

Lessons from a long-term Beer Game dataset played by natural resource managers: reinforcing systems education across disciplines

Benjamin Turner¹, Michael Goodman², Rick Machen³, Clay Mathis³, Ryan Rhoades⁴, Barry Dunn⁵

¹Assistant Professor, Texas A&M University-Kingsville; ²Partner, Innovation Associates Organizational Learning; ³Professor, Texas A&M University-Kingsville and King Ranch Institute for Ranch Management; ⁴Assistant Professor, Colorado State University; ⁵Professor, South Dakota State University

Abstract: Simple yet powerful, the Beer Game (BG) has been played by thousands of players, from students to experienced managers, introducing the fundamental principles of complex systems and system dynamics (SD). For managers in agricultural and natural resources (AGNR), where systems are inherently complex due to the biologic, geologic, economic, social, policy and climatic characteristics and where delays are just as powerful and often longer compared to corporate settings, possessing a systems-oriented mental model is purported while implemented strategies remain linear or symptom-driven. To better understand the mental models of AGNR managers, we analyzed 10 years of BG trials. We found that AGNR managers performed at least as bad, and often worse, than typical BG performances. However, we also found that younger players more readily adapted to surplus inventories, significantly reduced order rates and effective inventories compared to older players. We discuss the results in terms of the common learning disabilities and context of 21st century AGNR challenges, and conclude with encouraging AGNR disciplines to adopt SD education/collaboration across disciplines to better address complex AGNR challenges.

1. Introduction

For decades, the Beer Game has been played in corporate workshops and meetings as well as in management, business, or modeling classes to introduce practitioners and students to some important lessons about complex dynamic systems. Although simple, the game has provided important and lasting lessons to participants regarding the basic principles of systems. Designed to represent a simplified supply chain, the game has widely been commented on in supply chain and corporate management settings and has become the exemplary “go-to” educational tool used in engaging newcomers to system dynamics and encouraging them to adopt a systems-oriented (i.e., nonlinear system dynamics) perspective. For those managers in the agricultural and natural resources (AGNR) professions (e.g., agroecosystem management; cropping and livestock production; wildlife or soil and water conservation; etc.), where systems are inherently complex due to the biologic, geologic, economic, social, policy and climatic characteristics of the systems and where delays are just as powerful and oftentimes longer than in corporate settings, possessing a systems-oriented mental model is often purported while the implemented strategies remain linear and symptom-driven. The linear mental model has been perpetuated in the fragmented, siloed and disparate nature of AGNR education that is also observed in other fields of study.

Contemporary AGNR problems have been growing around the world and are increasingly affecting the livelihoods of people and the continuity and vigor of local communities as well as food systems in general. These 21st century challenges operate at local to global scales and include such problems as climate variability and change (Akerlof et al., 2012; Sheffield et al., 2012; Corlett and Westcott 2013; Trenberth et al., 2013), water resource scarcity and management (Taylor et al., 2012; Haddeland et al., 2014; Savenige et al., 2014; Walsh et al., 2015), soil erosion and land degradation (Seto et al., 2012; Nepstad et al., 2013; Van den Bergh and Grazi 2013; Mahmood et al., 2014), biodiversity loss (Bellard et al., 2012; Cheung et al., 2012; Pauls et al., 2013), and food security (Wheeler and von Braun 2013;; Shindell et al., 2012; Vermuelen et al., 2012; van Ittersum et al., 2013; Teixeira et al., 2013), among others. These complex problems overlap and feedback on one another making sustainable and regenerative management of resources even more challenging. Tomek and Robinson (2003) introduce the nature of these problems in the following:

“Although the agricultural sector is a declining component of most national economies, agricultural product prices remain important both economically and politically. They strongly influence the level of farm incomes, and in many countries the level of food and fiber prices are important determinants of consumer welfare and the amount of export earnings. A decline of only a few cents per pound in the prices of such internationally traded commodities as sugar, coffee, and cocoa can have serious political and economic repercussions in such countries as Mauritius, Colombia, and Ghana. As Deaton (1999) pointed out, inaccurate forecasts of commodity prices led to poor policy prescriptions for African nations. Even in the United States, a large drop in the farm price of hogs was reported as likely to drive 24,000 pork producers out of business (Wall Street Journal 1998)...

They then describe several reasons why such problems not only arise but continue to persist:

“The characteristics of agricultural product price behavior relate importantly to the biological nature of the production process. Significant time lags exist between a decision to produce and the realization of output, and actual production may exceed or fall short of planned production by a considerable margin. At least a year is required for producers to change hog production, 3 years to change the supply of beef, and 5 to 10 years for growers to increase the output of tree crops such as apples. Yields vary from year to year because of variability in weather conditions and the presence or absence of diseases or insect infestations...”

Besides the biological nature of production, farmers’ mental models are central to understanding the dynamics of these systems:

“Farmers’ production decisions are based partly on their expectations about future yields and prices (i.e., expected profitability) of the alternative commodities they might produce. Since these expectations are not always realized, price and yield risks exist in farming, and the way expectations are formed and acted on by farmers may affect a cyclical component to supply and prices.”

Lastly, the combination of economic and decision-making processes with the constraints of fixed resources converge to create complex dynamics that are difficult to respond to and manage, similar to the Beer Game:

“The nature of resources, like land and equipment, used in farming is such that producers cannot easily make major changes in production plans in response to expected price changes... [typical demand assumptions are] of one-way causation from prices to quantities, but this assumption is the opposite of a common view about the way that prices for agricultural commodities are thought to be determined. That is, current production of agricultural commodities is based on decisions made by farmers many months, even years, in the past. Thus, current supply may be influenced very little by current price. If so, then causality runs essentially from quantity to price; quantity is predetermined by prior events.” (Tomek and Robinson, 2003).

System dynamics (SD) has been interested in these issues from its adolescence, most notably described and discussed in the seminal work *The Limits to Growth* (Meadows et al., 1972). Since AGNR systems are complicated and with few incentives to adopt a SD mental model to better address the root causes of issues, most management efforts aimed at addressing the issues have used traditional, easily accepted methods promoted from within disciplinary silos (Beautement and Broenner 2011). As a result, many root causes of AGNR problems have continued to go unaddressed.

In order to improve and enhance the educational experience and outcomes of AGNR managers to better address these 21st century challenges, the King Ranch® Institute for Ranch Management (KRIRM) at Texas A&M University-Kingsville in 2003 implemented an innovative curriculum grounded in systems thinking and SD. The KRIRM mission is to educate professionals that will improve resource stewardship on working agricultural landscapes throughout the world. To achieve this, KRIRM offers an intensive two-year graduate degree program, a less intensive certificate program, and a distance education leadership program for those professionals in the field who cannot attend full time. At the core of each of these programs is a week-long lectureship in systems thinking that serves as a foundation to the curricula and a common framework through which student research projects are investigated. Noticing similar needs in undergraduate research programs, the South Dakota State University Honors College (SDSUHC) has implemented a systems thinking workshop to enhance the student’s experience prior to their beginning undergraduate research experiences. Unfortunately, no such “go-to” introductory educational tool based on a natural resource management example has been found to be nearly as powerful or insightful as the Beer Game. Therefore the Beer Game has been used by both the KRIRM and SDSUHC programs as the students’ first exposure to systems thinking and decision making in complex systems. Since the inception of these programs, we have kept records of Beer Game performances in an integrated database used for debriefing the new teams each year.

The purpose of this paper is to present the Beer Game database of decisions made by participants in KRIRM and SDSUHC classes over a 10 year period and what we can learn (and re-learn) from the results. First, we briefly review the Beer Game as well as the foundational literature regarding mental models and dynamic decision-making that has been conducted through use of the Beer

Game or management flight simulators inspired by the game. We then provide a profile of the participants that are represented in our database (experienced professional at KRIRM; undergraduate students at SDSUHC) and describe a Beer Game model that was used to facilitate analyses of team performances. Then the results are presented and discussed, including: a) how well our groups performed compared to the seminal work of Sterman (1989); b) the key points of the Beer Game in the context of AGNR systems; c) some interesting divergences observed between performances of the two groups (KRIRM and SDSUHC); and d) how overcoming the common learning disabilities seen in complex systems will require effective education regarding the dynamics of complex systems. We conclude with a challenge and encouragement for all SD practitioners to collaborate and support a wider array of researchers in AGNR industries due to the important role these industries and managers will have in addressing the complex 21st century challenges of our time.

2. Beer Game overview and decision-making research over time

2.1. Rules of the Beer Game

The Beer Game portrays a production-distribution system characterized by an inventory stock for each of four respective players: factory (or brewer), distributor, wholesaler, and retailer arranged in a linear distribution chain (Sterman 1989a; Sterman 1992). Pennies or plastic chips represent cases of beer. Flows influencing each of the stocks include the information flow of orders from the retailer toward the brewer and the physical transport of goods from one stock to the next through the supply chain, including a two-week delay. Exogenous to the system are consumer orders (delivered through a deck of cards at the retailer end of the chain) and the inputs to the brewer's inventory production (i.e., supply of chips the brewer has access to). For each week of the game, customers purchase from the retailer, who then orders from the wholesaler, who then orders from the distributor, who then orders from the factory, who then produces additional cases of beer to meet the anticipated demand. The objective of the game is to minimize total costs throughout the supply chain, where holding inventory costs each player \$0.50 per case per week or \$1.00 per case per week penalty when inventory is in backlog (representing lost revenues and discontent of customers when stockouts persist). Each player manages their respective inventory through forecasting demand, which is informed by the orders of their customer one-step down the supply chain, and placing orders to their suppliers one-step up the supply chain. However, only the retailer can see the actual demand from customer orders week by week. This information limitation is further constrained since players are not allowed to communicate with one another, therefore coordination is impossible (Sterman 1989a; Sterman 1992). Although simplified from real world systems, the game has revealed important characteristics and limitations in human decision-making (described below) and the principles of systems (section 4).

2.2. Research on decision-making in relation to the structures of complex systems

Research into human decision-making in complex systems has been aided by management flight simulators which have allowed system dynamicists to study an individual's decision-making within simple system structures, similar to that of the Beer Game (Table 1). Rouwette et al., (2004) summarized the key findings of the major works in this area based on characteristics of the system

complexity: delays, strength of feedback, and exogenous change. In general, the presence of time delays as well as the length of delays tends to reduce performance (Sterman 1989b; Diehl and Sterman 1995), stronger feedback loops tend to reduce performance (Sterman 1989b; Diehl 1989; Paich and Sterman 1993; Diehl and Sterman 1995; Young et al., 1997; Langley et al., 1998), while exogenous changes have shown either reduced (Sengupta and Abdel-Hamid 1993; Langley et al., 1998; Schultz et al., 2000), improved (Bakken 1992), or no effect on performance (Barlas and Özevin 2001).

Whenever the context of the decision situation changes (e.g., changing the decision interval in order to speed up or slow down experience generation; increasing the transparency and therefore knowledge of the model or task being managed; or the nature of the decision information such as interface design and types of feedback provided to the user), participants tend to do somewhat better (Young et al., 1992; O'Neill 1992; Machuca and Carrillo 1996; Größler 1998; Größler et al., 2000; Georgantzas 1990; Langley 1995; Howie et al., 2000), or at least no worse than their prior performance without the added transparency (Barlas and Özevin 2001; Ford and McCormack 2000; Young et al., 1992; Diehl 1998; Richardson and Rohrbaugh 1990; Sengupta and Abdel-Hamid 1993; Anderson et al., 1994; Ritchie-Dunham 2001). Lastly, personal characteristics that likely impact performance (e.g., stated goals; mental model or cognitive styles; the number of players in the simulator) have also shown mixed results, with improved performance resulting from whole system goals and a higher degree of similarity in participant mental model with the structure of the simulation (Yang 1996; Ritchie-Dunham 2001). Other studies have shown no effect on performance due to personal characteristics (Yang 1997; Georgantzas 1990; Scott-Trees et al., 1996; Schultz et al., 2000; Park et al., 1996).

More recent advances in decision making research using simulations of the Beer Game task have shown that coordination risk (the risk that individuals' decisions contribute to a collective outcome but the decision rules followed by each individual are not certain) which contributes to the bullwhip effect can be mitigated with coordination stock (holding additional on-hand inventory) but that the behavioral causes of the supply chain instability are robust (Croson et al., 2014). Sterman and Dogan (2015) show that because of this persistence of instability, individuals are likely to seek larger safety stocks (hoarding) or order more than what is demanded of them (phantom ordering). These irrational responses were shown to be triggered by environmental stressors, which overwhelmed individuals' rational decision-making abilities or when individuals inappropriately applied decision heuristics incompatible with effective performance in the game. Emotional, psychiatric, and neuroanatomical factors are also discussed in Sterman and Dogan (2015).

3. Natural resource managers (and students) play the Beer Game

Two notes made by Rouwette et al. (2004) are important for the present investigation and should be pointed out here: 1) although their paper only included system dynamics based studies, their study identified that there were few fundamental differences to none between SD-oriented tasks and performance task games from other social science disciplines; and 2) simulation players have primarily been sampled from university student populations, with professionals (e.g., university staff) being only a small subset of the participants. The first note is important because the SD based

research on decision-making, although rich, is relatively small compared to other social sciences like sociology and psychology. The fact that research findings between SD and other social sciences corroborate each other well provides additional confidence from which to apply and interpret SD based insights. The second note is important for several reasons: a) students, however knowledgeable, generally lack the real-world experience and accumulated wisdom that seasoned managers possess; and b) the simulated games could likely represent industries or management interests that differ from the students, making it more difficult for students to comprehend or self-motivate themselves to think systemically about their performance compared to those that begin with more accumulated experience and are therefore able to relate their task performance to real world experiences. In many industries, including AGNR, experience and wisdom can differentiate effective and successful managers from the pack, and these managers tend to think more systemically than others since their experiences over time and across industries has enhanced their ability to make inferences that oftentimes are not obvious to novices. We explored the above hypotheses by examining the Beer Game database created from over 10 years of classes at KRIRM and SDSUHC.

3.1. Participant profiles

Since the KRIRM programs target agriculture industry leaders, natural resource conservation professionals, and up-and-coming farm and ranch managers, the majority of participants already possess four year degrees (B.S. or B.A.), in some cases graduate degrees (M.S. or Ph.D.), and arrive with more than several years of professional experience in production agriculture or resources conservation. Generally, the systems these participants operate within are laden with time-delays. For example, crop and livestock production systems begin with producers making genetic selections on the types of production they wish to market, which take years to decades for payoffs to realize. Growing seasons generally allow only one harvest per year and replacement and maintenance efforts requiring two years or longer. Likewise, since participants come from agricultural businesses embedded with the agricultural food system, they have intimate knowledge of supply chain dynamics, since in the real world they sit at the producer, distributor, wholesaler, or retailer positions. We were highly interested in analyzing the decisions made in the Beer Game by these groups, since similarities between participant mental models and dynamic decision contexts have been shown to improve performance (Ritchie-Dunham 2001). Similarly, conservation professionals operate in organizations whose goals are inherently holistic and long-term, since ecosystem restoration or conservation operates at decadal to century's time scales. Graduate students in AGNR disciplines but without significant professional experience have also participated, however they have generally been no more than 20% of the participants. Few, if any, undergraduates have ever participated in the KRIRM classes. However, the attendees at the SDSUHC course has been weighted to Honors College students (many but not all of which having arrived from family farms or ranches) that are preparing for undergraduate research experiences ($\approx 80\%$). The SDSUHC class participants have also included faculty members throughout the College of Agriculture and Biological Sciences (e.g., animal scientists; wildlife and fisheries scientists; agricultural economists; $\approx 20\%$). In total, there were 56 KRIRM teams and 9 SDSUHC teams.

3.2. Database description

The data was compiled in Microsoft Excel™ beginning in 2004 (Figure 1), the first year the Beer Game was played at KRIRM. The SDSUHC games were added for the years 2012 through 2014. Orders for retailers, wholesalers, distributors and factors were kept on one sheet (Orders tab), with weeks 1 through 35 repeated for each position down the spreadsheet. A similar convention was used to record inventory or backlog (Inventory tab). The entries for each team stop at 35 weeks to reduce leveling effects (see Sterman 1989a). In total, the raw data included 64 teams.

Data for orders and inventories were obtained directly from each team's record sheets. For reference, each team's total cost was entered at the top of each tab (beginning in cell B5), however, costs at each player position were not recorded. Average costs of each year's teams were also calculated (e.g., cell C5). The remaining tabs of the database included graphs of trends-over-time of participant performances (e.g., average, best, and worst performances) used as visual aids in the debriefing sessions of participants in the KRIRM and SDSUHC classes (excluding the current year teams, which used the record sheets from their own performances).

The database includes two kinds of uncertainties that must be recognized. The first error common to all Beer Game results is due to human mistakes made playing the game (e.g., getting ahead or behind the weekly schedule; correctly recording inventory or backlog; correctly calculating costs; etc.). However, our database includes another risk of error due to the transfer of information contained on the record sheets into the Excel file. Since the games were played at the beginning of a week-long systems thinking class, results were generally entered by graduate students the following week, or about one week after the completion of the game itself, which precluded any clarification of results, inaccuracies, or illegible entries by participants, as well as the human error involved in the actual transfer of data.

Before proceeding to the analysis, these errors had to be reconciled in the database or teams simply removed from the analysis due to such large errors in effective inventory and costs. Sterman (1989a) identified that Beer Game teams with the highest costs were the most prone to accounting errors, and therefore, reduced that sample size from 48 to 11 teams, which were generally the best performing. However, successful teams can be just as susceptible to human accounting errors, since mistakes (or variance from optimum decision levels) made around a given average order quantity will not affect the overall rank of teams given lower average orders and therefore inventories. Besides addressing these errors, we also had to reconcile the costs of each player position, since only the total costs per team were entered in the database (which was critical if comparisons were to be made previous results presented in the literature). Rather than discarding only the poorer performing and analyzing the most successful, but not necessarily lesser flawed teams, we developed a beer game model to compare the observed team performances with expected performance given equal accounting standards. The model (described below) aided in identifying the teams with the greatest accounting errors that should be discarded from the final analyses as well as captured costs of each player position, allowing us to compare a more representative range of teams rather than only the most successful to previous research results.

3.3. Beer Game model

The Beer Game model was developed in Vensim™ (Ventana Systems, Inc.) in the same table top configuration of the Beer Game, with physical flows for inventories and information flows for orders (Figure 2; player positions are abbreviated to Retailer=r; Wholesaler=w; Distributor=d; Factory=F and are aggregately described in the italicized names that represent formulation for each of the positions). Equations were developed from previous beer game models (Sterman 1989a; Kirkwood 1998) and are provided in Appendix 1. Stocks of inventory (*inventory[*position*]*) are controlled by flows of cases (*in[*position*]* and *sold[*position*]*). Backlog stocks account for unmet demand along each position and are used to calculate effective inventory (*eff inv[*position*]*). *ORDER* represented customer orders, beginning at four cases and which steps to eight at the fifth week. Rather than using the decision order algorithms from either Sterman (1989a) or Kirkwood (1989) (i.e., a smooth function which provides first order exponential smoothing to represent an averaging process to place orders from each sector), observed orders made by our participants were input into the model (*import[*position*] placed orders*). This allowed for the least error prone data in the database, orders placed (where no calculations are required for record keeping), to be used to evaluate teams using equal accounting standards.

3.4. Data analyses

Initial data analyses consisted of screening the database entries for teams with the greatest errors by comparing observed total team costs and team ranks to the expected total costs and ranks given equal accounting standards. This was achieved via the Beer Game model. Outlier teams were identified and removed from the dataset. For the teams remaining, actual player orders and the modeled team and position costs were used for the analyses. Modeled costs were used to ensure comparison between teams was fair given the likelihood of unknown errors in the dataset and we attempted to minimize these through the screening process used to identify and discard teams. The remaining teams were then used to conduct two different analyses: 1) due to the similar background and interests of all participants in natural resources management, we examine the performance of all teams. Since no control and treatments were conducted in our Beer Game trials, we simply examine the participant performances in the database with the results of experiments presented in the SD literature ($H_0: \text{Database}_{tc} = \text{Reported}_{tc}$); and 2) due to the unique participant profiles at the two locations (mostly experienced professionals at KRIRM, mostly undergraduates at SDSUHC) we compared team performances between the more experienced and less experienced groups ($H_0: \text{KRIRM}_{tc} = \text{SDSUHC}_{tc}$).

4. Results and discussion

4.1. Model comparison of team performances

The database included both KRIRM ($n_1=55$) and SDSUHC ($n_2=9$) teams since 2004 (total $n=64$). Initial screening of the team performances revealed strong fit between total costs (Figure 3a) and team rank (Figure 3b). Overall, the expected costs and ranks of teams fit fairly well, r^2 values of 0.90 and 0.89 respectively (Table 2). Despite the overall strong correlations, we identified 20 teams that clearly did not fit the expected costs pattern between the modeled and observed costs of the majority of teams, resulting in a total n of 44 ($n_1=38$; $n_2=6$). Interestingly, the discarded teams were fairly normally distributed throughout the database, with three teams removed from the top

quartile, six from the third quartile, eight from the second quartile, and three from the bottom quartile. Removing these teams significantly improved the match between observed and expected team performances and ranks, with r^2 values of 0.97 and 0.98 (Table 2; Figure 3c and d). Average errors (in terms of total \$ costs, \$/week, or cases/week) decreased \$75 and \$2 respectively. Removing the teams with inconsistent costs relative to the remaining teams created a significantly improved fit in team ranks (e.g., from six to 11 exact rank matches, or from nine to 25 percent; only five teams with rank discrepancies greater than three positions, down from 55 to 11 percent). Utilizing the beer game model in this way allowed us to screen the database for the teams that most likely had the greatest accounting errors and gave added confidence that the remaining teams, that although not perfect in their accounting, were accurate enough to allow comparison across the dataset. The proportion of discarded teams due to likely accounting errors (31% of the original database) was therefore much smaller than Sterman (1989a), which discarded 75% of that database due to errors, indicating that players may do a better job of accounting than was previously expected.

4.2. Participants performance across the database

The team average total costs relative to the benchmark costs (identified in Sterman 1989a) are shown in Table 3. The average team cost was over 23 times the benchmark and twice the average reported in Sterman 1989a (Table 3; Figure 4), although that study only reported scores of the best performing teams. The wholesaler, distributor, and factory ratios of actual to benchmark costs were as high as 30 times greater than optimal cost levels, however, the retailers in our group performed similarly to other studies (Table 3). The differences in total costs and costs of each sector to the benchmark costs were all highly significant, and compared to Sterman (1989a), all sectors were significant except the retailer. To identify how well the best performing teams in our database performed relative to previous studies, total team and individual position costs were summarized into quartiles (Table 3). The top performing teams in Sterman (1989a), whose team average (\$2,028) and position average costs (retailer \$383, wholesaler \$635, distributor \$630, factory \$380) fell most closely between our third and fourth quartile of team performances, indicating similar performance between the above average teams.

Similar oscillations, amplifications, and phase lags were observed between our team performances and common beer game results (Table 4; Figure 5). Orders and inventories expressed large fluctuations, with average inventory recovery of 25.5 weeks. Backlogs of inventories migrate from the retailer to the factory similar to typical beer game results (Figure 5), with the peak order rate at the factory being over three times the peak order rate of the retailer. Closed loop gains ($\Delta[\text{factory orders}]/\Delta[\text{customer orders}]$) averaged nearly 1400%, or double that reported by Sterman (1989a). Maximum backlogs averaged 35 cases and occurred between 34 and 35 weeks (Table 4). As expected, inventories overshoot initial levels, peaking at week 35. Phase lags were more evenly distributed than typical beer game runs, however this was likely due to the larger sample size smoothing out the week of peak order rates. Participants' anticipated minimum inventory (date of minimum inventory minus date of week order rate) were generally delayed by one or two weeks, indicating reactive strategies that did not account for orders in the supply line and perpetuated extreme inventory levels later in the game.

Although the overall scores were poorer than team performances reported in the literature, the top 10 teams in our database (or $\approx 25\%$) performed better than the top 25% reported in Sterman 1989a (Table 5). Retailer, wholesaler, and distributor costs were all significantly lower (which contributed to an overall significantly lower team total cost), while the factory costs were significantly higher. Periodicity and phase lags were noticeably shorter and amplification lower than the Sterman 1989a teams. Of the top 10 teams of our database, the SDSUHC groups were disproportionately represented. Eight of the top 10 teams came from the KRIRM participants ($\approx 21\%$ of the KRIRM sample) while two teams came from the SDSUHC participants ($\approx 33\%$ of the SDSUHC sample).

4.3. Comparison of performances from more and less experienced participants

We hypothesized that the older, more experienced group (KRIRM) would perform better on the beer game task than the less experienced players, primarily undergraduate students (SDSUHC). We found no evidence to support this (Table 6), as neither the team total costs nor any of the player position costs were significantly different. This corroborates previous conclusions that management experience may not mitigate misperceptions of feedback (Paich and Sterman 1993). However, qualitative analyses of the trends in effective inventory and order rates tell a more interesting story (Figure 6). The SDSUHC teams appeared to achieve maximum inventory earlier than the KRIRM groups and by week 35 were reducing their overall inventory levels back toward the ‘anchored’ inventory level of 12. This was achieved through overall lower average order rates (Table 6; Figure 7). Although retailer orders were similar, wholesaler, distributor, and factory average order rates differed from as low as one to as high as six cases per week. After initial inventory recovery, discrepancies in order rates were even larger (up to 8 cases at the factory level) and were all statistically significant (Table 6). Based on the change in slope of order rates and effective inventories after week 29 for the SDSUHC teams, it appears the younger groups began accounting for cases in delivery much sooner than the KRIRM groups, whose maximum effective inventory levels continued to rise. It is possible that several interesting features are at work that created the divergence in trends of effective inventory between the participant groups.

First, the older KRIRM participants could have continued to order more cases after the initial inventory recovery as a way to accumulate “coordination stock” to hedge against the risk that customer orders will significantly change in the future (based on their perception of customer orders as well as experience in the real-world) or in case the other players deviate from the near equilibrium (but-suboptimal) position that the game reaches by week 30 (i.e., compensate for obvious weaknesses in their teammates) (Croson et al., 2004/2014). Relying on real-world experience requires participants to determine strategy via comparison of the game to previous experience by analogy, however decision makers who reason by analogy in complex dynamic situations have not performed as well as those who don’t (Georgantzis 1990).

Second, the older participants were likely less inclined to lower their order rates after inventory recovery, since the initial strategy (increase the order rate to get out of backlog) eventually paid off. In other words, so long as they achieved zero backlog, they were not as heavily anchored to the initial inventory level as the younger players. It’s been shown that experience with a particular set of behaviors improves performance, but that as opportunity costs of trying new strategies rises,

individuals will experiment with fewer decisions and are less likely to identify superior methods compared to their status quo (Levinthal and March 1981, Herriott et al., 1985). It is likely that the opportunity costs to change strategies appeared to be too high for the older players.

Third, the younger, less experienced players in the SDSUHC teams significantly lowered their order rates after inventory recovery compared to the KRIRM group (Table 6). Although inventories are effected by the choices of the other players, participants are forced to *discretely* place new orders based on each new inventory level, and new order rates represent desired change in the stock of the individual player. Therefore each choice in order is aimed at closing the gap between desired and actual states of inventory (albeit with the necessary receiving and shipping delays).

Research on dynamic decision making choices of younger versus older adults has shown that older adults (age 60-84) perform better on choice-dependent task, which require learning how previous choices influence current performance and making a new decision based on that knowledge (Worthy et al., 2011; Worthy and Maddox 2012). Older players in our sample were more heavily anchored to their previous strategy that worked (order more cases to get out of backlog), and because of that success continued to do so. Research on younger decision makers (age 18-23) has shown that they perform better on choice-independent tasks (where learning requires exploiting the options that give the highest reward on each new trial; Worthy et al., 2011; Worthy and Maddox 2012) and students have best learned the dynamic decision making in systems by ‘doing’ and ‘failure’ rather than ‘knowing’ or relying on experience (Young et al., 1992). Younger players in our sample were more heavily anchored to the initial inventory level and were therefore more responsive to escalating inventory levels (and therefore costs) by altering their order rates (Table 6).

Our older players increased order rates to get out of backlog, and rather than decreasing order rates once effective inventories recovered, continued to order at relatively high rates (i.e., they were heavily anchored to the choices that worked to get themselves out of backlog), while our younger participants made a more abrupt shift to lower order rates upon inventory recovery and escalation. Our results coincide with the behaviors observed in age-related studies (Worthy et al., 2011; Worthy and Maddox 2012) and likely explains the discrepancy in order rates between groups (Table 6; Figure 7). More importantly, the oscillatory dynamic behaviors in effective inventory results regardless of decision strategy (e.g., Sterman 1989a; Barlas and Özevin 2001), which only altered the minimum and maximum inventory points.

4.4. What we can learn and (re)learn from our study

There are a number of key lessons we can learn and re-learn as a result of this study. First, we observe that natural resource managers embedded in real-world systems with extremely long time-delays (e.g., year to decadal scales) are just as bad, if not worse, than managers from corporate contexts at identifying and managing the delayed-inventory management task illustrated in the Beer Game. Results closely corresponded to typical results seen in other Beer Game trials, indicating that our participants, despite intimate knowledge of natural and agricultural systems, had adopted a similar decision rule identified by Sterman (1989a), where participants anchor their

initial expectations to the starting inventory level that inevitably produces extremely poor results. This is due to the misperception of delayed feedback between placing and receiving orders and not fully accounting for cases in the supply line, both of which lead to over-ordering and instability in even the best performing teams (Figure 5). Even those that recognize and manage systems with many time delays that often vary from months to years in length, they still commit the same errors as ones without such experience with delays.

We can also relearn and reinforce the common beer game lessons identified decades ago which are now some of the major tenets of the SD perspective (Senge 1990). Different people in the same structure produce similar results and it is easy to look for external causes or blame others, but systems cause their own problems (the *structure drives behavior* tenet). Although structures can be defined by the basic interrelationships that influence, regulate, or control behavior (including external constraints), structure more importantly is the endogenous decision-making rules, operating policies, goals, and modus operandi, many of which are unwritten and embedded in the culture of organizations (the *structure is subtle* tenet). Lastly, we often don't recognize leverage in complex systems because we focus only on our own decisions rather than how our decisions affect others or how we are connected in the system as a whole. Failing to recognize or understand the interconnected decisions hinders our ability to effectively apply leverage to better cope with systemic problems (the tenet that *leverage often comes from new ways of thinking*; Senge 1990).

We may also be reminded of the common learning disabilities that are seen in humans across cultures and contexts (Sterman 2002; Senge 1990) and the barriers that impede our learning about complex systems (Sterman 1994). Almost regardless of history or experience, when inserted into a given position within a system or organization, the structure incentivizes that we “become our position”. When faced with unfavorable or detrimental circumstances, we point to others around us or to actors outside the system to assign blame. Too often we fail to recognize problems, and even when we do, our proactive decisions can sometimes do more harm than good. Since many externalities are never felt by those that made the decisions that created the problems (e.g., the soil and water externalities cited above), our ‘knee jerk’ reactions to are assign blame to others around us and we fail to effectively learn from experience and the collective wisdom of others in the system. Besides these learning disabilities, Sterman (1994) listed seven specific learning barriers that impede our learning about complex systems (each of which can be seen in the Beer Game), including:

- dynamic complexity (i.e., cause-and-effect are distant in time and space),
- limited information (i.e., we experience the world through filters such that our senses and information systems select only a fraction of possible data),
- misperceptions of feedback (i.e., we fail to recognize the complexity of the systems we are imbedded in),
- flawed cognitive maps of causal relations (i.e., very few of our mental models incorporate feedback loops),
- erroneous inferences about dynamics (i.e., people cannot mentally simulate the simplest feedback systems),

- defensive routines and interpersonal impediments to learning (i.e., defensive routines prevents learning by hiding important information from others),
- and implementation failure (i.e., imperfect implementation and irreversible actions with high stakes on performance often override the need to learn).

Although described and highly documented in the management and SD literature, similar learning disabilities and the consequences they exert on decision making have also been observed in the natural resource management arena (Mitchell et al., 2007; Lachapelle et al., 2003; Tàbara and Pahl-Wostl 2007).

To overcome these disabilities and barriers, the SD profession has prioritized and advocated for systems-based education from K-12th grade levels up to university graduate programs (see the Creative Learning Exchange at clexchange.org; Forrester 1965/1975; Forrester 1994; Stuntz et al., 2002; Forrester 2006; Booth Sweeney and Sterman 2007; Lyneis and Stuntz 2007; Forrester 2009). Given the results of our Beer Game database and experience in the agriculture and natural resource management professions, the need for systems education in these disciplines is as desperately needed as ever if effective change is to be expected and gaps in the 21st century challenge begin to sustainably close. Agriculture and natural resource professionals with systems education could likely achieve significantly different results compared to professionals without systems-oriented education. For example, thinking in *systems* forces us to recognize the interconnectedness and dynamic complexity of the problem at hand, the physical stocks and flows central to issue, and time-delays between decisions and results. Systems thinking and system dynamics modeling also encourages us to maintain an unwavering commitment to the highest standards and rigor of scientific method by recognizing and correcting our hidden biases and documenting and testing our assumptions about the problem. By doing so, we can explore a wider decision space for new or previously unrecognized leverage points to achieve our goals (Sterman 2002). Achieving the 21st century agriculture challenge requires input and collaboration across disciplines and cultures. System dynamics can provide a common unifying language to facilitate such collaboration.

4.5. Study limitations

There are a number of limitations of the study as presented. First, for general research purposes, our use of the traditional board, pencil and paper based Beer Game is unconventional, and has been noted that this use of the Beer Game is no longer acceptable because of the high rate of clerical and data recording errors. All modern Beer Game studies are expected to use a computer version of the game to prevent such errors and to offer tight controls on information availability among the players. Due to the structure of the lectureships where our participants played the traditional Beer Game and the available computer resources at the time, computer applications have been generally unfeasible. We attempted to limit clerical and data errors by screening the data with an application of a Beer Game model. Second, traditional incentive scheme in the board game version (\$1 entry fee and winner-take-all) is not consistent with current standards for experimental studies in economics, in which people are paid in proportion to their performance. Lastly, previous Beer Game studies have estimated an ordering decision rule to test the misperceptions of feedback hypothesis, which we have omitted here but is planned for in the future.

5. Conclusions

System dynamics has long been interested in dynamic decision making in complex, feedback driven systems. The classic Beer Game, played by thousands of players over many decades, has become a foundational learning tool that exposes key learning disabilities and barriers that, when used in decision making, create oscillatory behaviors and exuberant costs. We present an analysis of a Beer Game database developed using over 10 years of performances of agriculture and natural resource managers. One unique aspect of the database is that the majority of participants (87%) were working professionals in agricultural and natural resource systems with at least a B.S. degree and in many cases M.S. or Ph.D. degrees. Even players with extensive knowledge and experience working in and managing a complex system perform poorly, as seen by the performance of our agricultural and natural resource professionals. However, some evidence was found that younger players were more responsive to changed inventories and shifted their strategies sooner and with larger magnitudes than older players at the same time of the game. Given the increasing severity of 21st century issues, particularly those tied to Limits to Growth (e.g., food, agriculture, natural resource capacity, and pollution dynamics), managers in these diverse fields would benefit from a common language able to identify and describe the complexity of the problems around them. System dynamics can provide such a language. To better address today's evolving natural resource challenges, programs such as the King Ranch® Institute for Ranch Management and the South Dakota State University Honors College have adopted systems thinking and made it core curriculum for their respective programs to aid in students overcoming the common learning disabilities and barriers. Future directions for this project include ongoing replications of Beer Game-type studies investigating coordination risk and coordination stocks (Corson et al., 2014) and hoarding and phantom ordering (Sterman and Dogan, 2015), which have opened the door to exploring emotional and psychological underpinning to describe the behaviors seen in the Beer Game.

Acknowledgements: We'd like to thank 5 anonymous conference paper reviewers for their helpful comments and critiques. All were used to revise, remove, or improve specific sections of the paper.

References:

See Supplementary Material for full list of references.

Tables

Table 1. Summary of results from research regarding individual's abilities of task performance in dynamic decision making environments.

Dynamic decision making characteristic	Key findings (with sources)
Misperceptions of feedback	<ul style="list-style-type: none"> • Individuals are generally unable to account for delays and feedback effects because of highly simplified mental models and poor ability to infer correct behavior of simple feedback systems (Diehl and Serman 1995). • Strategies were insufficiently adjusted to account for strength of feedback between information and material flows (Diehl 1989). • Misperceptions of feedback have been attributed to poor recognition of delays and preference to maintain static decision rules (Serman 1989b; Young et al., 1992; Serman 2002; Rahmandad et al., 2009). • Misperceptions of feedback not only result in suboptimal performance but often lead to strategy and decision making that is in opposite direction of optimal changes in decision making (Serman 1989a; Serman 1989b; Langley et al., 1998; Paich and Serman 1993).
Types, transparency, and scale of models; types of feedback	<ul style="list-style-type: none"> • Transparency of model (including interface design) and prior knowledge of structural information can improve task performance in dynamic decision making contexts (Größler 1998; Größler et al., 2000; Howie et al., 2000; Machuca and Carrillo 1996). • Given an increasing time scale, participants are not likely to consistently or steadily improve performance (Ford and McCormack 2000). • Types of feedback (e.g., numerical, graphical, or both) are known to influence behavior (Serman 1989b; O'Neill 1992; Langley 1995; Sengupta and Abdel-Hamid 1993).
Environmental & Contextual Characteristics	<ul style="list-style-type: none"> • Coordination risk contributes to bullwhip effects since behaviors of other individuals is not known with certainty (Croson et al., 2014). • Stressors such as larger orders, backlogs, or late deliveries trigger hoarding and phantom ordering even those such behaviors are known to be irrational (Serman and Dogan, 2015).

Table 2. Comparison of observed (recorded) and expected (modeled) team costs and ranks before and after discarding outlier teams from the analysis.

	Full dataset (n=64)	Adjusted dataset (n=44)
<u>Costs</u>		
r ²	0.90	0.97
slope coefficient p-value	1.84E-32	1.21E-32
ave. error (\$)	414.3	338.4
ave. error (\$/week)	11.8	9.7
ave. error (cases/week)	15.8	12.9
root mean squared error	149.2	97.0
<u>Ranks</u>		
r ² (p-value)	0.89	0.98
slope coefficient p-value	4.1E-31	2.66E-35
exact ranks (% of n)	6 (9%)	11 (25%)
within 1 rank (% of n)	13 (20%)	13 (30%)
within 2 ranks (% of n)	8 (13%)	10 (23%)
within 3 ranks (% of n)	2 (3%)	5 (11%)
error > 3 ranks (% of n)	35 (55%)	5 (11%)

Table 3. Comparison of participant and benchmark costs identified by Sterman (1989a) as the upper bound estimate of optimal performance.

	Team total	Retailer	Wholesaler	Distributor	Factory
<i>Benchmark comparison</i>					
mean (n=44)	\$4,802	\$412	\$1,197	\$1,564	\$1,629
benchmark	\$204	\$46	\$50	\$54	\$54
ratio (mean/benchmark)	23.5	9.0	23.9	29.0	30.2
t-statistic	7.97	5.85	5.66	7.53	7.75
H ₀ : mean cost = benchmark	p < 0.00+	p < 0.00+	p < 0.00+	p < 0.00+	p < 0.00+
<i>Sterman 1988a comparison</i>					
Mean (n=11)	\$2028	\$383	\$635	\$630	\$380
ratio	2.4	1.1	1.9	2.5	4.3
t-statistic	4.81	0.46	2.77	4.65	6.15
H ₀ : mean cost = Sterman 1989a	p < 0.00+	p=0.64	p=0.008	p < 0.00+	p < 0.00+
<i>Quartiles</i>					
4th (top 25% of teams)	\$1,410	\$174	\$273	\$421	\$542
3rd (50-75%)	\$2,637	\$210	\$560	\$824	\$1,043
2nd (25-50%)	\$5,017	\$385	\$1,143	\$1,688	\$1,801
1st (bottom 25% of teams)	\$10,143	\$879	\$2,812	\$3,323	\$3,129

Table 4. Summary of experimental results from the database (n=44).

	Customer	Retailer	Wholesaler	Distributor	Factory
<i>Periodicity</i>					
Time to recover initial inventory	-	24	27	26	25
Date of minimum inventory	-	17	21	20	20
Date of maximum inventory	-	33	35	35	35
<i>Amplification</i>					
Peak order rate (cases/week)	8	16.9	23.8	39.0	59.5
Variance of order rate (cases/week)	1.6	14.4	42.1	138.1	293.7
Peak inventory (cases)	-	69.84	73.45	152.80	144.27
Minimum inventory (cases)	-	-4.02	-56.05	-64.27	-56.00
Range (cases)	-	73.9	129.5	217.1	200.3
Closed-loop gain*		3.23	4.95	8.74	13.89
<i>Phase lag</i>					
Date of peak order rate (week)	5	18	20	18	19
Anticipated minimum inventory		-1	1	2	1

Table 5. Evaluation of results from the top 10 performing teams in the database used for comparison to teams reported in the literature.

	Team total	Retailer	Wholesaler	Distributor	Factory
Top 10 performing teams (n=10)	\$1278	\$172	\$228	\$335	\$543
Sterman 1989a (n=11)	\$2028	\$383	\$635	\$630	\$380
t-statistic	-8.45	-22.69	-20.81	-6.67	2.59
H ₀ : mean cost = Sterman 1989a	p < 0.00+	p < 0.00+	p < 0.00+	p < 0.00+	p=0.02
Periodicity					
Time to recover initial inventory	-	22	22	22	22
Date of minimum inventory	-	12	16	16	14
Date of maximum inventory	-	24	27	30	27
Amplification					
Peak order rate (cases/week)	-	10.6	12.1	15.9	21.1
Variance of order rate (cases/week)	-	2.8	4.7	13.3	27.9
Peak inventory (cases)	-	14.2	24.2	37.1	41.1
Minimum inventory (cases)	-	-2.2	-3.4	-9.7	-12.7
Range (cases)	-	16.4	27.6	46.8	53.8
Closed-loop gain*	-	1.65	2.03	2.98	4.28
Phase lag					
Date of peak order rate (week)	-	17	13	16	17
Anticipated minimum inventory	-	-5	3	0	-3

Table 6. Comparison of King Ranch Institute for Ranch Management (KRIRM) and South Dakota State University Honors College (SDSUHC) class performances.

	Team total	Retailer	Wholesaler	Distributor	Factory
KRIRM (n=38)	\$4,988	\$400	\$1,227	\$1,644	\$1,717
SDSUHC (n=6)	\$3,623	\$488	\$1,008	\$1,057	\$1,070
t-statistic	0.809	-0.475	0.367	1.005	1.095
H ₀ : KRIRM = SDSUHC	0.423	0.637	0.715	0.321	0.280
<i>KRIRM</i>					
Inventory recovery time	-	24	27	26	25
Week of minimum inventory	-	17	21	19	20
Week of maximum inventory	-	35	35	35	35
Peak order rate (cases/week)	-	17.3	24.4	41.2	59.5
Mean order rate (cases/week)	-	9.5	11.9	16.9	21.2
Mean order rate post-inventory recovery	-	7.1 ^a	7.2 ^b	9.2 ^a	13.8 ^b
Variance of order rate	-	13.8	42.2	147.0	304.3
Peak inventory (cases)	-	67.8	76.4	163.4	148.5
Minimum inventory (cases)	-	-4.4	-54.7	-64.8	-54.6
Week of peak order rate	-	18	21	18	19
<i>SDSUHC</i>					
Inventory recovery time	-	23	27	26	25
Week of minimum inventory	-	13	20	22	20
Week of maximum inventory	-	29	35	30	32
Peak order rate (cases/week)	-	19.3	25.2	44.3	60
Mean order rate (cases/week)	-	9.6	10.7	13.2	15.9
Mean order rate post-inventory recovery	-	6.0 ^a	5.5 ^b	5.0 ^a	5.2 ^b
Variance of order rate	-	20.3	51.2	123.6	276.6
Peak inventory (cases)	-	95.3	54.5	98.3	143.5
Minimum inventory (cases)	-	-2.7	-65.5	-69.5	-64.7
Week of peak order rate	-	17	20	17	19

a- p-value < 0.10

b- p-value < 0.05

Figures

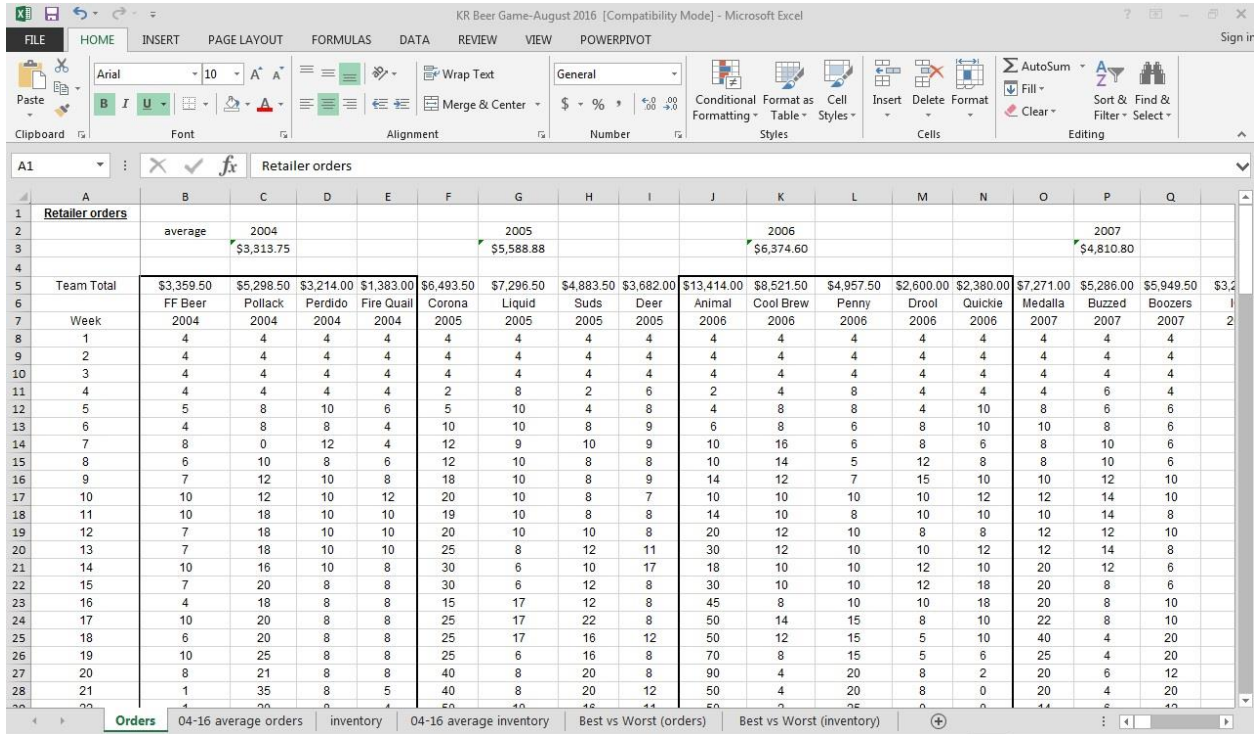


Figure 1. Image of the Beer Game database. Tabs of recorded data included participant orders (tab displayed) as well as inventory. Additional tabs include graphs of average, best, and worst performances, used as visual aids in the debriefing of participants in the KRIRM and SDSUHC classes.

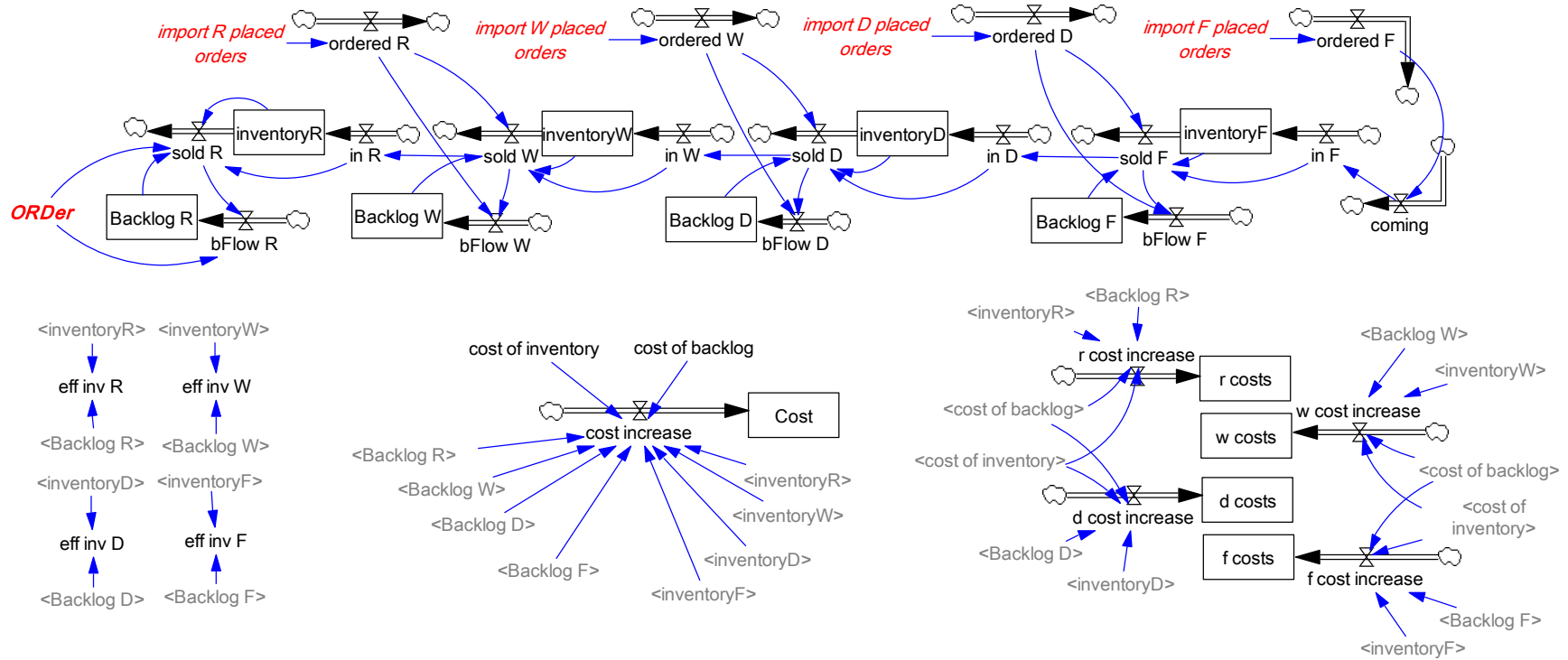


Figure 2. Beer Game model developed following Sterman (1989a) and Kirkwood (1998). The structure is identical except that the algorithm used to compute placed orders in the Kirkwood (1998) model formulation was replaced with observed placed orders of participants (indicated by the red italicized import variables) using the record sheets inside the Beer Game database. Equations to replicate the model are provided in the references cited above as well as the Supplementary Material.

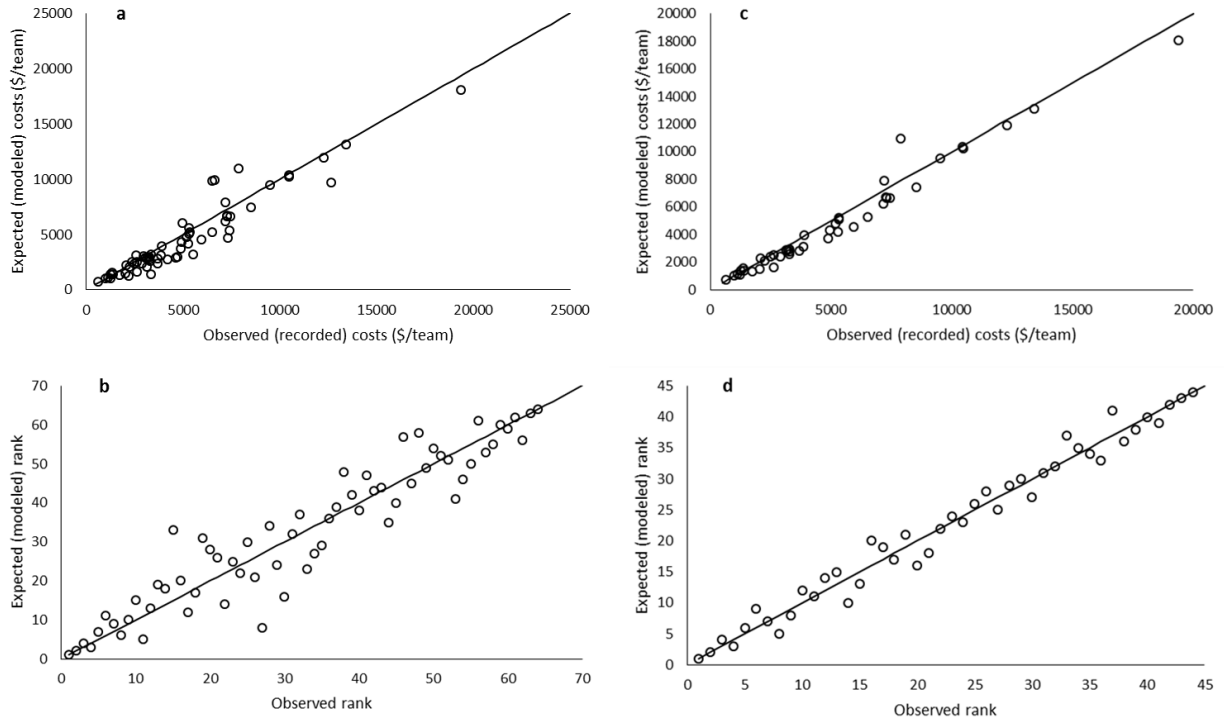


Figure 3. Comparative plots of observed and model expected team costs and ranks before discarding outlier teams from the dataset (panels a and b) and after (panels c and d). The black lines display $x=y$.

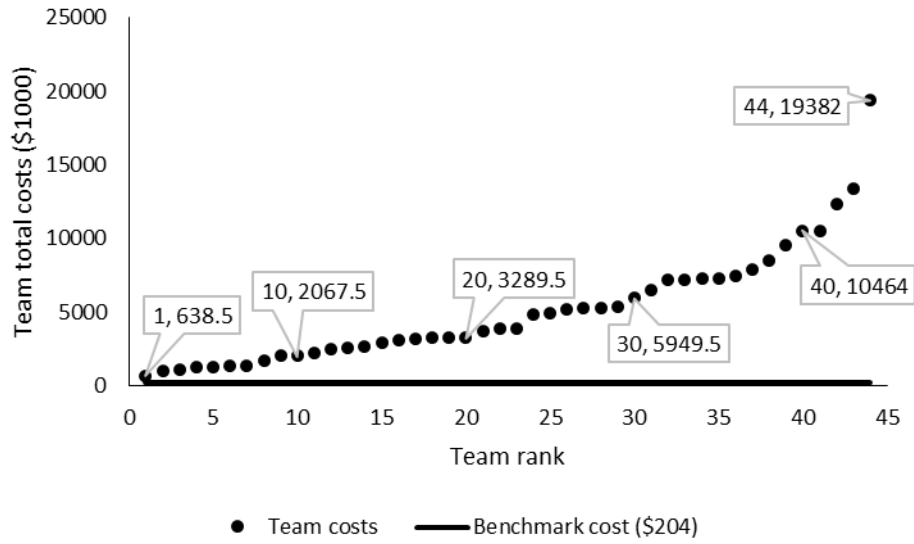


Figure 4. Distribution of team total costs from best (1) to worst (44) teams across the database relative to benchmark costs, with ranked teams 10, 20, 30, and 40 shown to illustrate the change in costs with progressively worse performances.

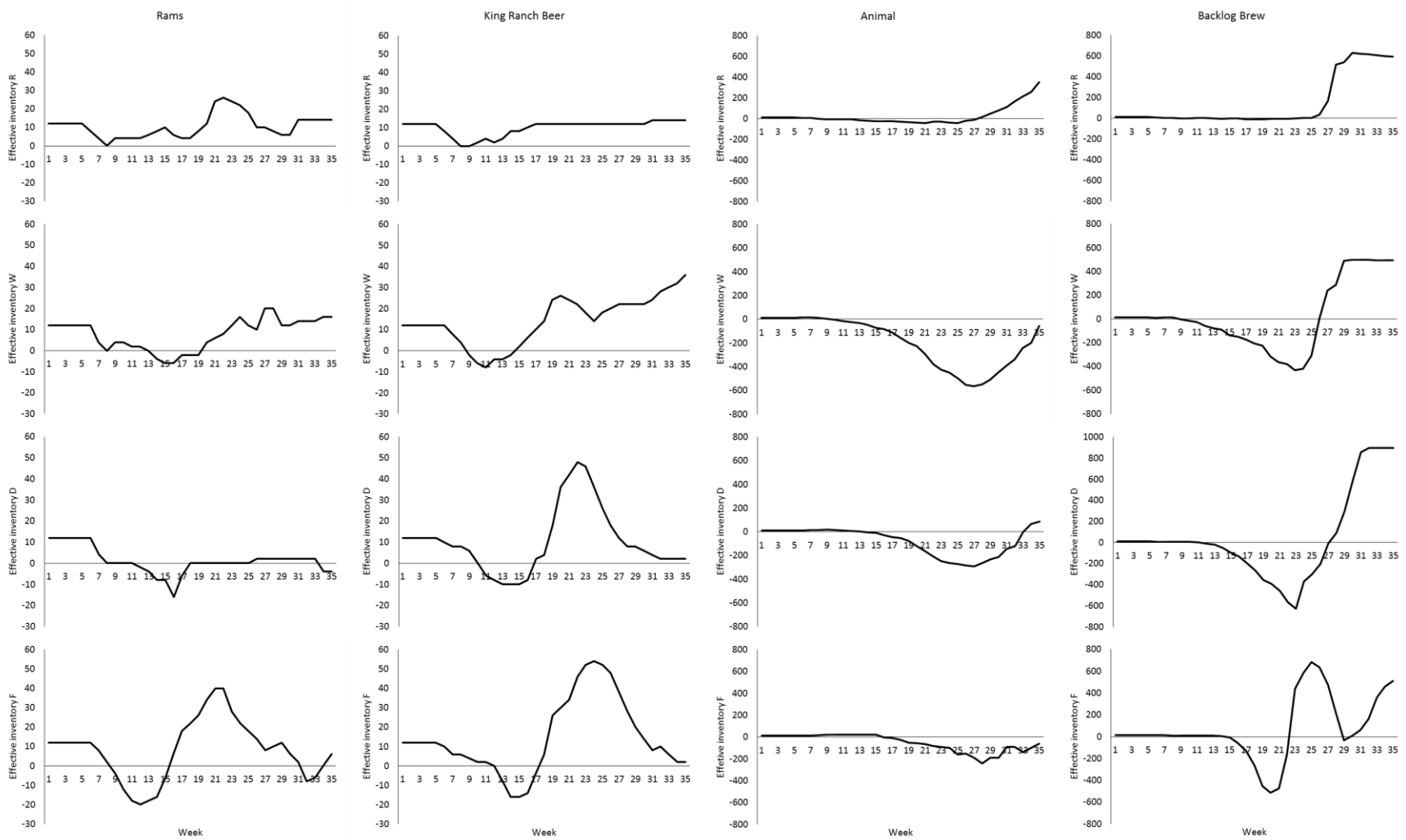


Figure 5. Illustration of effective inventories for the best two and worst two teams (R-retailer; W-wholesaler; D-distributor; F-factory) in the database (note the scale differences of nearly an order of magnitude).

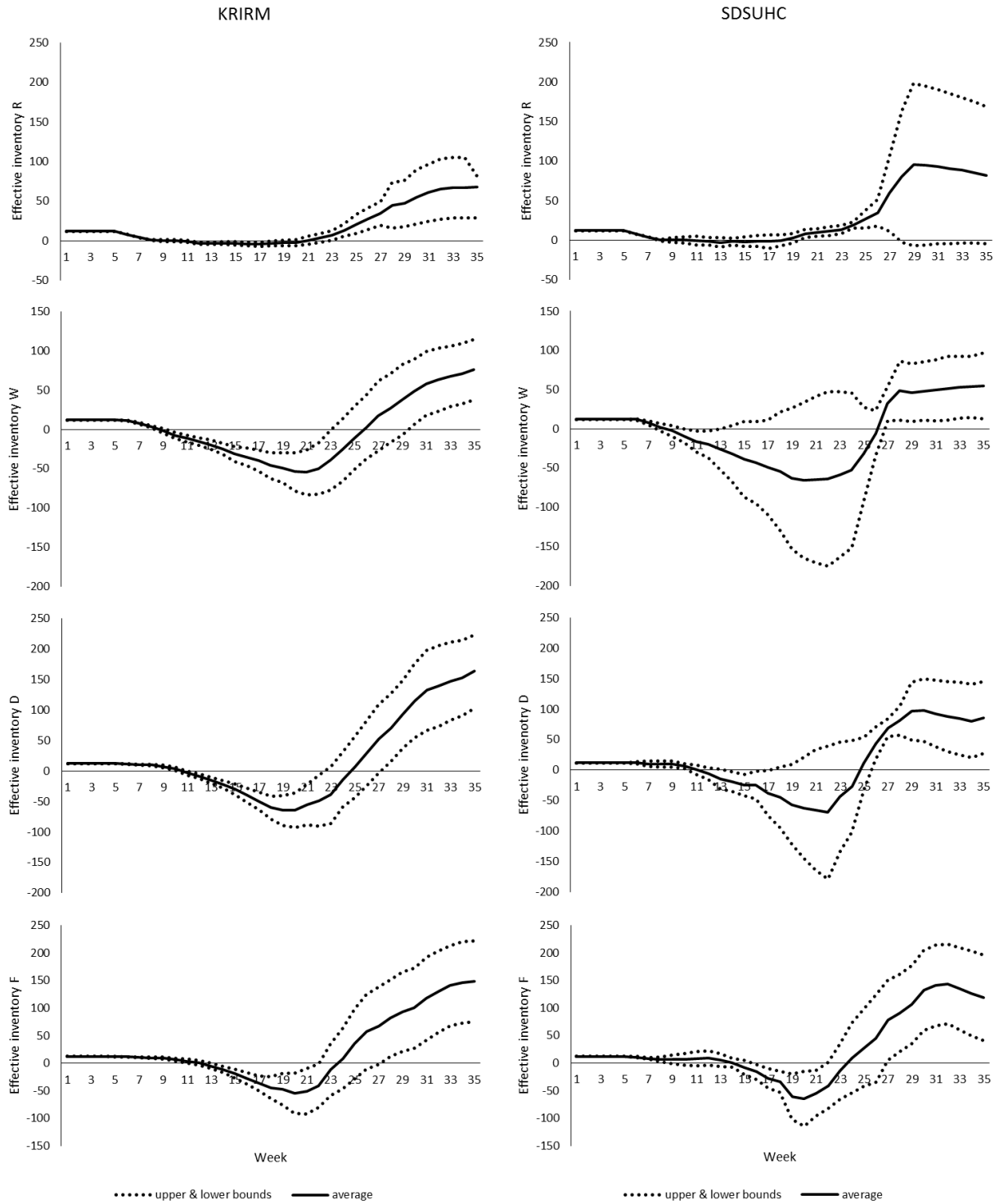


Figure 6. Comparison of mean effective inventories of King Ranch Institute for Ranch Management (KRIRM) and South Dakota State University Honors College (SDSUHC) participants. Upper and lower bounds represent 95% confidence intervals for each week.

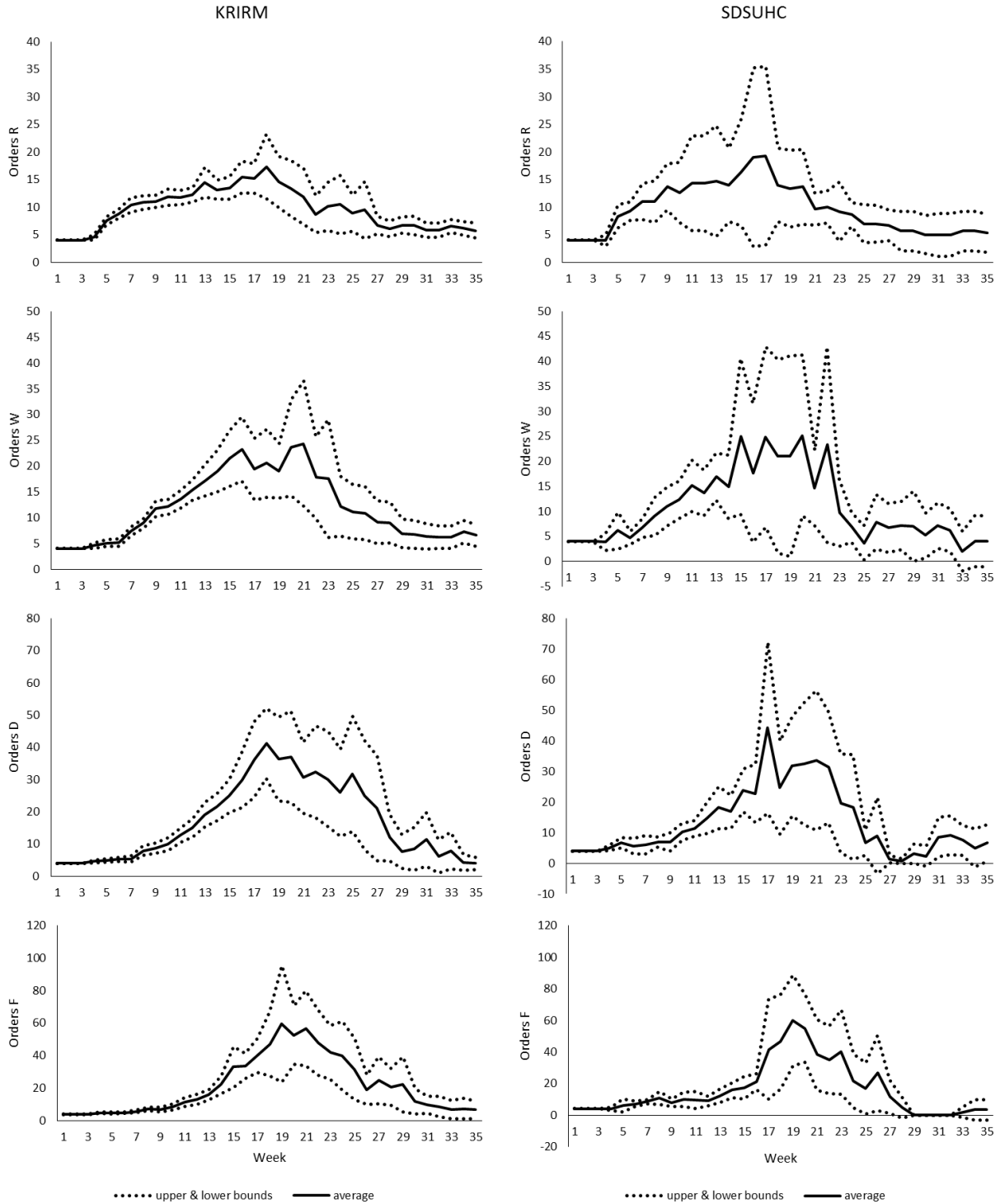


Figure 7. Comparison of mean order rates of King Ranch Institute for Ranch Management (KRIRM) and South Dakota State University Honors College (SDSUHC) participants. Upper and lower bounds represent 95% confidence intervals for each week.