Agent-Based Modeling of Human Decision-makers Under Uncertain Information During Supply Chain Shortages

Nutchanon Yongsatianchot
Northeastern University
Massachusetts, USA
yongsatianchot.n@northeastern.edu

Noah Chicoine
Northeastern University
Massachusetts, USA
chicoine.n@northeastern.edu

Jacqueline Griffin
Northeastern University
Massachusetts, USA
ja.griffin@northeastern.edu

Ozlem Ergun
Northeastern University
Massachusetts, USA
o.ergun@northeastern.edu

Stacy Marsella
Northeastern University
Massachusetts, USA
s.marsella@northeastern.edu

ABSTRACT

In recent years, product shortages caused by supply chain disruptions have generated problems for consumers worldwide. In supply chains, multiple decision-makers act on uncertain information they receive from others, often leading to sub-optimal decisions that propagate the effects of supply chain disruptions to other stakeholders. Therefore, understanding how humans learn to interpret information from others and how it influences their decision-making is key to alleviating supply chain shortages. In this work, we investigated how downstream supply chain echelons, health centers in pharmaceutical supply chains, interpret and use manufacturers' estimated resupply date (ERD) information during drug shortages. We formulated a computational model of a health center based on a partially observable Markov decision process that learns a manufacturer's information sharing tendencies through an observation function. To investigate the model and important factors influencing decisions and perceptions of ERD, we conducted a human experiment to study where subjects played the role of a health center during a drug shortage. They received ERDs from a manufacturer on a weekly basis and decided whether or not to switch to an alternative product (and pay additional costs) to avoid running out of stock. The results show that different manufacturers' sequences of ERDs and the accuracy of ERDs could impact subjects' decisions, beliefs, performance, and perception of the manufacturer. We also found that the subjective belief of ERDs is the best predictor of subjects' switching decisions. Lastly, we fit the observation function's learning rate and show that the model can predict subjects' decisions better than other baseline models in most conditions.

KEYWORDS

supply chain; information modeling; modeling; human experiment

ACM Reference Format:
Nutchanon Yongsatianchot, Noah Chicoine, Jacqueline Griffin, Ozlem Ergun, and Stacy Marsella. 2023. Agent-Based Modeling of Human Decision-makers Under Uncertain Information During Supply Chain Shortages. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 9 pages.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom. © 2023 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

1 INTRODUCTION

Recently, the United States and other nations witnessed global supply chain disruptions that led to shortages in valuable minerals, computer chips, and pharmaceutical products, among other important items [18]. Some of these disruptions were caused by the COVID-19 pandemic and resultant policy changes, while others were the results of more common disruptions, like manufacturer shutdowns. The problem of supply chain disruptions, however, is not new, and is only getting worse. Recent increases in globalization and complexity of supply chains has only increased the vulnerability of supply chains already experiencing frequent disruptions [10]. In fact, the National Academies of Science, Engineering, and Medicine have compiled work in an attempt to address the exact problem of resiliency in critical supply chains [32].

Supply chain disruptions are particularly troublesome in pharmaceutical supply chains (PSCs) [11], where demand is inelastic. In PSCs, the customers (patients) depend on life-saving medicines at health centers (HCs). If a drug is in short supply because of a supply chain disruption, patients may not receive the healthcare they need, or they may be treated with alternative medications that are more costly and/or more harmful. Unlike the products in other supply chains (e.g., clothes, produce, etc.), disruptions to PSCs have a direct, sometimes unavoidable, impact on people's well-being.

Unfortunately, drug shortages are too common [12], so HCs are constantly tasked with trying to mitigate effects of drug shortages on patients. In order to do this, pharmacists may decide to employ any number of different mitigation strategies within the HC to conserve inventory for the most high-risk patients. The most costly mitigation strategy a pharmacist could implement involves switching medical practices entirely over to an alternative product that is in good supply. Pharmacists usually try to avoid this mitigation strategy because of how costly it is to HC operations and because of potential health impacts on patients. However, it is not always avoidable when supply is low and there is no resupply in sight.

Ultimately, a pharmacist's goal is to avoid running out of stock of any particular medication. Interviews with hospital pharmacists of a major hospital system in the United States brought to light the uncertainty and unreliability of future supply information. In the PSC setting, HCs are given **estimated resupply dates** (ERDs), dates communicated by pharmaceutical manufacturers (MNs) to HCs indicating when the HCs should receive their next shipment of a specified product (which, usually, partially fills an existing back order). Though the intent of the information is to inform inventory management decisions at HCs, this information is often unreliable and changing, making the information difficult for pharmacists to reason with.

During a drug shortage, HC pharmacists have to consider their inventory, utilization, and the ERD information provided to them, and decide whether or not they will run out of inventory in the future (and they should implement mitigation strategies now) or whether their inventory will last them until replenishment arrives. This problem can be simplified into a choice between two possible decisions: **(1) "Switch now":** implement mitigation strategies now, at lower cost than switching products later, or **(2) "Wait and see":** wait and see if more supply will arrive (therefore, incurring no costs) at the risk of being forced to change products (incurring the maximum cost) if the inventory runs out. We call this "The Switching Problem". It will be the focus of the rest of this paper. Making the optimal decision that minimizes the cost to HCs is especially difficult when the ERD information is uncertain and subject to change during a shortage. This work examines and computationally models how human decision-makers address The Switching Problem given uncertain ERD information.

The Switching Problem is a Wicked Learning Problem, a problem involving learning in uncertain and changing environments [17]. HCs attempt to learn about the communication tendencies of the MNs as they experience additional shortages and receive new (uncertain) ERD information over time. How HCs learn from the ERD information, create mental models of the MNs, and then generate actions from their beliefs, are all important parts in understanding how HCs interact with the rest of the supply chain network. These decisions ultimately influence the costs experienced by different supply chain stakeholders, and the effectiveness of the overall system.

To study human behavior in this system, we formulated a model of a HC's decision-making process in The Switching Problem. Since a HC faces multiple sequential decisions and has to rely on partial information, we modeled the decision-making process using a partially observable Markov decision process (POMDP) framework, which is a framework for modeling sequential decision-making under uncertainty based on partial observation. As the HC interacts with the same MN over multiple ordering periods, the model maintains the beliefs about the MN and learns to adapt to the ERD behaviors of MN. Additionally, we explored a way to make the model more human-like by including the effect of memory limitation on learning to the model.

To investigate our model, we conducted a human subject experiment that simulated repeated trials of The Switching Problem. Subjects played the role of a HC and made switching decisions on a weekly basis based only on ERD messages from MNs and their own inventory information. We investigated multiple types of MNs defined by their ERD communication tendencies. Furthermore, we explored subjects' perceptions of different MNs, including willingness to work with them again, trust, and trust-related factors. The data from the experiment shows that different types of MN induce different patterns of decision-making in the participants, and influence participants' beliefs and perceptions of the MN. It also shows that participants' subjective ERD is a good predictor of their decisions, highlighting the importance of modeling HC's beliefs (that can keep changing). Finally, the results from model fitting reveal that a model that learns with limited memory can predict human decisions well across multiple types of MNs.

2 RELATED WORK

This work applies a POMDP framework to model decision-making in the context of disrupted supply chains with information sharing. Many studies in operations research and decision sciences literature already explore the effects of information and disruptions on overall supply chain performance [23, 40].

On the theoretical side, studies have shown how supply chain disruptions can cause the bullwhip effect when agents follow a typically optimal base stock policy [24]. Doroudi et. al. (2020)[9] show how such effects are exacerbated and prolonged when retailers in a multi-echelon supply chain change their ordering choices according to historical order fulfillment, mimicking a real-life scenario where retailers order more from retailers they trust to supply them.

In efforts to improve supply chain systems by incorporating information sharing structures, some studies were able to demonstrate the value of information sharing in improving overall supply chain profits. These studies have all considered several different kinds of information that could be shared between supply chain stakeholders, including manufacturing capacity [4], production yield [7], lead time estimates [6], demand projections [25], and disruption duration information [28]. Though these studies provide theoretical evidence for the value of information sharing, it is important to study empirically how humans utilize and act on this information in order to make inventory management decisions.

Previous studies have taken an empirical approach to understand how the behavior of human decision-makers changes in disrupted supply chain scenarios. Sakar and Kumar (2015)[36] showed in a replication of the beer game, that the bullwhip effect (i.e. players over-ordering supply) could be mitigated during a supply disruption by sharing knowledge of the disruption with other stakeholders. They found that informing others of the supply chain disruption decreases deviations from typical orders and strengthens the buyer-supplier relationship, improving the long-term efficiency of the system. Using an agent-based simulation framework [8], Mohadessi et al. (2020) were able to replicate the bullwhip effect in a study of human players in a simulated beer game experiment with supply disruption [30]. Unlike the findings in Sakar and Kumar (2015), participants in this study did not alter their behavior during the game despite being given addition information on suggested (optimal) order amounts. A follow-up study [29], however, did show evidence for supplier inventory information influencing players' ordering behaviors. In our study, players will be forced to use ERD information (as part of the only known information in the experiment) during the supply chain disruption to determine the best inventory management policy, granting us the ability to model the decision-making behavior more directly than in these studies.

Some studies have attempted to isolate the effects of information on human decision-making by focusing on supply and/or information uncertainty as the only caveat in the supply chain experiment, omitting the effects of supply chain disruptions. In 2013, Ancarani et al. [2] filled an existing gap in supply chain literature by examining the supplier-side uncertainty on supply chain efficiency. They used a beer game experiment to show that stochastic lead times from suppliers cause participants to increase the variability in their orders, exacerbating the bullwhip effect. Hofstra et al. (2022) [16] considered an additional type of uncertainty: inventory record inaccuracy. The authors showed in a human experiment of managing inventory in a supply chain, that participants were more likely to over-stock on products when they had inaccurate inventory records as compared to those who had uncertain supply lead times. These results suggest that participants are more likely to respond to and correct for uncertainty internal to their organization, as opposed to the uncertainty created by other entities.

In almost all of the aforementioned studies, researchers have mostly utilized the beer game or news-vendor problems to capture how ordering behavior changes with extra information and/or uncertain supply. These experimental designs assume stakeholders (including the players) are simply trying to optimize their inventory management, and that any unmet demand can be tracked as backlog. In this study, we expose a case of supply chain structure that is unique to the current literature: a case with inelastic demand, where if demand is not met, maximum penalties are incurred. Exploring the decision-making process of humans in this particular supply chain condition lends us the opportunity to uncover trends and behavior that may be useful knowledge in efforts to improve the operation conditions of supply chains with critical demand, like pharmaceutical supply chains.

2.1 Human Aspect

One important aspect to consider when it comes to human learning is memory [3]. People have limited memory, and the learning could be biased by the timing of the experiences, in particular, the primary effect in which the observations at the beginning of a sequence tend to be recalled very well and the recency effect in which the observations at the end of a sequence tend to be recalled more than the middle observations [31]. Therefore, time and order of observations could have an impact on human learning and memory, unlike a standard Bayesian update, which is part of POMDP, where the order of observations does not matter. As has been emphasized by bounded rationality, to create a human-like model, it is important to recognize human limitation [14, 38]. Thus, in this work, we attempt to incorporate the limitation of memory into the model.

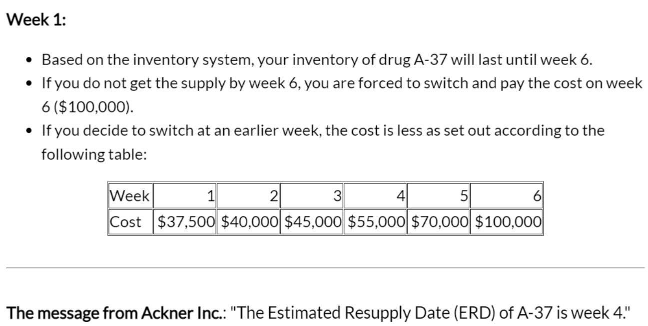
Another key concept in the ERD scenario is trust. Trust is essential for effective, cooperative multi-agent interaction where one agent relies on another agent for some risky decisions. A common definition by Mayer et al. defines "trust" as the willingness of an agent (the trustor) to become vulnerable to the action of another agent (trustee) based on the expectation that the other will perform the action important to the trustor without the ability to monitor or control that other agent. Researchers have also identified several trust dimensions, including competence (or ability), benevolence, integrity, and predictability [26, 27]. A few existing works have studied trust in the supply-chain context. A study by Kim (2009) [21] showed that trust-based relationships could decrease inventory variability. Another study by Jalbut and Sichman (2018) [19] found that trust-based relationships could thrive if communication is honest. Doroudi et al. (2020) [9] show that the timing and scale of the supply-chain disruptions could depend on the buyer's trust in a seller. As a study on trust in The Switching Problem is still lacking, we also explore how trust and trust-related factors could differ due to the types of MNs and their ERD communication tendencies.

3 THE ERD EXPERIMENT

In this section, we detail the ERD experiment to explore potential factors influencing subjects' learning and decision-making while facing The Switching Problem. The experiment structure closely follows actual HC's experiences when given ERD information, as discussed in Introduction. Subjects play the role of a health center pharmacy supply chain manager who orders drugs and manages drug shortages at a health center. The subject's job is to maintain a supply of different drugs at a minimum cost to the health center. The HC typically orders drugs from a primary MN (called Ackner Inc.). When a supply disruption occurs and a drug goes on shortage, Ackner Inc. will give an ERD to the HC, communicating when they expect to send the HC a replenishing shipment. We assume in the experimental design that the next shipment of products will replenish the HC with enough inventory for the foreseeable future. This way, subjects are only concerned about the timing of the shipment, not the quantity of supply that arrives. This also allows us to designate each order/shipment period as an individual trial for the purpose of having subjects learn MN's communication tendencies from trial to trial. Subjects receive an ERD update from Ackner Inc. each week in each trial. Subjects are also told that the MN does not know about their inventory nor when it will run.

Meanwhile, the HC is also aware of another MN (called Belltree Corp.) that sells a similar drug. The HC has the option to switch to Belltree Corp. as a MN for the alternative medication. If the HC decides to switch MN and use the alternative product, it will incur a one-time switching cost. The cost of switching MNs increases as the current inventory depletes, since having less inventory of the primary product often requires more fast-paced and widespread practice changes throughout a HC. If the HC runs out of inventory, the maximum penalty is incurred (symbolizing a forced practice change everywhere in the HC at the very last minute). This scenario is also provided to subjects as a cover story. All drugs and MNs are fictional and designed to be neutral. Subjects are provided with a starting fund to spend for switching costs, and the remaining fund is the measure of their overall performance.

Figure 1 shows an example of the experiment interface. For all trials, the week that the current inventory for the drug is anticipated to run out (i.e., runway) is week 6. Every week, subjects are reminded about the runway and the switching costs. There are a total of 7 trials, and each trial consists of up to 6 weeks, being the inventory runway. Each week, subjects have two choices: either wait for the product from the primary MN, or switch to the alternative MN and pay the switching cost for that week. If subjects decide to switch, they cannot change their mind, but they still observe the rest of the ERDs from Ackner Inc. communicated in that trial. If the subject does not switch, and a shipment doesn't arrive by week 6, the HC is forced to switch products and the maximum penalty is incurred. At the end of the trial, subjects receive a summary of the whole sequence of ERDs and the true resupply date (TRD).



Week 1:

- Based on the inventory system, your inventory of drug A-37 will last until week 6.
- If you do not get the supply by week 6, you are forced to switch and pay the cost on week 6 ($100,000).
- If you decide to switch at an earlier week, the cost is less as set out according to the following table:

Week	1	2	3	4	5	6
Cost	$37,500	$40,000	$45,000	$55,000	$70,000	$100,000

The message from Ackner Inc.: "The Estimated Resupply Date (ERD) of A-37 is week 4."

Figure 1: The experiment interface.

3.1 Measurement

On the 1st and 3rd week of every trial, after making a decision, we ask subjects two questions: (1) when (which week) do you think Ackner Inc. (the primary MN) will deliver the product? and (2) how likely (in percentage) do you think that Ackner Inc. will deliver the supply before the end of the runway (before inventory depletes)?. We refer to the answer of question (1) as the "subjective ERD" (not to be confused with ERD as told by MN), and the answer to question (2) as the "subjective probability" of receiving the product before the end of the runway. After the 1st, 3rd, and 6th trials, we also ask them to rate their perceptions of the primary MN on trust, benevolence, competence, and predictability[1]. After the experiment, there is a brief post-questionnaire asking them whether or not they would like to work with the primary MN again, in addition to questions collecting demographic information on the study participant.

3.2 Conditions

There are five conditions in our experiment. Each subject only experienced one condition. Thus this was a between-subjects experiment. The experimental factors we manipulated were the sequence of ERDs (which characterizes the type of MN), the switching cost, and the arrival time of the order. These factors align with elements of the POMDP model structure. Specifically, the order arrival time corresponds to the state, the sequence of ERDs corresponds to the observations, and the switching cost corresponds to the reward function of the POMDP. We outline how these three factors are changed to create the five experimental conditions, and explain their modeling implications in Section 4.

First, the type of MN is defined based on the characteristics of the ERD communications along two criteria: how many times the ERD changes from weeks 1 through 6, and the difference between the True Resupply Date (TRD), the eventual resupply date, and ERD (TRD - ERD). A positive difference means that the ERD is earlier than TRD, which will result in the order getting pushed back. A negative difference means ERD is later than TRD, which will result in push-forward. Based on these characteristics, we defined three distinct types of MN: (1) Push back (PB) MN, where the changes happen 1 to 3 times and the difference is +1 to +2 weeks, (2) Accurate (ACC) MN, where the changes happen 0 to 1 times and the difference is +1 week, and (3) Random (RND) MN where the changes happen 1 to 3 times and the difference is -2 to +3 weeks.

There are two conditions for switching costs: high (37.5k, 40k, 45k, 55k, 70k, 100k) and low (20k, 24k, 30k, 38k, 50k, 100k). The main condition is high so PB, ACC, RND have high switching costs. We only consider one condition with low cost, PB with low cost (PBL). In all cases, subjects are given $700,000 in funds and asked to minimize the overall cost of managing the shortages.

Additionally, there are two possible order arrival times: the product arrives before or after the end of the runway. If the product arrives before the runway, we classify it as a "Good" (G) trial, and if it does not, we classify it as a "Bad" (B) trial. As the experiment contains seven trials, we consider two sequences of seven order arrival times: the main sequence ($G_1G_2G_3B_1B_2B_3G_4$) and the reserved sequence ($B_1B_2B_3G_1G_2G_3G_4$), which swaps the first three G trials with the following three B trials. In the reversed sequence, the B trials come first, creating early negative experiences for subjects. Similar to switching cost, we only consider the reversed order for PB, Push Back Reverse (PBR). Further, subjects in all conditions see an equal number of G and B trials in the first six trials and they all experienced a G on the trial 7. Crucially, the ERD for week 1 of trial 7 is the same for all five conditions. We used this time period for comparison across conditions. In total, there are five conditions. Table 1 shows all the sequences of the five conditions.

Table 1: The sequences of ERDs for each condition from trials 1 to 7. If ERD is the same as the week, the product arrives on that week. For example, if a sequence is (3, 4, 4, 5, 5), the product arrives at week 5. The sequence conditions are as follows: ACC = Accurate, PB = Push Back, PBL = Push Back Low cost, PBR = Push Back Reverse, RND = Random.

Condition	Sequences: From trial 1 to 7
ACC	(5,5,6,6,6,6), (4,4,5,5,5), (6,6,6,6,6,6), (6,6,6,7,7,7), (7,7,8,8,8,8), (7,7,7,7,7,7), (5,5,5,6,6,6)
PB & PBL	(5,5,5,6,6,6), (3,4,4,5,5), (4,4,4,5,6,6), (5,5,5,6,7,7), (6,6,6,6,7,8), (6,6,6,6,6,7), (5,5,5,5,6,6)
PBR	(5,5,5,6,7,7), (6,6,6,6,7,8), (6,6,6,6,6,7), (5,5,5,6,6,6), (3,4,4,5,5), (4,4,4,5,6,6), (5,5,5,5,6,6)
RND	(5,5,7,7,6,6), (7,7,6,6,5), (5,5,5,6,6,6), (7,7,6,6,7,7), (6,6,7,7,7,8), (4,4,6,6,7,7), (5,5,5,7,7,6)

3.3 Data Collection

We recruited subjects from Prolific (www.prolific.co). We had 40 subjects for each condition, 200 subjects in total (gender: 52.4% female, age: mean 38.2, s.d. 13.5). The experiment lasted for approximately 15 minutes, and we paid subjects $2.50. Importantly, subjects could receive an additional $5 bonus based on their performance. Specifically, 20% of subjects were randomly chosen to receive the bonus, where the chance of being chosen was proportional to their performance in the experiment. The experiment was approved by Northeastern University's IRB. In our analysis, we excluded subjects who did not correctly answer the attention check questions in the tutorial and those who always chose to switch immediately

[1]We leave the analysis on the dynamics of these factors for future work.

in every trial. This left us with 164 subjects (ACC: 30, PB: 31, PBL: 35, PBR: 34, RND: 34).

4 MODEL

To model HC decisions in the Switching Problem, we use a partially observable Markov decision process (POMDP) framework [20], which is a framework for modeling sequential decision-making where the agent operates under uncertainty based on partial observation. At each time step, the agent receives an observation about the current situation, then chooses an action that could yield a reward, and then moves to the next step. This matches well with the ERD scenario because the HC does not know the TRD of the replenishment and has to rely on ERDs, which can be uncertain. In addition, the ERD scenario is a sequential decision-making process because the HC makes multiple decisions in one episode that lasts multiple weeks, and future decisions depend on earlier decisions. POMDPs also have been shown to be suitable for modeling human decision-making in other uncertain domains [33], [35], [44], [39].

POMDP $< S, R, \Omega, O, A, T >$ consists of the following elements.

State (S): A state $s \in S$ represents a possible situation in the scenario and exhibits a set of features. In our ERD scenario, a state consists of a TRD (week 1 to 10) and a current time/week t. Because the agent (playing the role of the HC) doesn't know the state for sure (doesn't know the TRD), we assume they maintain a probability distribution over all possible states, or a belief state $b = P(S)$, symbolizing when the subject believes the TRD will be. We can view it as HC's subjective ERD. This distribution is modeled as a Multinomial distribution, and the probability of outcomes (TRD) is modeled as the Dirichlet distribution, which is a conjugate prior to the Multinomial distribution. The Dirichlet distribution is necessary for learning which is detailed in the next section.

Action (A) and **Transition Function** (T): An action $a \in A$ represents an available action in the scenario and the transition function describes the dynamic of the action. There are only two actions: **wait** which will move the subject to the next time step and to the next ERD, and **switch** which will result in paying the switching cost and stopping the trial. If the next state is the runway, then the subject is forced to make the **switch** action unless the TRD is that period.

Observation (Ω) and the **Observation function** (O): An observation $o \in \Omega$ represents an observation that the agent can see. The observation function $O(\Omega|s, a) = P(\Omega|s, a)$ is the probability that the agent will see an observation $o \in \Omega$ given the state s and the action a. This observation function describes how the agent perceives the relationship between observations and states. In the ERD scenario, the observation function is essentially the subject's model of the MN.

For our ERD scenario, the natural observation is obviously the ERD from MN. However, the agent then has to maintain a distribution for every state (every TRD), which makes learning difficult given that the agent only observes a few sequences. Instead, we define the observation to be the difference between TRD and ERD, $P(o = \text{TRD} - \text{ERD}|s = \{\text{TRD}, t\})$. This way, the agent can generalize from one state to another. Nonetheless, the agent still needs to maintain a distribution for every week because ERDs can't be a time that is earlier than the current week. The observation function is also a Multinomial distribution where the probability is modeled by the Dirichlet distribution, similar to the belief state.

The probability of observing o given b and a is

$$P(o|b, a) = \sum_{s'} P(o|s', a) \sum_{s} P(s'|s, a) b(s). \tag{1}$$

After taking a and observing o, the new belief b' is

$$b'(s') \propto P(o|s', a) \sum_{s} P(s'|s, a) b(s), \tag{2}$$

Reward (R): Reward function $R(s) \in \mathbb{R}$ maps a state to a real number summarizing how good or bad a given state is. In this case, the reward is simply the switching cost. If the agent waits for the product and it arrives before the runway, the agent receives a reward of zero. If, instead, the product arrives after the runway, the agent receives a reward equal to the maximum switching cost. If the agent decides to switch, the reward is equal to the switching cost at that time.

4.1 Making Decisions

The utility or Q-value of an action for action selection can be expressed as a Bellman equation for a POMDP as follows:

$$Q^*(s, a) = \sum_{s', o} P(s'|s, a) P(o|s', a)(R(s, a) + \max_{a'} Q^*(s', a')), \tag{3}$$

where $Q^*(s, a)$ is the optimal expected cumulative rewards from state s and action a. The agent then selects the action with the highest Q-value. Alternatively, we can convert the Q-value into a probability using the softmax function. As this is a relatively small search space problem, the Forward Search algorithm, an online POMDP algorithm, was used to calculate the above equation [22].

4.2 Updating the Model

At the end of each episode (or trial in the experiment), the agent will know the TRD along with the sequence of ERDs so the agent can update their prior belief and observation function to be used in the next episode. Because the Dirichlet distribution is a conjugate prior to the Multinomial distribution, the update is simply adding the count of observations to the corresponding outcome [42].

Formally, given a multinomial distribution

$$y_1, \ldots, y_k \sim Multinomial(p_1, \ldots, p_k)$$

and the corresponding Dirichlet distribution

$$p_1, \ldots, p_k \sim Dirichlet(\alpha_1, \ldots, \alpha_k),$$

after observing an outcome y_j, the posterior would become

$$Dirichlet(\alpha_1, \ldots, \alpha_j + 1, \ldots, \alpha_k). \tag{4}$$

In a standard Bayesian update, the order of the data does not matter, and all the outcomes have the same weight of +1 for the update. However, humans have limited memory, and the timing could affect how we update our beliefs, like what has been observed in the primary and recency effects [31]. Therefore, in this work, we explored the effect of time on the update by parameterizing the count and making it depend on time. After observing an outcome $y^{(j)}$ at time t, the update is

$$\alpha_j = \alpha_j + \beta^t, \tag{5}$$

where $\beta \in [0, \infty)$. β can be thought of as a learning rate: one for the prior state β_s and another for observation function β_o. If $\beta > 1$, the recent observations influence the update more. If $\beta < 1$, the earlier observations influence the update more. The model simulation of the experiment, the experiment materials, data, and codes can be found at https://github.com/yongsa-nut/ERDstudy.

5 EXPERIMENTAL RESULTS

To investigate the difference between conditions across various factors, we use the brm package [5] to do Bayesian Hierarchical Regression, which allows us to deal with the issue of multiple comparisons effectively by partial pooling [13].

Figure 2 shows the estimated mean of the five conditions for six factors. The first factor is the decision shown in Figure 2.A, specifically, the probability of waiting at week 1 of trial 7. As mentioned earlier, this is the time that all conditions have the same ERD allowing us to compare between conditions. The results show that, on average, subjects in ACC and RND conditions were more likely to wait after seeing ERD = 5 than the other three conditions (ACC vs. PB: estimated difference in mean (est) = 1.43, 95% Confidence Interval (CI) = [0.38, 2.59]; ACC vs. PBL: est = 1.30, 95% CI = [0.27, 2.42]; ACC vs. PBR: est = 2.18, 95% CI = [1.05, 3.41])[2]. ACC and RND were not different, and neither was PB and PBL. On the other hand, subjects in the PBR condition were most likely to switch (mean = 0.29, se =0.07) with a lower probability of waiting than those in the PB condition (est = -0.75, 95% CI = [-1.78, 0.22]).

Secondly, Figure 2.B shows subjects' reported probabilities of receiving the product before the runway at week 1 of trial 7. Similar to their decisions, subjects in ACC and RND conditions reported a higher perceived likelihood of receiving the product before the end of the runway as compared to the other three conditions (ACC vs. PB: est = 18.17, 95% CI = [7.28, 29.07]; ACC vs. PBL: est = 15.64, 95% CI = [5.15, 26.40]; ACC vs. PBR: est = 23.82, 95% CI = [12.55, 35.01]), And, again, subjects in the PBR condition reported the smallest probabilities (mean = 36.17, se = 3.77) but were not different from subjects in the PB condition (est = -5.65, 95% CI = [-15.79, 4.31]).

Figure 2.C shows subject-reported (subjective) ERDs at week 1 during trial 7. We observed that subjects in the RND condition reported sooner ERDs than those in the three PB conditions (PB vs. RND: est = 0.63, 95% CI = [0.21, 1.07]; PBL vs. RND: est = 0.46, 95% CI = [0.06, 0.86]; PBR vs. RND: est = 0.84, 95% CI = [0.40, 1.28]). Interestingly, the responses in the ACC condition are only less than those of the PBR conditions (est = -0.60, 95 CI% = [-1.03, -0.18]). Subjects in the PBR conditions believed ERD to be the highest/latest (mean = 6.50, se = 0.15) compared to the other conditions. Nevertheless, the means of the subjective ERDs (again in Week 1) across all conditions are higher than the ERD from the MN in week 1, which is 5 across all conditions.

The next result is the overall adjusted performance in Figure 2.D. Since the performance depends on the switching cost, we adjusted the performance of the PBL condition by using the high switching cost instead. The results show that subjects in the RND condition performed the worst while those in the other four conditions performed about the same (ACC vs. RND: est = 0.79, 95% CI = [0.53, 1.05]; PB vs. RND: est = 0.71, 95% CI = [0.45, 0.96]; PBL vs. RND: est = 0.58, 95% CI = [0.33, 0.83]; PBR vs. RND: est = 0.79, 95% CI = [0.53, 1.04]).

[2]The differences are in the logit scale.

Figure 2.E shows subjects' reported likelihood of working with the primary MN again. We see that the PBR MN has the lowest likelihood, which was lower than other groups, except the PBL. All other groups had similar ratings (PBR vs. ACC: est = -1.00, 95% CI = [-1.62, -0.36]; PBR vs. RND: est = -0.99, 95% CI = [-1.60, -0.37]; PBR vs. PB: est = -0.73, 95% CI = [-1.32, -0.14]).

Finally, Figure 2.F displays the reported MN trust levels for each condition, on a Likert scale (1-5). The results show that subjects in the ACC condition trusted the MN more compared to other conditions, except for subjects in the RND condition. (ACC vs. PB: est = 0.38, 95 CI% = [-0.04, 0.82]; ACC vs. PBL: est = 0.52, 95% CI = [0.08, 0.97]; ACC vs. PBR: est = 0.72, 95% CI = [0.20, 1.20]; ACC vs. RND: est = 0.24, 95% CI = [-0.16, 0.66]) All other condition pairs were similar in trust ratings. The ratings of benevolence and competence were similar to trust. The results for subjects' rating of predictability unsurprisingly show that ACC has the highest predictability rating, higher than every conditions, while the other conditions have about the same rating. Complete results for benevolence, competence, and predictability as well as other comparisons can be found in the supplemental materials in Github.

5.1 Using Regression to Predict Switching Decisions

Table 2: Prediction results using features in the data. The target is the decision at week 1 of trial 7. Prob refers to the probability of receiving the product before the runway. T = trial. W = week. The Baseline predicts the majority (wait).

Feature	Accuracy	F-1 Score
Baseline	56.00%	.72
Condition	63.41%	.63
T7 W1 Prob	72.56%	.76
T7 W1 ERD	81.71%	.85
T6 Trust	61.59%	.69
T6 Competent	63.41%	.72
T6 Benevolent	60.37%	.69
T6 W1 ERD	64.02%	.72
T6 W1 Decision	59.76%	.61
T7 W1 ERD + Prob	82.32%	.85
T7 W1 ERD + Prob + Condition	79.88%	.82

Next, we investigate features that can predict the decision at week 1 of trial 7. Table 2 shows the prediction results based on logistic regression and leave-one-out cross-validation using brm package [5]. When considering only one feature, we see that the subjective ERD performs better than other features in the data on both accuracy and F1-Score (Accuracy = 81.7%, F1-Score = 0.85). Trust features and the earlier trial information achieve better performance than the baseline, but are still lower than the subjective ERD. The subjective ERD and probability together achieves the best accuracy (82.3%) but does not improve from the subjective ERD alone, while achieving a slightly lower F-Score (0.848). However,

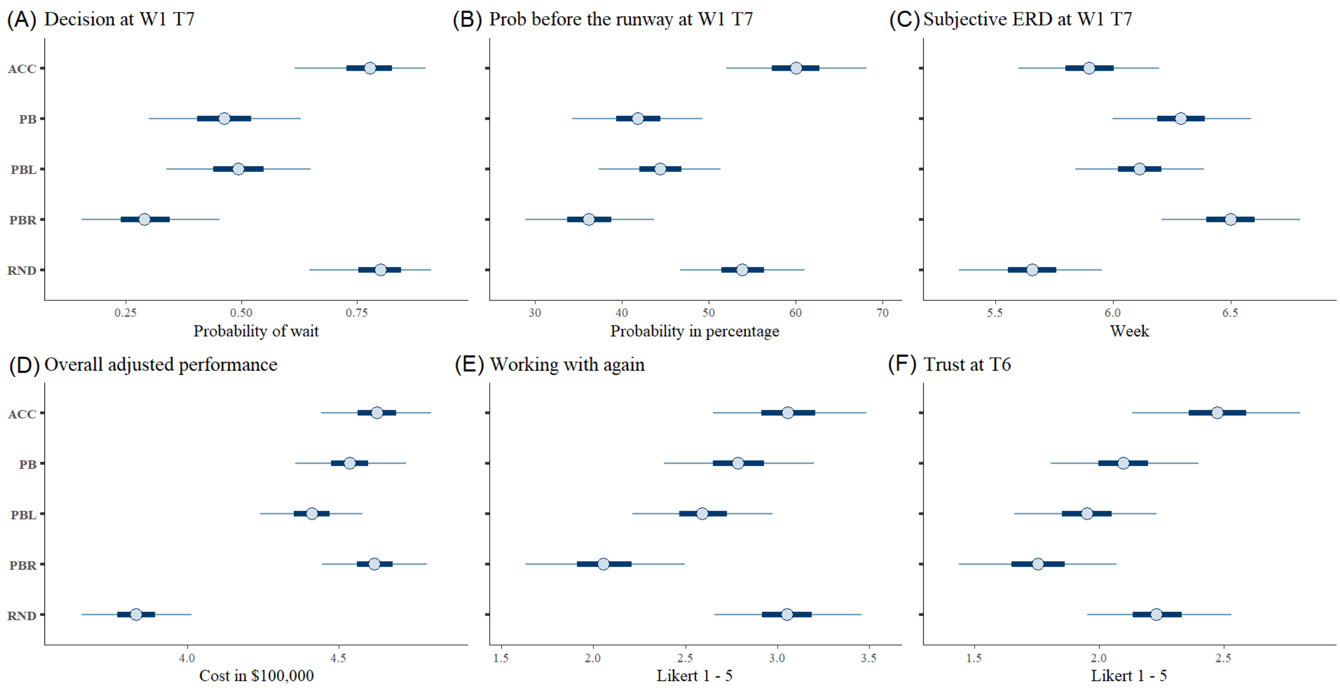


Figure 2: Estimated means of key factors: from top to bottom and left to right, decision, probability of receiving the product before runway, subjective ERD, overall adjusted performance, the likelihood of working with the MN again, and trust. The last two factors are on the scale from one to five. the shaded areas are 50 % CI and the lines are 95% CI.

the probability alone only achieves 72.6% accuracy. In sum, the results suggest that the subjective ERD is the best predictor.

Figure 3 shows the subjective ERD regression on decisions plot. When the subjective ERD is 5 or lower, subjects are most likely to wait in week 1; when the subjective ERD is 7 or higher, they are mostly to switch in week 1; and when the subjective ERD is 6, they are unsure what to do but slightly lean toward waiting.

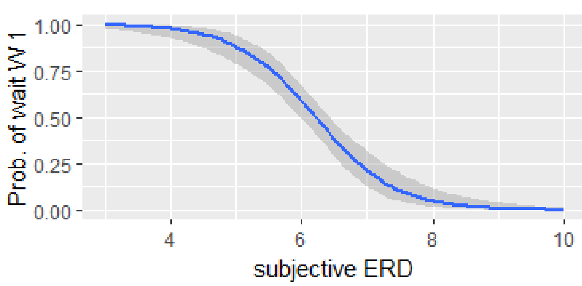


Figure 3: The plot between the subjective ERD and the decision (probability of waiting) at week 1 from the regression. The blue is the mean, and the shaded area is the 95% interval.

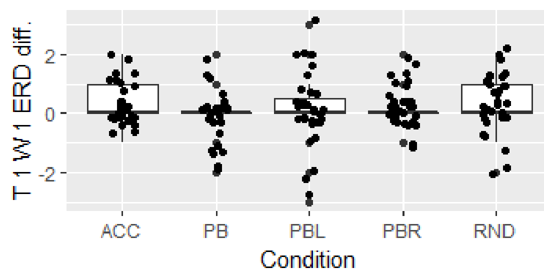


Figure 4: The differences between the subjective ERD and the ERD as reported by MN at week 1 of trial 1. Positive numbers mean that the subjective ERD is higher than the ERD.

Table 3: The fitting results. The table shows the absolute error for the different setups of the model across five conditions.

Model	ACC	PB	PBL	PBR	RND
No learning	**.154**	.502	.468	.689	**.130**
Default parameters	.173	.336	.463	**.149**	.163
Learn from earlier trials	.161	.342	.259	.710	.822
Learn from other cond.	.170	**.042**	**.031**	.571	.135

5.2 Fitting the Learning Rate of the Model

Lastly, we turn to the results of fitting the model from the sequence of ERDs. The setup for this was as follows. Unlike the previous prediction results, the input here is the entire sequences of ERDs from trial one to six, the ERD at week 1 trial 7, and the switching costs. The model was fit separately to each of the five conditions in the experiment. The prior belief distribution was a uniform distribution, and the prior observation function distribution was a distribution that weakly centers at zero, indicating that people believe ERD to be accurate, as we observe in the data at week 1 trial 1 (Figure 4).

The baseline models include the model with no learning, so the prior distributions are never changed, effectively predicting based on the ERD from the MN, and the model with default learning rate ($\beta_s, \beta_o = 1$) for both prior state and observation distribution. We considered two ways of training for the learning rate parameters: (1) train the learning rates within the condition (using that condition's earlier trials, 1 to 5, to predict week 1 trial 6), and (2) train the learning rates from other conditions. Again, in both cases, the test of the training is predicting week 7, trial 1. We used grid search to find the optimal parameters that minimize the absolute error. Aside from these two parameters, all models shared the same setup, including switching costs, search depth of one, prior state distribution, and prior observation function, as mentioned above.

Table 3 shows the model fitting results in terms of absolute error. The results show that the No learning baseline model achieves the best performance for RND and ACC, suggesting that the prior distributions capture the beliefs at week 1 trial 7 well. The model that was trained from the earlier trials performed the worst overall, suggesting it is crucial to take into account the outcome of the latest trial, trial 6, when estimating the parameters as it could have the most effect on the learning. On the other hand, the model that trained on other conditions achieved the lowest error for PB and PBL conditions and achieved similar results for RND and ACC conditions when compared to other baseline models. However, this model achieved higher error for PBR than the model with default parameters. One potential explanation is that PBR is quite different from other conditions due to starting the sequence with bad outcomes, so training on other conditions is not effective.

6 DISCUSSION

Overall, the results show that different types of MNs, as defined by their ERD communication behaviors, could lead to distinct patterns of decisions, beliefs, and perceptions of MN. We found that when the sequence of outcomes started out with bad experiences (PBR), subjects were more likely to switch early, had less trust, and were unlikely to want to work with the MN again. One possible explanation is that an early negative impression could have a long-lasting influence [1, 41]. However, we found that people in the RND condition performed the worst, and those in PBR did not perform differently from the other conditions. This suggests that people perform better when working with someone who they can predict ERD communication behaviors, regardless of whether the communications are accurate or the ERD is being consistently push back.

Nevertheless, people still trust and want to work with RND again more than PBR. One potential explanation is that people attribute the cause of push back to the MN themselves if it happened frequently, but in the case of random, people may attribute the cause to other external factors, as suggested by the fact that they still trust the MN. This is an important direction for future work to explore how people attribute the blame for delaying the products. On the other hand, we found that there is no clear difference between PB and PBL in all the factors that we looked at. The reason for this could be that from the subject's perspective, the switching costs between the two conditions are not different enough. Future work is needed to further investigate different switching costs as well as other costs beyond monetary, such as the cost of changing practices or side effects of a new drug.

The results also suggest that people can learn from limited interaction and observation and how they learn can be impacted by the sequence of observations. We also found the subjective ERD to be the best predictor for the switching decisions, and the model with the fitted learning rate can predict the decision better than the model without learning. Altogether, the results support the proposed model suggesting that it is crucial to incorporate subjective beliefs and how people update their beliefs about the state of the world. It is also important to consider the sequence of observations when updating the beliefs.

The experiment and the model can be extended in many ways to capture decision-making in this scenario better. For the experiment, there are several limitations that could be addressed in future work, including larger sample sizes, longer sequences, and additional types of MN. For the model, one important direction is to include Theory of Mind in the model that would allow the agent to reason about the causes of ERDs [15, 34, 43]. This means that HC will have a more complete model of MN and not just an observation function which would also allow us to model other HC's perceptions of MN, including blameworthiness, trust factors, and willingness to work with the MN again [15, 37]. Another future direction for both experiment and model is to extend the scenario to a larger supply chain structure, including inventory management, placing orders, and multiple echelons.

Beyond these results, this work also highlights the importance of information sharing from both MN's and HC's perspectives. From MN's perspective, it is crucial to consider that HCs could adapt to predictable patterns and that the early interactions could have a prolonged effect on their perceptions, especially negative experiences. Thus MN should consider making ERD predictable and consistent and providing reliable information regarding the causes of changes in ERD so HC can correctly understand the situation. From HC's perspective, this work suggests that there is a need for a better information sharing system across different echelons in the supply chain that provides accurate information for a better understanding of ERDs and for more efficient planning to avoid unnecessary operational costs.

To conclude, we proposed and investigated a model on how people deal with limited and uncertain information in supply-chain scenarios. In a human subject study, we collected data that showed people rely on their subjective beliefs and can learn to adapt to a specific MN. Further, we showed that the proposed POMDP model that learns the observation function of ERD could predict HC's decisions better than other baseline models in most conditions. This work is a crucial step toward a more realistic simulation of a supply-chain network which could help ease operational burdens during the supply-chain shortages.

ACKNOWLEDGMENTS

We would like to thank three anonymous reviewers for their helpful comments and suggestions. We would like to thank Larren Suh, PharmD, for participating in interviews that disclosed the difficulties of managing drug shortages at hospitals and inspired the design of the human experiment conducted in this study. This material is based upon work supported by the National Science Foundation under Grant No. 2028449.

REFERENCES

[1] Nalini Ambady and Robert Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology* 64, 3 (1993), 431.

[2] A. Ancarani, C. Di Mauro, and D. D'Urso. 2013. A human experiment on inventory decisions under supply uncertainty. *International Journal of Production Economics* 142, 1 (March 2013), 61–73. https://doi.org/10.1016/j.ijpe.2012.09.001

[3] John Robert Anderson. 2000. *Learning and memory: An integrated approach.* John Wiley & Sons Inc.

[4] İsmail Serdar Bakal, Nesim Erkip, and Refik Güllü. 2011. Value of supplier's capacity information in a two-echelon supply chain. *Annals of Operations Research* 191, 1 (Nov. 2011), 115–135. https://doi.org/10.1007/s10479-011-0937-9

[5] Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80 (2017), 1–28.

[6] Fangruo Chen and Bin Yu. 2005. Quantifying the Value of Leadtime Information in a Single-Location Inventory System. *Manufacturing & Service Operations Management* 7, 2 (April 2005), 144–151. https://doi.org/10.1287/msom.1040.0060

[7] Hyun-cheol Paul Choi, James D. Blocher, and Srinagesh Gavirneni. 2008. Value of Sharing Production Yield Information in a Serial Supply Chain. *Production and Operations Management* 17, 6 (Nov. 2008), 614–625. https://doi.org/10.3401/poms.1080.0068

[8] Rozhin Doroudi, Rana Azghandi, Zlatan Feric, Omid Mohaddesi, Yifan Sun, Jacqueline Griffin, Ozlem Ergun, David Kaeli, Pedro Sequeira, Stacy Marsella, and Casper Harteveld. 2018. An integrated simulation framework for examining resiliency in pharmaceutical supply chains considering human behaviors. In *2018 Winter Simulation Conference (WSC)*. IEEE, Gothenburg, Sweden, 88–99. https://doi.org/10.1109/WSC.2018.8632387

[9] Rozhin Doroudi, Pedro Sequeira, Stacy Marsella, Ozlem Ergun, Rana Azghandi, David Kaeli, Yifan Sun, and Jacqueline Griffin. 2020. Effects of trust-based decision making in disrupted supply chains. *PloS one* 15, 2 (2020), e0224761.

[10] Özlem Ergun, Jacqueline Griffin, Noah Chicoine, Min Gong, Omid Mohaddesi, Zohreh Raziei, Casper Harteveld, David Kaeli, and Stacy Marsella. 2022. Supply Chain Resilience: Impact of Stakeholder Behavior and Trustworthy Information Sharing with a Case Study on Pharmaceutical Supply Chains. In *Tutorials in Operations Research: Emerging and Impactful Topics in Operations*. INFORMS, 133–159.

[11] Erin R. Fox and Milena M. McLaughlin. 2018. ASHP guidelines on managing drug product shortages. *American Journal of Health-System Pharmacy* 75, 21 (Nov. 2018), 1742–1750. https://doi.org/10.2146/ajhp180441

[12] Erin R. Fox, Burgunda V. Sweet, and Valerie Jensen. 2014. Drug Shortages: A Complex Health Care Crisis. *Mayo Clinic Proceedings* 89, 3 (March 2014), 361–373. https://doi.org/10.1016/j.mayocp.2013.11.014

[13] Andrew Gelman, Jennifer Hill, and Masanao Yajima. 2012. Why we (usually) don't have to worry about multiple comparisons. *Journal of research on educational effectiveness* 5, 2 (2012), 189–211.

[14] Gerd Gigerenzer and Reinhard Selten. 2002. *Bounded rationality: The adaptive toolbox.* MIT press.

[15] Mark K Ho, Rebecca Saxe, and Fiery Cushman. 2022. Planning with Theory of Mind. *Trends in Cognitive Sciences* (2022).

[16] Nienke Hofstra, Eirini Spiliotopoulou, and Sander de Leeuw. 2022. Ordering decisions under supply uncertainty and inventory record inaccuracy: An experimental investigation. *Decision Sciences* (May 2022), deci.12564. https://doi.org/10.1111/deci.12564

[17] Robin M Hogarth, Tomás Lejarraga, and Emre Soyer. 2015. The two settings of kind and wicked learning environments. *Current Directions in Psychological Science* 24, 5 (2015), 379–385.

[18] Wallace J. Hopp, Lisa Brown, and Carolyn Shore. 2022. *Building Resilience into the Nation's Medical Product Supply Chains.* National Academies Press, Washington, D.C. https://doi.org/10.17226/26420 Pages: 26420.

[19] André Jalbut and Jaime Simão Sichman. 2018. Impact of trust on agent-based simulation for supply chains. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 45–58.

[20] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101, 1-2 (1998), 99–134.

[21] Whan-Seon Kim. 2009. Effects of a trust mechanism on complex adaptive supply networks: An agent-based social simulation study. *Journal of artificial societies and social simulation* 12, 3 (2009), 4.

[22] Mykel J Kochenderfer. 2015. *Decision making under uncertainty: theory and application.* MIT press.

[23] R. Sendhil Kumar and S. Pugazhendhi. 2012. Information Sharing in Supply Chains: An Overview. *Procedia Engineering* 38 (2012), 2147–2154. https://doi.org/10.1016/j.proeng.2012.06.258

[24] Hau L Lee, V Padmanabhan, and Seungjin Whang. 1997. The Bullwhip Effect In Supply Chains. *Sloan Management Review* 38, 8 (1997), 93–102.

[25] Hau L. Lee, Kut C. So, and Christopher S. Tang. 2000. The Value of Information Sharing in a Two-Level Supply Chain. *Management Science* 46, 5 (May 2000), 626–643. https://doi.org/10.1287/mnsc.46.5.626.12047

[26] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. *Academy of management review* 20, 3 (1995), 709–734.

[27] D Harrison McKnight and Norman L Chervany. 2000. What is trust? A conceptual analysis and an interdisciplinary model. (2000).

[28] Mili Mehrotra and William Schmidt. 2021. The Value of Supply Chain Disruption Duration Information. *Production and Operations Management* 30, 9 (Sept. 2021), 3015–3035. https://doi.org/10.1111/poms.13415

[29] Omid Mohaddesi, Jacqueline Griffin, Ozlem Ergun, David Kaeli, Stacy Marsella, and Casper Harteveld. 2022. To Trust or to Stockpile: Modeling Human-Simulation Interaction in Supply Chain Shortages. *arXiv:2201.02694 [cs]* (Jan. 2022). https://doi.org/10.1145/3491102.3502089 arXiv: 2201.02694.

[30] Omid Mohaddesi, Yifan Sun, Rana Azghandi, Rozhin Doroudi, Sam Snodgrass, Ozlem Ergun, Jacqueline Griffin, David Kaeli, Stacy Marsella, and Casper Harteveld. 2020. Introducing Gamettes: A Playful Approach for Capturing Decision-Making for Informing Behavioral Models. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–13. https://doi.org/10.1145/3313831.3376571

[31] Bennet B Murdock Jr. 1962. The serial position effect of free recall. *Journal of experimental psychology* 64, 5 (1962), 482.

[32] and Medicine and others National Academies of Sciences, Engineering. 2022. Building Resilience into the Nation's Medical Product Supply Chains. (2022).

[33] David Pynadath, Ning Wang, and Stacy Marsella. 2013. Computational models of human behavior in wartime negotiations. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 35.

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

[35] David V Pynadath, Heather Rosoff, and Richard S John. 2016. Semi-automated construction of decision-theoretic models of human behavior. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*. 891–899.

[36] Sourish Sarkar and Sanjay Kumar. 2015. A behavioral experiment on inventory management with supply chain disruption. *International Journal of Production Economics* 169 (Nov. 2015), 169–178. https://doi.org/10.1016/j.ijpe.2015.07.032

[37] Mei Si, Stacy C Marsella, and David V Pynadath. 2010. Modeling appraisal in theory of mind reasoning. *Autonomous Agents and Multi-Agent Systems* 20, 1 (2010), 14–31.

[38] Herbert A Simon. 1990. Bounded rationality. In *Utility and probability*. Springer, 15–18.

[39] Ryan Smith, Karl J Friston, and Christopher J Whyte. 2022. A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology* 107 (2022), 102632.

[40] Lawrence V. Snyder, Zümbül Atan, Peng Peng, Ying Rong, Amanda J. Schmitt, and Burcu Sinsoysal. 2016. OR/MS models for supply chain disruptions: a review. *IIE Transactions* 48, 2 (Feb. 2016), 89–109. https://doi.org/10.1080/0740817X.2015.1067735

[41] Philip E Tetlock. 1983. Accountability and the perseverance of first impressions. *Social psychology quarterly* (1983), 285–292.

[42] Stephen Tu. 2014. The dirichlet-multinomial and dirichlet-categorical models for bayesian inference. *Computer Science Division, UC Berkeley* 2 (2014).

[43] Andrew Whiten and RW Byrne. 1991. *Natural theories of mind: Evolution, development and simulation of everyday mindreading.* B. Blackwell Oxford, UK.

[44] Nutchanon Yongsatianchot and Stacy Marsella. 2021. A computational model of coping for simulating human behavior in high-stress situations. In *AAMAS Conference proceedings*.