexts. In the Literature Review section, I have illustrated the application of the model to revenue management. Below, I offer two extra application examples. Firstly, with a simple re-framing, our model naturally applies to the project management context, where the manager has a fixed budget B and allocates the budget to different requests/opportunities as the project proceeds. Second, as I noted above, in every decision period, the (sophisticated) decision maker calculates $V_{t+1}(b_t - c_t)$ and $V_{t+1}(b_t)$. Here, $V_{t+1}(\cdot)$ is derived from calculating the optimal expected value of continuing the project (with a given amount of budget). In the finance literature, this is equivalent to the calculation of the "real options" value. The finance literature has a long history of studying the real options value from a theoretical viewpoint, see Trigeorgis (1996) for a review. However, the study of how humans actually derive the real options value has been fairly limited.[2] In this chapter, by studying human subjects' behavior in our model, I am (indirectly) assessing their ability in deriving the real options value.

B.1.2 General Structure of the Optimal Policy

To optimally solve this problem, let $V_t(b_t)$ denote the value function, which is the expected continuation value when the optimal decision path is executed from period t to T and with a remaining budget b_t on hand at the beginning of period t. I have the following recursive expression:

$$V_t(b_t) = E[\max\{u_t + V_{t+1}(b_t - c_t),\ V_{t+1}(b_t)\}],\ \forall b_t \geq c_t, \tag{B.1}$$

with the boundary conditions of $V_t(0) = 0, \forall t,\ V_t(b) = V_{t+1}(b), \forall b < c_t$, and $V_{T+1}(b) = b, \forall b \geq 0$.

In every decision period, given the realization of the benefit u_t, the manager simply needs to compare $V_{t+1}(b_t - c_t)$ and $V_{t+1}(b_t)$. The manager should accept any opportunities with benefit $u_t \geq V_{t+1}(b_t) - V_{t+1}(b_t - c_t)$, i.e., the optimal policy is a threshold rule as a function of t and b_t. I call the value of $V_{t+1}(b_t) - V_{t+1}(b_t - c_t)$ the "optimal threshold" for the given state of the system

[2]Two papers have assessed human subjects' ability in calculating the real options value in financial investment contexts. Denison (2009) consider the effect of using real options calculation to prompt subjects to consider different possible scenarios in project management, particularly evaluating the value of early project termination. Oprea et al. (2009) consider a model where the real options value is calculated based on a project whose value follows a Brownian motion. They find that subjects can learn to estimate the real options value precisely when the value is low, but they systematically underestimate the value when the value is high. Both papers suggest that human subjects are capable of understanding the general logic in calculating the real options value. In our context, this suggests that (some) subjects may be able to perform similar calculations in deciding whether or not to implement an opportunity, but the extent to which they can act optimally requires studies.

(t, b_t). To explicitly calculate this optimal threshold, I only need to determine the values of $V_t(b_t)$, which can be derived recursively from the last period T.

B.1.3 Optimal Policy in the Two Conditions

As I note in Section 3.4, I consider two conditions of the above core model with different degrees of complexities: the Constant Cost condition and the Increasing Cost condition. In the Constant Cost condition, c_t is assumed to be fixed and known throughout the project. In the Increasing Cost condition, c_t is also known to the decision maker, and it is knowingly going to double when it reaches the second half of the project. As it turns out, this simple change in the cost structure leads to dramatically different optimal policies in the two conditions.

In the Constant Cost condition, the problem is much simpler since c_t does not change over time. The optimal threshold $V_{t+1}(b_t) - V_{t+1}(b_t - c_t)$ therefore has a simple and intuitive structure [Papastavrou et al. (1996)]: (1) fixing budget b_t, it is a nonincreasing function of t; (2) fixing t, it is a nonincreasing function of b_t. Combined with the fact that the total budget B is scarce, subjects should then be selective and set medium or high thresholds for the majority of the project.[3] On the other hand, in the Increasing Cost condition, the fact that c_t doubles later in the project suggests a dramatic change in the optimal thresholds as a function of time t, and the simple properties in the Constant Cost condition no longer hold. Now, in the first half of the project there is a tradeoff between capturing benefit and spending budget when the cost is low. Under the experiment parameters introduced in Section 3.4, this results in almost always choosing the low or medium benefit as the threshold for the first half of the project[4] ($t \leq 5$) and the high benefit as the threshold for the majority of the second half of the project ($t \geq 6$). Such a structure has an intuitive interpretation: Subjects should spend more budget in the first half when the cost is low; in the second half, use the remaining budget to capture the most valuable opportunities only.

B.2 Reported Decision Rules from Top and Bottom Performers in the Increasing Cost Condition

In this appendix, I provide examples of the top and bottom performers' reported decision rules and how they are connected to the "Decision Decomposition" heuristic and the Selective Heuristic.

[3]To see this, note that the smallest optimal threshold in the Constant Cost condition is achieved when it is in the last decision period ($t = T$) and subjects still have the full budget B. Then, the properties suggest that subjects should have a higher threshold when it is earlier in the project and for any budget smaller or equal to the full budget B.

[4]In the first half of the project, subjects should never choose the highest benefit as the threshold for the first 4 opportunities (regardless of their remaining budget); for the 5th opportunity, they should choose the high benefit as the threshold only when they are able to implement all of the first 4 opportunities.

I first attach two self-reported decision rules from bottom performers (bottom 1/3) in the Increasing Cost condition. These rules clearly point to the Selective Heuristic.

> **Bottom Performer** #1**:** *For the product development projects, I chose 7000 for most of the rounds. Earning 7000 rather than 3000 or 5000 in a round would result in more earnings, and there was no difference in the amount deducted in the earnings.*

> **Bottom Performer** #2**:** *Decision Rules I used were only accepting 7000 until the very end if I had payoff left. Tried to have 4000 left for the second half of the experiment. If there was 2000 left at the end, lowered minimum value to 3000.*

Next, I lay out the self-reported decision rules from the top 2 (No. 1 and No. 2 performing subjects) in the Increasing Cost condition. They both point to a specific way to decompose the problem through conducting budget planning based on whether or not the cost has increased: They allocate 4000 ECU to the first half (before the cost increase) and 2000 ECU to the second half (after the cost increase. They also try to organize the decision in every period to support this budget allocation plan.

> **Top performer** #**1:** *"Since the cost of investing in part 2 was double that in part 1, I wanted to invest in part 1 as much as possible. I aimed to invest in exactly 4/5 of the opportunities for part 1 to make sure to leave* \$*2000 to be able to invest in part 2. In stage 2, I would only invest if the payoff was high 7000. If a 7000 payoff did not occur by the last stage, I would invest my remaining budget into the minimum threshold."*

> **Top performer** #**2:** *"My primary goal was to ensure I got 5 product improvements for each project. I also wanted to avoid a lower payout if possible. However, since the cost of the improvement was less than the gain I wanted to ensure, I got 5 the maximum given the 6000 budget, and costs of improvements possible each time. To achieve this I first entered 5000 for the first round and all subsequent until I encountered a 3000 once this happened I switched to 3000 because I wanted to ensure I received 4 of the 5 improvements when they cost only 1000. Then in the second half I started by putting 7000 for the first three rounds, hoping to get it and then relaxed it, to ensure I received some benefit. In round 9 I put 5000 and then in round 10 I put 3000."*

I note that both top performers mention a specific way to allocate budget (the underlined text): spend twice as much budget in the first half than in the second half (i.e., 4000 in the first half and 2000 in the second half). In addition, I observe that both top performers mention how to set specific thresholds in every decision period, and do so in a way to support this budget allocation plan.

When I examine other top performers, I continue to find that an overwhelming majority of them either explicitly refer to this budget allocation plan or list it as one of two offered action plans. Meanwhile, none of the bottom performers form a explicit budget allocation plan into the two halves of the project.

I note that this is one specific way to implement the decision decomposition heuristic as it only includes *one way* to allocate the budget. In general, there can be at least 5 different ways to allocate the budget to the two halves of the project, and I include all of them when I conduct the decision mechanisms analysis in the Increasing Cost condition.

B.3 Instruction for Experiments of the Chapter "Human Decision-Making in Dynamic Resource Allocation"

In B.3.1, I provide the instructions for the increasing cost condition. The instructions for the constant cost condition are mostly identical, except for necessary adjustments made to reflect the fact that the cost is constant throughout; I therefore omit it here. In Sections B.3.2 and B.3.3, I provide the *additional instruction* I offer to subjects in the decision prompt treatment and the social learning treatment, respectively. Such additional instruction is complementary to the main instruction (the increasing cost condition instruction), and is presented to subjects at the beginning of each project to serve as the managerial intervention. Finally, in Section B.3.4, I offer the instructions for the diagnostic task of Hit-15 that is relatively new to the community of behavioral operations management. The instructions for the other two diagnostic tasks - risk preference measurement and cognitive reflection test - are standard and therefore omitted.

B.3.1 Instructions for the Increasing Cost Condition

Welcome. Thank you for joining the experiment.

Starting from this point, please don't talk to other participants or look at their screens. Please do not use other electronic devices during the whole experiment. During the experiment, you will be muted in the Zoom room. However, if you have any questions, please feel free to use your Zoom to text the host (experimenter). If you accidentally close the web tab for the experiment, simply click the link on your Zoom again to get reconnected. Please make sure that you have reviewed the informed consent form shared to you earlier. By checking the box on the screen, you will certify that you have read and agreed with the informed consent form.

Today's session is a study of product development management.

There are two stages of the main experiment. The first stage, which is the main part of the experiment, consists of managing 5 product development projects. The second stage consists of several decision tasks that are not directly related to the first stage; the instruction for the second stage will be given after I finish the first stage. There will be a short survey after the main experiment.

In the first stage, the 5 projects are independent of each other. For each product development project, you will be asked to make a series of decisions that will affect your payoff. The payoff is expressed in Experimental Currency Units (ECUs), with an exchange rate of 2000 ECU to 1 dollar. At the end of the experiment, I will randomly select one product development project and use your earned ECU in that project, plus payoff from stage 2 and 5 dollars show-up fee, to determine your final payoff. At the end of the experiment, you will also need to complete a survey that collects your payment-related information.

Stage 1: Making Sequential Decisions for One Product Development Project

The following instruction demonstrates how decisions will be made in managing one product development project. You are the program manager of a product development project. Your team is now developing a new product under a given financial budget. Your responsibility is to monitor the development process and allocate financial resources wisely.

The project will last for 10 periods. At the beginning of each period, you will be presented with a design opportunity to improve the design of your product. Each design opportunity, if implemented, will bring you certain benefits. In particular, the benefit takes three possible values: 7000 ECU, 5000 ECU, or 3000 ECU. You will not be able to see the benefit of each design opportunity until you have reached the beginning of the period. For example, if you are at the beginning of period 1, you will observe the benefit of the design opportunity for period 1, but you do not observe the benefit of the design opportunity for period 2, period 3, . . . , period 10. However, you do know that the three possible values happen with equal probability for each design opportunity.

Meanwhile, implementing design opportunities will incur financial costs, and you must decide how to spend your money wisely. For each project, you have a total budget of 6000 ECU. In period 1 to 5 (phase 1), each design opportunity costs you 1000 ECU. In period 6 to 10 (phase 2), it becomes more expensive to implement each design opportunity: Each design opportunity costs you 2000 ECU. Note that the increase in cost means that you need to have at least 2000 ECU in your budget to be able to handle any design change in period 6-10. In addition, suppose you enter period 6 with a remaining budget of 3000 ECU, then you can only implement one design opportunity in period 6-10.

There is no correlation between the benefits of these 10 design opportunities. That is, you cannot make inference for future benefits based on historical benefits. Also, you cannot reinvest the benefit you have collected for future design opportunities. In other words, the 6000 ECU is the only financial resource you can use to implement design opportunities. Your payoff for each product development project is determined by the total amount of benefits you collect, plus any remaining budget.

In summary, the sequence of events is as follows:

1. You enter a new period. You are presented with the information regarding your remaining budget and the history of your decision-making.
2. If you still have a remaining budget greater than or equal to the cost to implement a design change (1000 ECU in period 1-5, 2000 ECU in period 6-10), you will be asked to think carefully about your investment strategy for this period. Specifically, you will choose the minimum value of design benefit you are willing to accept.

   For example, if you choose 5000 ECU, then this means you are willing to implement the design opportunity only if its benefit is 5000 ECU or 7000 ECU in this period, and you will not want to implement it if its benefit is 3000 ECU. You have 1 minute to make this decision.

3. The actual benefit for the design opportunity of this period is shown to you. There are two possible outcomes:
   a. If the benefit is within what you have chosen to implement in step 2, then you will implement the design opportunity and earn the corresponding benefit, and your budget will be deducted by 1000 ECU if in period 1-5, or 2000 ECU if in period 6-10.
   b. If not, then the design opportunity is not implemented; no benefit is collected, and your remaining budget is unchanged.
4. You proceed to the next period.

At the end of the product development project (10th period), if you still have a remaining budget, it will be added to your total benefit collected. You will also see a final summary screen for this product development project. You will then proceed to a new product development project. All projects share the same decision sequence outlined above.

In the history table, you will be able to see your decisions from previous projects as well, but please be noted that projects are independent of each other. In other words, the sequence of realized benefits you observe in one project does not predict the sequence of realized benefits for another product development project. Finally, please be sure to make decisions in every period; skipping decision periods will make you lose valuable opportunities to earn payoffs.

Questions: Please enter the answers on your screen.

1. You are in period 4 of product development project 3. Your remaining budget is 5000 ECU. Suppose you choose 3000 ECU as the minimum benefit you are willing to accept for the design opportunity of this period. What will happen if the actual benefit of period 4 is 3000 ECU?

Q1.1 How much benefit will you be able to collect? Q1.2 How much will be deducted from your budget?

2. You are in period 8 of product development project 5. Your remaining budget is 3000 ECU. Suppose you choose 5000 ECU as the minimum benefit you are willing to accept for the design opportunity of this period. What will happen if the actual benefit of period 8 is 5000 ECU?

Q2.1 How much benefit will you be able to collect? Q2.2 How much will be deducted from your budget?

3. You are in period 5 of product development project 5. Your remaining budget is 4000 ECU. Suppose you choose 7000 ECU as the minimum benefit you are willing to accept for the design opportunity of this period. What will happen if the actual benefit of period 5 is 3000 ECU?

Q3.1 How much benefit will you be able to collect? Q3.2 How much will be deducted from your budget?

4. Suppose you collected 7000 ECU for your 1st design change opportunity, 5000 ECU for your 3rd opportunity, 5000 ECU for your 5th opportunity, and 3000 ECU for your 9th opportunity, what is your final total payoff (in terms of ECU)? (Hint: The cost of each design opportunity in the first 5 periods (phase 1) is 1000 ECU; the cost of each design opportunity in the last 5 periods (phase 2) is 2000 ECU. Your total budget is 6000 ECU. Any remaining budget will be added to your final total payoff.)

Q4.1 What is your final total payoff?

B.3.2 Additional Instruction in the Decision Prompt Treatment

I now provide a planning prompt to assist you in making decisions for the coming project. In this planning prompt, I ask you to think carefully about how you want to allocate your total budget to the first half and the second half of the coming project. Being able to allocate budget wisely has been observed to be associated with good performance in this decision task.

As a quick reminder for the decision-making in every project. You have a total budget of 6000 in every project, and you will face 10 decisions. In the first half (decision 1-5), the cost to implement each design opportunity is 1000; in the second half (decision 6-10), the cost to implement each design opportunity is 2000.

Below, please enter the amount of budget you want to spend on the first half and the second half, respectively. They should sum up to the total budget of 6000.

Note that this is only a planning prompt. In other words, I will NOT enforce the budget allocation for you. You can also deviate from the plan you lay out here when you actually make those 10 decisions.

B.3.3 Additional Instruction in the Social Learning Treatment

I now provide a suggestion to assist you in making decisions for the coming project, based on what has worked well for subjects in previous sessions. Because the two halves of the project have different costs, think carefully about a target amount that you want to allocate from your total budget to the first half and the second half of the coming project. Of course, the particular sequence of design opportunities may mean you want to do something different from your initial plan. Specifically, many subjects who have succeeded in this decision task report to us that they follow a relatively simple decision rule: They try to allocate 4000 ECU to the first half (period 1-5), and leave 2000 ECU to the second half. This is equivalent to implementing 4 design opportunities in the first half, and only implement 1 design opportunity in the second half.

The rationale behind this allocation rule, according to those successful subjects, is quite simple: The cost to implement any design opportunity doubles in the second half. Therefore, it is beneficial to spend more in the first half than in the second half.

Note that this is only a suggested budget plan. I will NOT enforce this budget allocation on you. You are free to make any decisions when you actually make those 10 decisions.

B.3.4 Instructions for the Hit-15 Task

Consider the following two-person game: There is a basket in which people place points. The two players take turns placing 1, 2, or 3 points in the basket. The person who places the 15th point in the basket wins a prize. Say you are playing and want to win the prize. You will answer two questions regarding the actions you will take in this decision task.

Please answer the following two questions. You will receive 1000 ECU for each question you get correct.

- Question 1. If you go first, how many points will you place in the basket? Please pick one of the answers below (1, 2, or 3).
- Question 2. If you go second and the other player has already put 2 points in the basket on her first turn, how many would you put in? Please pick one of the answers below (1, 2, or 3).

APPENDIX C

Appendix for "Procurement with Change Order Renegotiation"

Appendix - Proofs of the Results

Before I proceed to prove the two key results in Section 4.5.2, I first establish two Lemmas. Lemma 1 that describes the comparison between two mechanisms *given* the realization of types. Lemma 2 addresses properties of the suppliers' dropout bids.

Lemma 1. *Suppose the budget is chosen in such a way that supplier i can perform the design change while supplier j could not, for the given type realization. Then, the budget mechanism performs no worse than the simple mechanism if and only if supplier i wins the auction in the budget mechanism.*

C.1 Proof of Lemma 1

C.1.1 Preliminaries

I first present the suppliers' dropout bids and the OEM's total transfer in the two mechanisms under the situation specified in Lemma 1.

C.1.1.1 Dropout Bids in the Budget Mechanism

In this situation, I have the following relationship between the dropout bids in the two mechanisms:

$$DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_b^{\overline{m}} (m-b)dF(m);$$
$$DB_j^b(\theta_j) = DB_j^S(\theta_j) + \int_{\Delta C_j\theta_j}^{\overline{m}} (m-\Delta C_j\theta_j)dF(m).$$

This has important implications when I compare the budget mechanism with the simple mechanism. In the analysis below, I will first lay out the situation in the budget mechanism and then consider what would have happened if the budget were not imposed.

First, if supplier i wins in the budget mechanism, then she must also win in the simple mechanism: $D_i^b(\theta_i) < D_j^b(\theta_j)$ directly implies $D_i^S(\theta_i) < D_j^S(\theta_j)$ because supplier i's drop out bid increases more when I move from the simple mechanism to the budget mechanism. On the other hand, if supplier j wins in the budget mechanism, it will be unclear who wins in the simple mechanism. In this case, more careful analysis needs to be conducted. Below I discuss these two cases separately.

C.1.1.2 OEM's Total Transfer in the Two Mechanisms

Here I formally present the OEM's total transfer in the budget mechanism and the simple mechanism under the condition of Lemma 1.

In the budget mechanism, note that supplier j cannot handle the design change means I have $\Delta C_j \theta_j > b$, i.e., her change cost is larger than the budget b. With this, I have $\int_b^{\overline{m}}(m-b)dF(m) > \int_{\Delta C_j\theta_j}^{\overline{m}}(m - \Delta C_j\theta_j)dF(m)$. That is, the increase in supplier i's dropout bid is larger than that of supplier j's dropout bid. For a given realization in this situation, the OEM's total transfer in the budget mechanism is:

$$T^b = max\{DB_i^b, DB_j^b\} - \mathbb{I}_{\{DB_i^b < DB_j^b\}} \cdot \int_b^{\overline{m}}(m-b)dF(m).$$

That is, the OEM will only be able to benefit from the expenditure cut from the design change when supplier i wins; this is because of the assumption that only supplier i can handle the design change. In the simple mechanism, since no budget is imposed, both suppliers can handle the design change. As I have discussed at the beginning of Section 4.5, the OEM's total transfer is just the ending price of the auction:

$$T^S = \max\{DB_i^S(\theta_i), DB_j^S(\theta_j)\}$$

C.1.2 Case 1: Supplier i Wins in the Budget Mechanism

In this case, I have $DB_i^b < DB_j^b$. The OEM's ex-post total transfer is:

$$\begin{aligned} T^b(\text{i win}) &= DB_j^b(\theta_j) - \int_b^{\overline{m}} (m-b)dF(m) \\ &= DB_j^S(\theta_j) + \int_{\Delta C_j \theta_j}^{\overline{m}} (m - \Delta C_j \theta_j)dF(m) - \int_b^{\overline{m}} (m-b)dF(m) \\ &< DB_j^S(\theta_j). \end{aligned}$$

Recall that if supplier i wins in the budget mechanism, she must also win in the simple mechanism had the budget not been imposed. Therefore, the comparison benchmark in the simple mechanism is just supplier j's dropout bid, which is exactly $DB_j^S(\theta_j)$. Therefore, I find that the OEM's total transfer is lower in the budget mechanism compared to the simple mechanism for this given type realization, i.e., the OEM benefits from the budget mechanism.

C.1.3 Case 2: Supplier j Wins in the Budget Mechanism

I now consider the case where supplier j wins in the budget mechanism, i.e., $DB_i^b > DB_j^b$. As has been discussed above, in this case, it is unclear who will win the auction in the simple mechanism, had the budget not been imposed. Therefore, they need to be discussed separately.

Scenario 1: Supplier j Wins in the Simple Mechanism As Well

In this scenario, supplier j wins both under the budget mechanism and the simple mechanism for the given type realization $DB_i^S > DB_j^S, DB_i^b > DB_j^b$. The OEM's total transfer in the budget mechanism is:

$$T^b(\text{j win}) = DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_b^{\overline{m}} (m-b)dF(m).$$

The OEM's total transfer in the simple mechanism is:

$$T^S(\text{j win}) = DB_i^S(\theta_i).$$

Therefore, the OEM's total transfer is higher in the budget mechanism, by an amount independent of the type realizations. In fact, this is exactly the amount that the OEM thought he would have saved from cutting the budget. The intuition is that when supplier j wins the budget mechanism, and if supplier j cannot handle the design change, then the OEM suffers both in the auction stage and the post-auction design change stage in the budget mechanism. In the auction stage, the dropout bid is determined by supplier i, whose dropout bid increases due to the budget b. In the post-auction design change stage, supplier j cannot handle the design change, so there is no way for the OEM to benefit from it.

Scenario 2: Supplier i Wins in the Simple Mechanism

This scenario describes the situation where supplier i wins in the simple mechanism but loses in the budget mechanism because her dropout bid increases too much after the budget is imposed: $DB_i^S < DB_j^S, DB_i^b > DB_j^b$. The OEM's total transfer in the budget mechanism is still:

$$T^b(\text{j win}) = DB_i^b(\theta_i).$$

Now, the OEM's total transfer in the simple mechanism is:

$$T^S(\text{i win}) = DB_j^S(\theta_j).$$

Notice that I have the following relationship from the fact that supplier j wins in the budget mechanism:

$$DB_i^b(\theta_i) > DB_j^b(\theta_j) = DB_j^S(\theta_j) + \int_{\Delta C_j \theta_j}^{\overline{m}} (m - \Delta C_j \theta_j) dF(m) > DB_j^S(\theta_j)$$

Therefore, I still have the outcome that the OEM's total transfer is higher in the budget mechanism. Combining both scenarios together, I conclude that, under the situation specified in Lemma 1, the OEM is always worse off in the budget mechanism so long as supplier j wins the auction. Finally, combining with the results in Case 1, I conclude that the budget mechanism is beneficial if and only if supplier i wins the auction. Q.E.D.

Lemma 2. *Supplier's dropout bid in the budget mechanism is an increasing function w.r.t. her type. This is true for any value of budget b*

C.2 Proof of Lemma 2

I use supplier i for illustration, since the case for supplier j is analogous. If the budget is large such that all types of supplier i can handle the design change in the budget mechanism, then her dropout bid is:

$$DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_b^{\overline{m}} (m - b) dF(m).$$

It is assumed that $DB_i^S(\theta_i)$ is increasing in the type, i.e., the dropout bid is an increasing function in the simple mechanism. Therefore, $DB_i^b(\theta_i)$ is also increasing in the type.

For an intermediary value of b such that only the supplier with a type lower than the threshold $b/\Delta C_j$ can handle the design change, I analyze the case separately.

If she can handle the design change (has a type below the threshold $b/\Delta C_i$), then the dropout bid is also:

$$DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_b^{\overline{m}} (m - b) dF(m),$$

which is increasing in the type θ_i.

If she cannot handle the design change (has a type above the threshold $b/\Delta C_i$), her dropout bid is simply $C_j \cdot \theta_j$, which is in increasing in θ_j. Finally, I note that supplier j's dropout bid is continuous in the type, particularly at the threshold $b/\Delta C_j$. To see this, note that the dropout for the type below the threshold, the bid can be expressed as:

$$DB_j^b(\theta_j) = DB_j^S(\theta_j) + \int_b^{\overline{m}} (m-b) dF(m) = C_j \cdot \theta_j - \int_{\Delta C_j \theta_j}^{\overline{m}} (m - \Delta C_j \theta_j) dF(m) + \int_b^{\overline{m}} (m-b) dF(m)$$

Therefore, by letting θ_i equal to the threshold type $b/\Delta C_i$, this term exactly equals to $C_i \cdot \theta_i$. Therefore, supplier i's dropout bid is also increasing in the type θ_i.

Finally, for value low b value such that no type of supplier i can handle the design change, then the dropout bid function is simply $C_i \cdot \theta_i$, which increases in θ_i. Q.E.D.

C.3 Proof of Proposition 4

Under the situation specified in Proposition 4, neither supplier can handle the design change under the budget mechanism. Therefore, their dropout bids in the budget mechanism are:

$$DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_{\Delta C_i \theta_i}^{\overline{m}} (m - \Delta C_i \theta_i) dF(m);$$
$$DB_j^b(\theta_j) = DB_j^S(\theta_j) + \int_{\Delta C_j \theta_j}^{\overline{m}} (m - \Delta C_j \theta_j) dF(m).$$

The OEM's total transfer in the budget mechanism, for a given type realization satisfying the condition specified in Lemma 2, is:

$$T^b = \max\{DB_i^b(\theta_i), DB_j^b(\theta_j)\}$$

The OEM's total transfer in the simple mechanism is still:

$$T^S = \max\{DB_i^S(\theta_i), DB_j^S(\theta_j)\}$$

It is very straightforward to show that I always have $T^b > T^S$. To see this, suppose first that I have $DB_i^b(\theta_i) > DB_j^b(\theta_j)$, then:

$$DB_i^b(\theta_i) > DB_j^b(\theta_j) = DB_j^S(\theta_j) + \int_{\Delta C_j\theta_j}^{\overline{m}} (m - \Delta C_j\theta_j)dF(m) > DB_j^S(\theta_j).$$

Meanwhile:

$$DB_i^b(\theta_i) = DB_i^S(\theta_i) + \int_{\Delta C_i\theta_i}^{\overline{m}} (m - \Delta C_i\theta_i)dF(m) > DB_i^S(\theta_i).$$

Therefore, the budget mechanism always leads to a strictly higher total transfer compared to the simple mechanism. The situation where $DB_i^b(\theta_i) > DB_j^b(\theta_j)$ is analogous.

Finally, note that the above analysis is true for any given type realization (ex-post) such that neither supplier can handle the design change. Suppose this is the case for all the type combinations in the type space, which is the situation specified in Proposition 4, then from an ex-ante point of view, the budget mechanism must also perform worse. Q.E.D.

C.4 Proof of Proposition 5

Proposition 5 operates under the situation where the most efficient supplier i (type $\underline{\theta}$) is not as competitive as the least efficient supplier j (type $\overline{\theta}$) in the simple mechanism, in the sense that such supplier i has a higher dropout bid then such supplier j. In other words, supplier j always wins the auction in the simple mechanism, for any type realization. Now, I turn to the budget mechanism and consider what will happen when I keep decreasing the budget from some very large initial value.

When the budget is so high that all the types of both suppliers can handle the design change for all their types ($b > \Delta C_i\overline{\theta}, b > \Delta C_j\overline{\theta}$), then the budget mechanism performs the same as the simple mechanism for the OEM in terms of the total transfer; see Section 4.5.1.1 for the analysis. Moreover, note that as I decrease b, the dropout bids for both suppliers increase by the same amount $\int_b^{\overline{m}}(m - b)dF(m)$, for any given type realization.[1]

Now, consider the case where I have an intermediary b value such that all types of supplier i can

[1] One way to visualize this is that if I draw the two dropout bids w.r.t. the type realizations, then as b decrease both curves shift upward, but at the same rate.

handle the design change, but some types of supplier j cannot handle the design change.[2] Then, in the budget mechanism, it must be the case that the most efficient supplier i is not as competitive as the least efficient supplier j. To see this, I refer to Section C.1.1., which states that supplier i's dropout bid will increase more than supplier j's dropout bid for such a budget.

In addition, I note that the dropout bids for either supplier i or j are still monotonically increasing w.r.t. their types based on Lemma 2. Therefore, in the budget mechanism, supplier j always wins the auction for any type realization. With this, I can conclude from Lemma 1 that the budget mechanism always leads to worse outcomes for the OEM in terms of the total transfer for any type realization (ex-post). Finally, I note that if this is the case for any realization ex-post, then this must be true ex-ante as well. Q.E.D.

C.5 Proof of Proposition 6

Proposition 6 operates under the situation where the least efficient supplier i (type $\underline{\theta}$) is more competitive than the least efficient supplier j (type $\overline{\theta}$) in the simple mechanism, in the sense that such supplier i has a lower dropout bid then such supplier j. Define the difference to be Δ: $\Delta = DB_j^b(\overline{\theta}) - DB_i^b(\overline{\theta})$

I follow a similar narrative sequence as in Proposition 5, and first consider high budget value where both suppliers can handle the design change for all their types. In this case, the budget mechanism performs the same as the simple mechanism for the OEM in terms of the total transfer.

Next, consider an intermediary value of b such that all types of supplier i can handle the design change, but some types of supplier j cannot handle the design change. I now consider two different type combination cases. Suppose supplier j can handle the design change ($\Delta C_j \cdot \theta_j < b$), then still both supplier i and j can handle the design change; therefore, the budget mechanism still performs the same as the simple mechanism.

On the other hand, suppose supplier j *cannot* handle the design change, then from Lemma 1, I know that there is the opportunity for the budget mechanism to outperform the simple mechanism, so long as supplier i wins the auction in the budget mechanism. Specifically, such suppliers are in the interval $\theta_j \in [\frac{b}{\Delta C_j}, \overline{\theta}]$. The overall strategy of selecting b is to make sure that, in the budget mechanism, supplier j in the interval always loses to (any) supplier i.

Consider the case that b is set exactly at $\Delta C_i \cdot \overline{\theta}$ from the left, such that the length of this interval is 0. This b value is the smallest value such that all the type combinations of supplier i and j can still handle the design change. Note also that the gap between the least efficient supplier i and the least efficient supplier j's dropout bid is still Δ, because both of their dropout bids increase by the

[2]Specifically, supplier j with a type above the threshold $b/\Delta C_j$ cannot handle the design change; this certainly includes the least efficient supplier j who has the largest type.

same amount $\int_b^{\overline{m}}(m-b)dF(m)$.

Now, suppose I slightly further decrease the value of b. Then, I have a short interval $\theta_j \in [\frac{b}{\Delta C_j}, \overline{\theta}]$ where the supplier j in this interval cannot handle the design change. When I consider the supplier j at the two endpoints of the interval, the supplier j at the left end point clearly has a smaller dropout bid because her type is smaller: $DB_j^b(\frac{b}{\Delta C_j}) < DB_j^b(\overline{\theta})$, but the difference is small and close to 0. On the other hand, for supplier i, by slightly further decreasing b, her dropout bid increases a bit, but is still smaller than that of the least efficient supplier j: $DB_i^b(\overline{\theta}) < DB_j^b(\overline{\theta})$; in particular, the difference should be close to the original difference Δ. Therefore, I have the following relationship: $DB_i^b(\overline{\theta}) < DB_j^b(\frac{b}{\Delta C_j}) < DB_j^b(\overline{\theta})$. In other words, all the suppliers j in this interval is not as competitive as the least supplier i. Therefore, all the suppliers j in this interval always lose to (any) supplier i in the budget mechanism.

Putting the two cases together, I have identified a range of b values that can achieve the goal in Proposition 6: to make the budget mechanism perform no worse than the simple mechanism in any type realization. In short, the selection of b is to make sure that $DB_i^b(\overline{\theta}) < DB_j^b(\frac{b}{\Delta C_j})$ is satisfied. Note also that $DB_i^b(\overline{\theta}$ increases with b while $DB_j^b(\frac{b}{\Delta C_j})$ decreases with b. Therefore, the b such that $DB_i^b(\overline{\theta}) = DB_j^b(\frac{b}{\Delta C_j})$ is the smallest b value that can achieve the goal in Proposition 6, and generate the larger benefit for the OEM under the budget mechanism for all such b values. Q.E.D.

C.6 Proof of Proposition 7

The proof follows the standard approach in classical mechanism design literature; Börgers & Krahmer (2015) is an excellent reference. I focus on showing the result for supplier i; the analysis for supplier j is identical. I first show the necessity. With a bit abuse of notation, I use $U_i(\theta_i)$ to denote $U_i(\theta_i;\theta_i)$ - supplier i's expected utility when she truthfully reports her type. incentive compatibility (IC) requires:

$$\begin{aligned} U_i(\theta_i) &= \max_{\tilde{\theta}_i} U_i(\tilde{\theta}_i;\theta_i) \\ &= \max_{\tilde{\theta}_i} T(\tilde{\theta}_i) + W(\tilde{\theta}_i)[C_i\cdot\tilde{\theta}_i - C_i\cdot\theta_i + (\Delta C\cdot\tilde{\theta}_i - \Delta C_i\cdot\theta_i)\cdot G_i(\tilde{\theta}_i)]. \end{aligned}$$

For part 1, consider two arbitrary type θ_i, θ_2. Under IC constraint, I have:

$$U_i(\theta_1) > U(\theta_2;\theta_1);\ U_i(\theta_2) > U(\theta_1;\theta_2);$$

Adding these two inequalities together will give us part 1.

For part 2, note that the utility function $U_i(\theta_i;\tilde{\theta_i})$ is an affine function of the reported type $\tilde{\theta}_i$; therefore, it is convex almost everywhere, and the maximum of it ($U_i(\theta_i)$) is also convex. Convex functions are differentiable almost everywhere. Therefore, I can apply the Envelope theorem, which specifies the necessary condition for its form when IC is satisfied:

$$U_i'(\theta_i) = -W_i(\theta_i) \cdot [C_i + \Delta C_i \cdot G_i(\theta_i)].$$

By taking the integral, I have the following expression:

$$U_i(\theta_i) = U_i(\theta_i) + \int_{\theta_i}^{\overline{\theta_i}} W_i(x) \cdot [C_i + \Delta C_i \cdot G_i(x)]dx.$$

Note that $U_i(\theta_i) = T_i(\theta_i) - 0 = T_i(\theta_i)$. Therefore, necessity of part 2 is shown. Finally, to show sufficiency, I can simply compare $U_i(\theta_i)$ with $U_i(\theta_i\tilde{;}\theta_i)$ by plugging in the expression of $T_i(\theta_i)$ I derive above. It is easy to verify that I indeed have $U_i(\theta_i) \geq U_i(\tilde{\theta}_i;\theta_i), \forall \tilde{\theta}_i$. Q.E.D.

C.7 Proof of Proposition 8

For each supplier, with the result from IC, I can now consider the OEM's expected total transfer from selecting a supplier with type θ. Below use supplier i for illustration. For supplier i with type θ_i, by selecting it the OEM pays her $T_i\theta_i + W_i(\theta_i) \cdot C_i\theta_i$ at the procurement stage, and gains back $W_i(\theta_i) \cdot \int_{\Delta C_i\theta_i}^{\overline{m}}(m - \Delta C_i\theta_i)dF(m)$ at the post-auction stage.

Such a gain is due to the fact that the OEM induces the supplier to truthfully report her type θ_i and writes this into the contract; therefore, the OEM can collect all the benefit that exceeds the supplier i's design change cost $\Delta C_i\theta_i$. Note also that this term (without the expected winning probability) is exactly the same as $B_i^S(\theta_i)$, supplier i's expected gain from handing the design change. In quick summary, the OEM's expected total payment to supplier i with type θ_i, denoted as $L_i(\theta_i)$, is:

$$L_i(\theta_i) = T_i(\theta_i) + W_i(\theta_i)[C_i\theta_i - B_i^S(\theta_i)].$$

So far, the "expected" for supplier i is w.r.t. supplier j's type distribution. That is, $T_i(\theta_i) = \int_{\underline{\theta}}^{\overline{\theta}} t(\theta_i,\theta_j)dH(\theta_j)$, $W_i(\theta_i) = \int_{\underline{\theta}}^{\overline{\theta}} w(\theta_i,\theta_j)dH(\theta_j)$, where $t(\theta_i,\theta_j)$ and $w(\theta_i,\theta_j)$ are the transfer received and the winning probability for any given type realizations.

Now, to derive the OEM's expected total transfer from implementing the mechanism, for $L_i(\theta_i)$ I need to take the expectation of supplier i's type for the term $L_i(\theta_i)$ to determine the ex-ante payoff to supplier i. With the changing integration order technique, I can show that:

$$\mathbb{E}_{\theta_i}[L_i(\theta_i)] = \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{\theta}}^{\overline{\theta}} w(\theta_i,\theta_j) \cdot v_i(\theta_i)dH(\theta_i)dH(\theta_j),$$

where $v_i(\theta_i)$ is the virtue value term specified in the proposition. The analysis process for supplier j is the same. Therefore, to derive the minimum ex-ante total payment $\mathbb{E}_{\theta_i,\theta_j}[L_i(\theta_i) + L_j(\theta_j)]$, the OEM can simply select the supplier with the lower virtue value. Finally, I need to guarantee the monotonicity condition for the allocation rule is met, i.e., satisfying part 1 of Proposition 7. It is easy to verify that this condition will be met so long as the virtue values are increasing in their respective types. Q.E.D.

BIBLIOGRAPHY

Bajari, P., Houghton, S., & Tadelis, S. (2014). Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review*, *104*(4), 1288–1319.

Bajari, P., & Tadelis, S. (2001). Incentives versus transaction costs: A theory of procurement contracts. *Rand journal of Economics*, (pp. 387–407).

Bearden, J. N., Murphy, R. O., & Rapoport, A. (2008). Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management*, *10*(4), 625–636.

Becker-Peth, M., & Thonemann, U. W. (2018). Behavioral inventory decisions. *The Handbook of Behavioral Operations*, (p. 11393).

Bendoly, E. (2011). Linking task conditions to physiology and judgment errors in rm systems. *Production and Operations Management*, *20*(6), 860–876.

Bendoly, E. (2013). Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, *31*(1-2), 62–71.

Bolton, G. E., & Katok, E. (2008). Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, *10*(3), 519–538.

Bolton, G. E., Ockenfels, A., & Thonemann, U. W. (2012). Managers and students as newsvendors. *Management Science*, *58*(12), 2225–2233.

Börgers, T., & Krahmer, D. (2015). *An introduction to the theory of mechanism design*. Oxford University Press, USA.

Bostian, A. A., Holt, C. A., & Smith, A. M. (2008). Newsvendor "pull-to-center" effect: Adaptive learning in a laboratory experiment. *Manufacturing & Service Operations Management*, *10*(4), 590–608.

Brandts, J., & Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, *14*(3), 375–398.

Brown, A. L., Chua, Z. E., & Camerer, C. F. (2009). Learning and visceral temptation in dynamic saving experiments. *The Quarterly Journal of Economics*, *124*(1), 197–231.

Cadsby, C. B., Song, F., Engle-Warnick, J., & Fang, T. (2019). Invoking social comparison to improve performance by ranking employees: The moderating effects of public ranking, rank pay, and individual risk attitude. *Journal of Economic Psychology*, *72*, 64–79.

Carbone, E., & Duffy, J. (2014). Lifecycle consumption plans, social learning and external habits: Experimental evidence. *Journal of Economic Behavior & Organization*, *106*, 413–427.

Carpenter, J., Graham, M., & Wolf, J. (2013). Cognitive ability and strategic sophistication. *Games and Economic Behavior*, *80*, 115–130.

Charness, G., Frechette, G. R., & Kagel, J. H. (2004). How robust is laboratory gift exchange? *Experimental Economics*, *7*(2), 189–205.

Charness, G., Karni, E., & Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and uncertainty*, *35*(2), 129–148.

Charness, G., Karni, E., & Levin, D. (2010). On the conjunction fallacy in probability judgment: New experimental evidence regarding linda. *Games and Economic Behavior*, *68*(2), 551–556.

Charness, G., & Sutter, M. (2012). Groups make better self-interested decisions. *The Journal of Economic Perspectives*, *26*(3), 157–176.

Chen, F. (2007). Auctioning supply contracts. *Management Science*, *53*(10), 1562–1576.

Cooper, D. J., & Kagel, J. H. (2005). Are two heads better than one? team versus individual play in signaling games. *The American economic review*, *95*(3), 477–509.

Cox, J. C. (2002). Trust, reciprocity, and other-regarding preferences: Groups vs. individuals and males vs. females. In *Experimental Business Research*, (pp. 331–350). Springer.

Croson, R., & Donohue, K. (2006). Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management science*, *52*(3), 323–336.

Denison, C. A. (2009). Real options and escalation of commitment: A behavioral analysis of capital investment decisions. *The accounting review*, *84*(1), 133–155.

Donohue, K., Özer, Ö., & Zheng, Y. (2020). Behavioral operations: past, present, and future. *Manufacturing & Service Operations Management*, *22*(1), 191–202.

Duenyas, I., Hu, B., & Beil, D. R. (2013). Simple auctions for supply contracts. *Management Science*, *59*(10), 2332–2342.

Duffy, J. (2016). Macroeconomics: a survey of laboratory research. In *Handbook of Experimental Economics*, (pp. 1–90). Elsevier.

Enke, B., Gneezy, U., Hall, B. J., Martin, D., Nelidov, V., Offerman, T., & de Ven, J. v. (2020). Cognitive biases: Mistakes or missing stakes?

Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, *42*(1), 1–34.

Feri, F., Irlenbusch, B., & Sutter, M. (2010). Efficiency gains from team-based coordination—large-scale experimental evidence. *American Economic Review*, *100*(4), 1892–1912.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, *10*(2), 171–178.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25–42.

Gavirneni, S., & Xia, Y. (2009). Anchor selection and group dynamics in newsvendor decisions—a note. *Decision Analysis*, *6*(2), 87–97.

Greiner, B. (2004). An online recruitment system for economic experiments. Tech. rep., University Library of Munich, Germany.

Grushka-Cockayne, Y., Erat, S., Wooten, J., Donohue, K., Katok, E., & Leider, S. (2018). New product development and project management decisions. In *The Handbook of Behavioral Operations*, (pp. 367–392). nJ: Wiley.

Hafalir, I., & Krishna, V. (2008). Asymmetric auctions with resale. *American Economic Review*, *98*(1), 87–112.

Hart, O. D., & Tirole, J. (1988). Contract renegotiation and coasian dynamics. *The Review of Economic Studies*, *55*(4), 509–540.

Herweg, F., & Schwarz, M. A. (2018). Optimal cost overruns: Procurement auctions with renegotiation. *International Economic Review*, *59*(4), 1995–2021.

Hey, J. D. (1982). Search for rules for search. *Journal of Economic Behavior & Organization*, *3*(1), 65–81.

Hey, J. D., & Dardanoni, V. (1988). Optimal consumption under uncertainty: An experimental investigation. *The Economic Journal*, *98*(390), 105–116.

Hill, G. W. (1982). Group versus individual performance: Are n+ 1 heads better than one? *Psychological bulletin*, *91*(3), 517.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644–1655.

Kagan, E., Sahin, O., & Leider, S. (2020). Dynamic decision making in operations management.

Kocher, M. G., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. *The Economic Journal*, *115*(500), 200–223.

Kugler, T., Bornstein, G., Kocher, M. G., & Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic psychology*, *28*(6), 646–657.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, (pp. 159–174).

Laya, N., & Pavlov, V. (2015). Team decision-making and individual learning in the newsvendor problem.

Maskin, E., & Riley, J. (2000). Asymmetric auctions. *The review of economic studies*, *67*(3), 413–438.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, *6*(1), 58–73.

Oprea, R., Friedman, D., & Anderson, S. T. (2009). Learning to wait: A laboratory investigation. *The Review of Economic Studies*, *76*(3), 1103–1124.

Özer, Ö., Zheng, Y., & Chen, K.-Y. (2011). Trust in forecast information sharing. *Management Science*, *57*(6), 1111–1137.

Özer, Ö., Zheng, Y., & Ren, Y. (2014). Trust, trustworthiness, and information sharing in supply chains bridging china and the united states. *Management Science*, *60*(10), 2435–2460.

Papastavrou, J. D., Rajagopalan, S., & Kleywegt, A. J. (1996). The dynamic and stochastic knapsack problem with deadlines. *Management Science*, *42*(12), 1706–1718.

Ryan, N. (2020). Contract enforcement and productive efficiency: Evidence from the bidding and renegotiation of power contracts in india. *Econometrica*, *88*(2), 383–424.

Schunk, D., & Winter, J. (2009). The relationship between risk attitudes and heuristics in search tasks: A laboratory experiment. *Journal of Economic Behavior & Organization*, *71*(2), 347–360.

Schweitzer, M. E., & Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, *46*(3), 404–420.

Seale, D. A., & Rapoport, A. (1997). Sequential decision making with relative ranks: An experimental investigation of the" secretary problem". *Organizational behavior and human decision processes*, *69*(3), 221–236.

Seale, D. A., & Rapoport, A. (2000). Optimal stopping behavior with relative ranks: The secretary problem with unknown population size. *Journal of Behavioral Decision Making*, *13*(4), 391–411.

Shachat, J., & Tan, L. (2015). An experimental investigation of auctions and bargaining in procurement. *Management science*, *61*(5), 1036–1051.

Sheremeta, R. M., & Zhang, J. (2010). Can groups solve the problem of over-bidding in contests? *Social Choice and Welfare*, *35*(2), 175–197.

Shin, J., & Ariely, D. (2004). Keeping doors open: The effect of unavailability on incentives to keep options viable. *Management science*, *50*(5), 575–586.

Song, H., Tucker, A. L., Murrell, K. L., & Vinson, D. R. (2018). Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science*, *64*(6), 2628–2649.

Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management science*, *35*(3), 321–339.

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management science*, *36*(6), 643–660.

Thonemann, U., & Becker-Peth, M. (2018). Behavioral inventory decisions: The newsvendor and other inventory settings. *The Handbook of Behavioral Operations*.

Tirole, J. (1986). Procurement and renegotiation. *Journal of Political Economy*, *94*(2), 235–259.

Trigeorgis, L. (1996). *Real options: Managerial flexibility and strategy in resource allocation*. MIT press.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, *90*(4), 293.

Waehrer, K. (1995). A model of auction contracts with liquidated damages. *Journal of Economic Theory*, *67*(2), 531–555.

Wang, R. (2000). Bidding and renegotiation in procurement auctions. *European Economic Review*, *44*(8), 1577–1597.

Wu, T., & Seidmann, A. (2015). Are groups better than individuals at making decisions? an experimental study on newsvendor problem. In *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, (pp. 4252–4261). IEEE.