Measuring the Impact of Complexity on Operational Performance: Complexity Does Matter!

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

[1]*Kedge Business School, 680 Cours de la Libération, Talence, France*

[1]*Supply Chain & Complexity Management Lab., Supply Chain Center of Excellence (CESIT) @ Kedge-Bordeaux.*

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

mozart.menezes@kedgebs.com, kyleb.hyndman@utdallas.edu

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

Abstract: **Problem Definition** 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 chance to increase profits by increasing the number of markets served, can managers properly execute or does increased complexity become a burden?

Academic/Practical Relevance Managers are often pressed to achieve sustained profit growth, which often means increasing the number of SKUs offered or markets served. Although such efforts will, almost by definition, increase revenue, they also increase complexity, which may lead to a decay in profits. This tension between complexity and opportunity, and the factors that can dampen the adverse effects of complexity, such as experience and teamwork, merits rigorous study.

Methodology 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.

Results We show that as complexity increases, average performance often deteriorates and many subjects destroy value, despite the increased opportunities for profit. Both teamwork and experience increase performance and reduce the variance of earnings. Experienced teams make better investment decisions, while more inexperienced subjects appear focused on revenue, rather than earnings.

Managerial Implications Our results highlight the pitfalls of expanding product or market offerings, which simultaneously increase profit opportunities and complexity. Teamwork and experience partially mitigate the negative effect of complexity by improving investment decisions. When considering options to expand the portfolio of a business unit, we show that managers should take pause and develop a plan that explicitly acknowledges complexity and has solutions for how to maintain execution quality.

1 INTRODUCTION

In this paper we seek to shed light on losses of efficiency due to an increase in the combinations of stock keeping units (SKUs) and/or markets – which we will refer to generically as complexity. Some reasons for losing efficiency are well known and studied but not all of them are well understood. For example, production may need a machine setup every time a new order is produced resulting in loss of productive time and probably also loss of raw material while calibrating the machine for the new product. Hence, the more different products exist, the worse these losses will become (see, e.g., Hopp 2007, Cachon and Terwiesch 2013). The proliferation of markets and SKUs may also impact batch sizes; although large sizes dilute the setup time and cost, they increase costs of holding inventory hurting financial performance. All these losses, in the end, could be categorised as resulting from interactions of different products and markets using limited resources.

Unfortunately, we don't have a good understanding of the effects of "proliferation" on the quality of decision making, and therefore, on profitability. Our assumption is that when proliferation is not forced but an option, then it should be an advantage since it increases the solution space without any additional cost for that increase. Thus, we want to verify the existence of performance losses, beyond those induced by limitations on resources availability, due to non-forced proliferation upon managers. That is, is it possible that managers, when offered more opportunities to increase profit, make worse decisions and lose money instead? Do managers see the possibility of increasing the number of markets as an obligation rather than an option when the goal is to grow profits? Is an excess of a good thing, as discussed in Barnett and Freeman (2001), bad as some may claim?

In an industry-wide study, Barnett and Freeman (2001) conduct an empirical study of the effects of new product introduction on the likelihood of survival of U.S. semiconductor manufacturers. They find that under some conditions proliferation can increase substantially the chances of firms' failure; i.e., departure from the market. Hence, proliferation may lead to, or increase the chance for, bankruptcy. Bayus and Putsis (1999) study the personal computer industry and they 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 seems to avoid SKU proliferation beyond a certain point (Saeed and Young 1998).

Firms' operations are designed for maximum efficiency in the use of resources. However, marketing needs push for an increase on the number of SKUs, markets, and channels, which although increase revenues, may reduce the efficient use of resources. Assuming a profit maximising firm, those efficiency losses are not a problem if the balance between increased revenues and efficiency losses result in higher profits. Otherwise, the proliferation of SKUs, markets, and channels, is said to have gone too far as commonly mentioned in the literature. For example, Carrefour, a French supermarket holding, announced in June, 2018, a strategic shift from very large business units to smaller shops which are "less complex and more adapted to customers' behaviour". Lego, a Danish toymaker, announced in September 2017 major reduction in headcount, alleging bureaucracy and complexity as the root cause of their problems (see Milne (2017)). Several papers have been focusing on the issue of product proliferation (Adams et al. 2016, Mocker and Ross 2017, Fisher et al. 2017). The issue of product proliferation is particularly relevant because many practitioners cite it as an important factor that causes major losses in the industry (George Group 2006, Mariotti 2008).

Shifting from the analysis of an industry or from case studies based on logical reasoning limited data is difficult because the impact of complexity on performance at the firm level is not easy to measure. There are two main reasons for the difficulty in measuring that impact: the first reason is technical, since there are too many confounding variables that affect a firm's results; and second is related to business needs for confidentiality of firms' information, which refrain from breaking down some financial key performance indicators (KPIs) to the right level of granularity.

Ruiz-Hernández et al. (2018) propose a measure for complexity using Shannon's measure of information content – see Shannon (1948) – in a particular way that helps with the technical reason by allowing one to measure complexity in a holistic way. They, and henceforth we, call the measure as pars-Complexity (C_p). A *pars* is a unit representing the triple $\{$SKU, channel, market$\}$; each combination of the triple define a *pars* which is the basic set of information the company is organised around. The movement of *partes* (plural of *pars*) generates, in its turn, more information which reaches managers through the supply chain information flow; see Figure 1. The unit of this measure is *bit*.

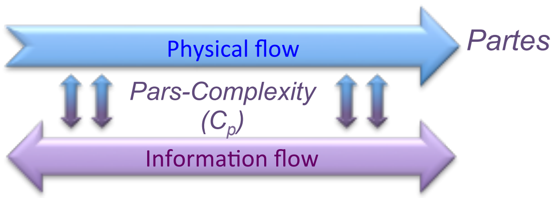


Figure 1: Schematic representation of information originating from delivering of *partes* (singular *pars*), magnified by subsequent managerial needs to keep the physical flow moving and eventually being incorporated to the information flow becoming basic input for management decisions. The quantity of information needed to describe the set of *partes* is called *pars*-Complexity and it is represented by the symbol C_p.

Menezes and Ruiz-Hernández (2018), working with 27 business units of a multi-billion dollar company with global footprint and using the measure proposed by Ruiz-Hernández et al. (2018), could verify that for each *bit* increase of *pars*-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 *bit* of increased structural complexity, losses would amount to 50% of average operating profits.

Menezes and Ruiz-Hernández (2018) obtain the results analysing the 27 business units of the focal firm within operating within the same economic-area, operating under the same rules and currency, serving the same economic area in which operations are located, and using the same technology with minor variation of products. The main differentiator between business units are the orders' profile coming from the local markets and the amount of markets served. We note that several business units are operating with losses. The losses seen imply that, in that particular firm, they **probably already have too much** of a good thing. In their annual meeting with investors on December, 2017, the firm has pledged to cut 35% of product variety and reduce capacity by nearly 30%.[1]

[1]Citation withheld to preserve confidentiality of the firm.

One of the authors of this paper is also working with a Fortune 500 company. In that work the author analyse six different product lines and four out of the six products are producing beyond the amount that would lead to optimal profits.

All these results suggest that indeed there is a crisis of proliferation as Mariotti (2008) says, and asks us for trying to understand better the role of (poor) decision making under complexity on financial results. In this paper, we take the view that if we have a perfect system where all decisions made are implemented without failure and all resources work as they should without ever malfunctioning or ill-performing, then in that perfect system, any variation of profits would be due to the decision making process. If in that perfect system the decision maker sells to market A then, when offered to sell to market A, or market B, or both, profits should only increase. If moreover, markets A, B, C, or D, or any combination of them is offered then we would expect an even higher profit than when the options were only two markets. Let $\Pi(\bullet)$ be the optimal profit obtained when markets $\bullet$ are offered. Then, it is straightforward to see that $\Pi(\{A,B,C,D\}) \geq \Pi(\{A,B\}) \geq \Pi(\{A\})$. Any departure from that relationship must result from poor decision making choices.

A simulated supply chain can create such perfect world. We conduct a human-subjects experiment designed to examine the role of complexity on operational performance and also to determine how complexity interacts with experience as well as the size of the decision making unit (i.e., individual versus teams). 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. Subjects are managers of a fictitious firm which operates in either 1, 2 or 4 regions depending on complexity. That is, we can offer subjects different worlds where each world will have markets in $\{A,B,C,D\}$, $\{A,B\}$, and $\{A\}$) to explore.

We present in this paper the results of a human-subjects experiment in which we vary the number of markets subjects can serve, which lead to more potential profit opportunities, but at the cost of additional complexity. In particular, we assume that information is a key input for any managerial decision making processes and that the more information content in the decision environment, the more sophisticated are the needs for properly executing in that environment. We measure the amount of information induced by the fundamental drivers of supply chain complexity (using Ruiz-Hernández et al. (2018)) and, through a computer-based business simulation exercise, compare the performance of managerial teams submitted to different levels of complexity. We then test the performance for different combinations of complexity level and sizes of management teams.

In our 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 and not made any decisions. The final conclusion suggests that offering more possibilities hurt profitability.

The negative effects of complexity were mitigated with both teamwork and experience so that, in most cases, subjects were not destroying value but they were still failing to capture the additional profit due to higher complexity. Another result of increased complexity is also increased variability of earnings. As with average earnings, variability of earnings are also reduced with experience and teamwork, primarily by reducing the likelihood of 'big' mistakes. Additionally, we also show that teamwork and experience are complementary, such that investment in fixed assets actually become a positive driver of performance, rather than a drag on performance when working individually or without sufficient experience.

Although both experience and teamwork improve earnings, we still document a "complexity gap", which is the difference in potential profits and realized profits after accounting for experience and teamwork. In our least complex environment, this gap is only about 3%, but jumps to more than 15% in our other two, more complex, settings.

In an attempt to understand how teams functioned, we also conducted an extensive survey regarding subjects' perceptions of their teammate and working in a team. Our results show that team frictions, tension, emotional conflict and the frequency of disagreement were all increasing in complexity. Although teams which reported more friction, tension or emotional conflict did not perform worse, we find that teams with more frequent disagreement (which was more frequent in more complex settings) did perform significantly worse. Regarding the survey, we also find evidence for apparent attribution bias in team performance. Specifically, players rated their contribution to the team significantly higher than that of their teammate and, moreover, the difference was largest when performance was high. Similarly, subjects were more likely to report that they emerged as a team leader and, the effect was particularly strong when team performance was high.

The overall message of the paper is that operational complexity is something that firms should take into consideration. Although both teamwork 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 active measures to ensure that it does not worsen performance.

2 MEASURING COMPLEXITY

This work is developed through analysis and base the calculations using Shannons measure of information Shannon (1948) which has a complete theory around it from which we can borrow for our work on complexity. Recall that our idea is to focus on each combination of SKU, market, and channel is called a pars. Let p_i to be the fraction of potential sales (it could be potential profit as well) induced by the i^{th} pars among a set of n partes. Then the following expression, due to Shannon (1948), defines the amount of information of the supply chain which is the complexity measure proposed.

$$\sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i} \tag{1}$$

The advantage of the approach of measuring pars is the focus on what has been often claimed by practitioners and suggested by academics to be a major source of all evils, if not the most important, in supply chain complexity de Leeuw et al. (2013), Mariotti (2008), Manuj and Sahin (2011), Milgate (2001), Vachon and Klassen

(2002). Using Shannons measure of information adds properties (e.g., additivity) that allows for an stable result independent of the scope chosen for the analysis. It also allows for analyzing the complexity across borders of supply chain partners.

In our context, we sell a same product in one, two or four regions. Setting demand for each product to be the same, then when there is one region the value of $C_p = 0$. Similar potential demands for each of two regions would induce a value of $C_p = 1$, and when four regions are managed the $C_p = 2$.

3 LITERATURE REVIEW

3.1 Team Decision Making

To the best of our knowledge, there is very little work in operations management on team decision-making, with Li et al. (2018) being the lone exception.[2] 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, while 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, while in their strategic environment they consider a forecast information sharing setting, where the sender has an incentive to inflate their forecast (and so the receiver should ignore the message). In contrast to the literature from behavioral economics, 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.

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 and, so it is not clear that teams will perform better than individuals.

Englmaier et al. (2018) is another interesting paper, which focuses on the role of incentives in non-routine,

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

analytical team tasks. Although not directly related to our study, because we do not manipulate the incentive scheme as a treatment variable, it is important nonetheless because it shows that strong incentives do encourage better team performance across subject pools with different characteristics.

3.2 Complexity in Operations Management

There is a similar dearth of research on how complexity affects operational performance. Early work by Rivkin (2000) shows the difficulty – because of both the number and inter-connectedness of decisions needed to implement any given plan of action – of using heuristics or learning-based approaches to successfully imitate complex strategies. This is clearly related to our work because, as we expand the number of regions under management, we increase both the number of decisions that need to be made and the possibility inter-dependencies of decisions.

In terms of experimental research, though very 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 contracts even though they are theoretically sub-optimal. Cui et al. (2018) argue that this preference may actually be optimal/rational 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 or ordering for 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.

4 THE EXPERIMENT

In this set of experiments we use the Supply Chain simulation game (SCSG), created by Chopra and Afeche (2016), in which subjects manage a fictitious firm's supply chain, Jacob Industries, consisting of between one and five regions. The headquarters is based in a region called Calopeia which already has an open production and a storage facility. For the other regions, subjects must decide whether or not to open such facilities in that region. In addition, with respect to all regions including Calopeia, subjects need to make decisions related to purchasing production capacity, set inventory control parameters, and making choices for the transportation mode.

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". Students report their initial and final cash position, sales, lost sales, production, inventory, and transportation costs, and report the amount invested in fix assets, and inventory write-offs at the end of the game. That information allow us to calculate performance measures that most interest us. The KPIs are net income, operating margins, return on assets, and return on investment but we will focus herein on total cash at the end of the game which, in this particular game, is equivalent to operating profits.

Subjects consist of approximately 120 students from two different Specialized Master classes and one Master of Science class. Subjects play the SCSG three times. In the first trial, which was common to all subjects and used only to familiarize subjects with both the software and the nature of their decision problem, subjects manage a single region. 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 $3 \times 2 \times 2$ experimental design. In one dimension we vary the supply chain complexity using values $X \in \{0.0, 1.0, 2.0\}$. The different level of complexities are obtained by increasing the number of markets subjects need to deal with from 1 to 2 to 4. The second dimension are management team size which can be either a single subject or a team with two subjects (i.e., $Y \in \{1, 2\}$). Finally, we account for learning by having subjects participate twice – once as a team of two and once as an individual.

Subjects are assigned to teams randomly where the number of teams for each value Y is the same (keeping one-third of subjects in a team of one and two-thirds in pairs). We keep a similar demand pattern to all subjects; that is, all demand streams have similar seasonality profile and same mean daily demand, and standard deviation of daily demand from a same demand distribution. Demand peaks are different in intensity. Note also that some subjects participated first as teams and some participated first as individuals. Additionally, some subjects saw complexity increase in the second trial while others saw complexity decrease. Importantly, subjects who were part of a team in the second trial had the same experience in the first trial. That is, if a team was managing x regions in the second trial, then both subjects in the team managed y regions in the first trial, where $y \neq x$. It is also important to note that a team playing two regions would see one of the regions' demand exactly as seen by those playing a single region; and those playing four regions would see two of the regions with exactly the same demand pattern as seen by those playing two regions.

4.1 Incentives

Subjects are rewarded based on performance with marks counting toward the course overall grade. In total 50% of the course marks are related to the experiment that is part of a supply chain management and design course. Of the 50% of the marks, half of the points are given based on a report students wrote explaining their performance and suggesting, in hindsight, corrections for decisions made during the simulation. The other half of the points are based on the quality of their strategies and the deployment of that strategy; points are given based on a linear combination of the profits obtained using a given 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) - the points for those two extreme points are 100 and 0 respectively. 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.

5 Results

We begin our analysis of the experimental data by providing summary statistics in Table 1 on earnings performance differentiating between our key treatment variables: Number of regions (i.e., complexity), trial number and whether subjects participated as individuals or as pairs. For a frame of reference, note that if subjects did nothing, then they would have earned 41.47 across all conditions. As can be seen, a substantial number of subjects actually destroyed value through their actions, and this is particularly pronounced in initial trials and when operating as an individual.

Table 1: Summary Statistics: Earnings By Regions, Trial and Group Size

(a) Group Size = 1				(b) Group Size = 2				(c) Overall Average			
	Trial				Trial				Trial		
Reg.	1	2	Tot.	Reg.	1	2	Tot.	Reg.	1	2	Tot.
1	41.20	42.58	41.82	1	43.42	43.81	43.65	1	41.84	43.07	42.45
	(5.00)	(2.58)	(4.10)		(2.37)	(1.61)	(1.93)		(4.47)	(2.30)	(3.60)
2	39.72	39.78	39.75	2	39.93	40.67	40.32	2	39.77	40.06	39.91
	(5.89)	(6.68)	(6.19)		(2.64)	(8.75)	(6.44)		(5.12)	(7.23)	(6.16)
4	31.63	42.82	37.36	4	40.24	43.64	42.03	4	34.30	43.08	38.84
	(9.93)	(7.02)	(10.17)		(4.95)	(5.94)	(5.62)		(9.51)	(6.60)	(9.20)
Tot.	37.73	41.72	39.63	Tot.	41.24	42.84	42.11	Tot.	38.74	42.10	40.41
	(8.16)	(5.96)	(7.44)		(3.77)	(5.83)	(5.02)		(7.30)	(5.90)	(6.84)

Note: Standard deviations in parentheses below.

At a high level, the results show that absolute performance is statistically better in the second trial than in the first trial (42.10 vs. 38.74; $p \ll 0.01$) and absolute performance is significantly better as part of a team than as an individual (42.11 vs. 39.63; $p = 0.023$). We also see that absolute performance is significantly worse when managing two regions than one region (39.91 vs. 42.45; $p = 0.006$) and four regions than one region (38.84 vs. 42.45; $p = 0.005$) but the difference between two and four regions is not statistically different ($p = 0.454$). If we use the previously defined notation, with simple names for each region, then we expect that $\Pi(\{A,B,C,D\}) \geq \Pi(\{A,B\}) \geq \Pi(\{A\})$. Let the subjects' expected profits be represented by $\pi(\bullet)$. As can be seen in Table 1, we find overall that $\pi(\{A,B,C,D\}) \leq \pi(\{A,B\}) < \pi(\{A\})$, which shows that subjects fail to exploit the increased profit opportunities.

Importantly, observe that Table 1 suggests that these mainline effects do not necessarily hold at finer levels of comparison. In particular, the effect of being part of a team (rather than an individual) is not significant when managing two regions ($p \gg 0.1$). Even more starkly, the learning effect is only significant when managing four regions. However, this is not surprising: the benefits of repeating a task or of working in a team should be most pronounced in more complex environments. 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 Table 1 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 ($p = 0.001$) and for the second trial than the first trial ($p = 0.046$). Thus repetition and working as a team reduce the variance of earnings, largely through reducing the likelihood of big mistakes.[3] 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, experience and teamwork both significantly reduce the variability of earnings.*

As noted, while subjects could always guarantee earnings of 41.47 by doing nothing, regardless of the number of regions managed, by increasing the number of regions, these subjects had a larger feasible set of earnings and a higher upper bound of possible earnings. While it is difficult to provide a precise upper bound based on "optimal" behavior, our naive heuristic[4] suggest that reasonable upper bounds for earnings were 45.18, 47.96 and 51.94 for 1, 2 and 4 regions, respectively.

Figure 2 provides another perspective on earnings and earnings variability. 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.

Table 2 replicates Table 1 but reports earnings as a fraction of the proposed upper bounds for each number of regions.

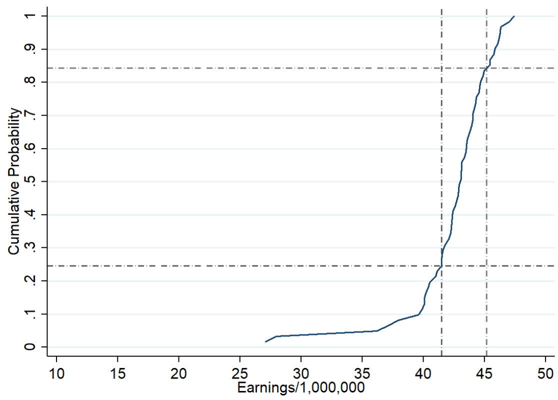
Looked at this way, we can see more starkly the negative effects of increasing complexity. For example, if we consider the first trial and a group size of 1, with one region subjects achieve 91.2% of the upper bound earnings, while this decreases to 82.8% for two regions and further decreases to 60.9% with four regions. Performance

[3] This follows because are data are generally negatively skewed.

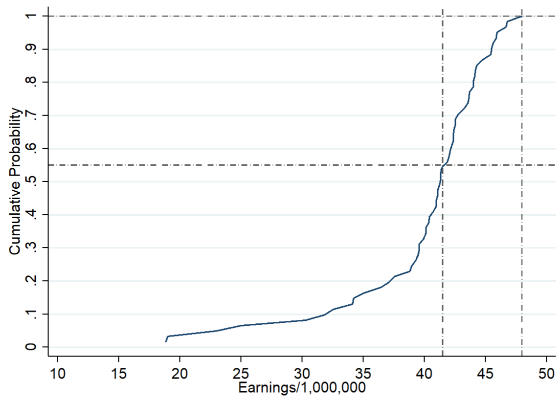
[4] The naive heuristic goes as follows: we assume demand is stable with annual demand equal to the average of previous years and coefficient of variation of $1/4$. We apply the newsvendor model for estimating capacity and the economic order quantity (EOQ) model for estimating production lot-size. The 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 of the "warm-up" game.

Figure 2: Empirical CDF of Earnings By Regions

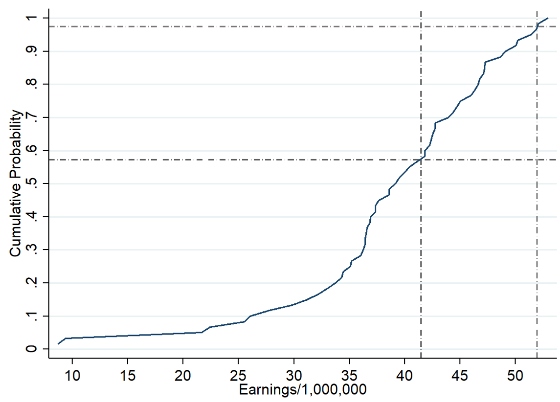
(a) 1 Region



(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.

increases with repetition and with teamwork such that on the second trial a pair working together to manage four regions achieves 84% of the upper bound earnings.

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.*

5.1 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:

$$\texttt{Metric}_i = \beta_0 + \beta_1 \texttt{Experience}_i + \beta_2 \texttt{Teamwork}_i + \beta_3 \texttt{Experience}_i \times \texttt{Teamwork}_i + \varepsilon_i, \quad (2)$$

where $\texttt{Metric}_i$ is either absolute earnings or earnings normalized by the best heuristic. We estimate this equation separately for each number of regions and also pooling over all regions and we use the estimated marginal

Table 2: Summary Statistics: Normalized Earnings By Regions, Trial and Group Size

(a) Group Size = 1

	Trial		
Reg.	1	2	Tot.
1	91.19	94.24	92.56
	(11.06)	(5.72)	(9.08)
2	82.82	82.95	82.88
	(12.29)	(13.92)	(12.91)
4	60.90	82.44	71.93
	(19.12)	(13.51)	(10.59)
Tot.	78.91	86.21	82.38
	(18.95)	(12.84)	(16.67)

(b) Group Size = 2

	Trial		
Reg.	1	2	Tot.
1	96.10	96.98	96.60
	(5.25)	(3.56)	(4.27)
2	83.26	84.79	84.07
	(5.51)	(18.24)	(13.43)
4	77.46	84.01	80.91
	(9.54)	(11.44)	(10.82)
Tot.	85.70	89.26	87.63
	(10.57)	(13.10)	(12.04)

(c) Overall Average

	Trial		
Reg.	1	2	Tot.
1	92.61	95.33	93.95
	(9.90)	(5.08)	(7.96)
2	82.92	83.52	83.21
	(10.68)	(15.08)	(12.85)
4	66.04	82.94	74.77
	(18.30)	(12.71)	(17.71)
Tot.	80.87	87.26	84.03
	(17.11)	(12.94)	(15.48)

Note 1: Maximum possible value in each cell is 100.

Note 2: Standard deviations in parentheses below.

effects to measure the performance gain due to experience and teamwork. 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 3. Using the estimated marginal effects, we see the proportion of the performance gain due to experience and 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 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 3. 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

Table 3: 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 Heuristic	Performance Gain			Complexity
Regions	Group Size = 1	Group Size = 2	(Norm.)	$(2)-(1)$	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.

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

5.2 Looking at Investment

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 4, 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 (even more so) teamwork reduce variability – here of investment.

Of course, these summary statistics on investment do not tell us anything about the relationship between

Table 4: 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.

investment and earnings, which we now turn to. In Table 5 we report regression results where the dependent variable is either earnings or normalized earnings. As explanatory variables we include investment and the number of regions. Since we are interested in the role of experience and teamwork, we also interact investment with indicator variables for the second trial and also for working as a team. 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. Consistent with the results presented earlier, subjects earn significantly less the more regions under operation.

More interestingly, the baseline effect (i.e., first trial, as an individual) of investment on earnings is significantly negative. 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.[5] In contrast, the coefficients on both teamwork and experience interacted with investment are significantly positive and larger in magnitude than the baseline negative effect. We also see evidence that, via that experience and 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. Therefore, while experience alone or teamwork alone is not enough to make increased investment a positive driver of earnings, the two factors combined do.[6]

[5] 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, we are not able to observe when investments were made. We conjecture that the heterogeneity in investment timing masks what would be an otherwise positive relationship between investment and earnings for small investments. Indeed, as the columns with a team interaction show, investment and earnings are positively related, up to a point, for teams, but not for individuals.

[6] 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.

Table 5: The Effect of Investment on Earnings

Parameter	Earnings		Norm. Earnings	
2 Regions	-2.346^{*}	(1.273)	-10.448^{***}	(2.530)
4 Regions	-4.423^{***}	(1.500)	-21.092^{***}	(2.982)
Investment	-0.347^{**}	(0.142)	-0.676^{**}	(0.283)
Inv.$\times\mathbf{1}$[Inv. >20 & 4 Reg.]	-0.218^{**}	(0.107)	-0.419^{*}	(0.213)
Inv.$\times\mathbf{1}$[Inv. >10 & 2 Reg.]	-0.412^{***}	(0.155)	-0.874^{***}	(0.307)
Inv.$\times\mathbf{1}$[Inv. >5 & 1 Reg.]	-0.324	(0.216)	-0.749^{*}	(0.430)
2^{nd} Trial $\times$ Investment	0.575^{***}	(0.097)	1.117^{***}	(0.194)
Team Trial $\times$ Investment	0.498^{***}	(0.143)	0.983^{***}	(0.285)
Team Trial $\times$ 2^{nd} Trial $\times$ Investment	-0.330^{*}	(0.189)	-0.647^{*}	(0.377)
Constant	43.226^{***}	(0.895)	95.765^{***}	(1.779)
R^2	0.391		0.538	
N	162		162	

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

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 and, when combined, increased investment becomes a significantly positive driver of earnings.*

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. In particular, as we show in the appendix (Table 9), investment is strongly, positively correlated with revenues. However, it also comes with the *opportunity cost* of foregone cash – which reduces interest earnings on cash balances. Additionally, investments in capacity also create additional hidden costs due to complexity because, with more levers of control, execution becomes more difficult. Therefore, balancing the gain in revenue due to an additional investment against these (less visible) costs, we obtain the result that investment is not a driver of earnings.

6 SURVEY RESPONSES, 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. We are interested in two things. First, do survey responses differ based on the number of regions that subjects managed when participating as a team? For example, were disagreements more difficult when managing four regions than only one? Second, do survey responses explain performance? For example, did teams with more perceived conflict perform worse or better? The precise questions are provided in Appendix A.

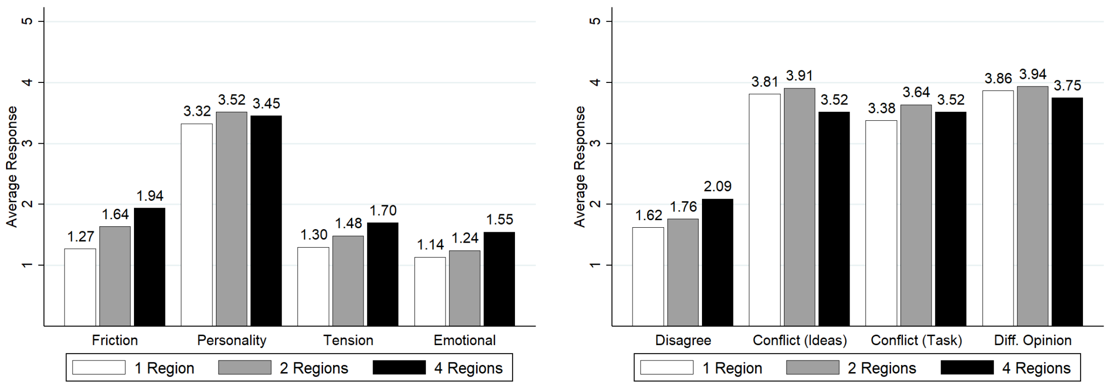
6.1 Survey Responses and Complexity

The survey consisted of six blocks of questions. We discuss four of these question blocks here and relegate two blocks to Appendix A.[7] Figures 3–6 show differences in the survey responses, for each the four main question blocks, broken up by number of regions.

First consider Figure 3, 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.

Figure 3: Conflict in Team Setting

(a) Questions 1–4 (b) Questions 5–8

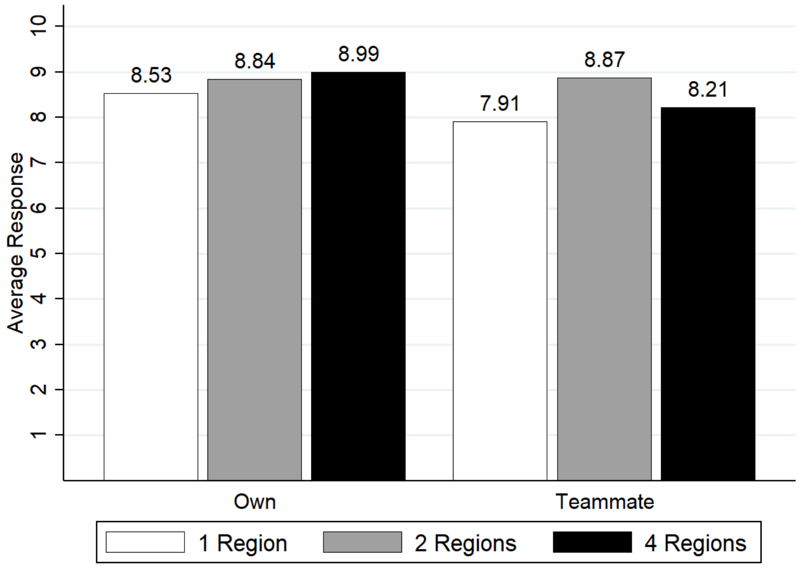


In Figure 4 we report results where subjects were asked to rate their own contribution and the contribution of their teammate. Subjects rated their own performance quite highly and their self-evaluation appears to increase as the number of regions under management increases, but this is not significant ($p = 0.135$). 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 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.

Another interesting finding is that people tend to rate their own contribution as significantly higher than their

[7] Information about the two relegated blocks are provided in Figures 7 and 8. 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 4: Rating Own and Teammate's Contribution



teammate's ($p = 0.014$), which suggests some over-confidence.[8] 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 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.

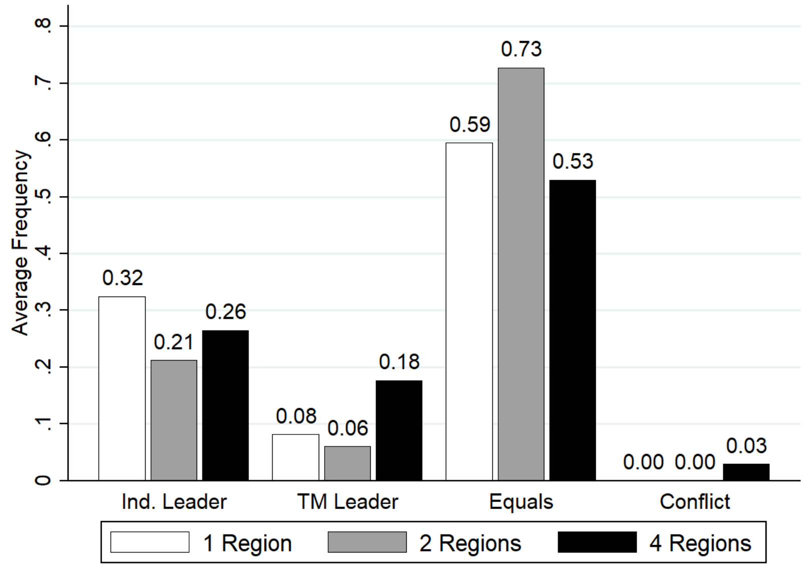
Figure 5 sought to uncover the emergence of a team leader in the team trial. First, as can be seen, there does not appear to be a difference in leadership emergence by the number of regions and most of the time, subjects report that they participated as equals.[9] Consistent with our previous result 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.

Finally, in Figure 6, 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

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

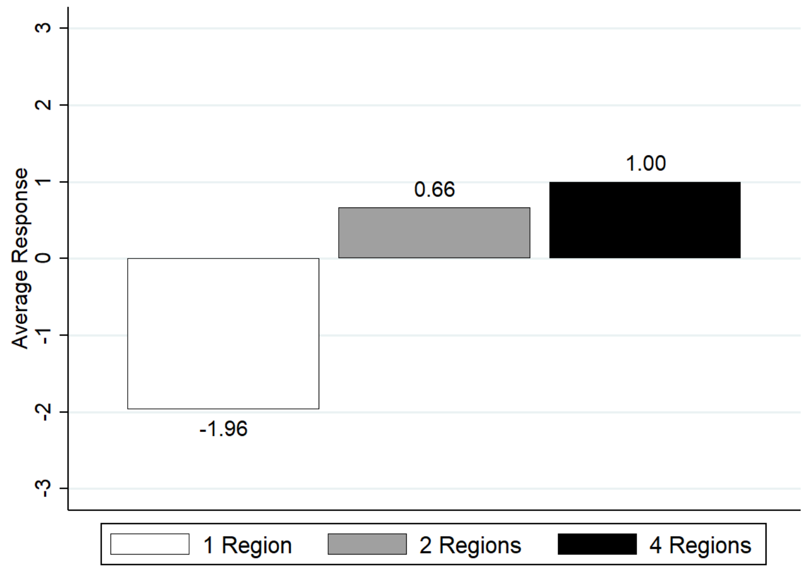
[9]The one marginally significant difference is that subjects reported beng more likely to participate as equals when managing two versus four regions ($p = 0.070$).

Figure 5: Did a Team Leader Emerge?



more difficult (but this is not significant; $p > 0.2$).

Figure 6: Relative Ranking of Team Versus Individual (Positive: Team Harder)



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 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.

We can summarize our survey results thus far as:

Result 6. *In the team trials, subject 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 over-weight their own contribution and their own leadership above that of their teammate, particularly when team performance was above the median.*

6.2 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 6, we report the estimated coefficients from regressions of the form:

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

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 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.[10] In Column (3), people who reported that "It was easier to achieve

[10]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.

Table 6: 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 * $p < 0.10$.

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.[11] As can be seen, these two variables maintain their sign and significance, though the coefficient on frequency of disagreements in reduced by about half.

We can summarize this analysis as follows:

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

7 DISCUSSION AND CONCLUDING REMARKS

In this paper we report on a series of experiments with human subjects who make typical supply chain decisions, which were distinguished by the degree of complexity. Although subjects were free to ignore the additional complexity by letting the business opportunities presented to them via the additional regions pass, or they could seek to exploit these additional opportunities for, presumably, a gain in profit. The results suggest that the option for ignoring these business opportunities were seldom exercised. Our results strongly suggest that subjects made active decisions to try to influence earnings and exploit the potential gains from improving the supply chain. However, our results also provide clear evidence that as complexity increases, many such attempts fail and subjects often destroy, rather than create, value.

One possible explanation which we previously hinted at 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 revenues. A naive view is then that, with more regions, investing in capacity to serve these regions will (and, as we show in the appendix, does) generate substantially higher revenues. 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 difficult managing capacity and foregone earnings on cash reserves), when final profits are calculated, the investments do not generally pay off. Thus, it appears that as complexity increases, subjects focus on inappropriate measures of performance, such as revenue, rather than the more appropriate – but difficult to forecast – earnings metric.[12]

We have also 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 reduce the variability of earnings, seem to reduce the likelihood of big mistakes and seem to lead to better investment decisions.

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

[12] Indeed, it is interesting to note that there is no correlation between earnings and revenue in our data set. This further supports our conjecture that as complexity increased, subjects focused on revenues rather than earnings.

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 – say by entering new regions or offering new products: conflict, which has a detrimental effect on profits may increase. 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 and failure is shared across the team. One wonders if this is part of the reason why conflicts within the team increase as complexity increases.

ACKNOWLEDGEMENTS

We thank the whole team at Responsive.net for their support during this experiment. In particular, we thank Samuel Wood, Mark Paul, and Jeff Kropp for sharing their experiences while we fine-tuned the parameters for this experiment and for making the whole experiment possible.

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.

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.

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.

de Leeuw, S., Grotenhuis, R., and van Goor, A. R. (2013). Assessing complexity of supply chains: evidence from wholesalers. *International Journal of Operations & Production Management*, 33(8):960–980.

Englmaier, F., Grimm, S., Schindler, D., and Schudy, S. (2018). The effect of incentives in non-routine analytical team tasks - evidence from a field experiment.

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

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

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

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.

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

Manuj, I. and Sahin, F. (2011). A model of supply chain and supply chain decision-making complexity. *International Journal of Physical Distribution & Logistics Management*, 41(5):511–549.

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. and Ruiz-Hernández, D. (2018). Am empirical analysis of an information-content based 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.

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

Ruiz-Hernández, D., Menezes, M. B. C., and Amrani, A. (2018). An information-content based measure of proliferation as a proxy for structural complexity. Mimeo.

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

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

Vachon, S. and Klassen, R. (2002). An exploratory investigation of the effects of supply chain complexity on delivery performance. *IEEE Transactions on Engineering Management*, 49(3).

APPENDIX A: 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 Results Note Reported in Main Text

In this section, we provide Figures 7 and 8, 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.

Figure 7: Trustworthiness and Reliability of Teammate

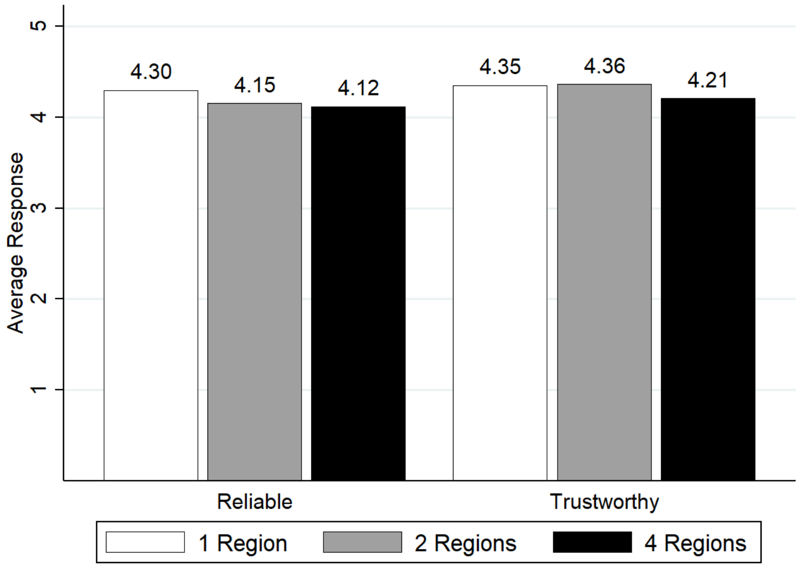
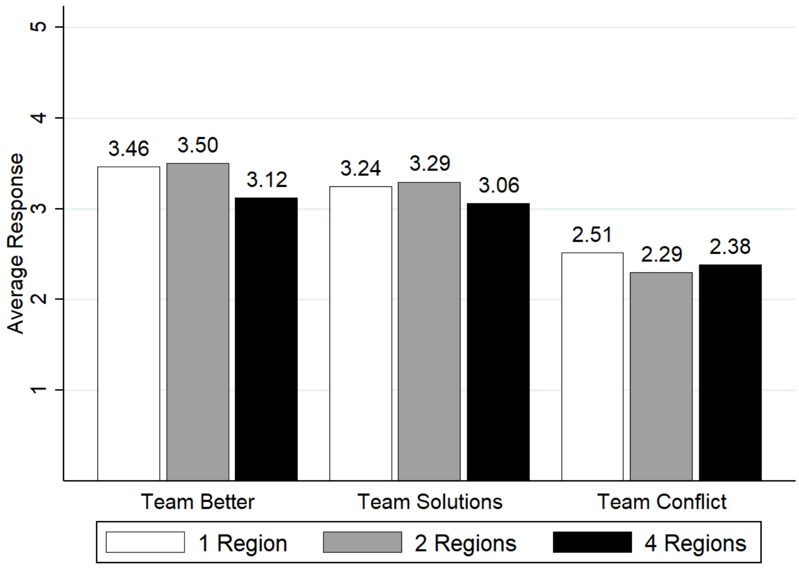


Figure 8: Team Solutions



Revenues, Complexity and Earnings

Here we provide some suggestive evidence that, at least some subjects may have focused more on revenues than on profits. The first evidence comes from Table 7, which shows that revenues are increasing in the number of regions, even though earnings are actually decreasing.

Table 7: 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

In Table 8, we report 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 and controls for the second trial and whether it was a team trial or individual trial. As can be seen, there is no relationship between earnings and revenue.

Table 8: Relationship Between Earnings and Revenue By Number of Regions

	1 Region		2 Regions		2 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 * $p < 0.10$.

Table 8 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 9, 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.

As Table 9 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).

Table 9: 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 * $p < 0.10$.

A plausible interpretation is that subjects focus on the large, and easily observable effect that investment has on revenues (i.e,, column (2) in Table 9) 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 strongly and so it more easily masks these hidden costs.