Introducing Gamettes: A Playful Approach for Capturing Decision-Making for Informing Behavioral Models

Omid Mohaddesi, Yifan Sun, Rana Azghandi, Rozhin Doroudi, Sam Snodgrass, Ozlem Ergun, Jacqueline Griffin, David Kaeli, Stacy Marsella, Casper Harteveld
Northeastern University, Boston, MA, United States
{mohaddesi.s, azghandi.r, doroudi.r, o.ergun, ja.griffin, c.harteveld}@northeastern.edu
{yifansun, kaeli}@ece.neu.edu; sam.psnodgrass@gmail.com; marsella@ccs.neu.edu

ABSTRACT
Agent-based simulations are widely used for modeling human behavior in various contexts. However, such simulations may oversimplify human decision-making. We propose the use of *Gamettes* to extract rich data on human decision-making and help in improving the human behavioral aspects of models underlying agent-based simulations. We show how Gamettes are designed and provide empirical validation for using Gamettes in an experimental supply chain setting to study human decision-making. Our results show that Gamettes are successful in capturing the expected behaviors and patterns in supply chain decisions, and, thus, we find evidence for the capability of Gamettes to inform behavioral models.

Author Keywords
decision-making; human behavior; simulation; agent-based model; gamette; supply chain; beer game

CCS Concepts
•**Human-centered computing** $\rightarrow$ **User studies; Empirical studies in HCI;**

INTRODUCTION
Understanding how humans make decisions is of interest to many different fields, from psychology to industrial engineering. A variety of methods are deployed to get such understanding, including simulation which is the second most widely used method for modeling and analyzing decision-making [2]. In our paper we focus on the use of agent-based modeling (ABM) among different simulation techniques—which we refer to as "agent-based simulations"—as it pertains to modeling behavior and simulating human systems [8, 36]. While simulations are practical for quantitative modeling and for providing qualitative insight into the dynamics of systems, it is much more difficult to model human behavior using these methods [32]. It is a difficult task because we typically do not have enough information about internal mechanisms that derive human decision-making. Because of that, human behavior is a black box whose internal mechanisms are modeled using a high level abstraction [40]. To address the oversimplification, abstraction, and simply the lack of behavioral aspects in traditional simulations, there is a growing trend toward a participatory design of simulations, and why we need to "put the human in the loop" [40], to be able to inform ABMs.

As part of this trend, Guyot and Honiden [23] introduced an approach called "agent-based participatory simulations", in which role-playing games and agent-based simulations are merged. The main idea is to have real humans play the role of simulated agents to design and then refine the behavior of agents through an iterative modeling process. This approach has some advantages over the companion modeling approach for coupling Role-Playing Games and Multi-Agent Systems (MAS/RPG) [6], including the possibility of recording all events, and analyzing behavior of participants in detail. This helps with understanding why and how the participants make decisions with respect to the information they gather during the simulation. Many other researchers have also tried to put the "human in the loop" for simulation and agent-based modeling, however, there is still room for improvement. Previous work includes studies that (1) created an interface to involve a human decision maker with a part of the simulation [42, 43], or (2) developed a serious game for the whole simulation [60, 29]. The shortcoming of (1) is that a simple interface will not create an authentic and immersive decision-making environment. While (2) addresses this, developing a complete game for the whole simulation requires more resources.

In this paper we introduce *Gamettes*, which are short and relatively simple games that immerse human players in a decision-making scenario and can be easily adapted to test different settings. Gamettes overcome the limitations of prior studies by combining (1) and (2) through providing a decision context for studying a short and specific part of the simulation while benefiting from the affordances of games. Similar to the gaming simulation introduced by Meijer [39] in which human participants enact a role in a simulated environment, our approach helps reduce the complexity of the environment. Of course, the validity of games for studying real world behavior is a point of discussion. According to Raser [47], "A model can be said to be valid to the extent that investigation of that model provides the same outcomes as would investigation in the reference system". Therefore, we validate Gamettes by considering a simulated environment for a drug delivery supply chain and comparing human behavior in Gamettes with expected behaviors and patterns in the supply chain literature. We provide evidence of how to practically use Gamettes for

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376571

modeling and understanding human behavior. In addition, we provide findings on the use of Gamettes for studying human behavior in interacting with different simulation agents.

The contribution of this work is in providing an empirical validation for using a new tool called "Gamette" for studying human decision-making. By using Gamettes we can put the human into the loop of an agent-based simulation and get insight into the black box of human behavior. For our validation, we compare the results obtained through the use of Gamettes in an experimental supply chain setting with the expected behaviors and patterns from the supply chain literature, and we compare how human behaviors may differ across different agent types. The first comparison (through applying Gamettes in the context of the well-known beergame setting) illustrates that behavior in Gamettes is similar to other (game) environments and the literature, providing evidence that Gamettes can be used as a research environment to capture decision-making. The second comparison (through comparing human behavior with two different agents and interface design) is to provide evidence that human behavior is consistent across settings and that, therefore, Gamettes are not sensitive to changes in the design to inform the development of behavioral models.

RELATED WORK

In our review of the related literature, we first focus on participatory approaches in simulations, then look into the use of games as research environments. We also review behavioral studies in supply chains, and the use of games for such studies.

Participatory Simulation

The idea of interacting with humans in the simulation process is not new. The focus of this involvement of humans is on modeling human behavior in a specific context, and as such validate the simulation or improve its calibration. It is explored under different terms such as "participatory approaches for simulation" or "modeling with stakeholders". While differences exist, we refer to all these approaches as *participatory simulations* (for a review, see [5, 61]).

Participatory simulation generally involves interactions between the simulation and participants at all or some stages of the modeling process, including design, implementation, running simulation, and outcome analysis. It has been studied through different techniques, including companion modeling [6, 4] and agent-based participatory simulation [23]. Companion modeling is usually associated with a process that involves a combination of multi-agent systems (MAS) and role-playing games (RPG), with an objective to increase the awareness of the stakeholders of the different viewpoints and the consequences of their actions. The idea behind this combination is the fact that both MAS and RPG involve implementing agents that interact within a shared dynamic environment [5]. This approach has been used at both the design stage with an aim to incorporate stakeholders' viewpoints in the model, as well as in model use for improvements and increasing its legitimacy. Prior work using companion modeling mainly involves environmental issues such as management of erosive runoff [54], forest management planning [52], fishery management [64]. Other researchers also used companion modeling to validate the agent-based models for policy making [34], and sustainable resource management [12, 21, 33].

A specific form of participatory simulation, and highly relevant for our work, is *agent-based participatory simulation*, which focuses on merging the RPG and MAS in the sense that human participants control some of the agents in the simulation [23]. This method mainly focuses on involving participants in the use of simulations rather than in the design process and has some advantages over the MAS/RPG for companion modeling. First, all interactions and events can be recorded and analyzed to improve the understanding of participants. In addition, merging of MAS/RPG helps in decreasing the gap between the agent-based model and the game the participants are involved in. Lastly, this approach allows for adding assistant agents with learning capabilities to the interface for improving the understanding of the model by the participants. An example of the assistant agents can be found in the work by Cacciaguerra and Roffilli [10], where they adopted a virtual player called "ghost player" with the aim at discovering emergent social behaviors through the simulation. Agent-based participatory simulation has also been used to help participants to improve understanding of complex systems components such as emergence, nonlinear interactions, action effects, adaptation, order, and tradeoffs [48, 37].

Our work can be considered an agent-based participatory simulation and is most similar to [1] who proposed a participatory approach for validating agents of an ABM via engaging humans with a simulation process. Our work builds forth on this and expands it to a much larger scale—with more participants and a more complex, interdependent system—and in the context of different agent models. In addition, our approach is different compared to [1] by using short game-based scenarios for immersing humans into the simulation process.

Games as Research Environments

The methodology we propose in this paper belongs to the sparse but increasing use of games and game-like environments as research environments [11, 3]. Games are especially attractive to study decision-making because they are essentially about making decisions [26]. More importantly, games and game-like environments offer affordances that make them appealing for use in research, including (1) the ability to immerse people in authentic situations, and observing phenomena that are difficult to study in reality; (2) retrieving rich behavioral data in a controlled but unobtrusive manner; and (3) engaging a broad global audience [27]. Immersing participants is important because participants may not respond authentically when they are put in artificial environments and are engaged with tasks that do not have any real consequences. A game, although virtual, is more ecologically valid because players have goals they need to achieve and their actions do have real consequences. Additionally, for the most part, individuals behave similarly in both virtual and real environments [7]. However, due to the virtual nature of games, one should be cautious about the validity of the outcomes. For this reason, part of this paper focuses on validating the use of Gamettes compared to the expected outcomes in the literature. As a research environment, Gamettes are short experiences designed by the researchers, rely on game telemetry data (and thus game analytics, see [20, 63]) for data collection, and this data collection needs to be mapped to a simulation framework. Like any method, Gamettes have their limitations. For example, we cannot collect any data that involves measuring internal cognitive processes ("thoughts") or affective states.

Supply Chains

Studying human behavior and decision-making is one of the most popular areas in the supply chain literature [2]. This is mainly because of the impact of individual decisions on supply chain performance as a whole [41]. Human behavior and the impact of individual decisions on the dynamics of supply chains has been explored using behavioral modeling and mainly through "beer game" studies [55, 16, 15, 13, 57]. The beer game is a role-playing experiment, where participants play the role of different entities within a beer supply chain. In this paper, we apply Gamettes to the context of drug delivery supply chains and use a setting similar to the beer game to validate the Gamettes by showing that our results are in line with the expected outcomes and behaviors according to prior research, such as the "Bullwhip Effect" [35]. Many researchers have validated their results by investigating the presence of this effect in their experiments [16, 41, 15].

GAMETTES

Gamettes are short game-based scenarios where individual decision makers are immersed into a specific situation and have to make decisions by responding to a dialog or taking actions. The term Gamette is a contraction of "game" with "vignette". Similar to a vignette, a Gamette aims to provide a *brief description* of a situation as well as to *portray* someone. In this study, we focused on modeling human behavior and decision-making in a drug delivery supply chain network. Based on our prior work [19], we designed an integrated simulation framework consisting of a *Flow Simulator* for simulating the dynamics of the supply chain, and a *Gamette* environment for immersing human decision makers into a specific role and particular state of the supply chain (see Figure 1).

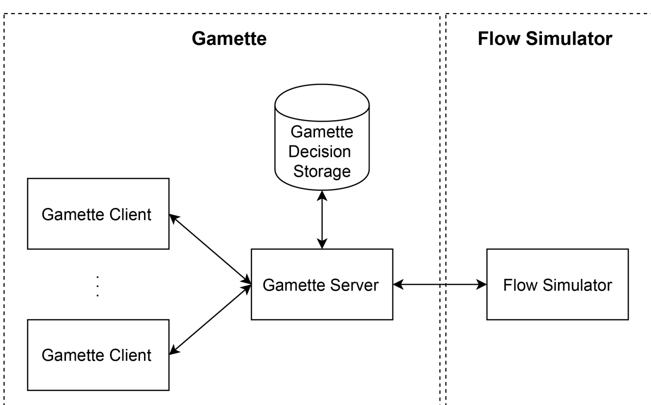


Figure 1: The integrated simulation framework. The Flow Simulator simulates the network and communicates with the Gamette Server to receive player input from the Gamette Client and send updated parameters to the players. Player decisions are also stored ("Gamette Decision Storage") for later gameplay analysis.

Drug Delivery Supply Chains

This study is part of a larger project on studying supply chain resiliency and the role of human behavior in drug shortages, motivated by the growing trend in reported drug shortages in the United States [28, 58]. Within this context, the role of human behavior is often under-explored, particularly when disruptions happen. Studying human behavior in the presence of disruptions is out of the scope of this paper, and is a point of interest in our future work. However, before using Gamettes to study human behavior with disruptions, we need to validate the use of Gamettes to explore human decision-making in supply chains as well as in interacting with agents.

Simulation Framework

The Flow Simulator is a multi-agent simulation environment designed based on Partially Observable Markov Decision Processes (POMDPs) [49]. A POMDP is a general framework to model sequential decision problems where the state of the system is not completely observable by the agent. Within this framework, an agent can estimate the outcome of its actions even when it makes incomplete observations of the system. A POMDP is represented as a tuple (S, A, T, R, Z, O) where:

- S is the set of all system states capturing all information relevant to the agent's decision-making process.
- A is the set of all possible actions.
- T is the transition function, where $T(s,a,s') = Pr(s'|s,a)$ represents the probability of ending in state s' if the agent takes actions a in state s.
- R is the reward function, where $R(s,a)$ is the immediate reward obtained by taking action a in state s.
- Z is the set of all possible observations.
- O is the observation function, where $O(s',a,z) = Pr(z|a,s')$ represents the probability of observing z if action a is executed and the system ends in state s'.

According to the POMDP framework, at any given time, the agent only has access to some observation $z \in Z$ that gives incomplete *information* about the current state. Thus, the agent takes an *action* considering a complete history of its past actions and observations, and then receives some *reward* according to R. Within our framework, the Flow Simulator is the central hub that simulates the information and physical flow in the drug delivery supply chain over time. The information and physical flows are driven by the decisions and actions taken by the agents of the system, such as manufacturers, distributors, and health centers. Such decisions and actions can be the result of running the Flow Simulator in a *standalone mode*. In that case, the Flow Simulator simulates the evolution of the supply chain system with predefined policies informing decision-making—therefore, without any human agents. Or the Flow Simulator can simulate by fetching information from the Gamette clients, which capture decisions of human players. Here, the evolution of the supply chain system is a result of human input. Figure 1 illustrates the architecture of our simulation framework and the interaction of its components.

Gamette Design

The aim of each Gamette is to provide a brief description based on a particular state of the supply chain network for a specific agent in the Flow Simulator in order to portray that agent's role. The design of the Gamettes follows the same idea of the POMDP framework by allowing players to collect <*information*>, take <*actions*>, and receive <*rewards*>. By using the drug supply chain as the decision context, we model an inventory management task where <*information*> maps to inventory, demand, backlog and shipment data; <*action*> to choosing order amount; and (negative) <*reward*> to inventory and backlog cost. This mapping between Gamettes structure and the simulation makes it easy to utilize Gamettes for a short slice of the simulation.

Gamettes are created with *StudyCrafter*, a playful platform where users can create, play and share gamified projects for behavioral and social science research [24, 25, 44].[1] Using this platform, Gamettes were constructed with *scenes* and *characters* for providing a setting to describe the state of the system and to ensure the authenticity of the decision-making process. A *scene* is a setting where a scenario unfolds, such as a hospital or an office, and includes objects that the player can interact with (e.g., computer, phone). We refer to these interactions with objects as *actions*. In addition to these interactions with objects, players (in the form of their assigned player-character) can interact with non-player characters (NPCs) through *dialog*. This dialog may involve the player-character *thinking* (via thought bubbles) and expressing beliefs about others. Each Gamette starts with a tutorial, which provides player's *reward function*, specifies player's objectives and performance evaluation criteria, and explains how to play. After this tutorial, players are presented with *choices* during several time periods as they take actions or when talking to NPCs, and their responses to these choices form the input for the agents in the Flow Simulator. Figure 2 shows screenshots of the tutorial, interaction with NPCs, and how decisions are made.

A unique feature of our approach is the tight coupling of the Flow Simulator with the Gamette. Each Gamette communicates closely with the Flow Simulator using a RESTful API to receive supply chain information and to send actions and decisions of players back to the Flow Simulator. The supply chain parameters (e.g., current inventory, received orders, on-order amounts, etc.) are received by the Gamette at the beginning of each time period. The player makes decisions and these decisions are posted to the Flow Simulator at the end of each period. The simulator then moves to the next time period and updates the network data.

Gamette Gameplay

The design of each Gamette consists of four phases: (1) briefing, (2) gameplay, (3) debriefing, and (4) survey. Each game starts with the *briefing phase*. In the first scene (Figure 2a) players first see their character, and then are informed about the task and the goal of the game. According to the beer game experiment, the goal of the game is framed as "minimizing the total cost". In addition, the difference between inventory cost and backlog (stockout) cost is explained. This information is delivered to players through dialog with a Non-Player Character (NPC). Players are also informed about the 2 buttons on the upper left corner of the screen. They can click on these at any time to check the goal as well as the current and cumulative total cost. The briefing continues in the second scene (Figure 4b) with a simple image of a supply chain network on a TV screen and an explanation of the key players' task: choosing an order amount to order a drug from the supplier. Next, players are provided with options to ask questions about lead-time, inventory and stockout cost, and how to place orders. The briefing ends after all these questions are asked and the NPC provided the necessary information.

[1]https://studycrafter.com

(a) Briefing and tutorial

(b) Non-Player Characters (c) Decision-Making

Figure 2: Gamette Design. (a) Briefing player about their role and their task in the game, (b) Interacting with NPCs, and (c) Interacting with a laptop computer and making decisions on how much to order.

After the briefing phase, the *gameplay phase* starts, where players can interact with a laptop computer to observe information and make decisions. By clicking on the laptop (Figure 2c) they will enter a management system where they can click on different buttons to look at the current inventory level, current demand, and shipments received from their supplier. Then, they can click on the "ordering" button to type-in an order amount and submit their orders. The gameplay phase continues for 20 weeks and each week players can review their inventory, demand, shipments and costs, and then place an order. After Week 20, the *debriefing phase* starts, where an NPC explains the purpose of the game, providing some background on the beer game, and showing cost and inventory level graphs to display their performance. In the last phase, the *survey phase*, players are asked to complete a short survey querying their interest in the game, knowledge of the context, and demographics.

METHODS

In this study, we used the well-known "Beer Distribution Game" or beer game [55] setting in a drug delivery supply chain to test the validity of Gamettes for supply chain decision-making. The beer game setting is well suited for examining behavioral factors in supply chain decision-making as it is simple enough to be quickly learned, while retaining key features of real supply chains [16]. Using the architecture of agents in the Flow Simulator, we designed an experiment similar to the beer game [16], which consists of four agents (health center, wholesaler, distributor, and manufacturer) each making decisions on how many drugs to order from their upstream link in a serial supply chain with patient demand at the health center (see Figure 3). In addition to validating the Gamettes compared to the outcomes of this traditional beer game (i.e., the bullwhip effect), our work seeks to validate if human behavior changes or stays the same contingent on the type of agents human agents/players are interacting with.

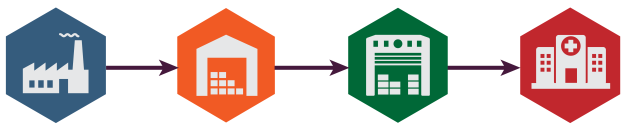


Figure 3: Supply chain network structure including manufacturer, distributor, wholesaler and health center. The wholesaler role is played by human players using Gamettes.

Hypotheses

Prior research on supply chains has observed that misunderstanding about the inventory and demand information results in a behavior that leads to the *bullwhip effect* [55, 35]. This effect is defined as the amplification of orders when moving up in the supply chain [35], and can be measured for each echelon through a metric called Bullwhip Effect Index:

$$BEI_t = \frac{Var(O_t)}{Var(D_t)} \tag{1}$$

where $Var(O_t)$ is the variance of orders placed to upstream supplier by time t and $Var(D_t)$ is the variance of downstream incoming orders by time t. Because this phenomenon has been frequently observed in supply chain experiments and particularly with beer game studies, different researchers try to validate their experimental results by providing evidence of the bullwhip effect [41, 15]. Thus, we form our first hypothesis by looking for evidence of a bullwhip effect to test the validity of using Gamettes in supply chain decision-making.

H1. *In Gamettes simulating the beer game in the context of the drug supply chain, human decision making results in the bullwhip effect.*

The bullwhip effect is caused by both the irrational behavior of managers such as over-ordering when they face low inventory [55], and their rational behavior such as order batching or hoarding [35]. These behaviors are due to judgmental biases that not only ordinary people but experts and professionals suffer from [56]. Therefore, if the bullwhip effect exists, we expect to observe similar behavior regardless if players interact with rational order-up-to-level (OUL) or irrational human-like agents (see Agent Design) as long as the participants are not informed about different agent types.

H2. *There is no significant difference in players' order amounts between players interacting with OUL or human-like agents.*

Prior work used the anchoring and adjustment heuristic to model human decision-making under uncertainty [55], and showed that people often make decisions by starting from an initial value (anchor) and then adjusting that to make a final decision [59]. If this initial value is suggested to the decision maker, their adjustment is expected to be insufficient and as such, their final decision to be biased toward the suggested anchor. Thus, we expect to see this anchoring in players' behavior when providing them with optimal order suggestions according to an optimal ordering policy. In fact, the question is if players change their behavior and if their decisions will be biased towards the optimal order amount.

H3. *Providing optimal order suggestions to players will bias their decisions towards the suggested order amounts.*

Agent Design

We designed Gamettes where the human players only play the role of the wholesaler. We made this choice because with the wholesaler we can directly observe the contributions to the bullwhip effect. In designing the other agents (health center, distributor, and manufacturer), we assume two different ordering behaviors. First, we apply the following decision rule [55, 16] for orders $O_{i,t}$ placed by agent in role i in period t:

$$O_{i,t} = Max\{0, CO_{i,t} + \alpha_i(S_i^{'} - S_{i,t} - \beta_i SL_{i,t}) + \varepsilon_{i,t}\} \tag{2}$$

where CO is the expected incoming customer orders, $S^{'}$ is desired inventory, S is actual on-hand inventory including backlog, and SL is the supply line of unfulfilled orders (on-order units). In fact, the order amount for each agent is modeled as the replacement of expected incoming orders modified by an adjustment to bring the inventory in line with the target, which essentially is an attempt to model the anchoring and adjustment heuristic [41, 59]. The parameters α and β are the fraction of the inventory shortfall or surplus ordered each week, and the fraction of the supply line that the decision maker considers (on-order units that are not yet received), respectively. We call the agents with this type of behavior "Human-Like" agents as they try to mimic human behavior.

Second, we consider a base-stock policy, which assumes that all agents place orders or produce enough to bring their inventory position to a predefined level (order up-to level) based on a periodic review policy with zero fixed costs [53]. This policy is optimal if (1) demands are stationary, (2) the lead-time is fixed and there is no production limit, and (3) there is no fixed ordering cost and no changes in the purchase cost of the product over time. According to this policy, we assume that each agent is a rational decision maker and uses an approximately optimal local policy to place orders. In fact, each agent considers the base stock (S_{it}) over time which corresponds to the anchoring heuristic, and the inventory position (IP_{it}) which is equivalent to the adjustment heuristic. Thus, the indicated order rate for agent i (IO_{it}) is a function of inventory level, demand, lead-time, service level, and the amount that has been ordered but has not been received (on-order):

$$IO_{it} = S_{it} + IP_{it}. \tag{3}$$

In addition, the base stock level at time t is given by

$$S_{it} = \mu_{it}^{L} + z_{\alpha}\sigma_{it}^{L} \tag{4}$$

where μ_{it}^{L}, and σ_{it}^{L} are the mean and standard deviation of the stationary demand for echelon i over the fixed lead-time L, respectively. In addition, α is the cycle service level and z_{α} is the corresponding value to α obtained from the normal distribution table. We call these types of agents “Order Up-to Level” agents or in short OUL.

Experimental Design

We set up an experiment with 3 conditions with 2 agent types (human-like vs. OUL) for the health center, distributor, and manufacturer roles, and 2 options for order suggestions (with vs. without suggestions). Order suggestions are only provided when the other agents are OUL agents because it is not feasible to provide order suggestions according to the order up-to level policy in the condition with human-like agents as they try to mimic human behavior (and thus do not follow an order up-to policy). In all 3 conditions participants played the wholesaler role. Table 1 summarizes the settings for each condition.

Table 1: Summary of the experiment settings for Condition 1, 2 and 3

	Condition 1	Condition 2	Condition 3
Manufacturer	Human-like	OUL	OUL
Distributor	Human-like	OUL	OUL
Wholesaler	Human Player	Human Player	Human Player
Health Center	Human-like	OUL	OUL
Order Suggestion	No	No	Yes

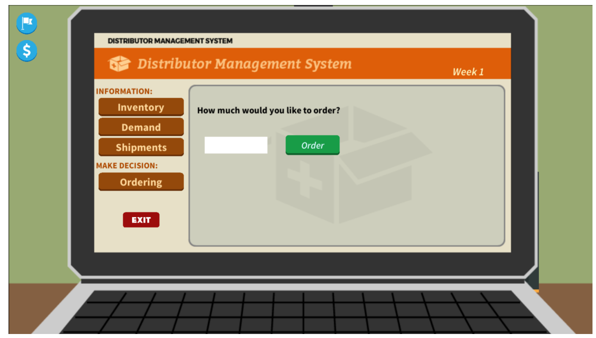
The demand distribution is the same for all 3 conditions and follows the beer game setting with a minor change to adapt it to the drug delivery context. We used a constant demand of 40 units until Week 4 and constant demand of 80 units after that.

Participants and Material

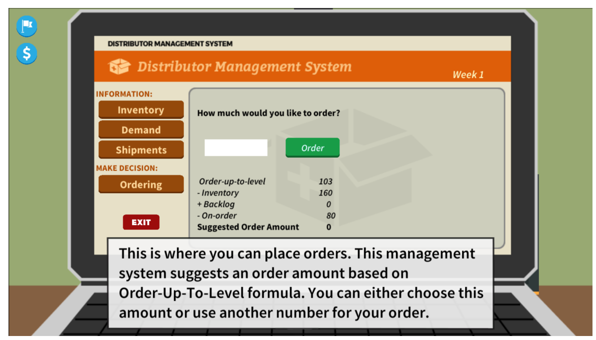
We recruited 68 voluntary participants (37 males, 10 females, and 21 not stated). Participants were full-time students enrolled in undergraduate and graduate level courses of Operations Research and Lean Manufacturing at Northeastern University. The age range is 19 to 35 years (M = 23.1, SD = 2.5). We created an online web page, which describes the purpose of the study and where the gamettes could be played on. Each participant used their own laptop to access this web page. At the end of each Gamette, we included the interest/enjoyment scale from the Intrinsic Motivation Inventory (IMI) [38]. In addition, we included three 7-point Likert scale items to test participants’ general knowledge of supply chains and a yes/no question asking if they had prior experience with any supply chain games. Twelve participants (18%) stated that they had experience playing the beer game or a similar supply chain game before. In addition, participants neither agreed nor disagreed with the statement “*I have great knowledge of supply chains in general*” (Mdn = 4, IQR = 2). Thus, participants were somewhat familiar with supply chains and some had prior experience playing these types of games.

Procedure

All participants were first formally briefed about the experiment and its purpose. Then, they were randomly assigned to one of the 3 conditions ($N_1 = 22$, $N_2 = 22$, $N_3 = 24$) and started playing. Each Gamette included 20 time periods of information gathering and decision-making. Gamettes in all 3 conditions looked the same in all scenes except for the ordering scene where the condition with order suggestions (Condition 3) included information about order up-to level and suggested order amounts (see Figure 4).



(a) Ordering scene in Condition 1 and 2



(b) Ordering scene in Condition 3

Figure 4: Ordering scenes in Gamettes for different conditions.

Participants played the role of a character named Karl who was hired as a supply chain director in a wholesaler company to administer saline inventory. Karl’s boss expresses the goal of the game as minimizing the cost of inventory and backlog while satisfying the health center’s incoming orders. Players are also informed about the lead time (2 weeks for orders and 2 weeks for shipments), inventory and backlog costs ($0.5 for each unit of inventory and $1 for each unit of unfulfilled demand per period) through dialog with the boss.

At each period, players first receive a shipment from their upstream distributor, and can then review the inventory, backlog, and incoming orders information. Next, they place an order to their upstream distributor. The Gamette sends these player decisions to the Flow Simulator, which moves the simulation to the next period and sends updated parameters to the Gamette for the next round (see Figure 1). After playing for 20 time periods, players complete the survey and are then presented with their results and debriefed about the experiment.

Figure 5: Average order and variation of orders for each role in Condition 1 (human-like agents vs. human player).

RESULTS

Evidence of Bullwhip Effect (H1)

We first looked for evidence of the bullwhip effect in Condition 1. Figure 5 shows the order amounts and variations for each agent in the condition where human players made decisions in the presence of human-like agents. The bullwhip effect is visible, with an increase in order quantities moving up the supply chain from the health center to manufacturer, which is expected and in line with prior literature [55, 15, 16]. In fact, these fluctuations in orders happens while the actual patient demand at the health center is constant. In addition, we calculated the BEI (see Equation 1) to find how much human players contributed to the bullwhip effect compared to human-like agents. Figure 6 illustrates the BEI score for each agent as well as human players. We can observe that human-like agents contributed to the bullwhip effect by showing fluctuations in their ordering behavior over time. Human players showed even more fluctuations and contributed to the bullwhip effect more than human-like agents. These observations support our first hypothesis (H1), indicating that Gamette players underweight the supply line and contribute to the bullwhip effect. To confirm H1 statistically, we used a non-parametric sign test [51]. For each adjacent agent in the supply chain, we considered an increase in the variance of orders as a success and a decrease as a failure. We tested the hypothesis

$$H_0 : \sigma^2_{upstream} = \sigma^2_{downstream}$$
$$H_1 : \sigma^2_{upstream} > \sigma^2_{downstream}$$

at the chance rate of 50%. Table 2 summarizes the sign test results for adjacent agents in Condition 1. Our results showed that 82% of the cases in Condition 1 have $H_1 : \sigma^2_{upstream} > \sigma^2_{downstream}$. Thus, we reject the null hypothesis and conclude that the bullwhip effect exist in Condition 1.

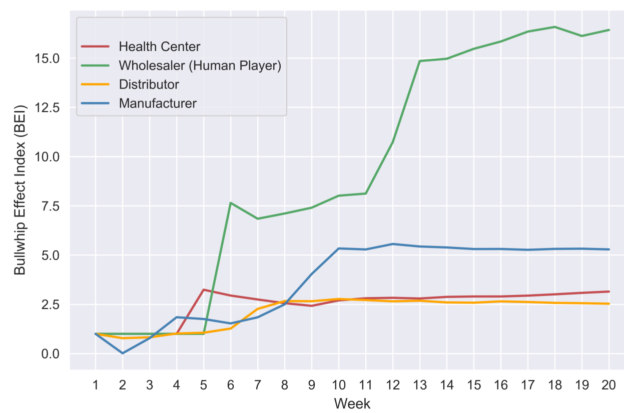


Figure 6: Bullwhip Effect Index for each role in Condition 1

Human Player vs. Agents (H2)

Next, we compared the ordering behavior of human players in Condition 1 with the other conditions. Figure 7 shows the order variations of human players in different conditions over time. Based on this figure, we can infer that human behavior is not affected by the behavior of agents. In fact, the presence of different agent types (i.e., human-like or OUL agent) did not have much impact on the ordering behavior of human players. We compared these observations by performing a one-way ANOVA for each time step to test the null hypotheses that or-

Table 2: Sign test results for comparison of order variances

	Condition 1
Success rate (%)	82
Overall p-value	$p < 0.001$
Health Center vs. Wholesaler (Player)	$p < 0.001$
Wholesaler (Player) vs. Distributor	$p = 0.022$
Distributor vs. Manufacturer	$p = 0.059$

der variations in different conditions have the same population mean for each time period. Table 3 summarizes the results, which show there is not a significant difference between order means of players in the different conditions. We also examined the data by estimating Bayes factors using Bayesian Information Criterion [62] to test the relative likelihood of the null hypothesis compared to the alternative hypothesis. Estimated Bayes factors suggest that positive evidence exists for supporting the null hypothesis (no difference) for most weeks (14/20; BF_{01} range: 3-10, see Kass and Raftery [31]). We believe that the weak evidence for some weeks (6/20; BF_{01} range: 1-3) is related to the dynamics of the simulation, not the agent types. For example, for Weeks 4-6, the sudden change in demand may have resulted in higher volatility in orders. Thus, the results support our second hypothesis (H2).

Table 3: One-way ANOVA and Bayesian analysis results for comparing means of orders in different conditions at each time period

Week	F-value	p-value	BF_{01}	Week	F-value	p-value	BF_{01}
1	0.26	0.765	6.65**	11	0.58	0.557	5.20**
2	1.24	0.295	3.19**	12	1.10	0.338	3.52**
3	0.81	0.448	4.42**	13	2.30	0.107	1.43
4	1.99	0.145	1.83	14	0.84	0.434	4.31**
5	1.60	0.209	2.45	15	0.94	0.394	3.97**
6	1.73	0.184	2.19	16	1.97	0.146	1.85
7	1.04	0.357	3.68**	17	4.00	0.022*	2.41
8	0.28	0.754	6.58**	18	0.37	0.685	6.10**
9	0.52	0.595	5.47**	19	1.11	0.334	3.50**
10	0.55	0.575	5.34**	20	0.50	0.603	5.53**

* Not significant at $\alpha = 0.01$
** Positive evidence for the null hypothesis (see Kass and Raftery [31])

We also evaluated if the overall supply chain cost for human players differs. Table 4 summarizes the average and standard deviation of overall cost, as well as the ANOVA results for comparing the average cost performance of human players in each condition. Because the results show a significant difference, and the homogeneity of variances was violated (Levene's test, $p = .002$), we performed the Games-Howell post hoc test. Results indicate that there is no significant difference between the cost performance in Condition 1 and 2 (human-like vs. OUL without suggestions) ($p = .053$), while the estimated BF suggests weak evidence for the alternative hypothesis ($BF_{10} = 2.99$). We suspect that this relates to our sample size and the huge variance in Condition 1. In contrast, between Condition 1 and 3 (human-like vs. OUL with suggestions) a significant difference exists ($p = .003$) and strong evidence for the alternative hypothesis ($BF_{10} = 85.01$). Lastly, between Condition 2 and 3, there is no significant difference ($p = .076$), and weak evidence for the alternative hypothesis ($BF_{10} = 2.69$). All this shows that while behavior patterns were similar in all conditions (see Table 3), players may have benefited from the optimal behavior of OUL agents (weak evidence), and clearly benefited from anchoring around optimal order suggestions (strong evidence).

Table 4: One-way ANOVA and Bayesian analysis results for comparing cost performance of players in different conditions

		M	SD	p-value	BF_{10}
(1) Human-like		3116	1681		
(2) OUL w/o suggestions		2153	780		
(3) OUL w/ suggestions		1756	238		
Between Conditions*				$< .001$	
Post Hoc Comparisons†					
(1) Human-like	(2) OUL w/o suggestions			.053	2.99
	(3) OUL w/ suggestions			.003	85.01**
(2) OUL w/o suggestions	(3) OUL w/ suggestions			.076	2.69

* One-way ANOVA
** Strong evidence for the alternative hypothesis (see [31])
† Games-Howell Post Hoc Test

Providing Order Suggestions (H3)

Similar to the previous section, and referring to Figure 7 and Table 3 we conclude that providing order suggestions to players in Condition 3 does not seem to have much impact on their ordering behavior, which contradicts our third hypothesis (H3). An interesting finding, however, is the deviation from suggested order amounts based on the order up-to level policy in Condition 3. Players who played against OUL agents and received order suggestions tend to deviate from those suggested order amounts although they have been informed about how the system calculates these order suggestions. Figure 8 shows the order deviations in the form of error bars on the average suggested order amount. The error bars show how much on average the players deviated from the order suggestions. Players tend to over-order at almost all time periods. More specifically, they tended to deviate more around Week 6 when they discovered the change in the demand and faced a stockout. This is in line with the results of the bullwhip effect: human decision makers start to over-order when they experience a stockout. We further see that the order deviations show convergence to the suggested orders towards the end, and even show under-ordering on Week 20, which indicates a behavioral adjustment to decrease inventory and reduce cost when human players are not facing a (sudden) change in demand.

Player Experience

The results from the interest/enjoyment scale of the Intrinsic Motivation Inventory (IMI) [50, 38] indicate that the majority of participants were interested and enjoyed playing the Gamette (M = 5.4, SD=1.1). As Figure 9 illustrates, 22 participants (32%) even strongly agreed on most statements and only very few participants disagreed or neither agreed or disagreed. The majority further agreed with the statements "*I think my decisions during the game mattered*" (Mdn = 6, IQR=1), and "*I believe lead time had a great impact on my orders*" (Mdn = 6, IQR=1). Therefore, players were engaged with the game and felt that the game was responsive to their decisions.

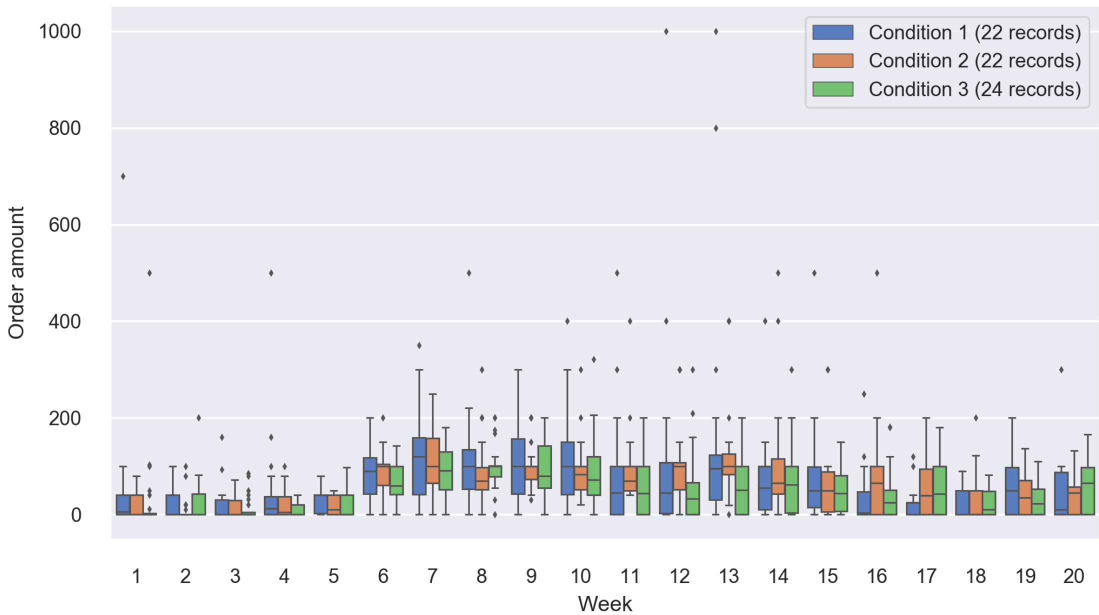


Figure 7: Order variations of human players in different conditions over time

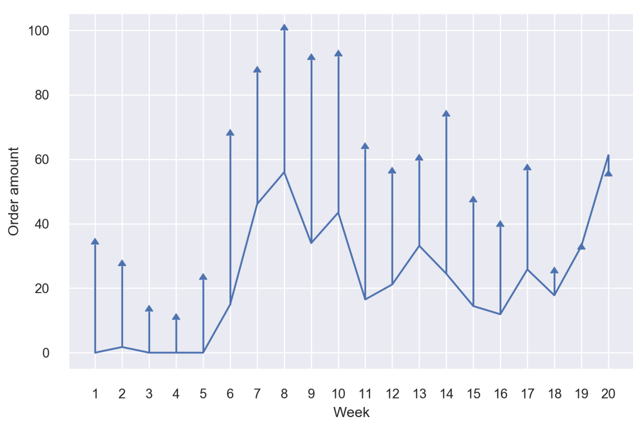


Figure 8: Player deviations from the order up-to level suggested order amounts. The line shows the average suggested amounts for all players in Condition 3 and the error bars show how much on average they deviated from the suggested amount.

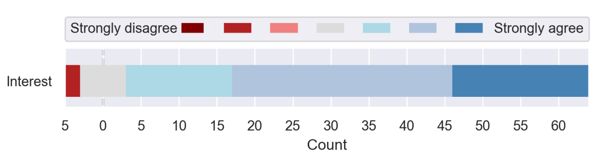


Figure 9: Interest/Enjoyment average score counts obtained from the interest/enjoyment scale of the Intrinsic Motivation Inventory (IMI).

DISCUSSION

Our study provides an authentic game-based approach called Gamettes for extracting data on human decision-making and for informing behavioral aspects of agent-based simulations. Below we describe how our approach and the results of our experiments are related to prior research, showing validity of Gamettes and their potential for informing behavioral models.

The Need for Gamettes

Prior research has described how participatory simulation can be used to put the human into the loop of a simulation process to validate simulation agents or improve the calibration of underlying models [23, 5, 1, 34]. While these approaches are useful, they do not necessarily provide an authentic setting for studying decision-making. In fact, Anand et al. [1] show that by using simulation games in an agent-based participatory framework, human participants have access to all aspects of the simulation which can make the process "unnecessarily complicated". This shows a trade-off between the accuracy of a simulation process and player immersion, and suggests to avoid details that have less impact on the player's decision-making process and make the game complicated [1].

Now one can argue that finding these unnecessary details may require iteration of the research environment to fine-tune the trade-off between accuracy and immersion. Gamettes serve a different purpose by portraying a short and specific scenario for a specific role and thus manage these complexities methodologically. In this way, we can make accurate models while immersing players into a specific scenario without overwhelming them with excessive information about all aspects of the simulation. In addition, we design each Gamette from the viewpoint of serious games and crowdsourcing in order to benefit from the affordances of player immersion, access to rich behavioral data, and engaging a broader audience [27]. According to Jager and van der Vegt [30], the use of serious games in agent-based simulation helps in building more intuitive and interactive simulation environments compared to the games by Anand et al. [1] and Guyot and Honiden [23]. Our results show that most players enjoyed our interface and were interested in playing with Gamettes, and thus, we have been successful in using serious games in our approach.

Prior studies have further shown that using serious games for agent-based simulation can provide a realistic experience to capture the main aspects of human behavior [30]. In this study, we showed that Gamettes can inform behavioral models by capturing aspects of decision-making that otherwise could not have been identified using previous methods. We also showed that human players behave consistently while interacting with different agents and do not change their ordering behaviors with different agent types. In fact, even when provided with optimal order suggestions, human players tend to deviate from these suggestions and behave irrationally in critical situations such as when facing a stockout or a sudden change in demand. Methodologically, Gamettes differ from the existing crowdsourcing and citizen science work [14, 22] by allowing for studying users' actual behavior rather than having them help us achieving some other task. With regard to gamification [17], in essence Gamettes are game-based surveys [25]. However, Gamettes are true games and use this medium to observe, in an unobtrusive manner, what people do rather than to suffer from the typical response bias of surveys.

Gamettes Are Valid Tools

The Gamettes are part of a larger project for studying human behavior when disruptions happen in a drug delivery supply chain [19]. This is motivated by the growing epidemic of drug shortages in the United States, and with a goal to study the understudied role of human behavior in such complex supply chains. We proposed Gamettes as a method to capture behavioral aspects of individual decisions, and to inform behavioral models in drug delivery supply chains. However, one could argue that the validity of such methods must be tested before applying them to new studies. According to Raser [47], "A model can be said to be valid to the extent that investigation of that model provides the same outcomes as would investigation in the reference system". This process, however, is not concerned with comparing two models, and in fact, is associated with validating through the use of the model [45].

Thus, we validated the use of Gamettes through an experimental setting replicating the well-known beer game, but changing the context to the drug deliver supply chain. We used the beer game setting as it is simple, has been explored in many studies [55, 16, 15, 41], and is considered as the state-of-the-art in supply chains. The bullwhip effect is one of the common patterns in beer game studies, which is the result of behavioral adjustment of supply chain decision makers to demand uncertainty [35]. Our results confirmed this behavior for individuals who played the wholesaler role in our experiment. We also observed that participants amplify their orders even more than human-like agents in our simulation. The first implication is that Gamettes are valid tools for behavioral studies. Second, we can conclude that Gamettes have the capability to provide rich behavioral data to inform simulation agents. The latter is confirmed through our results comparing players' actions with different agent types (i.e., human-like vs. OUL). In fact, we find that human decision makers do not make optimal actions even when playing with rational (OUL) agents.

Based on prior work, most real-world systems that deal with human decision-making include various aspects and characteristics—as a result of uncertainties in human behavior—that makes them complex [9]. As a result, ABMs tend to oversimplify the simulation and use high level abstractions [40]. We showed how Gamettes can be used to "put the human in the loop" of ABMs and provide insight into the black box of human behavior. This provides an opportunity to study more complicated systems (such as a more complex network structure with multiple competing roles), and to inform advanced ABMs for modeling human social behavior (such as models using Theory of Mind [46] for modeling trust and social interactions), which are the aims of our future research. For the latter, we will be focusing on using Gamettes in a multi-player setting with several decision makers.

Limitations

We observed that human behavior is similar across different conditions, however, this may not always be the case. In fact, other work has shown how even the player-character (see [18]) makes a difference in how people play. Therefore, further scrutiny is needed to understand under what circumstances design aspects may or may not make a difference. In addition, we worked with a specific audience. There is a question about how results would vary—in terms of engagement and behavior—across different audiences [22]. Also, we worked with simple supply chain system. An additional question is how the approach expands to more complex topics, which may require people to play longer, and cause further issues with recruitment and engagement. Lastly, we applied the Gamettes only in the context of supply chains. Behavior in supply chains may be very specific and have strong tendencies/patterns for people to behave in a particular way. For applying Gamettes to other contexts, it requires essentially to (1) map a simulation framework to a Gamette design (i.e., for us POMDPs) and (2) slicing a simulation into short and specific decision contexts that consist of <*information*>, <*actions*> and <*rewards*>.

CONCLUSION

In this paper, we proposed a new playful approach called Gamettes for studying decision-making and informing behavioral models for agent-based simulations. We provided empirical validation for the use of Gamettes in the context of drug delivery supply chains, by comparing our results with the expected behaviors and patterns in the well-known beer game studies. We showed that Gamettes are successful in capturing these expected behaviors and patterns. In addition, we provided evidence for the capability of Gamettes to inform ABMs and their underlying human behavioral models regardless of the changes in agent or interface design.

ACKNOWLEDGEMENT

This research was developed with funding from the National Science Foundation (NSF: 1638302). We further thank the StudyCrafter development team, the Drug Shortage team at Northeastern University, and everyone else who assisted us with the development and deployment of the gamettes.

REFERENCES

[1] Nilesh Anand, David Meijer, JHR Van Duin, Lóránt Tavasszy, and Sebastiaan Meijer. 2016. Validation of an agent based model using a participatory simulation gaming approach: the case of city logistics. *Transportation Research Part C: Emerging Technologies* 71 (2016), 489–499.

[2] Nasrin Asgari, Ehsan Nikbakhsh, Alex Hill, and Reza Zanjirani Farahani. 2016. Supply chain management 1982–2015: a review. *IMA Journal of Management Mathematics* 27, 3 (2016), 353–379.

[3] William Sims Bainbridge. 2007. The scientific research potential of virtual worlds. *science* 317, 5837 (2007), 472–476.

[4] Olivier Barreteau, Martine Antona, Patrick D'Aquino, Sigrid Aubert, Stanislas Boissau, François Bousquet, William's Daré, Michel Etienne, Christophe Le Page, Raphaël Mathevet, and others. 2003. Our companion modelling approach. (2003).

[5] Olivier Barreteau, Pieter Bots, Katherine Daniell, Michel Etienne, Pascal Perez, Cécile Barnaud, Didier Bazile, Nicolas Becu, Jean-Christophe Castella, William's Daré, and others. 2017. Participatory approaches. In *Simulating Social Complexity*. Springer, 253–292.

[6] Olivier Barreteau, François Bousquet, Michel Étienne, Véronique Souchère, and Patrick d'Aquino. 2014. Companion modelling: a method of adaptive and participatory research. In *Companion Modelling*. Springer, 13–40.

[7] Jim Blascovich and Jeremy Bailenson. 2011. *Infinite reality: Avatars, eternal life, new worlds, and the dawn of the virtual revolution*. William Morrow & Co.

[8] Eric Bonabeau. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences* 99, suppl 3 (2002), 7280–7287.

[9] Sally Brailsford, Tillal Eldabi, Martin Kunc, Navonil Mustafee, and Andres F Osorio. 2018. Hybrid simulation modelling in operational research: A state-of-the-art review. *European Journal of Operational Research* (2018).

[10] Stefano Cacciaguerra and Matteo Roffilli. 2005. Agent-based participatory simulation activities for the emergence of complex social behaviours. *Proceedings of Social Inspired Computing'05* (2005).

[11] Eduardo Calvillo-Gámez, Jeremy Gow, and Paul Cairns. 2011. Introduction to special issue: Video games as research instruments. (2011).

[12] PC Campo, François Bousquet, and TR Villanueva. 2010. Modelling with stakeholders within a development project. *Environmental Modelling & Software* 25, 11 (2010), 1302–1321.

[13] David E Cantor and John R Macdonald. 2009. Decision-making in the supply chain: examining problem solving approaches and information availability. *Journal of Operations Management* 27, 3 (2009), 220–232.

[14] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, and others. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466, 7307 (2010), 756.

[15] Rachel Croson and Karen Donohue. 2006. Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management science* 52, 3 (2006), 323–336.

[16] Rachel Croson, Karen Donohue, Elena Katok, and John Sterman. 2014. Order stability in supply chains: Coordination risk and the role of coordination stock. *Production and Operations Management* 23, 2 (2014), 176–196.

[17] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. ACM, 9–15.

[18] Ignacio X Domínguez, Rogelio E Cardona-Rivera, James K Vance, and David L Roberts. 2016. The mimesis effect: The effect of roles on player choice in interactive narrative role-playing games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3438–3449.

[19] Rozhin Doroudi, Rana Azghandi, Zlatan Feric, Omid Mohaddesi, Yifan Sun, Jacqueline Griffin, Ozlem Ergun, David Kaeli, Pedro Sequeira, Stacy Marsella, and others. 2018. An integrated simulation framework for examining resiliency in pharmaceutical supply chains considering human behaviors. In *Proceedings of the 2018 Winter Simulation Conference*. IEEE Press, 88–99.

[20] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. 2016. *Game analytics*. Springer.

[21] Michel Étienne. 2013. *Companion modelling: a participatory approach to support sustainable development*. Springer Science & Business Media.

[22] Chiara Franzoni and Henry Sauermann. 2014. Crowd science: The organization of scientific research in open collaborative projects. *Research policy* 43, 1 (2014), 1–20.

[23] Paul Guyot and Shinichi Honiden. 2006. Agent-based participatory simulations: Merging multi-agent systems and role-playing games. *Journal of Artificial Societies and Social Simulation* 9, 4 (2006).

[24] Casper Harteveld, Nolan Manning, Farah Abu-Arja, Rick Menasce, Dean Thurston, Gillian Smith, and Steven C Sutherland. 2017. Design of playful authoring tools for social and behavioral science. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*. ACM, 157–160.

[25] Casper Harteveld, Sam Snodgrass, Omid Mohaddesi, Jack Hart, Tyler Corwin, and Guillermo Romera Rodriguez. 2018. The Development of a Methodology for Gamifying Surveys. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 461–467.

[26] Casper Harteveld and Steven Sutherland. 2014. Finding the Game in Decision–Making: A Preliminary Investigation. *The Shift from Teaching to Learning: Individual, Collective and Organizational Learning Through Gaming Simulation* (2014), 199.

[27] Casper Harteveld and Steven C Sutherland. 2017. Personalized gaming for motivating social and behavioral science participation. In *Proceedings of the 2017 ACM Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces*. ACM, 31–38.

[28] Kristy L Hawley, Maryann Mazer-Amirshahi, Mark S Zocchi, Erin R Fox, and Jesse M Pines. 2016. Longitudinal trends in US drug shortages for medications used in emergency departments (2001–2014). *Academic Emergency Medicine* 23, 1 (2016), 63–69.

[29] Akhmad Hidayatno, Rachel Giovani Hasibuan, Girindra Chaska Wardana Nimpuno, Arry Rahmawan Destyanto, and others. 2019. Designing a Serious Simulation Game as a Learning Media of Sustainable Supply Chain Management for Biofuel Production. *Energy Procedia* 156 (2019), 43–47.

[30] Wander Jager and Gerben van der Vegt. 2015. Management of Complex Systems: Toward Agent-Based Gaming for Policy. In *Policy Practice and Digital Science*. Springer, 291–303.

[31] Robert E Kass and Adrian E Raftery. 1995. Bayes factors. *Journal of the american statistical association* 90, 430 (1995), 773–795.

[32] Jack PC Kleijnen. 2005. Supply chain simulation tools and techniques: a survey. *International Journal of Simulation and Process Modelling* 1, 1-2 (2005), 82–89.

[33] Christophe Le Page and Arthur Perrotton. 2017. KILT: a modelling approach based on participatory agent-based simulation of stylized socio-ecosystems to stimulate social learning with local stakeholders. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 156–169.

[34] Michela Le Pira, Edoardo Marcucci, and Valerio Gatta. 2017. Role-playing games as a mean to validate agent-based models: An application to stakeholder-driven urban freight transport policy-making. *Transportation Research Procedia* 27 (2017), 404–411.

[35] Hau L Lee, Venkata Padmanabhan, and Seungjin Whang. 1997. The bullwhip effect in supply chains. *Sloan management review* 38 (1997), 93–102.

[36] Charles M Macal and Michael J North. 2005. Tutorial on agent-based modeling and simulation. In *Proceedings of the Winter Simulation Conference, 2005.* IEEE, 14–pp.

[37] David Massey, Ola Ahlqvist, Kiril Vatev, and Johnathan Rush. 2019. A Massively Multi-user Online Game Framework for Agent-Based Spatial Simulation. In *CyberGIS for Geospatial Discovery and Innovation*. Springer, 213–224.

[38] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.

[39] Sebastiaan Meijer. 2009. *The organisation of transactions: Studying supply networks using gaming simulation*. Vol. 6. Wageningen Academic Pub.

[40] Fabien Michel, Jacques Ferber, and Alexis Drogoul. 2018. Multi-Agent Systems and Simulation: A Survey from the Agent Commu-nity's Perspective. In *Multi-Agent Systems*. CRC Press, 17–66.

[41] Arunachalam Narayanan and Brent B Moritz. 2015. Decision making and cognition in multi-echelon supply chains: an experimental study. *Production and Operations Management* 24, 8 (2015), 1216–1234.

[42] Joerg Nienhaus, Arne Ziegenbein, and Paul Schönsleben. 2006. How human behaviour amplifies the bullwhip effect. A study based on the beer distribution game online. *Production Planning & Control* 17, 6 (2006), 547–557.

[43] Afshin Oroojlooyjadid, MohammadReza Nazari, Lawrence Snyder, and Martin Takáč. 2017. A Deep Q-Network for the Beer Game: A Reinforcement Learning Algorithm to Solve Inventory Optimization Problems. *arXiv preprint arXiv:1708.05924* (2017).

[44] Nathan Partlan, Elin Carstensdottir, Sam Snodgrass, Erica Kleinman, Gillian Smith, Casper Harteveld, and Magy Seif El-Nasr. 2018. Exploratory Automated Analysis of Structural Features of Interactive Narrative. In *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.

[45] Vincent Peters, Geert Vissers, and Gerton Heijne. 1998. The validity of games. *Simulation & Gaming* 29, 1 (1998), 20–30.

[46] David V Pynadath and Stacy C Marsella. 2005. PsychSim: Modeling theory of mind with decision-theoretic agents. In *IJCAI*, Vol. 5. 1181–1186.

[47] John R Raser. 1969. Simulation and society: An exploration of scientific gaming. (1969).

[48] Christopher A Rates, Bridget K Mulvey, and David F Feldon. 2016. Promoting conceptual change for complex systems understanding: Outcomes of an agent-based participatory simulation. *Journal of Science Education and Technology* 25, 4 (2016), 610–627.

[49] Stéphane Ross, Joelle Pineau, Sébastien Paquet, and Brahim Chaib-Draa. 2008. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research* 32 (2008), 663–704.

[50] Richard M Ryan. 1982. Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of personality and social psychology* 43, 3 (1982), 450.

[51] Sidney Siegel. 1956. Nonparametric methods for the behavioral sciences. *New York* (1956).

[52] Christophe Simon and M Etienne. 2010. A companion modelling approach applied to forest management planning. *Environmental Modelling & Software* 25, 11 (2010), 1371–1384.

[53] Lawrence V Snyder and Zuo-Jun Max Shen. 2011. *Fundamentals of supply chain theory*. Wiley Online Library.

[54] Véronique Souchère, Laurent Millair, Javier Echeverria, François Bousquet, Christophe Le Page, and Michel Etienne. 2010. Co-constructing with stakeholders a role-playing game to initiate collective management of erosive runoff risks at the watershed scale. *Environmental Modelling & Software* 25, 11 (2010), 1359–1370.

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

[56] John D Sterman. 2006. Learning from evidence in a complex world. *American journal of public health* 96, 3 (2006), 505–514.

[57] John D Sterman and Gokhan Dogan. 2015. "I'm not hoarding, I'm just stocking up before the hoarders get here.": Behavioral causes of phantom ordering in supply chains. *Journal of Operations Management* 39 (2015), 6–22.

[58] K Thomas. 2017. A vital drug runs low, though its base ingredient is in many kitchens. *New York Times* (2017).

[59] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *science* 185, 4157 (1974), 1124–1131.

[60] Sean D Vermillion, Richard J Malak, Rachel Smallman, Brittney Becker, Michale Sferra, and Sherecce Fields. 2017. An investigation on using serious gaming to study human decision-making in engineering contexts. *Design Science* 3 (2017).

[61] Alexey Voinov and Francois Bousquet. 2010. Modelling with stakeholders. *Environmental Modelling & Software* 25, 11 (2010), 1268–1281.

[62] Eric-Jan Wagenmakers. 2007. A practical solution to the pervasive problems ofp values. *Psychonomic bulletin & review* 14, 5 (2007), 779–804.

[63] Günter Wallner. 2019. *Data Analytics Applications in Gaming and Entertainment*. CRC Press.

[64] Kobchai Worrapimphong, Nantana Gajaseni, Christophe Le Page, and François Bousquet. 2010. A companion modeling approach applied to fishery management. *Environmental Modelling & Software* 25, 11 (2010), 1334–1344.