Behavioral Pitfalls of Product Proliferation in Supply Chains: An Experimental Study*

Mozart B.C. Menezes[1], Kyle B. Hyndman[2]

[1]*NEOMA Business School, Paris, France*

[2]*Naveen Jindal School of Management, University of Texas at Dallas, U.S.A.*

mozart.menezes@neoma-bs.fr, kyleb.hyndman@utdallas.edu

Keywords: Proliferation, complexity, supply chain, information content, behavioral experiment, simulation tests, process performance.

Abstract: We study how increased complexity in terms of increased stock-keeping units (SKUs) and/or markets can affect operational performance, with an emphasis on managerial decision making. Specifically, when given the option to increase profits by increasing the number of markets served, we ask whether managers can increase profits by exercising this option or does increased complexity become a burden? We conduct a human-subjects experiment in which subjects manage a simulated supply chain across different levels of complexity, either as individuals or as part of a team. Subjects receive initial training in supply chain management and participate twice in the simulation – once as an individual and once as a team, while also varying complexity across trials. We show that as complexity increases revenues also increase. However, average performance often deteriorates and many subjects destroy value, despite the increased opportunities for profit. We argue that managers are tempted to chase new revenue sources without understanding the costs or risks to future profits. In a follow-up experiment, we show that when subjects are reminded about the importance of opportunity costs, revenue declines but earnings are the same or higher. Lastly, our experiments show that both teamwork and experience increase performance and reduce the variance of earnings. Experienced teams make better investment decisions. Less experienced individuals focus on revenue rather than earnings.

*We would like to thank the Editor (Xenophon Koufteros), a department editor, associate editor and two anonymous referees for their helpful comments, which have greatly improved the paper. We would also like to thank Sheen Levine for helpful conversations and Martin Willhoite for careful editing work.

1 INTRODUCTION

In today's business environment, executives and managers are under constant pressure to increase growth. In response, they are tempted to seek out new markets or to introduce new products as a way to attract new customers. We refer to this process as "product proliferation". Product proliferation results in not only new revenue streams, but also new costs. For example, production of different product lines may require lengthy machine setup or calibration for each new order, thereby resulting in lost productive time. Hence, as products proliferate, losses may accelerate (see, e.g., Buzacott and Shanthikumar 1993, Hopp 2007, Cachon and Terwiesch 2013). Product proliferation also impacts batch sizes. Although large sizes dilute the setup time and cost, they increase the costs of holding inventory, which hurts financial performance. There may be other non-operational direct and indirect costs such as capital investment, marketing expenses, debt servicing, etc.

However, these are not the only costs that may arise due to product proliferation. As the number of products/markets increases, so too does operational complexity. Common sense and experience tell us that too much complexity leads to poor decision making. Therefore, it is natural to expect that managers should be concerned about managing operational complexity.[2]

In this paper, we want to verify the existence of performance losses when managers are given the option but not the obligation to increase the number of markets served. In principle, this option – by expanding the choice set – should always weakly increase profits. Yet, managers often make decisions that increase operational complexity and costs beyond the rewards of new revenues. We seek to shed light on this tension caused by product proliferation. Specifically, we ask, do managers make good decisions and increase profits or does decision quality decline, causing managers to lose money when they increase their market presence? Put differently, can there be too much of a good thing (as discussed in Barnett and Freeman (2001))?

There is already some empirical evidence that product proliferation has negative consequences. In an industry-wide empirical study of U.S. semiconductor manufacturers, Barnett and Freeman (2001) find that, under some conditions, proliferation can substantially increase the chances of firms' failure. Bayus and Putsis (1999) study the personal computer industry and find that product proliferation does not help grow market share, nor does it help in deterring new competitors from entering the market. Suppliers suffer from product proliferation and even customers seem to avoid SKU proliferation beyond a certain point (Saeed and Young 1998).

We also see examples of this in the business press. For example, in June 2018, the French supermarket chain Carrefour announced a strategic shift from very large business units to smaller shops which are "less complex and more adapted to customers' behaviour". Additionally, in September 2017, the Danish toymaker Lego announced

[2]We follow Herbert Simon in using "complexity" to denote interdependency between elements of the problem. As he (Simon 1962, p. 468) puts it, "by a complex system I mean one made up of a large number of parts that interact in a nonsimple way. In such systems, the whole is more than the sum of the parts, not in an ultimate, metaphysical sense, but in the important pragmatic sense that, given the properties of the parts and the laws of their inter- action, it is not a trivial matter to infer the properties of the whole." Similar approaches have been taken by Ethiraj et al. (2008) and Sommer et al. (2020), among others.

a major reduction in headcount, alleging bureaucracy and complexity as the root cause of their problems (Milne 2017). Several academic papers have examined the issue of product proliferation (Adams et al. 2016, Mocker and Ross 2017, Fisher et al. 2017). Moreover, many practitioners cite product proliferation as an important factor that causes major losses (George Group 2006, Mariotti 2008).

Our interest in studying operational complexity due to product proliferation is also driven by interviews with supply chain management executives at national and multinational companies and non-governmental organizations. These interviews were conducted between January 2019 and September 2020 and focused on the subject of complexity and risk. The interviewees suggested that the number of combinations of products and destination points were behind the loss of efficiency of the supply chain. Moreover, they argued that part of the efficiency loss could be linked to poor decision making due to an excessive number of product/market combinations. These conversations motivated our desire to conduct an experiment where decision makers would need to make decisions around facilities, inventory, transportation and information – which are considered the main drivers of supply chain performance (Chopra and Meindl 2016) – but where we could control and vary complexity.

We conduct two human-subjects experiments designed to examine the role of complexity – as captured by product/market proliferation – on operational performance. Over the course of a two-hour simulation, subjects must manage production, inventory, market allocation, and shipping of their product with demand arriving in real time. They may also make fixed investments in production capacity and must make many other decisions, as would a real-world manager.

In our experiments, we vary complexity by varying the number of markets subjects can serve. Subjects with more potential markets have more potential profit opportunities, but these opportunities come at the cost of additional complexity. In particular, the complexity arises because subjects must manage and react to more information flows about supply and demand patterns in the additional markets. Beyond this, they must also consider making additional investments to better serve the markets. Through our computer-based business simulation exercise, we compare the performance of managerial teams submitted to different levels of complexity. In the first experiment, we compare performance for different combinations of complexity and sizes of management teams.

Given our hypothesis that increased operational complexity will erode performance, we are also interested in studying potential moderators of the relationship between complexity and performance. The two that we focus on are experience and the size of the decision-making unit. To this end, in our first experiment, subjects participate in the simulation twice, so we can also gain insight in the role of experience. They participate once as an individual and once as a part of a team to help us understand the value of larger decision-making units.

In our first experiment, although managing more regions is more complex, it comes with the potential for higher earnings. Yet, one of our first results is that not only do subjects not reap these higher earnings that arise from increased complexity, but they actually *destroy* value. That is, on average their profits were lower than if they would have simply left the system untouched (i.e., not made any decisions). We also show that increased complexity is associated with increased variability of earnings.

While product proliferation hurts profitability and increases variation, we show that the negative effects are mitigated with both teamwork and experience so that experienced subjects working as a team generally created value, even if they still failed to capture a significant portion of the additional potential profit due to being able to service more regions. Variability of earnings are also reduced with experience and teamwork.

Despite the beneficial effects of teamwork and experience, we still document a "complexity gap" – defined as the difference in potential profits and realised profits after accounting for teamwork and experience. In our least complex environment, this gap is only about 3%, but jumps to more than 15% in our other two, more complex, settings.

It is not necessarily surprising that performance suffers as complexity increases and that teamwork and experience can help to alleviate such complexity-induced suffering. It is of more interest to understand the underlying mechanisms for this result. Using data from the first experiment, we document that subjects' investment decisions are a key driver of their poor performance. Larger investments, especially beyond a certain threshold, are associated with lower earnings. However, once accounting for teamwork and experience, investment in fixed assets actually become a positive driver of performance.

In addition to this, we speculate that subjects may suffer from a salience bias (Kahneman et al. 1982, Loewenstein 1996, Hirshleifer 2008). Specifically, we argue that revenues are a highly salient metric and the link between revenue and key operational decisions (e.g., investment) is easy for subjects to both understand and observe. Moreover, this effect is stronger when subjects can service more regions, because there are more potential sources of revenue to tap. In partial support of this conjecture, we show that firms operating in more regions had substantially higher revenues, but earnings did not increase (or even decreased, depending on experience and team size). Thus, in chasing revenue, subjects created a level of complexity that was more than they could manage efficiently.

To further test this conjecture, we conducted a smaller follow-up experiment. Unlike our first study, all subjects worked in teams and had the option to operate in up to four regions. Instead, our treatment variation was designed to reduce the salience of revenues. We accomplished this by adding an additional paragraph of instructions for some subjects which reminded them of the *opportunity cost* of making additional investments and of the difference between revenues and profits. Our results show that teams which received these additional instructions invested significantly less and had significantly lower revenues but that their earnings were modestly higher, though not statistically different. Moreover, earnings were (weakly) significantly *less* variable.

In addition, because we collected richer data in this follow-up, we also present evidence which suggests that, not surprisingly, the timing of investments matters. Specifically, the later in the simulation that subjects made their initial investment, the lower were their earnings. We also show that earnings are significantly negatively related to the number of operational decisions that subjects made during the simulation. This is yet another potential channel for complexity to negatively affect earnings potential: in more complex environments, subjects may feel the need constantly tweak the system. Our results show that this is counter-productive.[3]

The overall message of the paper is that firms should consider operational complexity. Although both team-

[3] To be sure, this channel is conjectural because we did not vary complexity in this experiment.

work and experience can offset the worst effects of complexity, there is still a large complexity gap as soon as complexity is higher than the lower bound. Furthermore, complexity spills over into team dynamics and leads to more friction within the team, which may require redress to ensure performance does not suffer.

2 LITERATURE REVIEW

2.1 Complexity in Operations Management

The study of complexity in operations management has a long history and is very broad in scope (see e.g., Fonseca 2001, Blecker and Kersten 2006, to appreciate the breadth of scope). In our experiment, we view complexity through the lens of product (or market) proliferation. Even within this more narrow scope, there are several interesting and relevant papers.

First, when it comes to building a supply chain, Fisher (1997) argues that different products/product-lines may need different supply chain designs depending on the best trade-off between cost-efficiency capabilities and the level of responsiveness sought. How firms should balance these two priorities depends on whether (and to what extent) the product is innovative or functional. A functional product is one that typically has a low number of variants, while innovative products have significantly more variants. A supply chain designed for functional products (as is the case in our experiment) may not perform well when used for delivering innovative products. In our experiment, when subjects increase the number of markets served, the fundamentally functional nature of the products still suggests that the supply chain should be organized for efficiency. However, growing the number of markets – with their differing demand patterns – may tempt managers to focus on responsiveness. In this case, small mistakes may lead to substantial profitability losses. For example, different demand profiles may lead to difficulty in maintaining a high capacity-utilization, which is necessary for a cost-efficient supply chain.

Fisher and Ittner (1999) provide empirical evidence of some of the costs of high product variety in an automobile assembly setting. In particular, they argue that high product variety leads to high variability in work requirements, which necessitates having slack in labor. Nevertheless, they show via simulations that, in a properly calibrated work environment, higher product variety does not necessarily mean higher labor costs. However, from a practical perspective, finding this optimal calibration is made more difficult the greater is product variety. This is one of the main lessons from our experiment as well. In a more recent study, Shah et al. (2017) present evidence that product variety per assembly facility contributes to an increased number of product recalls, which ultimately increases costs.

Barnett and Freeman (2001) argue that, while some product proliferation can be beneficial, too much can be harmful. As they state, there are several factors that may cause disruption and efficiency losses, and that "[w]hen only a single new product is introduced, an organization's systems need to adjust only to this product's requirements, and so the resulting disruption is likely to be relatively minor. When multiple products are introduced all at once, however, each will require an adjustment of organizational systems . . . More generally, we expect that problems of coordination uncertainty increase with the number of products that are introduced simultaneously by an organization." Indeed, we believe that subjects in our experiment suffer from precisely these organizational problems, and that this contributes to the efficiency losses as complexity increases. Our results also show that teams are better-suited to overcome these organizational challenges.

Several other papers have linked complexity to reduced operational performance. Milgate (2001) observed a negative relationship between supply chain complexity and delivery speed and its reliability. Bozarth et al. (2009) argue that manufacturing complexity is negatively related to manufacturing cost-performance, among other things. Wan and Sanders (2017) suggested that increasing the product variety likely results in higher inventory levels due to large number of SKUs and low-quality of forecasting. The result is poor operational performance.

Tackling the same question, but from the perspective of product line optimization (see e.g., Tang 2010, for a survey), several recent papers have studied specific product lines and supply chains, with a particular focus on complexity/product proliferation and simplification. Rivkin (2000) argues that this is particularly challenging because of both the number and inter-connectedness of decisions needed to implement any given plan of action, and because of the need to use heuristics or learning-based approaches to successfully imitate complex strategies. Despite these challenges, some progress has been made. For example, Shunko et al. (2018) reports an interesting case study in which they highlight their efforts at restructuring a large manufacturing firm's product line. Like others, they note that product proliferation can be used as a competitive tool to increase customer satisfaction, but it comes with the cost of higher inventory requirements, difficulty in forecasting demand accurately for each variant and reduced opportunities for learning. By modeling and estimating consumers' preferences as well as the cost of complexity, they were able to reduce the number of product configurations from nearly 40,000 to just 135 while, at the same time, increasing sales by nearly 7%.

Menezes et al. (2020) applied the measure of complexity proposed by Ruiz-Hernández et al. (2019) to 27 business units of a large multinational corporation.[4] They verified that for each unit increase of complexity of a business unit, there is a loss of nearly 2.5 percentage points of operating profits. In an industry where the average operating profit is around 5%, this means that for every unit of increased structural complexity, losses would amount to 50% of average operating profits.

To gain a level of control not possible with empirical work, we decided to adopt an experimental approach. In terms of experimental research, researchers have looked at complexity, but often in a different context from ours. For example, Lee and Siemsen (2017) show that decomposing complex newsvendor ordering decisions into simpler elicitations of point forecasts, uncertainty estimates, and service-level decisions can lead to substantial performance improvements. Although different in scope, Kalkanci et al. (2011, 2014) have studied how contractual complexity affects supply chain performance. These papers show that (powerful) suppliers prefer simpler

[4]The measure of complexity developed by Ruiz-Hernández et al. (2019) is based on Shannon's (1948) measure of information content. It considers triples {SKU, channel, market} as the basic set of information the company is organised around. The precise definition is not central to our study, so we refer the interested reader to Ruiz-Hernández et al. (2019) for more details.

contracts even though they are theoretically sub-optimal. Cui et al. (2018) argue that this preference may actually be optimal if the retailer is boundedly rational.

The most closely related paper to us that studies complexity is Chen and Li (2018). They study how performance is affected when a manager must make multiple, simultaneous newsvendor decisions. They show that performance deteriorates when the manager is in charge of ordering two products, rather than one, but only when the product margins differ substantially because of what they term "cross-over" effects of a dynamic adjustment process. We too are interested in how performance changes when the number of products/regions under management varies. We differ in that the subjects in our simulation are placed in a much more immersive and continuous process rather than making repeated newsvendor ordering decisions.

2.2 Team Decision Making

To the best of our knowledge, there is very little work in operations management (OM) on team decision-making, though there are organizational behavior (OB) studies, with implications for the present research. Although teams often perform better than individuals, this is not universally true across all task environments (see, e.g., Hill 1982, Mueller 2012). One important mechanism for why teams may not perform better is the notion of *process loss* (Steiner 1972). Specifically, team performance may suffer – especially in larger teams – due to reduced motivation of individuals and worse coordination by team-members. This could be particularly relevant for us as our experiment requires subjects to make many different tasks and it is natural to expect some division of labor. However, these tasks are often interconnected and require some degree of coordination, which may not occur in a team setting. Despite these challenges, we show that teams do outperform individuals.[5]

Turn now to operations management. Li et al. (2018) is an exception to the dearth of OM research on team decision making.[6] Building on insights from behavioral economics (see, e.g., Cooper and Kagel 2005, Charness and Sutter 2012, for early work and a survey, respectively), they study team decision making in tactical and strategic environments. Tactical environments are typically problem-solving settings which do not involve strategic interactions. Strategic environments require one to consider other players' actions/reactions and incentives. In both environments, teams have been shown to perform better than individuals. In their tactical environment, Li et al. (2018) have teams of two make standard newsvendor ordering decisions. In their strategic environment they consider a forecast information sharing setting, where the sender has an incentive to inflate their forecast. In contrast to the behavioral economics literature, Li et al. (2018) show that teams actually perform worse in the newsvendor task and display a greater pull-to-center bias. On the other hand, in the strategic task, sender teams behave more strategically than individuals and inflate their forecasts more, while the receiver teams behave no differently from individual receivers and generally do not account for forecast inflation by senders.

[5] This result may be partially explained by the fact that our teams were quite small – indeed, pairs. It would be interesting to study the role of team size in complex operational decision environments.

[6] We keep our discussion of experimental research on teams brief. Li et al. (2018), provides a more detailed review of this literature.

Teams have been shown to be particularly effective when the problem under consideration rests on the discovery of a hidden insight that can be easily communicated to one's teammate. Such decision problems have been called "Eureka-type" problems (Li et al. 2018, Cooper and Kagel 2005). Our decision environment is very dynamic in nature, providing almost continuous feedback on sales revenues, lost sales, production and transportation costs and relies on making decisions in distinct domains such as transportation mode, production batch size, and ordering point. In addition, some actions, such as investment, are irreversible, which limits subjects' ability to take corrective action (Kleinmuntz 1985). Finally, investment decisions are made based on one's prediction of its impact on profitability. Therefore, we would not expect our problem to be a Eureka-type problem. Hence, it is not clear that teams will perform better than individuals.

Another reason why teams may be better than individuals is due to the so-called "wisdom of the crowd". For example, Sniezek (1989) showed that four different techniques for aggregating individual forecasts led to improved accuracy. Since forecasting demand is an important task in our experiment, if true, this could be a mechanism for improved team performance. However, it could also be a source of conflict, depending on the process teams use to aggregate forecasts. DeWees and Minson (2018) present evidence that people are less receptive to the information of others if they have already formed their own judgement.[7] Beyond this, Prelec et al. (2017) demonstrate that simple aggregation methods may not improve accuracy, especially when the task at hand requires specific expertise that only a minority of individuals may have. Despite these caveats, we do show that teams perform significantly better than individuals.

At the intersection of OM and OB, Huckman and Staats (2011) study team performance and, in particular, how diversity of experience affects the team's ability to adapt in response to changes in the task. They show that greater diversity of experience leads to *reduced* performance. However, this is mitigated when the team has greater familiarity with each other. On the other hand, "intrapersonal diversity" within a team – i.e., are the team members generalists or specialists? – actually *increases* performance in response to task changes. While we focus on the effect of teams (vs individual) in determining performance, we are unable to distinguish whether inter- or intra-personal diversity played a role in team performance. That said, we were interested in whether complexity played a role in the group dynamics. To gain some insight, we conducted a survey of subjects conducted after our first experiment, focusing on conflict, cooperation and leadership within the group. The results suggest that increased conflict within a team – which was more pronounced in more complex settings – negatively affected performance. Jehn et al. (2008) provides further insight from an organizational behavior perspective into how conflict affects team performance.

[7] We present survey evidence that conflict does increase within teams when complexity is high, although we cannot attribute the source of increased conflict to any specific aspect of the task, such as forecasting.

3 EXPERIMENT #1

We use the Supply Chain Simulation Game (SCSG), created by Chopra and Afeche (2016), in which subjects manage a fictitious firm's supply chain. The firm, Jacob Industries, services a varying number of regions (depending on the treatment). The headquarters is based in a region called Calopeia which already has an open production and a storage facility. During the experiment subjects must decide whether or not to open such facilities in the other region(s) that they are managing.

In the SCSG, subjects begin managing the supply chain at the beginning of "Year 3" and must do so for two "years", at which point the product line is discontinued and all remaining inventory is lost. For "Years" 1 and 2, the supply chain was operating automatically based on a pre-programmed strategy. Before subjects start the game, they have access to relevant financial and operational information for this initial period, and they have time to review this information. All information, including the way and time at which the experiment ends are known to subjects in advance.[8] From a practical perspective, a day in the simulation lasts approximately 11 seconds in real-time so that the entire simulation lasts two hours and 15 minutes.

While participating in the SCSG, subjects need to make several different types of decisions. In particular, key strategic decisions that need to be made are (i) whether (and if so, where and at what capacity) to invest in new manufacturing facilities; and (ii) where to open new warehouses. In addition to these high-level strategic decisions, subjects must also make several tactical/operational decisions, such as:

- which warehouses will serve which market;
- which manufacturing facility will serve which warehouse;
- the production order size (batch size);
- the re-order point (continuous replenishment system); and
- the transportation mode (mail or truck).

The SCSG strives for realism in many dimensions. For example, new manufacturing facilities and warehouses have a lead time before they are available for use. On the other hand, some decisions are (almost) immediately executed. For example, a change in the re-order point is implemented after the current order is produced, and changes in transportation mode are only implemented on new shipments and not those already in transit. As in practice, subjects make decisions through policies (i.e., set rules) which are automatically implemented by the system. For example, subjects set the re-order point and batch size such that every time the inventory of the product falls below the re-order point an order for producing the corresponding batch size is either immediately executed or enters a queue for later production if the manufacturing facility is busy with another order.

While the SCSG puts subjects in a very rich and dynamic business environment, it does make necessary and helpful simplifications. For example, the SCSG does not produce unexpected shocks. Additionally, there are no

[8] Chopra and Afeche (2016) have graciously allowed us to provide some screens that subjects saw in the experiment. This is provided in an Online Appendix. For more details on the SCSG, we refer the interested reader to `http://responsive.net/supply.html`.

disruptions such as a truck accident or a malfunction at a manufacturing facility, which do happen in practice, but would obscure the connection between decision making and performance. In the simulated environment any loss or gain can be only attributed to the decision making process. Furthermore, in the simulation, the only random component is demand, and subjects are given sufficient information to produce a reasonable forecast. Every other process is deterministic and known. Again, this helps make the connection between decision making and performance stronger.

While playing the SCSG, subjects are faced with many different tasks: demand forecasting, design cost, operational costs, and logistics cost evaluation. The objective of the simulation game is "to maximize cash position at the end of the game." During the simulation, subjects have access to information on logistics and manufacturing activities through reports that are similar to those presented by Enterprise Resource Management (ERP) systems widely used in practice. At the end of the simulation, students report their initial and final cash position, sales, lost sales, production, inventory, and transportation costs, and report the amount invested in fixed assets, and inventory write-offs at the end of the game. Our primary focus is on total cash at the end of the game, which is equivalent to operating profits.

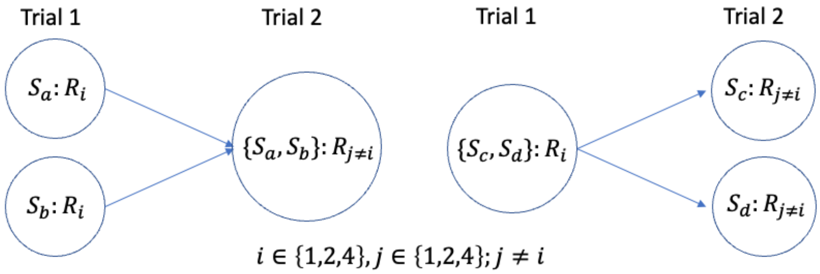
Approximately 120 students from two different Specialized Masters classes and one Master of Science class participated in our experiment as subjects. They played the SCSG three times. The first trial – managing a single region – was common to all subjects. It was used only to familiarize subjects with both the software and the nature of their decision problem, and will not be discussed further. For the next two trials, we implemented a randomized block design experiment in order to understand the connection between performance and complexity and how they are moderated by teamwork and experience.

We perform a 3×2 experimental design. In one dimension we vary the supply chain complexity by increasing the number of markets subjects need to deal with from 1 to 2 to 4. The second dimension is the way in which subjects participated in the next two trials: {Team First, Individual First}. In the *Team First* variation, subjects participated first as part of a team and then as an individual, while in the *Individual First* treatment, the reverse was true. We considered all possible progressions through the experiment with two exceptions: (i) subjects could not manage *the same* number of regions when working as an individual and when working as a team; and (ii) any pair of subjects who work together as a team would, when working as individuals, manage *the same* number of regions as each other. That is, if subjects a and b managed $i \in \{1, 2, 4\}$ regions together as a team, then they would *both* manage $j \neq i$ regions when working as individuals. This was true regardless of whether they started or ended the experiment as a team. Figure 1 illustrates this.[9]

To make the data more comparable across subjects and number of regions, the demand patterns for each region were the same. That is, the demand pattern for Region 1 (when there was only one region to manage) was the same as the demand pattern for Region 1 when there were two or four regions to manage. Similarly, the demand pattern for Region 2 was the same whether there were two or four regions to manage. Moreover, when

[9]Though we present our design as 3×2, we will also consider the learning dimension in our analysis – that is, changes between Trial 1 and Trial 2. Additionally, we will contrast team vs individual behavior.

Figure 1: Illustrative Progression of Subjects Through The Experiment



Note: S_m, $m \in \{a,b,c,d\}$ indexes four generic subjects in the experiment, while R_i/R_j : $i, j \in \{1,2,4\}$, $j \neq i$ indexes the number of regions a subject controlled in the trial. For example, S_a : R_1 indicates that subject a participated as an individual and managed one region. On the other hand, $\{S_a, S_b\}$: R_2 indicates that subjects a and b participated as a team and managed two regions. The only possible progressions through the experiment that we did not consider were (i) a subject could not manage *the same* number of regions both as an individual and as a team; and (ii) two subjects, when working together as a team could not manage a *different* number of regions when working as individuals.

there was more than one region, the demand patterns followed different time-paths, but they had the same mean and standard deviation.

3.1 Incentives

The experiment took place in a classroom environment and was part of the regularly scheduled classroom activities.[10] To provide strong incentives, 50% of their course grade was related to their performance in the SCSG and a written report reflecting on their experience in the simulation. Subjects also used the written report to transmit key decision and performance data to us.[11] Both performance and the written report were given equal weight. For the performance component, points were given based on a linear combination of the profits obtained using a pre-specified near-optimal strategy and the profits obtained by a "do-nothing" strategy (i.e., assuming that no active decision is made during the whole game). Performance equal to the "do-nothing" strategy was given 0 points and performance equal to the near-optimal strategy was given 100 points. Finally, the top-performing individual/group would also receive a 10% bonus on their overall course grade. Therefore, we believe that subjects in the experiment were highly motivated to perform well in the simulations. While we admit that our choice of incentives is non-standard, other studies in OM have used course credit – arguably a weaker incentive than in our study. Narayanan and Moritz (2015) is one such paper that uses course credit as an incentive. This paper also includes a nice discussion using non-monetary incentives and points out that there is often no clear difference in behavior with or without monetary incentives.

[10]That is, even if we did not pursue this research, the students' experience in the classroom would have been exactly the same. For this reason, university administration officials informed us that we did not need a formal approval from an institutional review board (IRB) and, as such, none was sought.

[11]While all subjects correctly reported performance data to us, they did not all accurately or completely transmit decision data. This imposes some limits on our data analysis in Experiment #1. Experiment #2, discussed below, sought to remedy this issue to allow us to better exploit the data on decisions.

3.2 Remarks on the Design

While our design allows us to gain some insight into the role of experience it is not perfect. As pointed out by a referee, there is also merit in studying behavior in treatments where subjects participate twice but both times as either an individual or as part of a team. This would allow us to identify the pure effect of experience for the individual and team environments separately. Given the limits on our sample size, we believed this to be a necessary compromise but we note that it is an interesting avenue for future research.

Before proceeding to the results, note that we are aware that our experiment deviates somewhat from traditional BOM experiments. As noted, rather than a more traditional, but more abstract experiment in which subjects make repeated, similar decisions from round-to-round, ours is a near continuous time experiment in which subjects must make numerous decisions on different aspects of supply chain management. These decisions dynamically affect subjects' revenue, costs and other performance metrics and can be revised periodically at the discretion of the subjects. Thus, our environment more closely embeds participants in a realistic business environment and allows us to study the connections between operational complexity, experience and team performance, which were at the heart of the problems that our interviewees (discussed in the introduction) struggled with. This decision comes at the cost of some loss of control and limits to the data we can analyze, but we hope that we achieve much greater external validity.

For Experiment #1, because subjects participated in the experiment twice – as an individual and as a team – we have to be somewhat careful in hypothesis testing. Given our assignment of subjects to treatments, standard hypothesis tests (i.e., parametric $t-$tests) are valid at the region level. That is, for any number of regions, $i \in \{1,2,4\}$, we can test for performance differences between the first and second trials and also between individual and team trials. This is because the subjects in each sample are different. Strictly speaking, if we pool across regions and conduct similar tests, then standard $t-$tests are not applicable because the samples are no longer independent. That said, for ease of exposition we report $t-$tests even on pooled data.

Of course, by conducting tests separately for each region, the question of multiple hypothesis testing arises. In particular, if we adopt the 5% level of significance and apply a simple Bonferroni correction, with three hypothesis tests (one for each number of regions, 1, 2 or 4), then the threshold for significance for each individual test would be $p^* = 0.0167$. With this threshold, as will be seen below, some of our results are only marginally significant.

However, despite these qualifications, we remain confident that the effects we report are real. As will be seen below, our results, at the region level, all point in the same direction. Moreover, fixed-effects regressions, which do account for the lack of independence in the data, and can collapse the tests of interest to a single hypothesis test, lead to the same conclusion and surpass the 5% level of significance. Appendix A contains the results of these regressions for earnings and normalized earnings (Table 10).

4 Results of Experiment #1

We begin our analysis of the experimental data by providing summary statistics in Figure 2 on earnings (interchangeably, profits) performance differentiating between the number of regions and whether subjects participated as individuals or as pairs. We will discuss learning below. For a frame of reference, note that if subjects did nothing, then they would have earned a profit of 41.47 across all conditions.[12] Panel (a) focuses on average earnings and so it provides a measure of performance in some absolute sense.

However, as noted, when the number of regions increases both the complexity and the *opportunity* for profit increase. That is, the maximum possible earnings are higher with four regions than with two or one. Therefore, panel (a) does not tell us whether subjects are "living up to their potential" with the differing number of regions. In panel (b) we seek to capture this by comparing behavior relative to a heuristic which approximates optimal behavior. To this end, panel (b) reports average normalized earnings, where we divide earnings by 45.18, 47.96 and 51.94 for 1, 2 and 4 regions, respectively.[13]

Consider first panel (a). As can be seen, when participating as an individual, when managing two or four regions, on average, subjects *destroyed* value through their actions. When participating as part of a team, only when subjects managed two regions did they destroy value on average (see also Figure 3 and the surrounding discussion below). Comparing earnings between individual and team trials, overall earnings are significantly higher when participating as a team (42.11 vs 39.63; $p = 0.009$, two-sample $t-$test with unequal variance). When comparing team versus individual earnings for the individual regions separately, the difference is statistically significant for 1 and 4 regions (R_1: $p = 0.021$; R_4: $p = 0.027$) but not for 2 regions ($p = 0.758$).[14] We also see that absolute performance is significantly worse when managing 2 regions than 1 region (39.91 vs. 42.45;

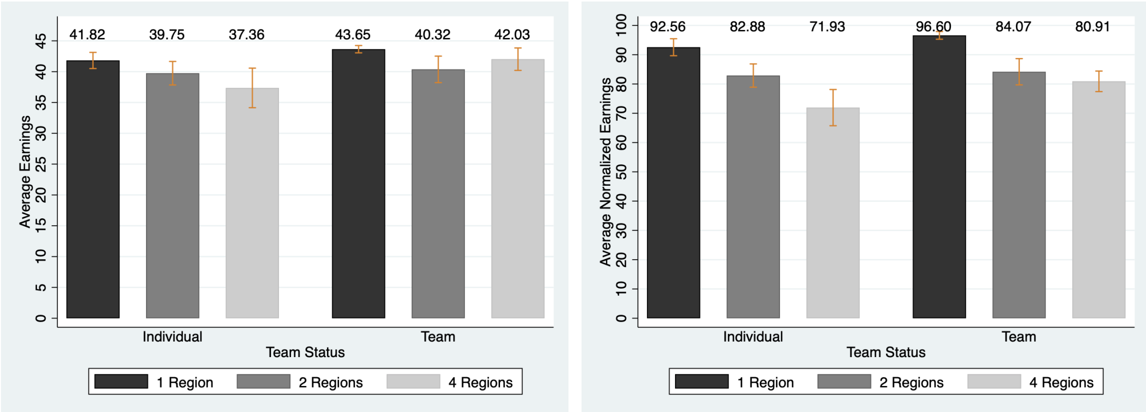
[12]As indicated above, subjects start the experiment at the beginning of "Year 3", where the system has been operating according to a pre-programmed strategy for the first two "years". This strategy was programmed to manage only one region – regardless of the number of potential regions available – and to follow a particular rule with respect to ordering and delivery. Additionally, this default strategy would not make any investments, either to expand capacity or the expand, if possible, the number of regions served. Therefore, if subjects, in any condition, did not make any changes, the default, pre-programmed strategy would be maintained and it would result in a profit of 41.47.

[13]Given the complexity of the environment, it is difficult to provide a precise upper bound based on "optimal" behavior. Instead, our naive heuristic goes as follows: we assume demand is stable with annual demand equal to the average of previous years and coefficient of variation of $1/4$. We apply the newsvendor model for estimating capacity and the economic order quantity (EOQ) model for estimating production lot-size. The reorder point is set to a very large number, which is never reached, so production does not stop. At the end of the game, we stop producing when inventory is equal to expected demand until the end of the simulation game. When playing multiple regions, individual capacities are decided as above but total capacity is capped based on aggregate demand when applying the newsvendor model. The heuristic is shown in the classroom at the debrief session following the "warm-up" game.

[14]The results are somewhat weaker if we use non-parametric Mann-Whitney tests, which are essentially a test of the difference in medians. In particular, for 1 region, the difference is significant at $p = 0.094$, while for 4 regions the difference is significant at $p = 0.105$. The pooled difference is still significant at $p = 0.029$. We do not feel like this detracts from our results because it also highlights, as we also show more directly, that teams reduce extreme behavior.

Figure 2: Summary Statistics: Profits By Regions and Group Size

(a) Average Earnings (b) Average Normalized Earnings



Note: The number on the top of each bar is the average of the statistic and the bars represent the 95% confidence interval.

$p = 0.007$) and 4 regions than 1 region (38.84 vs. 42.45; $p = 0.006$) but the difference between 2 and 4 regions is not statistically different ($p = 0.456$). This in contrast to what is theoretically possible where profits should increase in the number of regions served.

Although not visible in the figure because it pools across trials, our results show that absolute performance is statistically better in the second trial than in the first trial (42.10 vs. 38.74; $p \ll 0.01$). However, the learning effect is only significant when managing four regions. This is not surprising: the benefits of repeating a task or of working in a team should be most pronounced in more complex environments.[15]

We summarize this as follows:

Result 1. *Overall performance is (i) significantly higher when working as a team; (ii) significantly higher with previous experience and (iii) decreasing as the number of regions (i.e., complexity) increases. The effect of learning is strongest when complexity is highest and the effect of teamwork is not uniform as complexity varies.*

The other interesting feature that we see in Figure 2 is that not only do the treatment variations affect average performance, but the variance of earnings are also affected. At a high level, we see that the standard deviation of earnings is significantly lower for groups than for individuals (1 Region 1.9 vs 4.1, $p \ll 0.01$; 2 Regions 6.4 vs 6.2, $p = 0.80$; 4 Regions 5.6 vs 10.2, $p = 0.009$; Overall 5.0 vs 7.4, $p = 0.001$). We also see that variability of earnings increases in the number of regions being managed (one vs. two: $p \ll 0.01$; two vs. four: $p = 0.002$). Thus we have:

Result 2. *As complexity increases so too does the variance of realized outcomes. However, teamwork significantly reduces the variability of earnings.*

Turn now to panel (b) which reports earnings as a fraction of the proposed upper bounds for each number of regions. Looked at this way, we can see more starkly the negative effects of increasing complexity. As can

[15] Note that since earnings are simply normalized by a region-specific constant, the hypothesis tests comparing Trial 1 vs Trial 2 or team vs individual performance are exactly the same as for average earnings.

be seen, regardless of whether participating as an individual or as a team, as complexity increases, normalized earnings decrease. This is true even though, for example, absolute earnings increase for teams when moving from 2 to 4 regions. However, similar to earnings, normalized earnings are significantly higher for teams than for individuals.

Result 3. *Results 1 and 2 continue to hold when analyzing earnings normalized by an upper bound on earnings for each level of complexity. As a fraction of profits in our best heuristic, the effect of complexity is even larger.*

As noted, it is apparent that some subjects are actually destroying value through their actions and that earnings variability differs for teams than for individuals. Figure 3 provides another perspective on these points. Specifically, for each of our three region settings, we plot the empirical CDFs of earnings. The dashed lines in each panel highlight the earnings achievable by making no active decision or by following the naive heuristic. The greater variability in earnings as the number of regions under management increases is readily apparent in the figure. However, we also see that a non-negligible fraction of subjects actually destroy value – that is, they earn less than had they left the system unchanged. With one region, the fraction of subjects destroying value is approximately 25%, while this number rises to over 50% when there are two or four regions under management. Again, this is all the more surprising because with more regions, the potential to earn profits increases.

4.1 Quantifying the Cost of Complexity and the Benefits of Experience and Teamwork

The results presented above clearly show that increased complexity has a cost and also that both experience and teamwork are beneficial. We now seek to try to quantify these effects. Specifically, we estimate the following simple model at the region level, and then pooled:[16]

$$\texttt{Metric}_i = \beta_0 + \beta_1 \texttt{Experience}_i + \beta_2 \texttt{Teamwork}_i + \beta_3 \texttt{Experience}_i \times \texttt{Teamwork}_i + \varepsilon_i, \tag{1}$$

where $\texttt{Metric}_i$ is either absolute earnings or earnings normalized by the best heuristic. We use the estimated marginal effects to measure the performance gain due to experience and teamwork.[17] Specifically, the overall performance gain is captured by the difference between average performance in trial 2 when working as a team and average performance in trial 1 when working as an individual. This is given by the column $(2)-(1)$ in Table 1. Using the estimated marginal effects, we decompose the performance gain to the component due to experience and the component due to teamwork, which are shown in the next columns. Finally, we say that the *complexity gap* is given by the the difference between the upper bound of performance (column (3)) and performance in the second trial as part of a team (column (2)).

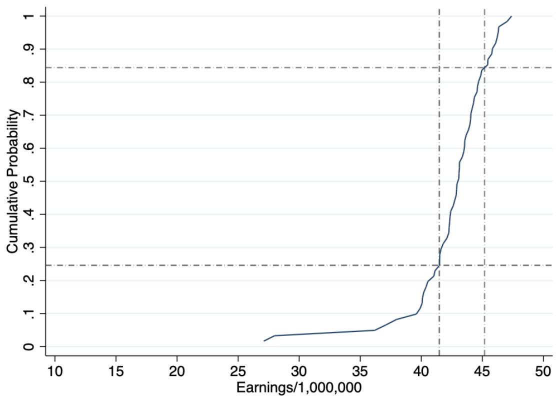
First, observe that the complexity gap is relatively small for the least complex environment (i.e., one region) and then it jumps substantially for two and four regions. Interestingly, the difference in complexity gap between two and four regions is quite small, which appears to be due to the observation that performance only increased modestly as subjects gained experience and worked in teams. Second, observe that we can never reject the

[16]As discussed in Section 3.2, these regressions are statistically valid at the regions level.

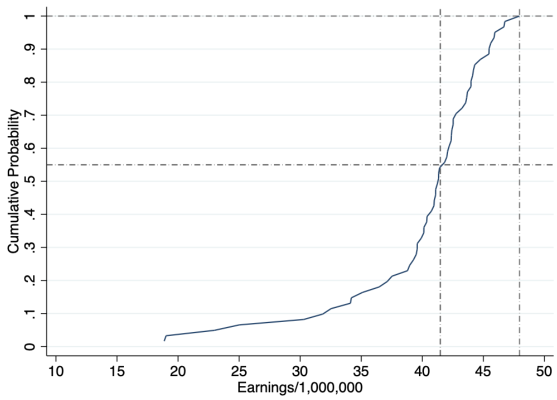
[17]The underlying regressions can be found in Table 11 in Appendix A.

Figure 3: Empirical CDF of Earnings By Regions

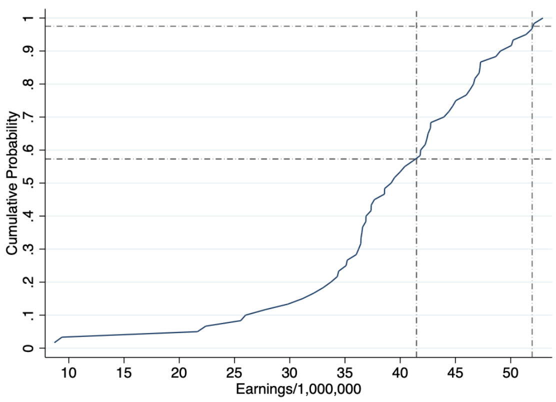
(a) 1 Region



(b) 2 Regions



(c) 4 Regions



(d) Note

On each plot, the left vertical line indicates the profits obtainable if subjects made not active decisions, while the right vertical line represents the profits obtainable in our heuristic solution. The horizontal lines are are meant to highlight important regions. The lower horizontal line gives he fraction of teams that destroyed value, the difference between the upper and lower horizontal lines gives the fraction of teams who created value, but less than the heuristic solution, while 1 minus the upper horizontal line indicates the fraction of teams that exceeded even the heuristic solution.

hypothesis that the marginal effects on experience and teamwork are the same (in all cases, $p > 0.1$). In fact, it appears that teamwork is slightly more important for one and two regions, but experience dominates for four regions (and the difference is closest to significant of all comparisons: $p = 0.168$ for both metrics). Finally, it is important to note that the coefficient on the interaction term, β_3, is negative in six of the eight regressions that help us generate Table 1. The two exceptions are the two regions case of each performance metric. However, in these cases, β_3 is not statistically distinguishable from zero ($p \gg 0.1$). This suggests that experience and teamwork are **substitutes** for each other. We can summarize this discussion as follows:

Result 4. *As complexity increases, so too does the complexity gap. Both teamwork and experience have approximately equal effects on performance gains and the two are substitutes for each other.*

The fact that experience and teamwork are substitutes for each other suggests that, when starting a new (complex) venture it may be better to build a team as teams better-suited (than an individual) to getting the project off to a good start. However, as the organization gains more experience, it is possible that an experienced

Table 1: Quantifying the Cost of Complexity and the Benefits of Experience and Teamwork

(a) Actual Profits

	(1)	(2)	(3)				
	Trial 1	Trial 2	Best	Performance Gain			Complexity
Regions	Group Size = 1	Group Size = 2	Heuristic	$(2)-(1)$	Experience	Teamwork	Gap $(3)-(2)$
1	41.20	43.81	45.18	2.61	0.98	1.63	1.37
2	39.72	40.67	47.96	0.95	0.31	0.64	7.29
4	31.63	43.64	51.94	12.01	7.87	4.14	8.30
Tot.	37.73	42.84	48.40	5.11	2.97	2.14	5.56

(b) Normalized Profits

	(1)	(2)	(3)				
	Trial 1	Trial 2	Best	Performance Gain			Complexity
Regions	Group Size = 1	Group Size = 2	Heuristic	$(2)-(1)$	Experience	Teamwork	Gap $(3)-(2)$
1	91.19	96.98	100	5.79	2.17	3.62	3.02
2	82.82	84.79	100	1.97	0.63	1.34	15.21
4	60.90	84.01	100	23.11	15.15	7.96	15.99
Tot.	78.91	89.26	100	10.35	5.73	4.62	10.74

Note: The part of the performance gain allocated to Experience and Teamwork comes from the estimated marginal effects of each variable from the regression: $\texttt{Metric} = \beta_0 + \beta_1 \texttt{Experience} + \beta_2 \texttt{Teamwork} + \beta_3 \texttt{Experience} \times \texttt{Teamwork} + \varepsilon$. Observe that in all cases (except with 2 regions), the coefficient on the interaction term, β_3, was always negative. This indicates that Experience and Teamwork are at least partial substitutes for each other.

individual could take primary responsibility for the project, freeing the other team members to develop and implement other projects.

4.2 A Deeper Exploration of Drivers of Behavior

In this section, we now seek to look a little deeper into some of the drivers of behavior and how complexity interacts with these drivers. We first look at investment. This is a prime candidate for a behavioral driver that interacts with complexity because, the more regions available to be served, the more investment is required to efficiently serve them. Next we will identify apparent "revenue chasing" behavior, which we show appears to occur as complexity increases. This also serves as our motivation for the second experiment, which tries to show how a fixation on revenue in complex environments may lead to worse performance.

4.2.1 The Impact of Investment on Performance

One of the main channels to increase earnings in this simulation is to make an appropriate (in terms of time and size) investment in capacity. Too little capacity and the manager misses opportunities to profit, while too much capacity is simply wasteful. As can be seen in Table 2, subjects chose higher investments when responsible for more regions, which is, in principle, rational. There is no consistent evidence that average investment is different by group size or by trial number. However, as can be seen in the table, there is some evidence that the standard deviation of investments is lower in later trials and when part of a team. This reinforces an earlier result that experience and teamwork reduce variability – here of investment.

Table 2: Capacity Investment By Regions, Trial and Group Size

(a) Group Size = 1

	Trial	
Regions	1	2
1	4.11	3.87
	(4.05)	(2.84)
2	7.19	8.39
	(5.89)	(2.99)
4	12.54	14.68
	(8.91)	(8.24)

(b) Group Size = 2

	Trial	
Regions	1	2
1	4.72	3.52
	(1.88)	(2.12)
2	7.96	6.86
	(6.69)	(3.40)
4	9.26	13.66
	(7.08)	(6.39)

Note: Standard deviations in parentheses below.

Of course, these summary statistics on investment do not tell us anything about the relationship between investment and earnings. We turn to this now. In Table 3 we report regression results where the dependent variable is either earnings (the results are identical if we use normalized earnings). To ensure statistically valid results, we report regressions for each number of regions separately, though we also report a pooled regression. As explanatory variables we include investment, as well as interactions for team trials and the second trial. Additionally, visual inspection of the data suggested that beyond a certain amount, additional investment is detrimental. The inflection point appeared to be at investment levels of 5, 10 and 20 in each of the 1, 2 and 4 region cases. Thus we include interaction terms to capture this.

More interestingly, the baseline effect (i.e., first trial, as an individual) of investment on earnings is negative for all regions and significantly so when four regions are managed. The same is true in the pooled regression. That is, higher investments actually lead to lower earnings. Furthermore, the negative effect becomes even stronger once investment exceeds the aforementioned threshold for each number of regions under management.[18] In contrast, the coefficients on both teamwork and experience interacted with investment are positive (and significant for 1 and 4 regions) and larger in magnitude than the baseline negative effect. We also see evidence that experience and

[18]A priori, we expected more of an inverted-U relationship, because initial investments should expand the range of possible earnings. However, these investments must also be done in a timely manner. That is, an investment of x at time t is not the same as an investment of x at time $t' \gg t$. Unfortunately, in this experiment, we are not able to observe when investments were made. In our follow-up experiment, we do observe the timing of investment and show that the later is the initial investment made, the lower are earnings.

teamwork are **complementary**. Specifically, although the coefficient on Team Trial $\times$ 2^{nd} Trial $\times$ Investment is negative, the magnitude of the coefficient is smaller than the positive individual effects.[19]

Table 3: The Effect of Investment on Earnings

Parameter	1 Region		2 Regions		4 Regions		Pooled	
2 Regions							-2.346^{*}	(1.273)
4 Regions							-4.423^{***}	(1.500)
Investment	-0.204	(0.250)	-0.037	(0.293)	-0.481^{**}	(0.233)	-0.347^{**}	(0.142)
$Inv. \times \mathbf{1}[Inv. > 5$ & $1\ Reg.]$	-0.378^{*}	(0.195)					-0.324	(0.216)
$Inv. \times \mathbf{1}[Inv. > 10$ & $2\ Reg.]$			-0.538^{**}	(0.217)			-0.412^{***}	(0.155)
$Inv. \times \mathbf{1}[Inv. > 20$ & $4\ Reg.]$					-0.188	(0.161)	-0.218^{**}	(0.107)
2^{nd} Trial $\times$ Inv.	0.301^{*}	(0.170)	0.092	(0.205)	0.745^{***}	(0.158)	0.575^{***}	(0.097)
Team Trial $\times$ Inv.	0.423^{*}	(0.213)	0.261	(0.252)	0.615^{**}	(0.253)	0.498^{***}	(0.143)
2^{nd} Trial $\times$ Team Trial $\times$ Inv.	-0.300	(0.316)	0.086	(0.372)	-0.481	(0.320)	-0.330^{*}	(0.189)
Constant	43.351^{***}	(0.722)	40.815^{***}	(1.736)	39.050^{***}	(2.317)	43.226^{***}	(0.895)
R^2	0.373		0.312		0.411		0.391	
N	57		50		55		162	

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

We can summarize this discussion as follows:

Result 5. *Inexperienced subjects, working as individuals, make relatively poor investment decisions, with higher investments leading to lower earnings. Both experience and teamwork lead to better investments such that increased investment becomes a significantly positive driver of earnings.*

4.2.2 Revenue Chasing and Complexity

Given that subjects make such poor investment decisions the question becomes, "Why?" We speculate that part of the reason is that subjects do not properly account for all costs and benefits when making investments and that they unduly focus on revenues, perhaps due to a salience bias (Kahneman et al. 1982, Loewenstein 1996, Hirshleifer 2008). That is, because revenues – and the connection between revenue and investment – are easily observable, subjects place undue emphasis on revenues. The mistake is to assume that increasing revenues must necessarily increase earnings.

Consider first revenues. Table 4 shows that revenues are increasing in the number of regions, even though earnings are actually decreasing. Beyond this, Table 5 shows that there is no relationship between earnings and revenue. More specifically, this table reports three separate linear regressions (one for each number of regions under management), where the dependent variable in all cases is earnings and we include revenues, controls for

[19]Furthermore, although not shown in the regressions, the negative effect of investment beyond a threshold also goes away when working as a team or with experience. This is partially driven by the result that teamwork and experience lower the variance of investment – particularly, reducing the frequency of extremely high investments.

the second trial and whether it was a team trial or an individual trial. As can be seen, the coefficient on revenue is never statistically significant.

Table 4: Earnings and Revenue By Number of Regions

Regions	Revenue	Earnings
1	55.38	42.45
2	88.66	39.91
4	156.48	38.84

Table 5: Relationship Between Earnings and Revenue By Number of Regions

	1 Region		2 Regions		4 Regions	
Revenue	0.002	(0.017)	0.014	(0.021)	0.001	(0.014)
2^{nd} Trial	0.698	(0.844)	0.277	(1.818)	8.633***	(2.565)
Team Trial	1.292	(0.898)	0.314	(1.877)	5.097**	(2.362)
Constant	41.802***	(1.070)	38.519***	(2.017)	32.414***	(2.361)
R^2	0.057		0.014		0.280	
N	57		52		55	

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Consider next the relationship between earnings, investment and revenue. Table 5 shows that there is no direct, unconditional, relationship between revenue and earnings. However, once we control for investment and interest earnings on cash reserves, we do see, in Table 6, that the relationship between earnings and revenue is positive and significant. Specifically, holding all else constant, every \$1 of revenue increases earnings by \$0.142.

Table 6: Relationship Between Earnings, Revenue, Interest and Investment (Seemingly Unrelated Regression)

	(1) Earnings		(2) Revenue		(3) Interest	
2^{nd} Trial	−0.257	(0.640)	32.671***	(4.723)	−0.223	(0.154)
Team Trial	0.241	(0.608)	9.647*	(5.055)	0.073	(0.165)
2 Regions	−1.952***	(0.713)	5.638	(5.974)	−0.176	(0.195)
4 Regions	−3.605***	(0.853)	26.792***	(6.765)	−0.568**	(0.221)
Revenue	0.142***	(0.009)				
Interest	3.041***	(0.287)				
Investment	−0.509***	(0.124)	8.276***	(0.420)	−0.316***	(0.014)
Constant	−24.830***	(6.247)	2.932	(5.075)	21.648***	(0.166)
R^2	0.731	N	162			

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

As Table 6 also shows, investment is the main driver of revenue. Therefore, rather than looking at the direct effect of revenue on earnings, it makes sense to examine the direct and indirect effects of an additional dollar of investment on earnings. Column (1) shows that the direct effect of and additional dollar invested is to reduce earnings by \$0.509. However, there are also two indirect effects: first, columns (1) and (2) shows that it increases earnings via revenue by $8.276 \times 0.142 = \$1.175$; second, columns (1) and (3) shows it decreases earnings via reduced interest receipts by $0.316 \times 3.041 = \$0.961$. Summing these three effects shows that the total effect of an additional \$1 of investment is -\$0.295. This is identical to the effect if we simply regress earnings on investment (controlling for trial, number of regions and teamwork).

A plausible interpretation is that subjects focus on the large, and easily observable – that is, salient – effect that investment has on revenues (i.e., column (2) in Table 6) while ignoring the opportunity cost of foregone interest earnings. Beyond this, there are also hidden costs that arise if the additional capacity brought online from investment is not utilized properly. Put differently, the link between investment and revenue is both easily understood and large, while the link between investment and earnings is difficult to properly score. These effects are exacerbated when subjects operate in a more complex environment (due to more regions) because the impact of investment on revenues is even stronger and so it more easily masks these hidden costs. We will return to this below in our discussion of our follow-up experiment. However, we first turn to an analysis of a post-experiment survey.

4.3 POST-EXPERIMENT SURVEY ON COMPLEXITY AND TEAM PERFORMANCE

As noted, at the end of the experiment, we conducted a survey designed to assess how subjects viewed their interactions with their teammates and whether this was related to complexity. We were interested in two things. First, do survey responses differ based on the number of regions that subjects managed when participating as a team? Second, do survey responses explain performance? Because these questions are not related to complexity per se but rather the more specific impact of complexity in a team setting, we provide only a brief summary of the results here. The precise questions and a more detailed analysis are provided in Appendix B. From this analysis, we can draw two fairly intuitive conclusions:

Result 6. *In the team trials, subjects reported significantly more friction, tension, emotional conflict and a higher frequency of disagreement as the number of regions under management increased. Subjects also tended to overweight their own contribution and their own leadership above that of their teammate, particularly when team performance was above the median.*

Result 7. *The frequency of disagreement between team members (which increases as complexity increases) and the relative perceived difficulty of the team trial are associated with significantly worse team performance.*

That is, as complexity increases, so too does conflict within a team, and this is associated with worse performance. That said, as our previous results have shown, team decision making units still substantially outperformed individual decision making units, even (indeed, especially) in the most complex environment.

5 EXPERIMENT #2

Experiment #1 showed that performance – especially compared against potential profit – actually declines as complexity increases. Section 4.2 presented evidence that subjects did not make optimal investment decisions and appear to be chasing revenue. To further test our conjecture about revenue-chasing, we conducted a small follow-up experiment. Unlike the first study, all subjects participated as part of a team (of 4) and subjects only participated in one trial. Furthermore, all subjects were responsible for managing four regions, which is where complexity (and so the temptation to chase revenue) is maximal.

We ran two treatments: the No Frame Treatment (19 teams) and the Frame Treatment (30 teams). In both treatments, all subjects received identical training as in Experiment #1. However, in the Frame Treatment, subjects were also told:

- Based on past experience, we find that students do not fully evaluate all of the costs and benefits of making particular investments. For example, money invested in operations will be money that won't earn interest, and the interest rate is 20%.
- It is in your benefit to make sure that if a unit of capacity is acquired then it should (produce and) sell enough to at least break-even considering the alternative gain: the interest on unused capital.
- Recall that revenues are not the same as profits.

For the Frame treatment, we were also able to collect detailed information on the timing and number of decisions that subjects made.[20]

5.1 Descriptive Analysis

Table 7 shows average earnings, revenue, interest income and investment for Experiment #2. As can be seen, whether or not subjects were framed to think about opportunity costs, average earnings are approximately the same (and, indeed, the difference is not statistically significant; $p > 0.1$). Therefore, our framing intervention did not increase subject performance on average. However, there were large and significant effects on revenues, interest and investments. Specifically, both revenues and investment were reduced by more than half when subjects were framed to think about opportunity costs (in both cases $p \ll 0.01$) and interest income was approximately 20% higher ($p \ll 0.01$).

Furthermore, there is also suggestive evidence that earnings were less variable when subjects were framed to pay attention to opportunity costs. Specifically, we can reject that the standard deviations are the same, in favor of the *one-sided* alternative that the standard deviation is lower when subjects are framed at $p = 0.052$. This suggests

[20]We classified subjects' decisions into one of three broad categories: strategic, tactical and operational. We classify strategic decisions as opening a new plant, increasing plant capacity, and opening a new warehouse; tactical decisions are warehouse to market allocations and changes in transportation mode; finally, operational decisions are those affecting the order quantity or order point. The key distinguishing feature of strategic decisions is that they come with a three-month lead-time, whereas other decisions are implemented almost immediately.

Table 7: Earnings, Revenue and Investment

Metric	No Frame		Frame	
Earnings	39.53	(6.93)	40.17	(4.95)
Revenue	198.54	(77.10)	88.81	(76.61)
Interest	15.91	(2.73)	19.22	(1.84)
Investment	17.68	(7.30)	7.63	(6.52)

Note: Standard deviations in parentheses to the right of the reported mean.

Table 8: Relationship Between Earnings, Revenue, Interest and Investment (Seemingly Unrelated Regression; Exp. #2)

	(1) Earnings		(2) Revenue		(3) Interest	
Frame Treatment	27.276^{*}	(15.166)			-1.032^{**}	(0.498)
Revenue	0.192^{***}	(0.016)				
Frame × Revenue	-0.058^{***}	(0.020)				
Interest	4.262^{***}	(0.491)				
Frame × Interest	-1.053	(0.700)				
Invest	-0.567^{***}	(0.130)	10.186^{***}	(0.643)	-0.354^{***}	(0.025)
Frame × Invest					0.092^{***}	(0.032)
Constant	-56.371^{***}	(11.061)	13.956	(9.127)	22.219^{***}	(0.463)
R^2	0.845		0.837		0.904	
N			49			

that the additional investment introduces risk into earnings and, we argue, the channel for this additional risk is due to the increased complexity. That is, the more investment, the more manufacturing and warehouse facilities must be managed. Consequently, there are greater demands on subjects, and this may introduce additional risk in overall performance.

Table 8 examines the relationship between earnings and revenue, interest, investment and framing in a seemingly unrelated regression framework to control for potential correlation of error terms. As was the case in Table 6, we see that, there is a strong, positive, relationship between revenue and investment, and a strong, negative relationship between interest earnings and investment. Likewise, controlling for investment and interest, there is a positive and significant relationship between earnings and revenue.

Of most interest to us is the role of framing. We found no evidence that framing effected the relationship between investment and revenue, and so it was omitted from the regression. On the other hand, framing does have an effect on the relationship between interest and investment. Specifically, when framed to think about opportunity costs, each additional unit of investment reduces interest by about 26% (i.e., $^{0.092}/_{0.354}$) less than when subjects are not framed. Even more importantly, we see from column (1) that while framing significantly weakens the relationship between earnings and revenue, the coefficient on the frame treatment dummy is positive and (marginally) significant. We argue that this beneficial effect of framing is due to the reduction in structural complexity that subjects faced in the framed treatment because of their different approach to investment. Thus, we have:

Result 8. *Reducing the salience of revenue and reminding subjects to consider the opportunity cost of their actions led to significantly lower investment. As a result, revenues were substantially lower; however, profits were modestly higher.*

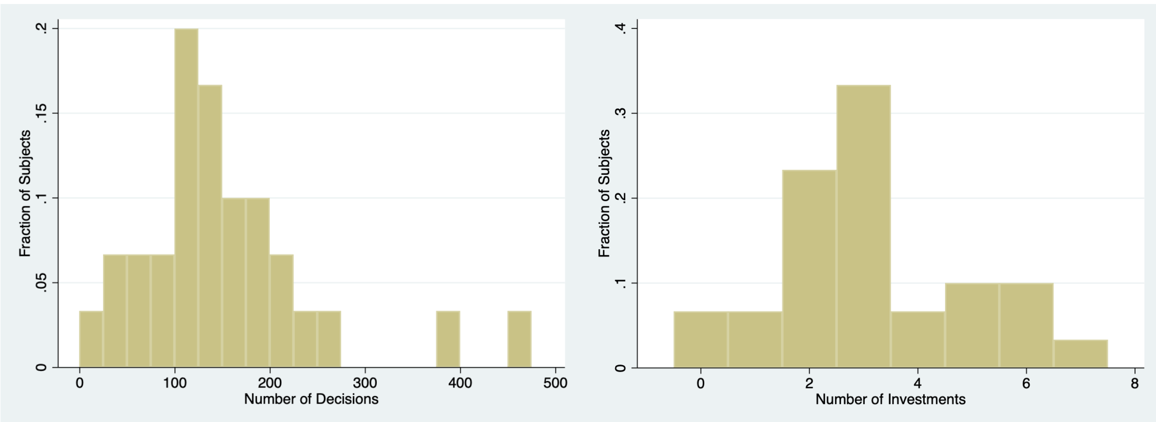
5.2 Richer Analysis of Decisions in Frame Treatment

As noted, in the Frame treatment, we have detailed data about the timing and number of decisions, which we have classified into strategic, tactical and operational (see Footnote 20 for how these are defined). In Figure 4 we plot histograms for the number of decisions overall (panel (a)) and the number of investments (panel (b)), which is a key strategic decision. As can be seen, there is a great deal of heterogeneity across subjects. The number of decisions overall – of which nearly 70% are classified as operational – varies from near 0 to nearly 500, while the number of investments varies between 0 and 7.

Figure 4: Tabulating the Frequency of Decisions

(a) Histogram of Number of Decisions

(b) Histogram of Number of Investments



It is then natural to ask if there is a relationship between these decisions and earnings. To this end, in Table 9 we report the results of a series of regressions of earnings on key variables such as the timing of the first investment, the number of decisions made, as well as the overall magnitude of investments in manufacturing (Mag. Investment) and opening warehouses (Mag. Warehouse). As can be seen, in all specifications – which change by gradually dropping insignificant variables – there is a significantly negative relationship between earnings and the timing of the first investment, and also a significantly negative relationship between earnings and the number of operational decisions. Thus, subjects who wait too long to make key investments suffer. This is intuitive because investments involve up-front costs and benefits that accrue over time and only after the lead-time to build-out these investments. This logic also suggests that it would likely hold even in less complex environments than the four-regions case considered in this treatment.

Table 9: Regression Analysis: Operational Decisions and Earnings (Dep Var: Earnings)

	(1)		(2)		(3)		(4)		(5)	
Time 1^{st} Inv.	-0.028^{***}	(0.010)	-0.028^{***}	(0.009)	-0.025^{***}	(0.009)	-0.026^{***}	(0.008)	-0.026^{***}	(0.008)
#Strategic	-0.388	(1.109)	-0.523	(0.505)						
#Tactical	-0.027	(0.057)	-0.027	(0.056)						
#Operational	-0.024^{**}	(0.010)	-0.023^{**}	(0.009)	-0.023^{**}	(0.009)	-0.023^{**}	(0.009)	-0.023^{**}	(0.009)
# Investments	-0.190	(1.376)								
Mag. Investment	-0.153	(0.173)	-0.157	(0.167)	-0.132	(0.159)	-0.087	(0.128)		
Mag. Warehouse	1.545	(1.849)	1.666	(1.589)	0.599	(1.215)				
Constant	67.050^{***}	(9.615)	66.803^{***}	(9.225)	62.645^{***}	(7.964)	64.444^{***}	(6.967)	63.383^{***}	(6.717)
R^2	0.430		0.430		0.398		0.392		0.380	
Observations	28		28		28		28		28	

In contrast, although we cannot definitively say because we did not vary complexity, we suspect that the result that subjects who make too many operational decisions suffer lower earnings is related to complexity. In more complex environments, there are more levers for the manager to "tinker" with and so much more temptation to do so. In contrast, in a less complex setting, the manager has fewer levers and can remain more focused on those that do remain. This may allow the manager to develop a clear strategy that does not need continual adjustment.

6 DISCUSSION AND CONCLUDING REMARKS

In this paper we report on two experiments with human subjects who make typical supply chain decisions, which varied in complexity via changes in the number of potential markets subjects could serve. Subjects were free to ignore the additional complexity by simply choosing not to service the new markets. However, by exercising the option – in theory and with optimal decision making – they should be able to increase their profit. The results suggest that most subjects actively tried to exploit these new market opportunities when they were available. However, our results also provide clear evidence that as complexity increases, many such attempts fail and subjects often destroy value.

One possible explanation which we have already discussed in the paper is that, when faced with a complex situation, it is difficult to estimate the gain in profit from making a particular investment or decision. Therefore, subjects focus on metrics which are more easily forecasted, such as revenue. A naive view is then that, with more regions, investing in capacity to serve these regions will generate substantially higher revenue. Moreover, these revenues are visible in real-time, which could reinforce in subjects' minds that they are making good decisions. However, by ignoring the hidden costs of investment (e.g., increased difficulty managing capacity and foregone earnings on cash reserves), when final profits are calculated, the investments tend not to pay off. Thus, it appears that as complexity increases, subjects focus on inappropriate measures of performance. The end result is poor performance in complex environments.

To put this conjecture to the test, our second experiment included a treatment in which we specifically reminded to think carefully about the opportunity cost associated with investments, and that revenues were different from profit. Our results clearly show that when subjects received this simple reminder, they invested less and revenues declined substantially. However, despite this, actual earnings were modestly *higher*, though the difference was not significant.

While this paper was primarily focused on the impact of complexity on performance, we also studied potential moderators of this relationship. To this end, we found that experience and team size can both reduce the negative impact of increased complexity. Experience seems to work better at higher levels of complexity and teamwork is most impactful at middle levels of complexity. More importantly, both teamwork and experience (i) reduce the variability of earnings; (ii) reduce the likelihood of big mistakes; and (iii) lead to better investment decisions.

Finally, using the survey, we showed that conflict within a team setting increases in complexity and, at least for some measures, higher conflict was associated with lower performance. This is an additional risk factor to consider when considering an increase in complexity: additional complexity may lead to increased conflict, which may harm performance. We also saw evidence that subjects viewed their contribution as being more important than their teammate when performance was good and that they reported that *they* were the leader when performance was good. This suggests that subjects may suffer from a self-attribution bias where they attribute success to their performance, while failure is shared across the team. One wonders if this is part of the reason why conflicts within the team increase as complexity increases.

From a managerial point of view, our results suggest that careful attention must be paid when deciding whether to enter new markets. In particular, it is important that the different functional areas of the business have voices which are heard. While the sales/marketing area will often find it tempting to expand into new products or new markets, the operations side of the business must also be heard to ensure that all costs associated with such an expansion are properly accounted for. The fact that more decisions do not necessarily lead to more profit also suggests that decision support tools could be valuable. Such a system could make recommendations about the need for change or could ask the decision maker to provide justification for the change. This would help ensure that any changes that are implemented have been more carefully analyzed. Our results also suggest that product-line rationalization strategies, which take operational complexity into account, are potentially valuable exercises for business to undertake.

REFERENCES

Adams, C., Alldredge, K., Mueller, C., and Whitmore, J. (2016). Simpler is (sometimes) better: Managing complexity in consumer goods. (McKinsey&Company).

Barnett, W. P. and Freeman, J. (2001). Too much of a good thing? product proliferation and organizational failure. *Organization Science*, 12(September-October):523–659.

Bayus, B. L. and Putsis, W. P. (1999). Product proliferation: An empirical analisys of product line determinants and market outcomes. *Marketing Science*, 18(2):137–153.

Blecker, T. and Kersten, W., editors (2006). *Complexity Management in Supply Chains: Concepts, Tools and Methods*, volume 2 of *Operations and Technology Management*. Erich Schmidt Verlag.

Bozarth, C., Warsing, D., Flynn, B., and Flynn, E. (2009). An information flow model for conflict and fission in small groups. *Journal of Operations Management*, 27(1):78–93.

Buzacott, J. A. and Shanthikumar, J. G. (1993). *Stochastic Models of Manufacturing Systems*. Prentice-Hall, Inc.

Cachon, G. and Terwiesch, C. (2013). *Matching Supply with Demand: An Introduction to Operations Management*. McGraw-Hill Education, third edition.

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

Chen, K.-Y. and Li, S. (2018). The behavioral traps in making multiple, simultaneous newsvendor decisions.

Chopra, S. and Afeche, P. (2016). Supply chain simulation game. Simulation game by Responsive.net.

Chopra, S. and Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation, 6th Edition*. Pearson.

Cooper, D. J. and Kagel, J. H. (2005). Are two heads better than one? Team versus individual play in signaling games. *American Economic Review*, 95(3):477–509.

Cui, T. H., Kong, G., and Pourghannad, B. (2018). Is simplicity the ultimate sophistication? Wholesale pricing vs. non-linear pricing.

DeWees, B. and Minson, J. A. (2018). Yhe right way to use the wisdom of crowds.

Ethiraj, S. K., Levinthal, D., and Roy, R. R. (2008). The dual role of modularity: Innovation and imitation. *Management Science*, 54(5):939–955.

Fisher, M., Gaur, V., and Kleinberger, H. (2017). Curing the additction to growth. *Harvard Business Review*, January-February.

Fisher, M. L. (1997). What is the right supply chain for your product? *Harvard Busienss Review*, (March/April).

Fisher, M. L. and Ittner, C. D. (1999). The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science*, 45(6):771–786.

Fonseca, J. (2001). *Complexity and Innovation in Organizations*. Routledge, 1st edition.

George Group (2006). Unraveling complexity in products and services. (Wharton, University of Pennsylvania).

Hill, G. W. (1982). Group versus individual performance: Are $n+1$ heads better than one? *Psychological Bulletin*, 91(3):517–539.

Hirshleifer, D. (2008). Psychological bias as a driver of financial regulation. *European Financial Management*, 14(5):856–874.

Hopp, W. (2007). *Supply Chain Science*. McGraw-Hill Irwin.

Huckman, R. S. and Staats, B. R. (2011). Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management*, 13(3):310–328.

Jehn, K. A., Greer, L., Levine, S., and Szulanski, G. (2008). The effects of conflict types, dimensions and emergent states on group outcomes. *Group Decision and Negotiation*, 17:465–495.

Kahneman, D., Slovic, P., and Tversky, A. (1982). *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, UK.

Kalkanci, B., Chen, K.-Y., and Erhun, F. (2011). Contract complexity and performance under asymmetric demand information: An experimental evaluation. *Management Science*, 23(2):269–284.

Kalkanci, B., Chen, K.-Y., and Erhun, F. (2014). Complexity as a contract design factor: A human-to-human experimental study. *Production and Operations Management*, 57(4):689–704.

Kleinmuntz, D. N. (1985). Cognitive heuristics and feedback in a dynamic decision environment. *Management Science*, 31(6):680–702.

Lee, Y. S. and Siemsen, E. (2017). Task decomposition and newsvendor decision making. *Management Science*, 63(10):3226–3245.

Li, J., Beil, D., and Leider, S. (2018). Team decsion making in operations management.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3):272–292.

Mariotti, J. (2008). *The Complexity Crisis: Why too many products, markets, and customers are crippling your company.* Adams Media, Copyright by John L. Mariotti, 57 Littlefield Street, Avon, MA 02322.

Menezes, M., Ruiz-Hernández, D., and Yen-Tsang, C. (2020). On the validity and practical relevance of a measure for structural complexity. Mimeo.

Milgate, M. (2001). Supply chain complexity and delivery performance: an international exploratory study. *Supply Chain Management: An International Journal*, 6(3):106 – 118.

Milne, R. (2017). Lego must press the 'reset button' to regain growth.

Mocker, M. and Ross, J. W. (2017). The problem with product proliferation. *Harvard Busienss Review*, May-June.

Mueller, J. S. (2012). Why individuals in larger teams perform worse. *Organizational Behavior and Human Decision Processes*, 117:111–124.

Narayanan, A. and Moritz, B. B. (2015). Decision making and cognition in multi-echelon supply chains: An experimental study. *Production and Operations Management*, 24:1216–1234.

Prelec, D., Seung, H. S., and McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541(7638):532–535.

Rivkin, J. W. (2000). Imitation of complex strategies. *Management Science*, 46(6):824–844.

Ruiz-Hernández, D., Menezes, M. B. C., and Amrani, A. (2019). An information-content based measure of proliferation as a proxy for structural complexity. *International Journal of Production Economics*, 212:78–91.

Saeed, B. and Young, D. (1998). Managing the hidden costs of complexity. (Boston Colsulting Group).

Shah, R., Ball, G. P., and Netessine, S. (2017). Plant operations and product recalls in the automotive industry: An empirical investigation. *Management Science*, 63(8):2439–2469.

Shannon, C. (1948). The mathematical theory of communication. *Bell System Technical Journal*, 27:379–423.

Shunko, M., Yunes, T., Fenu, G., Scheller-Wolf, A., Tardif, V., and Tayur, S. (2018). Product portfolio restructuring: Methodology and application at caterpillar. *Production and Operations Management*, 27(1):100–120.

Simon, H. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106(6):467–482.

Sniezek, J. (1989). An examination of group process in judgmental forecasting. *International Journal of Forecasting*, 5:171–178.

Sommer, S. C., Bendoly, E., and Kavadias, S. (2020). How do you search for the best alternative? experimental evidence on search strategies to solve complex problems. *Management Science*, 66(3):1395–1420.

Steiner, I. D. (1972). *Group Process and Productivity*. Academic Press.

Tang, C. (2010). A review of marketing-operations interface models: From co-existence to coordination and collaboration. *International Journal of Production Economics*, 125(1):22–40.

Wan, X. and Sanders, N. R. (2017). The negative impact of product variety: Forecast bias, inventory levels, and the role of vertical integration. *International Journal of Production Economics*, 186:123–131.

APPENDIX A: SUPPLEMENTAL MATERIALS

In Table 10 we report the results of fixed effects regressions where the dependent variable is either earnings or normalized earnings. As can be seen, the results are qualitatively similar to those reported in the main text based on simpler $t-$tests.

Table 10: Fixed Effects Regression of Earnings on Treatment Variables

Parameter	Earnings		Normalized Earnings	
2 Regions	-2.741^{***}	(0.971)	-0.112^{***}	(0.019)
4 Regions	-2.438^{**}	(0.958)	-0.169^{***}	(0.019)
Team Trial First	0.610	(2.707)	0.020	(0.054)
Trial #	2.888^{***}	(0.708)	0.056^{***}	(0.014)
Team Trial	1.718^{**}	(0.709)	0.034^{**}	(0.014)
Constant	39.940^{***}	(1.362)	0.887^{***}	(0.027)
R^2	0.245		0.490	
Observations	239		239	

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

In Table 11 we provide the underlying regressions that are used to generate the data reported in Table 1. Remember that, within a given region structure $i \in \{1,2,4\}$, there is only one observation per subject, making a simple regression statistically valid. For simplicity and ease of comparison, we report the same pooled regression.

Table 11: The Effect of Experience and Teamwork

(a) Absolute Earnings

Parameter	1 Region		2 Regions		4 Regions		Pooled	
2^{nd} Trial	1.378	(1.122)	0.062	(1.948)	11.187***	(2.425)	3.984***	(1.186)
Team	2.222	(1.397)	0.208	(2.615)	8.604***	(3.115)	3.512**	(1.530)
2^{nd} Trial×Team	−0.983	(1.920)	0.674	(3.657)	−7.786*	(4.312)	−2.384	(2.117)
Constant	41.198***	(0.753)	39.721***	(1.328)	31.632***	(1.735)	37.732***	(0.818)
R^2	0.084		0.003		0.324		0.091	
N	61		60		60		181	

(b) Normalized Earnings

Parameter	1 Region		2 Regions		4 Regions		Pooled	
2^{nd} Trial	3.050	(2.484)	0.130	(4.061)	21.537***	(4.668)	7.304***	(2.270)
Team	4.917	(3.093)	0.434	(5.452)	16.565***	(5.997)	6.791**	(3.510)
2^{nd} Trial×Team	−2.175	(4.249)	1.405	(7.625)	−14.991*	(8.302)	−3.745	(4.855)
Constant	91.186***	(1.667)	82.821***	(2.770)	60.900***	(3.341)	78.907***	(1.876)
R^2	0.084		0.003		0.324		0.067	
N	61		60		60		181	

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

APPENDIX B: SURVEY QUESTIONS AND SUPPLEMENTAL ANALYSIS

Here we document the main survey questions that we asked:

1. Reliability and trustworthiness of teammate (5 point Likert Scale; 1 = Strongly Disagree):
   (a) I could rely on my teammate.
   (b) Overall my teammate was trustworthy
2. Participation in the team (5 point Likert Scale; 1 = None at all; 5 = A great deal):
   (a) How much friction was there in your team?
   (b) How much were personality conflicts evident in your team?
   (c) How much tension was there in your team?
   (d) How much emotional conflict was there in your team?
   (e) How often did you and your teammate disagree about opinions regarding the task?
   (f) How frequently were there conflicts about ideas in your team?
   (g) How much conflict about your task was there in your team?
   (h) To what extent were there differences in opinion in your team?

3. When you participated in a team of 2, did a team leader emerge in your group?
   - Yes. I was the leader.
   - Yes. My teammate was the leader.
   - No. We participated as equals.
   - No. We had conflict about who should lead.
4. Own and teammate's contribution to the simulation ($[0,10]$, 1 decimal)
   (a) My own contribution to the team simulation.
   (b) My teammates contribution to the team simulation.
5. Team versus individual simulation (5 point Likert Scale; $1 =$ Strongly Disagree):
   (a) It was easier to achieve high performance in the team simulation.
   (b) The team came up with solutions that I did not think of working on my own.
   (c) Having to agree with my teammate on a plan made the task more difficult.
6. In comparison to the individual simulation, how difficult was the team simulation. For example, if you found the team simulation more difficult, then move the slider to the right. If you found the team simulation easier, then move the slide to the left. ($[-10,10]$, 1 decimal; $0 =$ individual and team same difficulty)

Survey Responses and Complexity

The survey consisted of six blocks of questions. We discuss four of these question blocks here and relegate two blocks to Appendix A.[21] Figures 5–8 show differences in the survey responses, for each the four main question blocks, broken up by number of regions.

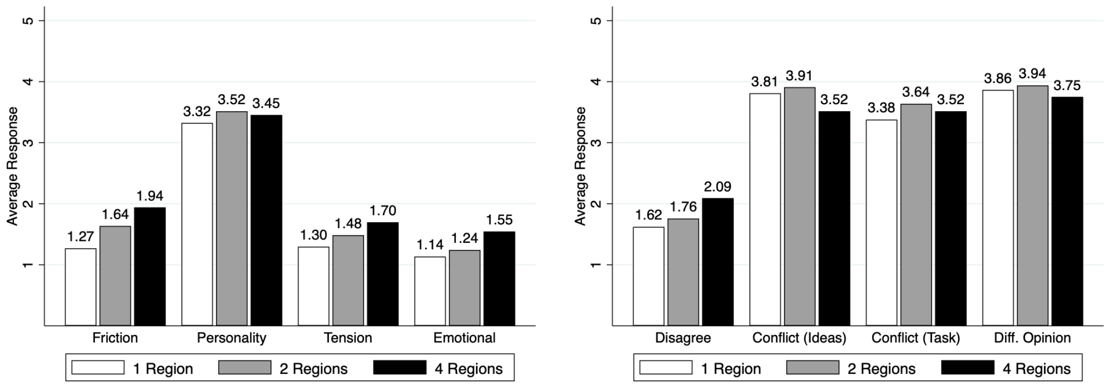
First consider Figure 5, which provides results for the question block concerning conflict in the team setting. As can be seen, some differences emerge depending on the number of regions. Specifically the questions involving friction ($p = 0.003$), tension ($p = 0.060$), emotional conflict ($p = 0.014$) and the frequency of disagreement ($p = 0.030$) were all significantly more frequent when the team managed four regions than one region.

In Figure 6 we report results where subjects were asked to rate their own contribution and the contribution of their teammate. Subjects rated their own performance quite highly and their self-evaluation appears to increase as the number of regions under management increases, but this is not significant ($p = 0.135$). For the subjective evaluation of one's teammate's contribution, there appears to be a non-linear relationship. In particular, subjects rate the teammate more highly when managing two versus one region ($p = 0.048$) and there is no difference between one region or four regions ($p = 0.518$). We also see that subjects' self-assessment of their own performance

[21] Information about the two relegated blocks are provided in Figures 9 and 10. The first concerns the trustworthiness and reliability of the teammate and the second concerning subjects' perceptions regarding whether the team came up with better solutions than the individual. For these two question blocks, there were no differences in survey responses by complexity (in all cases, $p > 0.2$).

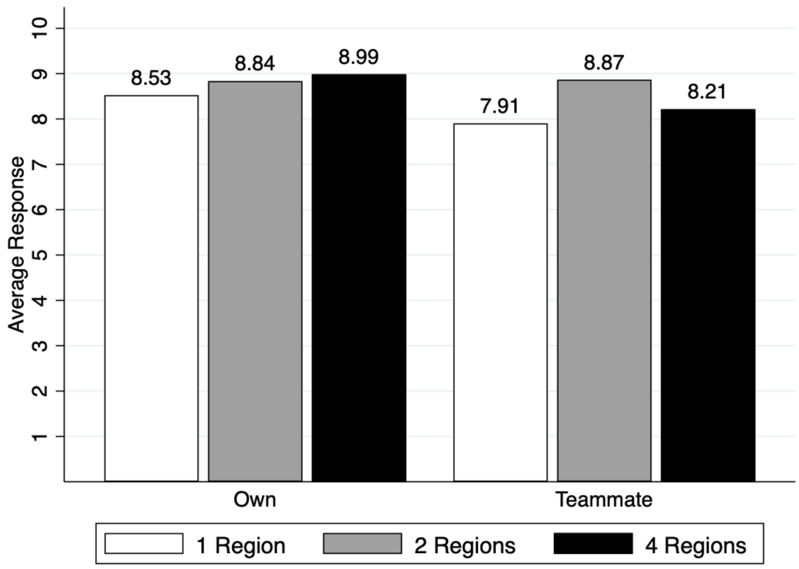
Figure 5: Conflict in Team Setting

(a) Questions 1–4 (b) Questions 5–8



is increasing in complexity, while the relationship between one's assessment of their teammate and complexity is non-monotonic. Subjects appear to rate their teammate's contribution as being relatively low in the least and most complex environments. It would be interesting to see if such patterns hold more broadly as they may hint at sources of tension in complex environments.

Figure 6: Rating Own and Teammate's Contribution



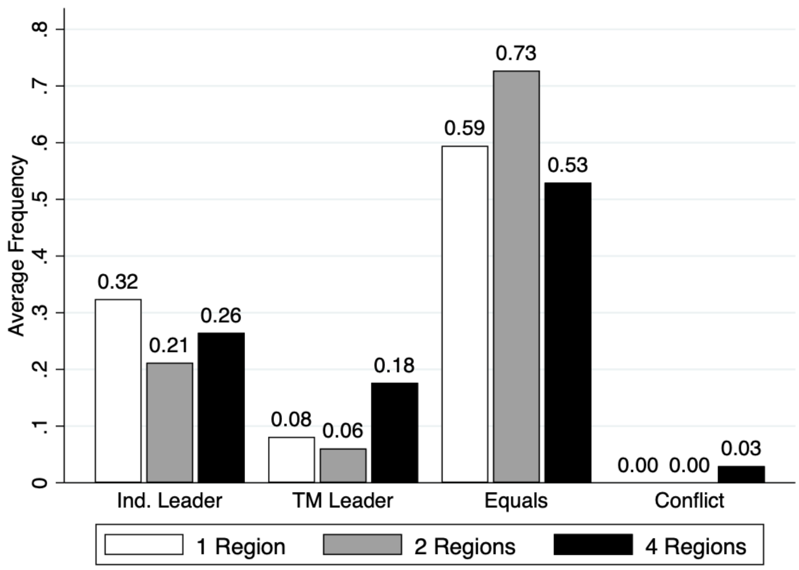
Another interesting finding is that people tend to rate their own contribution as significantly higher than their teammate's ($p = 0.014$), which suggests some over-confidence.[22] Moreover, a deeper inspection shows that they rate their performance higher than their teammate's only when the team performance was above the median. This

[22] It is, however, interesting to note that a subject's self-evaluation of their performance in the team trial is positively correlated (and marginally statistically significant) with their performance during the individual trial. In contrast, there is

suggests a self-attribution bias in that they attribute good team performance to their own contribution, but when team performance is poor, they rated their contribution as equal to their teammate.

Figure 7 sought to uncover the emergence of a team leader in the team trial. First, as can be seen, there does not appear to be a difference in leadership emergence by the number of regions and most of the time, subjects report that they participated as equals.[23] Consistent with our previous result that subjects rated their contribution higher than their teammates, the frequency that subjects reported that they were the team leader is significantly higher ($p = 0.006$) than the frequency that they reported that their teammate was the team leader. Given that they were also more likely to say that they were the leader when team performance was good, this further reinforces our belief that subjects attribute the success of the team to their leadership and that subjects may suffer from a self-attribution bias.

Figure 7: Did a Team Leader Emerge?



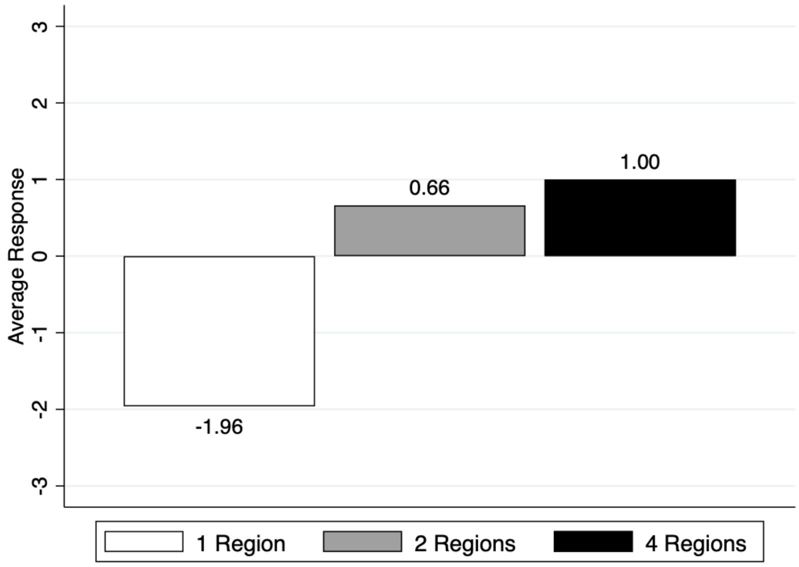
Finally, in Figure 8, we asked subjects to rate the relative difficulty of the team versus individual trials. High numbers indicated that the team trial was more difficult. As can be seen, there appear to be clear differences between number of regions, with subjects managing one region as a team finding the team trial to be easier (and this is significant at $p = 0.024$), while subjects who managed four regions as a team found the team trial to be more difficult (but this is not significant; $p > 0.2$).

Remark 1. *Before we conclude that team environments are more challenging the more complex the environment, we should note that, due to the limitations of our sample size, someone who participated in the team trial with*

virtually no relationship between the evaluation of one's teammate and their teammate's performance during the individual trial.

[23]The one marginally significant difference is that subjects reported being more likely to participate as equals when managing two versus four regions ($p = 0.070$).

Figure 8: Relative Ranking of Team Versus Individual (Positive: Team Harder)



four regions under management participated as an individual with either one or two regions under management. Therefore, since four regions is more difficult to manage than one or two, it should not be surprising that they found the team trial more difficult. Similarly, for someone who participated in the team trial with one region under management, they participated as an individual with either two ro four regions under management. Hence, it should not be surprising that they found the team trial easier.

Survey Responses and Performance

We now turn our attention to whether the survey responses are associated with performance in the team trial. We just showed that higher complexity leads to more friction, tension, emotional conflict and disagreements. It is interesting to see if these variables are also associated with worse performance as it would suggest an additional mechanism for why firms should be cautious to increase complexity. Specifically, in Table 12, we report the estimated coefficients from regressions of the form:

$$\begin{aligned}\texttt{Performance} = \beta_0 + \beta_1 2^{nd}\ \texttt{Trial} + \beta_2 \texttt{2 Regions} + \beta_3 \texttt{4 Regions} + \beta_4 2^{nd}\ \texttt{Trial} \times \texttt{2 Regions} \\ + \beta_5 2^{nd}\ \texttt{Trial} \times \texttt{4 Regions} + \beta_6 \texttt{Own Individual Performance} \\ + \beta_7 \texttt{Teammate's Individual Performance} + \vec{\beta}\,\overrightarrow{\texttt{Question}}\ \texttt{Block N} + \varepsilon\end{aligned}$$

The coefficients β_1–β_5 capture the main treatment interactions, while β_6 and β_7 are meant to capture the effect of ability (as measured by the subject's and his/her teammate's performance when they participated as individuals). Finally, the vector of coefficients, $\vec{\beta}$, captures the effect of each survey question from the given question block.

Columns (1)–(6) show results for the above regressions where we always keep the same treatment and individual measures but separately include the different questions blocks. As can be seen, in all specifications, the

Table 12: The Effect Survey Responses on Team Performance

(a) Block 1

Parameter	(1)		(2)		(3)		(4)	
2^{nd} Trial	4.923	(3.333)	1.500	(3.349)	4.871	(3.280)	4.377	(3.376)
2 Regions	-14.942^{***}	(3.464)	-16.855^{***}	(3.508)	-14.107^{***}	(3.379)	-15.536^{***}	(3.524)
4 Regions	-20.322^{***}	(3.328)	-21.698^{***}	(3.363)	-20.361^{***}	(3.273)	-20.856^{***}	(3.459)
2^{nd} Trial $\times$ 2 Regions	1.155	(4.736)	4.584	(4.733)	-0.371	(4.661)	1.673	(4.764)
2^{nd} Trial $\times$ 4 Regions	2.775	(4.665)	6.227	(4.955)	3.185	(4.594)	3.348	(4.775)
Own Ind. Perf.	0.180^{**}	(0.070)	0.208^{***}	(0.069)	0.232^{***}	(0.071)	0.187^{**}	(0.072)
TM Ind. Perf.	0.147^{**}	(0.065)	0.119^{*}	(0.065)	0.114^{*}	(0.065)	0.144^{**}	(0.069)
Rely on Teammate	1.720	(2.457)						
Teammate Trustworthy	-0.156	(2.633)						
Team Friction			1.678	(1.528)				
Personality Conflict			2.333	(1.524)				
Team Tension			0.684	(1.671)				
Emotional Conflict			1.109	(2.165)				
Freq. Disagreements			-5.560^{***}	(1.826)				
Freq. Conflict (Ideas)			-0.393	(1.309)				
Freq. Conflict (Task)			-1.148	(1.429)				
Diff. Opinion			-0.320	(1.392)				
Easier Team					2.201^{**}	(0.902)		
New Ideas Team					-0.477	(0.897)		
Need to Agree Made Difficult					-0.119	(0.800)		
Participated as Equals							1.254	(9.868)
I was the leader							0.420	(10.077)
Teammate was the leader							-0.118	(10.239)
Constant	61.670^{***}	(8.383)	73.710^{***}	(8.724)	61.135^{***}	(7.764)	67.457^{***}	(12.283)
R^2	0.477		0.545		0.494		0.460	
N	98		95		99		98	

(b) Block 2

Parameter	(5)		(6)		(7)	
2^{nd} Trial	4.364	(3.403)	4.533	(3.326)	3.476	(3.313)
2 Regions	-15.601^{***}	(3.509)	-12.635^{***}	(3.564)	-13.018^{***}	(3.514)
4 Regions	-21.219^{***}	(3.386)	-20.216^{***}	(3.443)	-20.102^{***}	(3.391)
2^{nd} Trial $\times$ 2 Regions	1.552	(4.771)	0.496	(4.812)	1.607	(4.767)
2^{nd} Trial $\times$ 4 Regions	3.712	(4.778)	3.765	(4.830)	5.728	(4.895)
Own Ind. Perf.	0.175^{**}	(0.071)	0.225^{***}	(0.071)	0.217^{***}	(0.070)
TM Ind. Perf.	0.151^{**}	(0.066)	0.130^{*}	(0.068)	0.122^{*}	(0.067)
Subjective Own Contrib.	0.513	(0.807)				
Subjective TM Contrib.	0.208	(0.545)				
Relative Difficulty of Team			-0.711^{***}	(0.209)	-0.600^{***}	(0.212)
Freq. Disagreements					-2.425^{**}	(1.152)
Constant	62.714^{***}	(8.904)	65.029^{***}	(6.946)	70.963^{***}	(7.397)
R^2	0.463		0.523		0.548	
N	97		87		86	

Note: Standard errors in parentheses. Significance given by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.10$.

Region indicator variables are all negative and significant. While the coefficients on 2^{nd} `Trial` are not individually significant, if we consider the marginal effect of having a second trial, it is always positive and significant. We also see that the individual performance measures do capture something like ability. Those subjects who earned more as individuals also earned more in the team trial.

As can be seen, very few of the survey questions significantly impact performance. In Column (2), we see that subjects who reported more frequent disagreement had significantly lower earnings. However, none of the other variables – let alone friction, tension and emotional conflict, which were all shown to increase with complexity – have a significant impact on performance.[24] In Column (3), people who reported that "It was easier to achieve high performance in the team simulation" had significantly higher earnings, while in Column (6), subjects who rated the team simulation as relatively more difficult had significantly lower earnings.

Finally, in Column (7), we report a regression where we include only the frequency of disagreement and the relative difficulty of the team simulation questions.[25] As can be seen, these two variables maintain their sign and significance, though the coefficient on frequency of disagreements in reduced by about half.

Other Survey Results

In this section, we provide Figures 9 and 10, which show the relationship between survey responses and number of regions for the question blocks on trustworthiness/reliabilty of one's teammate and whether the team came up with better solutions than when participating as an individual. As noted in the main text, responses are not influenced by the number of regions under management.

[24] Indeed, the coefficients on team friction, personality conflict, team tension and emotional conflict are all individually positive (but not significant). However, if we consider the sum of these four variables then the effect is actually positive and significant. This suggests the counterintuitive result that more conflict/tension might actually be good for teams, though outright disagreement appears to be bad.

[25] We do not include the significant coefficient from Column (3) because it is highly correlated with the relative difficulty variable.

Figure 9: Trustworthiness and Reliability of Teammate

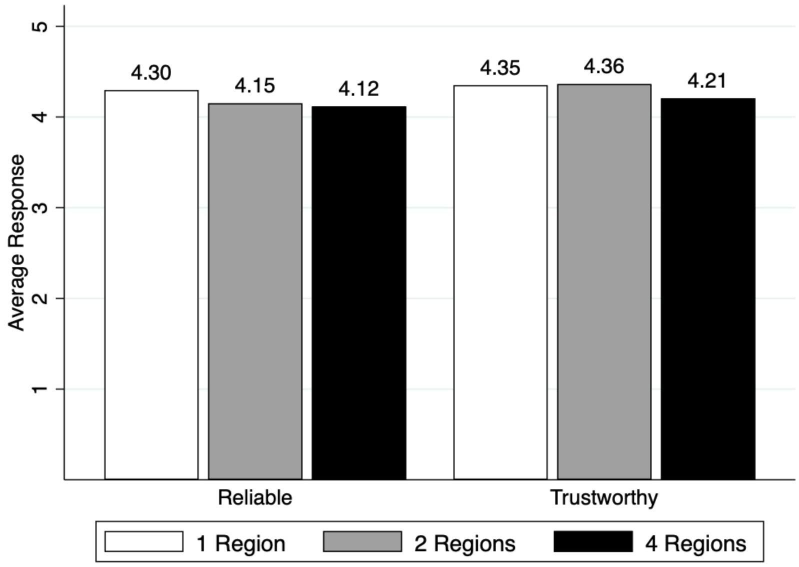


Figure 10: Team Solutions