Semi-Automated Construction of Decision-Theoretic Models of Human Behavior

David V. Pynadath
Institute for Creative Technologies, University of Southern California
pynadath@usc.edu

Heather Rosoff
Sol Price School of Public Policy & CREATE, University of Southern California
rosoff@usc.edu

Richard S. John
Department of Psychology & CREATE, University of Southern California
richardj@usc.edu

ABSTRACT

Multiagent social simulation provides a powerful mechanism for policy makers to understand the potential outcomes of their decisions before implementing them. However, the value of such simulations depends on the accuracy of their underlying agent models. In this work, we present a method for automatically exploring a space of decision-theoretic models to arrive at a multiagent social simulation that is consistent with human behavior data. We start with a factored Partially Observable Markov Decision Process (POMDP) whose states, actions, and reward capture the questions asked in a survey from a disaster response scenario. Using input from domain experts, we construct a set of hypothesized dependencies that may or may not exist in the transition probability function. We present an algorithm to search through each of these hypotheses, evaluate their accuracy with respect to the data, and choose the models that best reflect the observed behavior, including individual differences. The result is a mechanism for constructing agent models that are grounded in human behavior data, while still being able to support hypothetical reasoning that is the main advantage of multiagent social simulation.

General Terms

Algorithms

Keywords

multiagent social simulation, POMDPs, disaster response

1. INTRODUCTION

Understanding hypothetical outcomes is critical for making good policy decisions. A city government trying to ensure its readiness to respond to a disaster does not have the benefit of repeated practice. It must somehow make policy decisions that affect and are affected by thousands to millions of people.

To understand the implications and impact of such policy decisions within a complex social environment, the standard approach is to have social scientists who are experts in the domain gather human subject data, either from real or hypothetical disasters (e.g., [10]). Surveys ask questions of survivors of real disasters or experimental participants faced with a hypothetical one. The gathered data contains their answers as to what they believed about the situation, what they did (or would do), and what the important factors were in that decision. Social scientists use statistical analyses of these data to make policy recommendations. Unfortunately, the gathering and analysis of such data is costly, and the conclusions drawn are often applicable to only the specific scenario studied. To understand a different situation, one must gather more data, which is costly and labor-intensive, when it is possible at all.

Multiagent social simulation takes an alternative approach to answering the hypothetical questions asked by policy makers [2, 7, 11, 12, 18]. Such simulations instead represent people with autonomous agents that reflect individuals' or groups' decision-making perspectives and behavior. For example, we can use decision-theoretic models to capture people's decision-making processes, in the form of beliefs, choices, preferences, etc. [4, 5, 13, 15, 20]. By representing these relatively persistent characteristics, the agent can make decisions that are aligned with the corresponding real people in hypothetical situations of interest.

Of course, social simulation relies on building an accurate model. Manual construction of such models, even when grounded in well-established social science theories, is a time-consuming, labor-intensive process that often requires a great deal of trial-and-error iterations. Multiagent researchers have developed automated methods that are sometimes capable of learning such models from data. Unfortunately, in rarely occurring scenarios like disasters, we do not have observations of the transitions, as is typically required for learning decision-theoretic models of human behavior [5]. Furthermore, from the decision-making perspective, the real-world transition likelihoods are not as important as what each person *perceives* them to be (e.g., the likelihood of finding another job if s/he moves to a different city). Data on such subjective perceptions of transitions are even harder to find than data on the real transitions themselves.

In this work, we seek to exploit what subjective data we *do* have, namely surveys of beliefs, preferences, and decisions in response to hypothetical disasters (illustrated in Section 2). We present a framework for modeling the decision-making process as a Partially Observable Markov Decision Problem (POMDP) in Section 3. We exploit a factored representation and additional structural assumptions to arrive at a modeling language that is more restrictive than a general POMDP, but is more amenable to input from domain experts. This

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.), May 9–13, 2016, Singapore.*

language allows us to elicit a hypothesized model from such experts that we can potentially use within a multiagent social simulation.

However, manual model construction is no guarantee of validity, regardless of the experts' knowledge. We must instead measure the model's accuracy against human behavior observed in our data. Section 4 describes how we can quantify a model's accuracy by identifying what subset of individuals behave consistently with the model's predictions. Comparing these predictions at the individual level allows us to identify systematic errors in the model and suggest improvements. We can elicit such improvements from domain experts, allowing us to generate a hypothesis space of possible variations on our model. By identifying the individuals consistent with each of these variations, we can quantify their accuracy and make informed choices about which models to use in the eventual multiagent social simulation.

Importantly, we do not need to simply find a model that covers the most people possible. We instead select *multiple* models to cover the diverse decision-making perspectives represented in the data. We are thus able to arrive at multiple agent models that reflect the individual differences among our surveyed population. At the same time, each model is consistent with a significant number of that population, so we do not sacrifice generality to achieve this diversity.

The result is a general human-in-the-loop methodology for arriving at a set of representative agent models that are usable in simulation. This methodology offers a great deal of room for further refinement and extensions: in the space of hypothesized models, their evaluation mechanism, their selection for use in simulation, etc., as discussed in Section 5. We thus believe that this line of investigation provides a promising direction for researchers to take in pursuing rich agent-based models of human decision-making.

2. DISASTER RESPONSE SCENARIO

The scenario under investigation in this work is a hypothetical anthrax attack in Seattle [16]. The original investigation of this scenario was motivated by an exercise exploring methods for urban recovery from such an attack. As such, it was critical to understand how the public would react both to the attack itself and to official announcements and recommendations in its aftermath. Policy makers need to understand how the public will assess such situations and formulate plans of action in response. The investigators of this scenario were particularly desirous of insight into whether residents would remain in the city, leave temporarily, or leave permanently. Such behavior would have enormous long-term implications for the viability of Seattle as an urban center.

Understanding such behavior, as well as how to influence it, requires an understanding of the public's concerns in that regard. What are their primary concerns: their personal health or safety, their long-term financial stability, etc.? How deeply would the attack affect their feelings of well-being, and how willing would they be to change their behavior based on any long-term damage done to the city? Answers to these questions would help guide the city government in choosing how to prioritize their preparations and how to direct any needed post-disaster response.

It is unfortunately impossible to answer these questions using real-world data, as there fortunately has not been any such attack in Seattle. The investigation of these questions instead used a hypothetical simulation of such an attack through a video simulation of local news reports which immersed participants in an unfolding timeline of events. There were five episodes, corresponding to different points of time, from hours after the attack, to two years afterward.

After watching each of the five videos, respondents answered a series of questions. Some of these questions pertained to demographic details, such as "What is your current employment status (paid opportunities only)?" Others asked respondents to indicate a level of agreement from 1 (Strongly Disagree) to 5 (Strongly Agree) with statements like:

- "I believe the anthrax attack poses a serious risk to me."
- "If I miss ONE paycheck, I would not have enough money to buy food and pay bills."
- "I believe the government will provide emergency response and health services I might need, for example antibiotics."
- "I would LEAVE the Seattle area for at least a short time after watching the earlier news video."
- "I would change my daily routine to avoid exposure to anthrax."
- "I would obtain the antibiotic and take it as directed."
- "I would continue going to work or school if open."

Another portion of the survey asked respondents to rank personal objectives:

- "Personal safety (from crime)"
- "Financial stability (being financially stable/secure, general affordability of necessities)"
- "Health safety and personal anthrax survival (effectiveness and access to antibiotic)"
- "Family or friend anthrax survival"
- "Other"

3. AGENT MODELS

We wish to construct a social-simulation model of the residents of Seattle within a multiagent framework like PsychSim [15]. To do so, we must represent the decisions expressed in the survey within a POMDP [8]. In precise terms, a POMDP is a tuple, $\langle S, A, P, \Omega, O, R\rangle$. This section presents the components of the POMDP and how we model Section 2's disaster response survey within them.

3.1 Actions, A

A is the set of actions available to the agent, so it corresponds to the possible decisions that an individual faces in the scenario to be modeled. The behavior of interest in the survey is whether and when residents would leave Seattle, and if so, whether and when they would return. Both options become possible actions for our agent:

LeaveSeattle: "LEAVE the Seattle area for at least a short time after watching the earlier news video."

ReturnSeattle: "RETURN to the Seattle area at this point in time"

These two actions also imply a third option, which would be to stay in the person's current location:

Stay: Remain in current location, either in Seattle or beyond the Seattle area.

The survey also contains questions about behavioral choices if staying in Seattle:

ChangeRoutine: "Change my daily routine to avoid exposure to anthrax."

ContWorkSch: "Continue going to work or school if open."

OutdoorPrecaution: "Take necessary precautions to avoid anthrax exposure when outside."

TakeAntiBiotic: "Obtain the antibiotic and take it as directed."

A person can choose to perform any combination of these four actions, so that these questions imply 16 possible actions,

The survey also contains questions about employment choices upon leaving Seattle:

TempWork: "Look for temporary work."

PermanentMove: "Move out of the Seattle area (e.g. look for permanent employment, school, housing)."

These two actions are mutually exclusive.

To structure the decision problem, we divide the person's decisions into two phases, "where" and "how". In the "where" phase, the person chooses between "LeaveSeattle" and "Stay" if in Seattle and between "ReturnSeattle" and "Stay" if not. In the "how" phase, the person chooses among the 16 behavioral options if in Seattle and between "TempWork" and "PermanentMove" if not.

3.2 States, S

The state space, S, represents the key features of the agent's operating environment, both observable and hidden, both external and internal. These features can capture both objective facts and subjective perceptions about a person's decision-making context. We use the individual survey questions to constitute a factored state space, $S = S_0 \times S_1 \times \cdots \times S_n$ [1]. Demographic responses become state features, where it is safe to assume that the person has knowledge of the true values. The questions that begin "I believe" become state features, where the person may have only uncertain beliefs. There are 13 such features, of which the following are a representative sample:

JobFulltime: "My current employment status is: full-time (35 or more hours per week)"

RiskMe: "The anthrax attack poses a serious risk to me."

MissPay1: "If I miss ONE paycheck, I would not have enough money to buy food and pay bills."

GovtProvides: "The government will provide emergency response and health services I might need, for example antibiotics."

We define each such feature as a binary variable, valued as 1 if the statement is true, and 0 if it is false. We could treat these features as more fine-grained (e.g., different levels of risk). However, the survey asks the respondent's to express a degree of agreement with the statement, so we represent that degree within the beliefs (see Section 3.5), rather than within the true state of the world. There is no fundamental obstacle to enriching the domains of these state features, but in this case, there is little incentive to do so, as we do not have more fine-grained information from the respondents.

We also need state features to represent the degree to which the person's objectives are satisfied. We ignore the miscellaneous "Other" objective, leaving us with 4 new state features.

PersonalSafety: "Personal safety (from crime)"

FinanceStable: "Financial stability (being financially stable/secure, general affordability of necessities)"

PersonalSurvival: "Health safety and personal anthrax survival (effectiveness and access to antibiotic)"

FriendSurvival: "Family or friend anthrax survival"

We define these features as binary variables as well, valued as 1 if the objective is satisfied, and 0 otherwise. This is obviously a gross oversimplification, in that "PersonalSurvival" cannot distinguish between dying from anthrax and becoming infected with anthrax but surviving. We choose to start with the simplest possible model here, but we can always add additional levels to the domains of these state features later without changing our methodology.

As described in Section 3.1, certain choices depend on the person's location and on the current phase of decision-making. We therefore introduce additional state features to keep track of these conditions:

location: Either "Seattle" or "beyond", indicating whether the person is within or beyond the Seattle area.

phase: Either "where" or "how", depending on the current phase of decision-making

These two features potentially eliminate certain actions from consideration. For example, the action "LeaveSeattle" is considered only when "location" is "Seattle". Likewise, "PermanentMove" is considered only when "location" is "beyond".

3.3 Reward, R

The agent's reward, R, represents the objective function that it is seeking to maximize, so it makes a natural mechanism for representing people's preferences. As described in Section 3.2, a subset of the state features in S represent the objectives from the survey. We limit the reward function to concern only this subset of state features, and specify it according to the priorities expressed in the respondent's ranking. In particular, we define the reward function as a weighted sum of the values of the objective state features. The survey asks people to rank the objectives from 1 to 5, so we weigh each state feature by (6 − its rank) within R. This translation allows us to treat people's ranking of the objectives as a direct expression of their reward function.

The ranking allows us to identify how important the satisfaction of these objectives is to people, but it does not give us any information about the degree to which they believe them to be satisfied. We may get indirect information about these beliefs from the other questions, but we make a simplifying assumption that there is no such information. We instead treat the initial value of the objective state features as if everyone had responded with a 3 to questions about their belief. This is not a critical assumption at this point, as it is the *change* in objective satisfaction that drives behavior, and the starting value has less impact.

3.4 Transition Probability, P

The transition probability, P, represents the probabilistic effects of actions on the state of the world. Representing such effects allows us to capture the way in which people can anticipate the expected outcomes of their possible decisions. The transition probability for our non-survey state features ("location" and "phase") is straightforwardly deterministic. If the person chooses, "Stay", then "location" does not change. If the person chooses "LeaveSeattle", then "location" becomes "beyond". If the person chooses "ReturnSeattle", then "location" becomes "Seattle". The transition probability for "phase" is similarly deterministic, in that when it is "where", it becomes "how" after performing the chosen action, and vice versa.

The transition probability for the other state features in Section 3.2 is not as simple. There is no a priori obvious definition of the effects of (for example) "TakeAntiBiotic" on "RiskMe" and "PersonalSurvival". Likewise, there are no explicit questions about these effects in the survey that can inform such a definition. Therefore, we treat the specification of this transition probability as the heart of the modeling task.

To both constrain the potential search space and to simplify the elicitation of expert knowledge, we restrict the structure of the transition probability function. We start from the standard factored POMDP's use of Dynamic Bayesian Networks [9] and influence diagrams [6] to exploit conditional independence in modeling the effects of actions [1]. We can thus express dependencies among our states and actions as links among the nodes of a dynamic influence diagram [19], as in the example model visualized in Figure 1. The ovals on the left represent the state values at time t, the rectangles in the middle represent the possible action choices at time t, and the ovals on the right represent the state values at time $t+1$. The colored nodes represent states and actions specific to the person, while the uncolored nodes represent global states (e.g., if the government provides health care, it applies to every resident). The links from "location" and "phase" to the action nodes indicate that the available choices depend on those variables.

3.4.1 Action Effects

The other links represent dependencies encoded in this particular model of the person's decision-making. In this illustrative model, "RiskMe" is affected by the possible "how" actions in Seattle (e.g., "TakeAntiBiotic"). In general, the dependency expressed on the links could be an arbitrary conditional probability table across the combinations of parent node values. To simplify the model specification, we instead specify the dependency on each link independently.

For the links from action nodes to subsequent state features, we specify a magnitude and direction of the dependency for each possible prior value of the state feature (`True` or `False`). We express the magnitude and direction by -1, 0, or 1, representing that the performance of the action has a negative, neutral, or positive effect, respectively, on the likelihood of the state feature being `True` afterward. For example, we can specify a value of 1 for the effect of "ContWorkSch" on "RiskMe" (whether starting at `True` or `False`) to represent the increased risk incurred by continuing to go to work or school. Likewise, we can specify a value of -1 for the effect of "LeaveSeattle" on "JobFulltime" (whether `True` or `False`) to represent the challenge of finding another full-time job after leaving Seattle. We do not currently allow these links to be contingent on other state features (e.g., going to work may have less impact on the anthrax risk if "Decontaminated" is `True`). However, this is again a trivial relaxation that would only change the search space, and not any of the methods.

We translate the links from actions to a state by aggregating the -1 and 1 values on those links. We first compute the minimum and maximum possible incoming weights by looking at the possible combinations of actions (e.g., all of the "how" actions simultaneously) and counting the -1 and 1 values separately. For our initial model, a person choosing "ContWorkSch" would incur a maximum value of 1 over "RiskMe"'s incoming links, while a person choosing "ChangeRoutine", "OutdoorPrecaution", and "TakeAntiBiotic" would incur a minimum value of -3. We then normalize the total values for each possible action choice across this range (in this example, [-3,1]) and map it to a Likert scale of 1–5. We then translate the result into a probability of the state feature being `True` using a table of our design: $1 \to 10\%$, $2 \to 25\%$, $3 \to 50\%$, $4 \to 75\%$, and $5 \to 90\%$.

3.4.2 Interdependency among States

We make a similar simplifying assumption for links incoming to objective nodes. We label each link from a state node to an objective node with two numbers on a 1–5 scale, with the two values representing the conditions where the parent node is `True` or `False`. A value of 1 (5) indicates that the parent node's being `True`/`False` strongly decreases (increases) the likelihood of the objective node being `True`. For example, our initial model has a link from "RiskMe" to "PersonalSurvival". This link is 1 for `True`, indicating that if there is a risk to the person, then the survival chances go down. This link is 5 for `False`, indicating that survival chances increase if there is little to no risk.

We fill in the conditional probability table over all of the incoming links at an objective node using a noisy OR. Based on the `True`/`False` values of the parents, we translate the corresponding link values into probabilities, p_i, using the same mapping as in Section 3.5. We then use a standard noisy OR formula, $\Pr(\text{child}|\text{parents}) = 1 - \prod_i (1-p_i)$. As in the action dependencies, we are making a strong assumption of independence among these effects on the objectives. Relaxing this assumption is also trivial, in that it only reduces the space of possible dependency definitions, and does not affect any of the algorithms.

3.4.3 Initial Model

The examples in this section are drawn from the transition probability function used in our initial model. We elicited the links in this initial model from the social scientists who conducted the survey and had performed some preliminary statistical analyses of the data. The resulting transition probability function included 6 non-zero links from actions

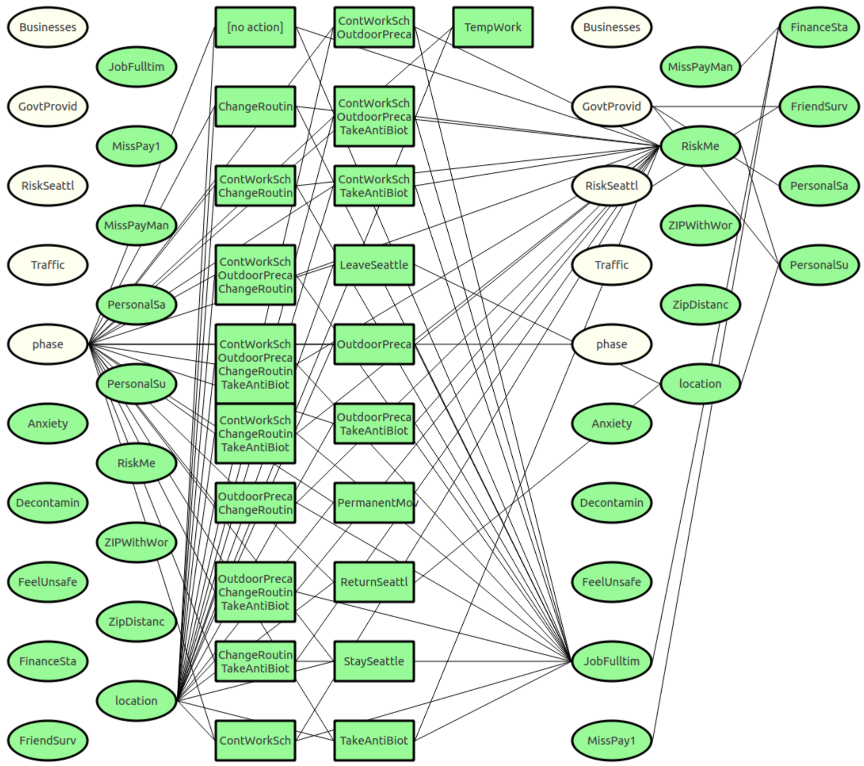

Figure 1: Influence diagram visualization of P.

to "RiskMe" and 1 to "JobFulltime". It also included 3 links from states to "PersonalSurvival", 3 to "PersonalSafety", 4 to "FinanceStable", and 5 to "FriendSurvival".

3.5 Observations, Ω and O

Ω is a set of possible observations that the agent may receive, while the function, O, represents the likelihood of receiving specific observations as a function of the current state of the world. We can use the combination of Ω and O to capture the information (possibly incomplete and noisy) that people receive about the state of the world, especially in terms of how that information shapes their subjective perceptions. Within our disaster-response scenario, we take two state features, "location" and "phase", to be fully observable. Therefore, these components of Ω have the same domain as their corresponding state variables in S. We do not model any other observations in the current investigation.

In general, we can define Ω just as we do state features, and the observation function, O, just as we do the transition probability, P. For example, if we did not limit ourselves to only the first decision stage, we could expand Ω to include observations like the number of deaths mentioned in a video broadcast. We could then add potential links from state features (e.g., "RiskMe") to these observation variables, with the same coarse positive/negative influence labelings we use within P. We likewise use the same noisy OR function to combine the incoming influences into a single probability of a given observation in O.

The current investigation does not explore this observation space, because we do not model the belief update that will occur due to future videos, nor the expectations of the decisions that will be made after those videos. We do still account for such partial observability, but we take the person's starting beliefs directly from the survey responses. The responses range from 1–5, where 1 (5) indicates strong disagreement (agreement) with the statement. We therefore translate the responses into probabilities of the statement being `True` using the same mapping from Section 3.4.1: $1 \rightarrow 10\%$, $2 \rightarrow 25\%$, $3 \rightarrow 50\%$, $4 \rightarrow 75\%$, and $5 \rightarrow 90\%$. This translation allows us to treat people's responses as a direct expression of their subjective belief state.

4. DATA-DRIVEN MODELING

Our initial model provides a good starting hypothesis, partially informed by both domain knowledge and data analysis. However, we need to make fuller use of the data to both validate and refine the model. Even after all of our restrictive assumptions made on the transition probability in Section 3.4, there is still an enormous space of possible models. There are $21 \times 13 = 273$ possible links from actions to states, with 9 possible specifications each ("increase", "decrease" or "stay the same" for `True` or `False` values of the original state). Similarly, there are $13 \times 4 = 52$ possible links from state features not related to objectives to state features that are tied to objectives, with 25 possible specifications each (1–5 Likert scale for `True` or `False` values of the original state). There are thus $9^{273} + 25^{52} = 3.2 \times 10^{260}$ possible models. This space of models would get only larger if we expand the domain of our variables beyond our current binary values. It is therefore infeasible to perform anywhere near an exhaustive search of this space.

4.1 Executing an Agent Model

Regardless of how we search the space of possible models, we first need a method to evaluate a candidate model. The POMDP represents a decision model that can compute the expected reward, $E[R]$, of each candidate action in A. We start from a belief state, b, to be provided from the survey responses, and consider each possible "where" action, a_W. We then apply the transition probability function to proceed to the "how" phase. At this point, we could also use the POMDP to generate "how" decisions, but we instead use the survey responses, translating the 1–5 responses into a probability of action choice, $\pi_H(a_H)$, using the same mapping from Section 3.5. We again apply the transition probability function to arrive at a final state. We apply the reward function to the states to arrive at a final value, as expressed in Equation 1:

$$V(b, a_W) = \sum_{s\in S} b[s] \sum_{s'\in S} \Big(P(s, a_W, s')R(s') + \sum_{a_H} \pi_H(a_H) \sum_{s''\in S} P(s', a_H, s'')R(s'')\Big) \quad (1)$$

Rather than optimize this POMDP and choose the action, a_W, for which V is the highest, we instead use a *softmax*, which uses a Boltzmann-like distribution to make the likelihood of an action choice dependent on its relative value:

$$\pi_W(b)[a_w] = \frac{e^{r\cdot V(b,a_w)}}{\sum_{a'} e^{r\cdot V(b,a')}} \quad (2)$$

where r is the reciprocal of the temperature constant in a Boltzmann distribution. In this case, r controls a degree of certainty or rationality by which the agent follows its value function. The higher this value, the more likely the agent is to choose the optimal action. We start with $r = 10$ based on experience with prior agent modeling in PsychSim.

4.2 Evaluating an Agent Model

A candidate POMDP model thus generates a policy of "where" behaviors that we can now compare against the data. We focus here on the "where" decisions made in the first stage of the survey, when everyone's "location" is "Seattle". We examine each individual response in the survey data, extract the person's beliefs, b, and "how" behaviors, π_H, and plug them into Equations 1 and 2 to generate a probability of "LeaveSeattle". We then map this probability to our 1–5 scale (inverting the mapping from Section 3.4.1) and compare it against the person's actual response for "LeaveSeattle".

Our initial model generates the correct "LeaveSeattle" response for 127 of the 433 people in the survey. This is not bad, considering the minimal dependencies included in our initial model. However, there is obviously much room for improvement. We can compute a confusion matrix for our model, looking more closely at the specific mistakes it is making. Figure 2 shows the counts of individuals under all combinations of actual and predicted responses. The bold entries along the diagonal indicate the numbers of individuals for whom the model makes an accurate prediction.

Survey Response	Model Prediction 1	2	3	4	5
1	**0**	0	9	3	0
2	0	**0**	43	24	0
3	0	0	**92**	32	1
4	0	0	85	**33**	0
5	0	0	68	15	**1**

Figure 2: Prediction matrix from initial model.

4.3 Model Search

By comparing our model's predictions against individual responses, we can more easily see systematic flaws in the model. In particular, we observe that our initial agent model predicts a much greater willingness to leave Seattle (higher values are stronger agreement with leaving). There are many possible disincentives for leaving that we could add to the model. Adding such disincentives would lead to models that are able to capture the individuals who are less willing to leave—in other words, those that responded with 1 or 2, all completely missed by our initial model in Table 2.

For example, upon leaving, the agent can choose a "how" behavior of "PermanentMove", which, in our initial model, increases the likelihood of achieving "JobFulltime" (value of 1) which in turns increases the likelihood of satisfying "FinanceStable" (value of 5). However, many people may be less optimistic about their chances of finding a full-time job upon moving, which we could capture by making the value on the former link 0. The resulting agent model would see less expected reward from choosing "PermanentMove" than our initial model, which would in turn lower the expected reward of "LeaveSeattle".

We can formulate other such hypotheses as well. Trying to leave Seattle may be dangerous, due to crowds of other people doing the same. We can represent this possible effect with a link from "LeaveSeattle" to "PersonalSafety" with a value of -1 for `True` and 0 for `False`. In other words, people who start off being safe would feel that leaving Seattle would slightly endanger their safety, while those who start off being unsafe would not have their safety affected by leaving.

By considering all possible combinations of these link modifications, we accumulate a set of possible variations on our initial model. We can then execute each of these new models across all of the individual survey responses, following the behavior generation procedure of Section 4.2. In this investigation, we generated a set of 10 possible variations, leading to 1024 possible models for consideration.

Comparing the output of these models against the survey responses allows us to clearly see the accuracy of potential hypotheses with respect to the data. For example, although leaving Seattle had an implied negative impact on financial stability (due to potential unemployment), we also hypothesized that perhaps adding a direct negative link from "LeaveSeattle" to "FinanceStable" might be more accurate. However, introducing this link did not change the predictions, and comparison against the survey responses provided concrete evidence of its redundancy.

4.4 Model Choices

Having compared the behavior generated by all 1024 models across the individuals in the survey, it turns out that our initial model's match of 127 respondents is the highest match count achieved. There are three other models that match an equal number of people, and, in fact, they match the same 127 individuals. These four models are identical except for the absence or presence of two links. We can therefore treat them as equivalent, and we use a bias toward fewer links to choose the model that has both links absent.

While having a model that matches 127 out of 433 people may not be very impressive, it is important to note that we are not done. Our goal is a multiagent social simulation, with multiple agents representing the population of interest. Therefore, there is no requirement that all of the agents in the simulation use the same model. We instead seek to find a *set* of agent models that covers the range of behaviors observed in the data.

The freedom to choose multiple models gives us an ability to capture individual differences within our data. The survey does capture those differences in beliefs and preferences, but there are likely to also be differences in each person's view of cause and effect. In other words, it is unreasonable to expect that everybody is going to forecast the effects of leaving in exactly the same way. By allowing ourselves to choose multiple agent models, we can arrive at a population of agents that share the same diversity of perspective as the people they are representing.

We have a variety of options in how to choose the set of models. In this work, we take a greedy approach, where we first choose the model that covers the largest number of people. We then add the model that covers the largest number of the people not covered by the first model. We proceed until we have either covered everybody, or have no models that match the remaining people.

When we hit the last case, it gives us an opportunity to re-examine our hypothesized model variations. Beyond looking at just the incorrect predictions, we can also examine the beliefs expressed by the individuals for whom our models are inaccurate. For example, our models failed to match the subset of respondents who expressed a strong willingness to perform all of the safe "how" behaviors (e.g., take antibiotics), but still wanted to leave, despite the diminished risk entailed. One hypothesis is that these people had a lower belief in the efficacy of these methods, when compared against the dependencies in our models. By varying our model so that these methods were not effective (i.e., had 0 effect on "RiskMe"), we arrive at a new hypothesis model that generates the correct staying behavior.

At the other end of the spectrum were people who responded in ways consistent with wanting to leave (e.g., high belief in "RiskMe", high rank for "PersonalSurvival") but were strongly against leaving. Some of these respondents also expressed a strong willingness to adopt the safe "how" behaviors. We therefore hypothesized that they may have a stronger belief in the efficacy of such behaviors, even stronger than in our initial models. We relaxed our restriction on the values of links from these actions to states, so that we used a value of -3 instead of -1. This variation was able to explain most of these people, who appeared as outliers with respect to our original models.

We have focused so far on variations that modify only the links within our transition probability function. This restriction is self-imposed, because we are free to consider any variation on our models that still supports the execution of the underlying POMDP. For example, we observed that there were cases when our models would be correct with respect to whether a person was favoring leaving or staying, but would be overly strong in that sentiment. In such cases, the model might be correctly capturing the tradeoffs being considered, but incorrectly capturing the confidence the person has in the resulting valuation. We can potentially represent such a lack of confidence by modifying our softmax parameter, r. We tried lower values for r, both 5 and 2, instead of 10.

4.5 Results

Our greedy search begins by making the following sequence of model choices:

Model variation	Matches
Original	127
"LeaveSeattle" decreases "PersonalSafety"	77
"PermanentMove" no effect on "JobFulltime"	44
Both of the previous two variations	43

The next selections include too many variations to list explicitly, but they add the following numbers of individuals to the covered set: 35, 23, 15, 13, 9, 5, 4, 3, 2, 2, 2, 2, 2, 1. We can therefore cover 407 of the 433 people (94%) using 17 different agent models. There is an obvious diminishing of returns for the later models added, and it is not clear how generally useful a model that covers only one or two people might be. However, there is still a good deal of generality with the most "useful" models. With the first 10 models, we cover 391 of the 433 people (90%), and with just the first 5 models, we cover 326 people (75%).

Of the 26 people who remain unmatched by any of the models, 19 of them favor staying. We are therefore still missing a disincentive for leaving. For example, many people are likely to have an emotional attachment to their homes, an attachment that would not show up in the survey itself, let alone in our current POMDP expected reward calculation. Again, while we would prefer being able to find models that cover all of the people, even such mismatches are informative when we examine the unmatched responses within a POMDP decision-making framework.

While the selected models demonstrate a certain degree of accuracy with respect to the surveyed Seattle residents, one might question how well these models will generalize to other scenarios. A subsequent survey placed 466 residents of the San Francisco area in a similar hypothetical anthrax scenario [17]. The variables in our models map to a subset of questions asked in this more recent survey, so it is straightforward to apply our agent model to the new data. Encouragingly, our initial model, which captured 127 (29%) of our Seattle residents, also captures 139 (30%) of our San Francisco residents. However, our secondary model choices do not generalize as well, as the first 5 models cover only 306 (66%, instead of 75%) of our San Francisco residents. By examining differences in the modeling features in more depth, both within and across scenarios, we can potentially gain further insight into the different dependencies that are guiding the thinking of different subpopulations.

5. DISCUSSION

In this investigation, we made several restrictive assumptions in developing a proof-of-concept for our general methodology. In this initial attempt, we were able to follow a human-in-the-loop search method to arrive at agent models (in POMDP form) that accurately represent our human behavior data. There are a wide number of possible extensions that can build upon this methodology and expand its power in useful and interesting ways.

Within the modeling assumptions made, there are many candidates for relaxation. It would be obviously easy to extend the potential dependency hypotheses to be more fine-grained, as we did in our one case of using -3 instead of -1 on a link from "how" actions to "RiskMe". As already mentioned, we could also allow links to be contingent on other state features. The domains of the variables themselves are also subject to expansion, as nothing in our method requires our current choice of binary domains.

We could also relax other aspects of the POMDP model, just as we did for the rationality parameter, r. For example, we could allow for an expanded horizon, rather than restricting all of the expected reward calculations to consider only two time steps. In general, every aspect of the POMDP model is subject to variation to generate novel hypotheses to consider. The evaluation of the hypotheses is even more general, in that the comparison against individual data and choices of model coverings would apply to non-POMDP agent models as well.

The current evaluation of our models is limited to the responses from the first stage of the survey, but the survey includes additional questions at four subsequent stages, after watching additional news broadcasts that reflect the further passage of time. These "longitudinal" questions would provide a more stringent validation of our models. However, the subsequent surveys do not have all the same questions, so we cannot extract the same belief states at those stages. However, in future work, we can incorporate a belief update stage that would allow us to include hypotheses about those missing future beliefs, as well as the ensuing decisions. This would greatly expand the hypothesis space, but structural assumptions related to the observation function should be able to keep this expansion within feasible and descriptive limits. The additional cost might be worth incurring, because agent models that correctly capture an individual's decisions over these multiple stages would have passed a higher bar of validation and would engender even more confidence in their accuracy.

In this investigation, we consider a relatively small number of modeling hypotheses, 10 possible variations, which allows us to exhaustively evaluate all possible combinations of them. With larger hypothesis spaces, we will not have that luxury, in which case search control over these combinations arises as a new challenge. One possibility is to consider one variation at a time to perform a local search through the model space, along the lines of gradient descent or evolutionary search. Such a search would be directed toward models that cover more and more people, although we would of course lose the guarantees of finding the models with the largest coverage.

In this work, we also had the luxury of domain experts to suggest variations to explore. It is easy to envision a completely automated approach to hypothesis generation instead. However, it is less clear how best to generate informed hypotheses. One possibility is to exploit piecewise linear representations, which have shown to provide some degree of invertibility within decision-theoretic models like Bayesian networks [3] and PsychSim itself [14]. Exploiting such representations would potentially allow one to identify and modify the links that are contributing the most error in the value function.

Having arrived at an evaluation of a satisfactory set of hypothesized models, there still remains the question of how to choose an appropriate subset of them to use in simulation. We use a greedy algorithm as a first approach, but there are undoubtedly better methods. An exhaustive search of all possible coverings may be intractable, but, even if it were feasible, it is not clear what optimality criterion we would use to make our choice. One possibility is to choose the covering that uses the fewest number of models. Another possibility is to minimize the overlap among the models chosen, so that the models essentially partition the population. It is an open research question as to what scoring function leads to modeling choices that provide the highest fidelity in social simulation.

Finally, we apply our simulation methodology here to a survey data set that was never intended for such a purpose. While we were regardless able to gain considerable modeling leverage from this data set, this exercise also informs us as to potentially valuable refinements to the survey instrument for future iterations. For example, even within our limited hypothesis space, there is considerable overlap where an individual's behavior is consistent with multiple models. By examining the points at which these models *differ*, we can potentially arrive at questions to ask in future surveys to disambiguate such cases. Thus, in addition to survey data informing our modeling, we return the favor by using our modeling to inform future data gathering.

6. CONCLUSION

By leveraging a structured POMDP representation, we arrive at a reduced space of possible modeling hypotheses. By comparing the predictions of candidate models against the responses of individual people, we get more detailed feedback on the accuracy of those models than we would from an aggregate statistical analysis. Finally, by choosing multiple agent models, rather than a single best model, we are able to capture individual differences, while still generalizing enough that each model can capture a significant number of people.

The overall impact of this methodology is that we leverage the POMDP's general decision-making framework to capture persistent and shared components of human decision-making (in particular, the cause-and-effect model within the transition probability). However, because our goal is not a single general claim about human behavior, but rather a multiagent simulation of a population, we are free to consider sets of models. We are thus able to include sufficient diversity within our agent models to better cover the whole population. This is especially important in social simulation, where understanding the "atypical" behaviors (e.g., looters after a disaster) is just as critical, if not more so, as understanding the typical ones. Thus, this work demonstrates a path toward enabling researchers, both in computer *and* social science, to more easily build and validate decision-theoretic multiagent models of human behavior.

Acknowledgments

This work was sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). The content does not necessarily reflect the position or policy of the Government. No official endorsement should be inferred.

REFERENCES

[1] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning: Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 11(1):94, 1999.

[2] K. M. Carley, D. B. Fridsma, E. Casman, A. Yahja, N. Altman, L.-C. Chen, B. Kaminsky, and D. Nave. BioWar: scalable agent-based model of bioattacks. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 36(2):252–265, 2006.

[3] H. Chan and A. Darwiche. When do numbers really matter? In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 65–74, 2001.

[4] A. S. Goodie, P. Doshi, and D. L. Young. Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, 25(1):95–108, 2012.

[5] J. Hoey and J. J. Little. Value-directed human behavior analysis from video using partially observable markov decision processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1118–1132, 2007.

[6] R. A. Howard and J. E. Matheson. Influence diagrams. In *The Principles and Applications of Decision Analysis*. 1984.

[7] *The Journal of Artificial Societies and Social Simulation*. http://jasss.soc.surrey.ac.uk/JASSS.html.

[8] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101:99–134, 1998.

[9] U. Kjaerulff. A computational scheme for reasoning in dynamic probabilistic networks. In *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, pages 121–129, 1992.

[10] M. K. Lindell, K. J. Tierney, and R. W. Perry. *Facing the Unexpected:: Disaster Preparedness and Response in the United States*. Joseph Henry Press, 2001.

[11] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan. MASON: A multiagent simulation environment. *Simulation*, 81(7):517–527, 2005.

[12] *Proceedings of The International Workshop on Multi-Agent-Based Simulation*. http://www.pcs.usp.br/~mabs/.

[13] P. Paruchuri, N. Chakraborty, G. Gordon, K. Sycara, J. Brett, and W. Adair. Inter-cultural opponent behavior modeling in a POMDP based automated negotiating agent. In *Models for Intercultural Collaboration and Negotiation*, pages 165–182. Springer, 2013.

[14] D. V. Pynadath and S. C. Marsella. Fitting and compilation of multiagent models through piecewise linear functions. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems*, pages 1197–1204, 2004.

[15] D. V. Pynadath and S. C. Marsella. PsychSim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1181–1186, 2005.

[16] H. Rosoff, R. John, W. Burns, R. Siko, et al. Structuring uncertainty and conflicting objectives for life or death decisions following an urban biological catastrophe. *Journal of Integrated Disaster Risk Management*, 2(1):49–69, 2012.

[17] H. Rosoff, R. Siko, R. John, and W. J. Burns. Should I stay or should I go? an experimental study of health and economic government policies following a severe biological agent release. *Environment Systems & Decisions*, 33(1):121–137, 2013.

[18] R. Sun. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*. Cambridge University Press, 2006.

[19] J. A. Tatman and R. D. Shachter. Dynamic programming and influence diagrams. *IEEE Transactions on Systems, Man and Cybernetics*, 20(2):365–379, 1990.

[20] N. Wang, D. V. Pynadath, and S. C. Marsella. Subjective perceptions in wartime negotiation. *IEEE Transactions on Affective Computing*, 6(2):118–126, 2015.