Satisficing Models of Bayesian Theory of Mind for Explaining Behavior of Differently Uncertain Agents

Socially Interactive Agents Track

Jan Pöppel
CITEC, Bielefeld University
Bielefeld, Germany
jpoeppel@techfak.uni-bielefeld.de

Stefan Kopp
CITEC, Bielefeld University
Bielefeld, Germany
skopp@techfak.uni-bielefeld.de

ABSTRACT

The Bayesian Theory of Mind (ToM) framework has become a common approach to model reasoning about other agents' desires and beliefs based on their actions. Such models can get very complex when being used to explain the behavior of agents with different uncertainties, giving rise to the question if simpler models can also be satisficing, i.e. sufficing and satisfying, in different uncertainty conditions. In this paper we present a method to simplify inference in complex ToM models by switching between discrete assumptions about certain belief states (corresponding to different ToM models) based on the resulting surprisal. We report on a study to evaluate a complex full model, simplified versions, and a switching model on human behavioral data in a navigation task under specific uncertainties. Results show that the switching model achieves inference results better than the full Bayesian ToM model and with higher efficiency, providing a basis for attaining the ability for "satisficing mentalizing" in social agents.

KEYWORDS

Theory of Mind; Action Understanding; Reasoning; Mental Models

ACM Reference Format:
Jan Pöppel and Stefan Kopp. 2018. Satisficing Models of Bayesian Theory of Mind for Explaining Behavior of Differently Uncertain Agents. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 9 pages.

1 INTRODUCTION

Inferring the mental state of others is a hallmark of our human capability for social understanding and interaction. We see somebody walking towards us and immediately think about a number of possible intentions or goal destinations she might have. This mentalizing is called Theory of Mind (ToM) [22], underlining that we are constructing and using a theory of the mental state of other agents in order to explain their actions. This enables us to prevent others from harming us or allows us to successfully work together with them without the need of overly explicit communication.

A common assumption is that humans strive to uncover other people's goals [11], their inner beliefs [34] and even social contexts [28] to a certain extent. However, research has shown that some mental states are harder to infer than others [32, 33]. While people can quite easily detect the desires or goals behind actions of other people, it is harder to infer their beliefs which potentially differ from one's own. Inferences about another agent's perceptual access, false beliefs, or emotions are likely to be even more difficult. It has also been argued that we do not always use our ability for mentalizing to its fullest extent [16] and that in some situations it is preferable to go with a lower order of ToM with less recursive mental states [9, 10]. Furthermore, some even argue that humans use a "minimal theory of mind" due to their bounded cognitive resources [6]. We assume that it is crucial for humans to employ "satisficing mentalizing" when inferring other's mental states. That is, as long as easily inferred information (like another's likely goal) is sufficient for one's current purposes, one will rarely invest additional cognitive effort to infer additional mental states such as the other's current beliefs. Yet, it is not clear what kind of ToM is needed to "satisficingly" explain different behavior of different agents.

While some work was directed at simulating mentalizing in a variety of agent-agent interaction scenarios (e.g. negotiation [17], or interactive narrative [27]), others have tried to employ such models to increase the abilities of conversational agents or social robots to interact with and adapt to human users [5, 26]. An approach that has become increasingly popular in such interactive systems, but also in computational cognitive science [2], are probabilistic (Bayesian) models of ToM. These models come with high computational costs needed to form and maintain beliefs. Consequently, models are usually optimized to a specific scenario but it is not clear if alternative models could have achieved equally satisficing performance. Only few have looked at comparing different ToM models, e.g. with respect to the modeling agent's actions [23], but not with regard to an accuracy-efficiency trade-off and to enable an agent to choose a suitable ToM model itself.

In this paper, we present work towards social agents that can switch Bayesian models of ToM to reason about the mental states of other agents in different situations in a satisficing way, i.e. with sufficient accuracy *and* reasonable costs. We start from a complex Bayesian ToM model and show how simpler models can be constructed, each suited to reasoning about agents with different uncertainties about their environment, that can achieve similar accuracy while being considerably more efficient. The next section provides background information on mentalizing in humans, the general Bayesian ToM framework and modeling attempts in social agents. Then we describe our proposed method and present results from a comparison of the different models on human behavioral data from a navigation task inducing different kinds of uncertainties.

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

2 BACKGROUND

Mentalizing, or Theory of Mind [22] has been studied for many years now. Recent research increasingly suggests that ToM is an adaptive ability and can incur different amounts of effort. During acquisition, there seems to be an order of difficulty from inferring desires, over differing beliefs, perceptual access, false belief up to emotions [32]. Even after acquiring the skills to infer complex (e.g. recursive) mental states, humans do not always use these skills to the fullest [9, 16]. While this partly can be due to the cognitive effort of the underlying mental inferences, it is not always necessary or even useful to infer too complex mental states [10]. One way to regulate the degree of mentalizing might be some measure of surprise (or predictability). Generally, surprising events are found to trigger a series of mental steps, starting from recognizing an event as surprising and ending in revising one's beliefs [19]. Recent work shows that surprising actions of other agents cause longer reaction times in humans, which could correlate to this adaptive revision of one's beliefs [20].

2.1 Bayesian Theory of Mind

In recent years a lot of research has considered the problem of inferring the mental states of agents through probabilistic inverse planning [2, 3, 7, 11, 14, 29, 30], usually in a Bayesian framework that was termed Bayesian Theory of Mind (BToM) by Baker et al. [1]. The general idea is to model causal relations between mental states such as desires, beliefs, preferences and actions. Bayes' rule is used to invert the model in order to infer the agent's desires and beliefs based on observed actions:

$$P(d, b|a) = \frac{P(a|d, b)P(d, b)}{P(a)} \quad (1)$$

where $P(a|d, b)$ represents the modeled likelihood of an action given a desire d and a certain belief b. $P(d, b)$ represents the prior knowledge about the joint desire and belief distribution. The probability of the action $P(a)$ serves as a normalization constant. Computing this normalization constant is often difficult since it requires to marginalize over all desires and belief states.

Baker et al. [3] point out that the inverse planning problem is ill-posed and "requires strong prior knowledge of the structure and content of agents' mental states, and the ability to search over and evaluate a potentially very large space of possible mental state interpretations".

One way of enabling the search over large mental state spaces is by use of sampling, which has even shown to provide better fits to human data in some conditions [31]. Sampling circumvents the need to marginalize over possibly infinite domains by approximating the inference and is the usual approach whenever the probabilistic models become too complex. However, since we are interested in a comparative analysis and sampling can be used on simple as well as complex models, we are not considering the potential added value of sampling for the remainder of this paper.

Apart from sampling, one generally limits the model to the belief states and desires one is interested in within a given scenario. Previous research has shown, that such simplified models can explain human behavior on a level comparable to humans [2, 3]. However, different scenarios require one to specify a different set of possible discrete belief states, thus a new model.

2.2 Artificial Social Agents

Modeling Theory of Mind in artificial systems is not only useful from the perspective of social psychology but also vastly important in artificial social agents: Inferring other agent's plans allows a better predictability of future actions [11, 25], better adaptation to other agents as well as collaboration [15]. ToM is also used to perform social simulations of multiple artificial agents [24]. Overall these capabilities can help build systems which behave and reason more like what humans are used to, which is very important for the acceptance of such systems [8].

Pynadath and Marsella [23] proposed an interesting approach to form minimal mental models in a multi-agent simulation scenario. Several models of another agent are clustered and considered equivalent based on different criteria (e.g. the expected utility or the resulting behavior of the modeling agent) as well as discrete assumptions about its belief states. Their results, where even the simplest of those equivalent models performed similarly to complex ones, show that not all possible belief states need to be considered.

3 OVERALL APPROACH

In order to develop a framework for "satisficing mentalizing" in social agents, we adopt a twofold approach. First, we construct different BToM models of different complexity and investigate their efficiency and accuracy in explaining behavior of different agents. Importantly, these models rest on different assumptions about the kinds of uncertainties an agent might have. To that end, we create simpler models by making discrete assumptions about certain beliefs in a complex model that contains the full belief states we might be interested in. Each set of assumptions yields a different simplification of the complex model, suitable to explain behavior under certain conditions. For example, in the simplest case one can assume that an observed agent holds a true belief about its environment. In this case, one can omit inferring belief states entirely and solely focus on inferring the agent's desire (as shown successfully in [3]). A competing assumption is that the modeled agent only knows part of its environment, e.g. being uncertain about the exact configuration of possible goal states.

Secondly, we propose a method for combining multiple simpler models when explaining the behavior of agents whose unknown mental states may require different kinds of assumptions. In order to satisfy the requirements for low computational costs, our approach will be model-switching, i.e. we look for a method to determine how well a current model can explain an observed behavior and to switch to another more appropriate, potentially more costly model only if needed. Our current approach to find the most satisficing model is to start by using the simplest model, i.e. making strong assumptions about the modeled agent. This corresponds to assuming initially that the modeled agent is fully competent and has true beliefs about the environment (e.g. how goal states can be achieved). This model will be able to explain rational goal-directed behavior but will make wrong predictions for behavior of agents holding differing beliefs. To enable this model switching, we are computing a surprise measure for each observed behavior. If this surprise measure becomes too large, the framework will re-evaluate its current assumptions, adapting them if necessary, analogous to what has been found in surprise theory in humans [19]. The adapted set of assumptions corresponds to a new model that explains the observed behavior better. Note that this surprise measure can be computed in a number of ways. We will discuss two previously proposed methods of computing surprise in our example below.

In the remainder of this section we present a first detailed implementation of this approach in the domain of explaining goal-directed navigation behavior. Section 4 will then describe in detail a maze-like environment that we used to gather human behavior data, under different conditions to induce different kinds of uncertainties in the participants. These data will be then be used for a comparative evaluation of the BToM models in section 5.

3.1 BToM in a Navigation Task

Our example domain is similar to the foodtruck domain of Baker et al. [1] in that agents are navigating an environment with multiple possible goal locations of which the agent will want to reach exactly one. We add two different sources of uncertainty: First, the agent may or may not know where the desired goal is. Similar to Baker et al. they may know potential locations of the goals, but may be uncertain about the exact mapping between goals and locations. We will refer to this knowledge as the *goal belief* of an agent. The second source of uncertainty, which goes beyond the foodtruck domain of Baker et al., is that agents may or may not know how to navigate the maze in order to reach a goal location. We will refer to this possibly uncertain knowledge as the agent's *world belief.*

The full model: Taking these two sources of knowledge into account, the full BToM model a modeling agent could construct for observing a behavior a_{t+1} at the next timestep $t + 1$ given the previously observed behavior $\boldsymbol{a}_t = a_1, .., a_t$ is given by:

$$P(a_{t+1}|\boldsymbol{a}_t) = \sum_{\substack{g \in G \\ b_g \in B_g \\ b_w \in B_w}} P(a_{t+1}|g, b_g, b_w, \boldsymbol{a}_t)P(b_w|\boldsymbol{a}_t)P(b_g|\boldsymbol{a}_t)P(g|\boldsymbol{a}_t) \tag{2}$$

where G, B_g and B_w represent the potential goals, the goal beliefs, and the world beliefs, respectively. This factorization assumes that the beliefs b_g and b_w are independent of each other[1]. Furthermore, we are assuming that our environment fulfills the Markov property in that future actions are independent of previous actions given an agent's goal intention g as well as its beliefs b_g and b_w which simplifies the likelihood to $P(a_{t+1}|g, b_g, b_w)$. This assumption is not always valid, but is an often used simplification which we will also employ in this example.

For the likelihood, we assume rational agents that usually pick the optimal actions given their goal and beliefs, but have a small probability of performing suboptimal actions (cf. [1]):

$$P(a_{t+1}|g, b_g^*, b_w^*) = \frac{exp(\beta U(a_{t+1}, b_g^*, b_w^*, g))}{\sum_{a_i \in A} exp(\beta U(a_i, b_g^*, b_w^*, g))} \tag{3}$$

where $U(a_i, b_g^*, b_w^*, g)$ represents the expected utility of the agent after executing action a_i with regard to goal g and the beliefs b_g^* and b_w^*. This dependents on the agent's position within the maze and may be learned through reinforcement learning or computed on the fly if the environment is known. The expected utility greatly depends on the scenario and the chosen reward function. Following [3] we model equal costs for all available actions, which results in utilities correlating to the (estimated) remaining distance towards the goal instead of potentially different rewards for different desires. In this example we concentrate on agents having exactly one desire, albeit unknown to the observer. β controls the optimality of the agent's actions. As β increases, the agent becomes more rational, as the probability of choosing suboptimal actions decreases.

The simplest model: To derive a highly simplified model from this full BToM, we make discrete assumptions about both our belief variables, i.e. we are assuming true goal beliefs and world beliefs. This simplifies equation 2 to:

$$P(a_{t+1}|\boldsymbol{a}_t) = \sum_{g \in G} P(a_{t+1}|g, b_g^*, b_w^*)P(g|\boldsymbol{a}_t) \tag{4}$$

since both belief distributions have their probability mass concentrated only on a single outcome b_g^* and b_w^*.

Intermediate models: An alternative, less confining assumption is either over the goal beliefs b_g or the world beliefs b_w, resulting in more complex but also more powerful models. The model for only assuming a true world belief would look like this:

$$P(a_{t+1}|\boldsymbol{a}_t) = \sum_{\substack{g \in G \\ b_g \in B_g}} P(a_{t+1}|g, b_g, b_w^*)P(b_g|\boldsymbol{a}_t)P(g|\boldsymbol{a}_t) \tag{5}$$

Basic reasoning works equally in all models. Inference about different mental states can be performed by rearranging the likelihood using Bayes' rule as already shown in the general BToM model in equation 1. Belief updates can be performed by using the posterior after one observation as prior for the next observation, or alternatively through more complex mechanism such as the BeliefUpdate in [4] which takes additional information into account.

Note that more complex models are more expressive as they can explain behavior in additional ways. However, in the Bayesian framework, more alternatives which are equally likely result in lower probabilities overall.

3.2 Measures of Surprise

Computing surprise values for events can be done in many different ways. Several measures were based on the probability of the observed event (cf. [18]), while others have taken the difference between a prior and a posterior distribution (after observing some event) of a current hypothesis as measure of surprise [12]. The latter Bayesian surprise is not applicable in our scenario as we do not have a single clear "model posterior". Instead, we are considering the following two methods:

S1: The surprise due to an action is the KL divergence between the probability our model predicts for the next action and the probability of actually observing that action (omitting the conditioning on the past for brevity):

$$S_1(a) = KL(Q(a)||P(a)) = \sum Q(a) \log(\frac{Q(a)}{P(a)}) \tag{6}$$

[1]In general, other factorizations with potential dependencies are also possible.

where $Q(a)$ is the probability of having observed action a and $P(a)$ is the probability our model predicted action a occurring. Assuming a perfect observer, i.e. $Q(a) = 1$ only for the behavior that actually occurred, this reduces to:

$$S_1(a) = -log(P(a)) \tag{7}$$

That is, the surprisal of an action is equivalent to the negative log of its predicted probability, a measure also known as the self-information of an action in information theory. This value can also be understood as a measure for the prediction error (the more certain the model is in predicting the action observed, the lower the surprise). This form was also used as in [18]. While an interpretation as error measure is appealing, this surprise measure is not intuitive in situations in which multiple actions are equally likely to achieve an hypothesized goal. In this case, the probability for the observed action is lowered since the probability mass is split upon all equally likely possibilities. Lower probabilities will result in larger surprise values, even though the observed action might have been perfectly reasonable. This motivates the second method:

S2: The surprise due to an action is the log of the difference between the probability of the most likely action and the probability of the observed action:

$$S_2(a) = log(1 + P_{max} - P(a)) \tag{8}$$

where P_{max} represents the probability for the most likely action in that situation. This measure was also used to explain human empirical data [18]. Further, it does not have the problem mentioned above and results in lower surprise values overall. However, this measure requires the computation of the maximum probability of an action in that situation. This computation requires evaluating all possible actions, which can quickly become infeasible.

We can easily extend both surprise measurements to entire sequences of actions by stating that the surprise for a sequence of actions is the sum of surprises of each action:

$$S_1(a_1, .., a_t) = \sum_{i=1}^{t} -log(P(a_i)) \tag{9}$$

$$S_2(a_1, .., a_t) = \sum_{i=1}^{t} log(1 + P_{max_i} - P(a_i)) \tag{10}$$

While the S_1 for sequences directly follows from logarithmic identities, we define S_2 for sequences this way mainly for convenience. Computing P_{max} for the entire sequence would be infeasible in all but the simplest cases.

As both measures have their advantages and disadvantages, in the evaluation (see next section) we applied both measurements and did not find any substantial differences. As S1 provides the better computational performance, we will thus report only results obtained using S1 in the following.

3.3 Model Switching

Given a number of alternative, differently complex models and a measure of surprise to assess the explanatory quality of a given model, it is now possible to realize the ability for autonomously switching models. Our current, first approach is to continuously check whether a surprise threshold γ is exceeded by the current model and, if so, to switch to the model that yields the lowest surprise. After switching, the threshold is increased by 50% to prevent the model from constantly switching back and forth when explaining difficult behavior.

This is obviously a very simple strategy that will run into problems when the number of alternative models (sets of assumptions) to evaluate increases. As shown in the next section, in the present example with only three distinct sets of assumptions, evaluating all models is still a lot more efficient than evaluating the full complex model. Yet, more sophisticated ways of switching models are obviously possible. For example, switching could follow a hierarchy of increasingly more complex models, ideally inspired by findings from human psychology regarding the ToM scales [32]. Actually learning and adapting such a hierarchy of suitable assumptions from past observations is subject of our own ongoing work.

4 SCENARIO

To test how well the different models along with their corresponding discrete assumptions are suited to explain real behavior, we collected behavioral data of humans navigating within a simple environment. We created six different mazes of equal size (20 by 13 blocks). A block in the world was either a free space, a wall or a goal position. Goals were of different color which was however unknown unless an agent created a line of sight with the goal. Otherwise, each goal looked identical. The agent, represented as a smiley, could move over free blocks as well as goal positions but not through walls. Such a simplified environment was used in related studies [2, 13, 28, 34] for it can be realistically tackled using BToM models, while still eliciting sufficiently complex intentional behavior and corresponding mental states in the participants.

Our main goal is to use this dataset to compare the simplified models, the full BToM model, and the combined model that rests on the switching strategy described above. Crucially, we wanted to get realistic data of behavior produced by humans with different kinds of uncertainties about their task environment. To that end we designed three different task conditions that induce different kinds and degrees of uncertainties, and that would correspond to different discrete assumptions within our modeling framework.

4.1 Conditions of Uncertainty

No uncertainty: In this first condition, we revealed the entire maze and showed the participants a single goal location, which they were told to reach. As a results, participants should have had full knowledge about the goal position (true goal belief) as well as any paths leading to it (true world beliefs) without any uncertainties regarding their task (see left maze of Figure 1 for an example).

The corresponding model for this condition is the simplest and also the one our observer will start with when considering the switching model. It is summarized by equation 4 above.

Goal uncertainty: In the second conditions, participants only had a true world belief. Uncertainty in the goal beliefs was introduced by showing the participants four goals, which all looked the same until they created a line of sight with them. The participant's instruction read "Find your way to the X exit" where X was replaced with the color of the goal they were supposed to reach. Since each participant saw each individual maze only once they had no way of knowing the position of their desired goal beforehand. An example of a situation in this condition can be seen in the middle of Figure 1: The agent currently has a line of sight to the blue target (B), it can therefore rule out all possible goal locations where the blue target is not at the seen location.

This condition corresponds to the intermediate model already presented in equation 5. We update the belief about potential goal locations similar to [1]:

$$P(b_g|\boldsymbol{a}_{t+1}) \propto P(o_t|s_t, b_g)P(b_g|\boldsymbol{a}_t) \quad (11)$$

where $P(o_t|s_t, b_g)$ represents the probability of actually observing what the agent sees at time t given his current belief in the goal positions b_g and the agent's current state (i.e. position within the maze) s_t. This is modeled by checking if the believed goal position is visible from the agent's current position and if it matches the assumed goal. We assign a probability of θ to scenarios that match reality and $1 - \theta$ to scenarios that do not match, e.g. seeing a red goal when one is assuming a blue one. θ represents the probability of making correct observations and is set to 1 for the results reported below. However, allowing a slight observation error does not change the results significantly. Lastly, $P(b_g|\boldsymbol{a}_t)$ is our previous belief about goal positions.

World uncertainty: In the last condition, participants hold a true belief about goal positions but had uncertainty regarding the world structure. This was achieved by only showing a single goal and instructing the participants to "Find your way to the shown exit". Therefore, there was no ambiguity of where they would need to go in order to reach their goal. However, as can be seen on the right of Figure 1, participants were only able to see up to three blocks of the maze around them. Again, since they never saw the same maze twice, they had no way of knowing their way to the goal location.

This, again, corresponds to an intermediate model as in equation 5 but with the fixed belief swapped around:

$$P(a_{t+1}|\boldsymbol{a}_t) = \sum_{\substack{g \in G \\ b_w \in B_w}} P(a_{t+1}|g, b_g^*, b_w)P(b_w|\boldsymbol{a}_t)P(g|\boldsymbol{a}_t) \quad (12)$$

Since we are assuming that agents have no knowledge about the mazes' structure, we cannot compute the true remaining distance to the goal positions for the utilities in this case. Theoretically, one would need to represent all possible worlds in B_w, computing the remaining distance to the goal positions for each of these potential worlds. Already with the very simple 2D environment of size 13 by 20 there are $2^{13*20} = 1.85\times10^{78}$ different wall configurations, many of which would not even contain traversable paths towards the goal. Even if one would a priori select only worlds with reachable goals, the number of possible goals would still be intractable and it does not seem likely that humans are considering multiple potential wall placements in an unknown maze. Therefore, we are employing the "free space assumption" that any field which has not been seen yet is traversable. Under this free space assumption, we can compute the distance to the goal with standard search algorithms like A* and use this as the expected utility.

The actual belief distribution simplifies greatly under this assumption:

$$P(b_w) \propto \begin{cases} 0 & \text{, if } b_w \text{ contains walls in unknown areas} \\ \frac{1}{|w|} * f & \text{, if } b_w \text{ contains false beliefs about already seen areas} \\ 1 & \text{, if } b_w \text{ corresponds to the seen environment with no walls in the unknown area.} \end{cases} \quad (13)$$

where $|w|$ represents the number of mistakes in b_w and f represents the probability of forgetting past observations. Assuming further that agents have perfect memory (i.e. setting $f = 0$) allows one to reduce the domain of world belief B_w to only a single entity in each situation, thus greatly simplifying the resulting inference. Since both the specialized and the complex models use this simplification, the comparison between these models remains unaffected.

In this work we are not directly interested in making inferences about the actual world structure, since we do not believe that humans are likely to make such inferences either. Instead we are interested in inferring whether or not agents hold a true and complete belief over the world structure or not, in which case we are making the free space assumption. The full model therefore treats B_w as a binary variable deciding between true world belief and free space assumption. This belief is updated similar to the belief about goal desires, i.e. by inverting the likelihood using Bayes' rule.

4.2 Procedure

We had 122 participants for our web-based study from Crowdflower who received a €0.2 upon completion. We did not restrict the country of origin of our participants. Of these 122 participants, 110 completed all six mazes while the remaining 12 stopped after 1 to 5 mazes. Overall, we collected data from 687 completed trails. The raw data as well as the scripts used to produce the results reported here can be found in [21]. Participants first saw a short text telling them that they will be asked to complete a series of different scenarios. Upon acknowledgement, the participants saw their first maze with instructions about which goal to reach on top of the maze as an overlay, which stayed until the participant performed an action (which was also instructed). Participants controlled the agent using either the keyboard or labeled buttons on the sides of the maze. The instructions were additionally shown below the maze throughout the trail along with the control instructions. Each action was logged by our server. Upon completion of the task, the next maze was presented in the same way. After completing the sixth and last maze, participants were given a code that they could enter in the Crowdflower study for being compensated. The conditions were randomly combined with either of two variants of the six mazes, only ensuring that each condition appeared twice for each participant and that each participant was presented with six different mazes. The order in which the mazes where presented to the participants was also randomized.

5 EVALUATION

As mentioned before, in this work we are less interested in how well any particular model predicts human behavior. Previous research has shown that the BToM framework is suitable to explain an agent's behavior in such navigation tasks. In fact, our simplest

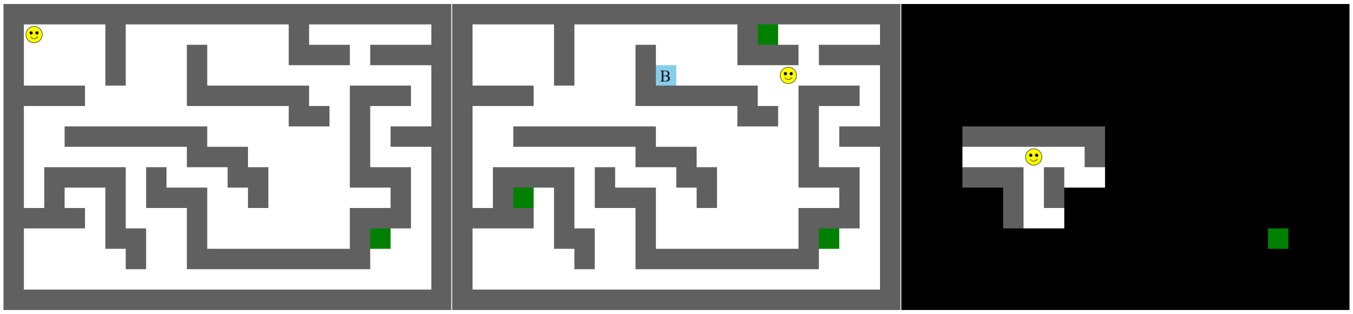

Figure 1: The three different conditions on the same map as seen by the participants: Left - *No Uncertainty*, Middle - *Goal Uncertainty*, Right -*World Uncertainty*. The yellow smiley represents the agent controlled by participants, the green squares represent possible goal locations. As can be seen in the middle, an actual goal color is only revealed when an agent has a line of sight towards the goal position. Multiple goal locations were only shown in the *True World Belief* condition.

model and the *No Uncertainty* condition are very similar to previous research [3]. Here, we are instead interested in comparing the three simpler models (*True World Belief, True Goal Belief, True World and Goal Belief*), the *Switching* model, and the full model making *No Assumption* (eq. 2)[2]. For the switching model we computed the initial threshold γ_0 based on the average surprise value on part of the data from the *No Uncertainty* condition. For the results reported below, we rounded the obtained threshold to 20 (although the results are stable across a range of different values). Likewise, the optimality parameter $\beta = 1.5$ was determined by testing the fit of all models on a number of samples from each condition. We will first consider objective statistics across all the recorded data, before discussing some concrete examples of the behavior of the different models.

5.1 Accuracy

To evaluate how well each of the five models can explain the human behavior, we use the negative log likelihood of the behavior produced by the models, which in this case corresponds to our S1 surprise measure for entire trajectories (see Table 1). The average is taken over all trials from the respective condition, or all trials overall respectively. As can be seen, the *Switching* model outperforms any single model in each of the different conditions. The differences are significant ($p<0.005$ according to Wilcoxon signed-rank test) for all other models and conditions except for the *Goal Uncertainty* condition where the test yields a p-value of 0.0058[3]. The full model, which makes none of the simplifying assumptions of the other models, performs better than the specialized models overall. In the *Goal Uncertainty* condition it achieves similar performance (no significant difference according to the Wilcoxon signed-rank test) to the *True World Belief* model. This is due to the fact that both models make the exact same inferences (the Kullback-Leibler divergence between the goal belief distributions taken after every action is effectively 0). In the *No Uncertainty* and the *World Uncertainty* conditions, the models making the corresponding assumptions achieve better results than the full model. Note that the large standard deviations in all but the *No Uncertainty* condition are due to large behavioral variations in the data.

5.2 Efficiency

Since we are interested in satisficing mentalizing, accuracy is only one of two important criteria. The other one, efficiency, refers to the fact that mental inferences need to be performed relatively quickly, especially in social interactive agents that have to act towards other agents in a timely manner. Table 2 reports the relative time needed by a model to evaluate a single action in the different conditions, normalized to the simplest and quickest model (*True World and Goal Belief*)[4]. Note, that the *Switching* model needs to perform occasional re-evaluations of all models, which is why we also report the average number of re-evaluations per trial.

5.3 Behavior Processing Examples

Due to the great variability of the human data across differently complex mazes, aggregated statistics can only convey so much information. To illustrate how the different models process human behavior, we analyze two actual trials from the *Goal Uncertainty* and the *World Uncertainty* conditions and present them in Figures 2 and 3, respectively. The red trajectory in the maze represents the actions participants performed, and the green squares represent the potential goal locations. In the *World Uncertainty* condition, participants only saw the goal, marked with a T. The modeling agent, however, has to consider any of the four locations as potential intended goals of the agent. The bottom part of the Figure shows the resulting surprise values computed by the different models against the action number. The vertical lines indicate when the current assumptions in the *Switching* model exceeds the surprise threshold and a re-evaluation resulting in a potential switch was done.

In the *Goal Uncertainty* example (Fig. 2), the participant searched each of the four possible goal locations before finding her desired goal. The *True World Belief* model and the full model behave similarly, with the complex model being better able to explain the

[2]This does not mean that there are no assumptions overall, cf. the likelihood function.
[3]For complete statistics as well as access to the raw data see [21]

[4]We report relative timings in order to highlight the models' differences, not tied to any experimental system. For completeness, the average times of the quickest model in seconds on a 3.5Ghz Xeon machine for *Overall, No Uncertainty, Goal Uncertainty* and *World Uncertainty*, respectively, are: 7.63e-5, 5.57e-5, 8.15e-5, 9.65e-5

Table 1: Average surprise values and their standard deviations for the different models (rows) applied to the three conditions (columns); numbers in bold represent the lowest values (best). * signals significant difference to the *No Assumption* model (p<0.005) according to Wilcoxxon signed-rank test. A more complete analysis can be found in [21].

Model	Overall	No Uncert.	Goal Uncert.	World Uncert.
No Assumption	43.6975 (41.98)	29.1360 (11.04)	48.8868 (38.29)	53.1374 (58.02)
True World and Goal Belief	47.8146 (59.76)	13.8305 (12.54)	75.3903 (56.71)	54.5828 (73.36)
True World Belief	49.0316 (44.56)	32.9640 (13.26)	50.8797 (39.45)	63.2752 (61.28)
True Goal Belief	45.6111 (51.67)	22.8657 (15.75)	72.0591 (47.27)	42.2535 (65.58)
Switching	**34.1854** (46.29) *	**13.0823** (10.38) *	**48.7999** (40.96)	**40.8646** (62.75) *

Table 2: Relative time needed by a model to evaluate an action, normalized to the most efficient model (second row). In the bottom row, the average number of model switches per trial is given in parentheses.

Model	Overall	No Uncert.	Goal Uncert.	World Uncert.
No Assumption	553.17	758.09	519.25	437.12
True World and Goal Belief	1	1	1	1
True World Belief	5.19	6.78	4.94	3.88
True Goal Belief	24.25	35.11	23.38	18.94
Switching	26.41 (1.42)	9.47 (0.29)	32.19 (2.41)	23.86 (1.57)

observed behavior. The *True Goal and True World Belief* model as well as the *True Goal Belief* model start with lower surprise values. The first jump in surprisal with the 19th action is due to the agent initially making to moves upwards at the intersection before turning around and going to the lower left goal, When turning around in the 34th action after seeing the lower left goal, the surprise for the *True Goal and World Belief* model and the *True Goal Belief* model increases quickly: an agent knowing where its goal is would not inspect a goal and turn away from it after seeing it. However, the *Switching* model does not switch models since their surprisal is still lower than the *True World Belief*'s until the 64th action.

In the *World Uncertainty* example (Fig. 3), the participant had limited sight but first moved directly in the direction of the goal before discovering the wall. After exploring the lower side for a path to the other side of the wall, the participant turned around and moved around the top side of the maze before reaching her goal. In this case, all models start off very similarly, mainly because the first actions are equally likely for all goals. In fact the first 17 actions are optimal for reaching the true goal location. However, starting with the 18th action, the simplest model can no longer explain the behavior well enough, resulting in a switch after the 19th action to the *True Goal Belief* model.

Thus, after a few deviations from the optimal behavior in both examples, the *Switching* model detects that the simplest model cannot explain the behavior and switches to a better model. In the *Goal Uncertainty* condition, the model first switches to the *True Goal Belief* model and only later switches to the best model.

6 DISCUSSION

Our evaluation study yielded several interesting insights. First, simpler models can outperform the complex full model in certain situations with regard to both accuracy and efficiency. This is not surprising as the simpler models were designed to correspond to a specific condition – an approach commonly taken when employing the BToM framework [2, 14]. A complex model, in contrast, may

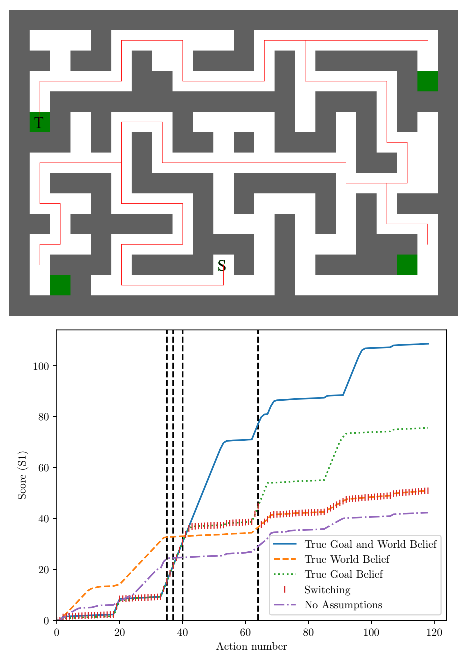

Figure 2: Top: Trajectory of a participant in the *Goal Uncertainty* condition (start position marked with an S, the actual target indicated with T). Bottom: Resulting surprise values calculated by the different models (vertical lines indicate re-evaluations in the *Switching* model).

consider too many possibilities instead of focusing on the important parts, a result in line with the findings in [10, 23].

Secondly, our results stress that while simpler BToM models can perform well in certain scenarios, they will often perform much

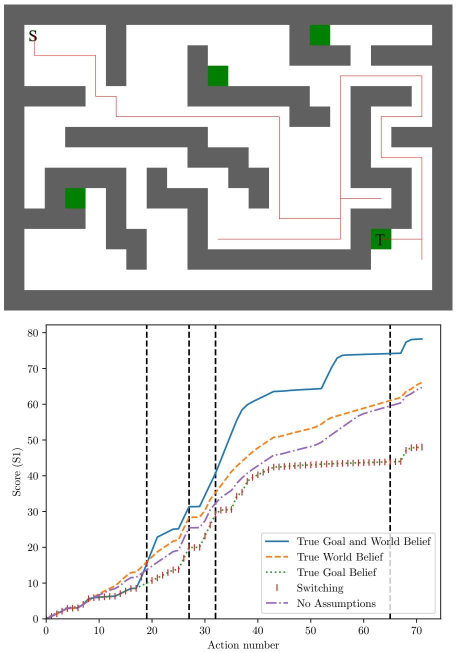

Figure 3: Top: Trajectory of a participant in the *World Uncertainty* condition (start position marked with an S, true target indicated with T). The participant had a limited field of view as shown in Fig. 1 (right) and only got to see the true goal. Bottom: Resulting surprise values calculated by the different models (vertical lines indicate re-evaluations).

worse in different scenarios. This speaks to the necessity of switching models in a way that guarantees that currently observed behavior is explained sufficiently. Interestingly, our straight-forward approach to model-switching shows both the performance of specialized models – in fact, it even outperforms the specialized models in accuracy under all conditions – as well as the flexibility required to adapt to qualitatively different mental states of the modeled agents. The fact that the *Switching* model achieves a better performance than the single specialized models points to the variability of human behavior in this experiment. A number of participants managed to perform objectively optimal actions even in the conditions with uncertainty, either by luck or due to the constraints of the mazes. Such cases of successful behavior under incomplete knowledge are arguably not unusual and the combined model will switch to the assumptions that are objectively wrong but correspond to the observed behavior. Likewise, the relatively large standard deviations can be explained as follows: On the one hand, we aggregated the data from six mazes of different complexity, due to space constraints in this paper. Some of the mazes allowed for many equally optimal actions, while others were a lot more constrained (just compare the mazes in Figures 2 and 3). Therefore, the average surprise value can vary greatly between the different mazes. On the other hand, the dataset comprises a number of "outliers", participants which performed action sequences that none of the models assuming rational actions could explain.

Thirdly, the simple models, albeit varying in computational complexity, are orders of magnitude more efficient than the full model. Crucially, even with its quite naive strategy of evaluating all simple models in case the surprisal exceeds a current threshold, the *Switching* model is still somewhat comparable to the *True Goal Belief* model. This is because the additional re-evaluations are only performed a few times per trail. Indeed, one could easily add another ten models of similar complexity to the *True Goal Belief* model, and would still outperform the full model.

In sum, our approach to construct a limited set of simpler models that correspond to relevant but qualitatively different mental states, and to combine them by way of a rather simple switching strategy fulfills the requirements for a satisficing model of ToM. The *Switching* model achieves the best accuracy and is efficient enough to evaluate the behavior online, i.e. while the participants are navigating the maze which take on average ~0.59s per action. Improving upon this naive switching strategy, e.g., by learning potential assumption hierarchies from past observations may be possible and is something we are currently working on.

Finally, we note that these results were obtained with specific surprisal measures. As already mentioned above, the S1 score reported here is not very intuitive in cases when different possible behaviors are observed. However, the S1 score corresponds to the log likelihood of the respective models given the observed trajectory, which is sufficient for assessing the performance of these models. S2 results in the same relative ranking of the models and yields more understandable absolute values, but at higher computational costs. In future work we will consider further alternative scores that might be useful for evaluating and switching models.

7 CONCLUSION

In this paper, we have considered the question what kind of Bayesian models of Theory of mind can be satisficing. We have compared the performance of simpler, specialized BToM models with a more complex general model on behavioral data from humans subjected to different uncertainties in a navigation task. The results show that simpler models are not only computationally more efficient, but can even outperform the full model in the situations that they were designed for. However, in other situations, the simpler models usually fail to perform well, which makes it hard to apply them to a wider variety of tasks. In order to avoid having to evaluate a complex model, which can be computationally intractable, we proposed the use of a surprisal measure in order to decide when to switch from one specialized model, mirroring certain assumptions about the observed agents, to another. We have shown that such a switching model can, on average, explain behavior even better than any simple or the complex model, without the computational cost of the full complex model. This suggests that switching between simplifying assumptions can yield satisficing models of Theory of Mind, taking advantage of many of the shortcuts humans usually employ when reasoning about the behavior of others.

ACKNOWLEDGMENTS

This research was supported by the Cluster of Excellence Cognitive Interaction Technology "CITEC" (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG).

REFERENCES

[1] Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. 2011. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the Cognitive Science Society*, Vol. 33.

[2] Chris L. Baker, Julian Jara-Ettinger, Rebecca Saxe, and Joshua B. Tenenbaum. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour* 1 (2017), 0064.

[3] Chris L. Baker, Rebecca R. Saxe, and Joshua B. Tenenbaum. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349. https://doi.org/10.1016/j.cognition.2009.07.005

[4] Chris L. Baker and Joshua B. Tenenbaum. 2014. Modeling human plan recognition using Bayesian theory of mind. *Plan, activity, and intent recognition: Theory and practice* (2014), 177–204.

[5] Hendrik Buschmeier and Stefan Kopp. 2014. A dynamic minimal model of the listener for feedback-based dialogue coordination. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue (SemDial)*. Edinburgh, UK, 17–25.

[6] Stephen A. Butterfill and Ian A. Apperly. 2013. How to construct a minimal theory of mind. *Mind & Language* 28, 5 (2013), 606–637.

[7] Eugene Charniak and Robert P. Goldman. 1993. A Bayesian model of plan recognition. *Artificial Intelligence* 64, 1 (1993), 53–79.

[8] Maartje de Graaf and Bertram Malle. 2017. How people explain action (and AIS should too). In *Proceedings of the Artificial Intelligence for Human-Robot Interaction (AI-for-HRI) fall symposium.*

[9] Harmen De Weerd, Rineke Verbrugge, and Bart Verheij. 2013. How much does it help to know what she knows you know? An agent-based simulation study. *Artificial Intelligence* 199 (2013), 67–92.

[10] Harmen de Weerd, Rineke Verbrugge, and Bart Verheij. 2015. Higher-order theory of mind in the tacit communication game. *Biologically Inspired Cognitive Architectures* 11 (2015), 10–21.

[11] Andreea O. Diaconescu, Christoph Mathys, Lilian A E Weber, Jean Daunizeau, Lars Kasper, Ekaterina I. Lomakina, Ernst Fehr, and Klaas E. Stephan. 2014. Inferring on the Intentions of Others by Hierarchical Bayesian Learning. *PLoS Computational Biology* 10, 9 (2014). https://doi.org/10.1371/journal.pcbi.1003810

[12] Laurent Itti and Pierre Baldi. 2009. Bayesian surprise attracts human attention. *Vision research* 49, 10 (2009), 1295–1306.

[13] Julian Jara-Ettinger, Chris L. Baker, and Joshua B. Tenenbaum. 2012. Learning what is where from social observations. In *Proceedings of the Cognitive Science Society*, Vol. 34.

[14] Alan Jern, Christopher Lucas, and Charles Kemp. 2017. People learn other people's preferences through inverse decision-making. (2017).

[15] Gal A. Kaminka, David V. Pynadath, and Milind Tambe. 2002. Monitoring teams by overhearing: A multi-agent plan-recognition approach. *Journal of Artificial Intelligence Research* 17, 1 (2002), 83–135.

[16] Boaz Keysar, Shuhong Lin, and Dale J Barr. 2003. Limits on theory of mind use in adults. *Cognition* 89, 1 (2003), 25–41.

[17] Jennifer Klatt, Stacy Marsella, and Nicole Krämer. 2011. Negotiations in the context of AIDS prevention: an agent-based model using theory of mind. In *Intelligent virtual agents*. Springer, 209–215.

[18] Luís Macedo, Rainer Reisezein, and Amilcar Cardoso. 2004. Modeling forms of surprise in artificial agents: empirical and theoretical study of surprise functions. In *Proceedings of the Cognitive Science Society*, Vol. 26.

[19] Wulf-Uwe Meyer, Rainer Reisenzein, and Achim Schützwohl. 1997. Toward a process analysis of emotions: The case of surprise. *Motivation and Emotion* 21, 3 (1997), 251–274.

[20] Andrew E. Monroe and Bertram F. Malle. 2017. Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *Journal of Experimental Psychology: General* 146, 1 (2017), 123.

[21] Jan Pöppel. 2018. Data and Analysis for Satisficing Models of Bayesian Theory of Mind for Explaining Behavior of Differently Uncertain Agents. (2018). https://doi.org/10.4119/unibi/2918529

[22] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behavioral and brain sciences* 1, 4 (1978), 515–526.

[23] David V. Pynadath and Stacy Marsella. 2007. Minimal mental models. In *AAAI*. 1038–1044.

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

[25] Charles F. Schmidt, NS Sridharan, and John L. Goodson. 1978. The plan recognition problem: An intersection of psychology and artificial intelligence. *Artificial Intelligence* 11, 1-2 (1978), 45–83.

[26] Thorsten Schodde, Kirsten Bergmann, and Stefan Kopp. 2017. Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. *Proceedings of ACM/IEEE HRI 2017* (2017).

[27] Mei Si and Stacy C Marsella. 2014. Encoding theory of mind in character design for pedagogical interactive narrative. *Advances in Human-Computer Interaction* 2014 (2014), 10.

[28] Yvonne Teoh, Emma Wallis, Ian D. Stephen, and Peter Mitchell. 2017. Seeing the world through othersâĂŹ minds: Inferring social context from behaviour. *Cognition* 159 (2017), 48–60.

[29] D. Ullman, Tomer, Chris L. Baker, Owen Macindoe, Noah D. Goodman, and Joshua B. Tenenbaum. 2009. Help or Hinder: Bayesian Models of Social Goal Inference. *Nips* (2009), 1–9. https://papers.nips.cc/paper/3747-help-or-hinder-bayesian-models-of-social-goal-inference.pdf

[30] Joey Velez-Ginorio, Max Siegel, Joshua B. Tenenbaum, and Julian Jara-Ettinger. 2017. Interpreting actions by attributing compositional desires. (2017).

[31] Edward Vul, Noah D. Goodman, Thomas L. Griffiths, and Joshua B. Tenenbaum. 2014. One and done? Optimal decisions from very few samples. *Cognitive science* 38, 4 (2014), 599–637.

[32] Henry M Wellman and David Liu. 2004. Scaling of theory-of-mind tasks. *Child development* 75, 2 (2004), 523–541.

[33] Evan Westra and Peter Carruthers. 2017. Pragmatic development explains the Theory-of-Mind Scale. *Cognition* 158 (2017), 165–176.

[34] Yang Wu, Chris L. Baker, Joshua B. Tenenbaum, and Laura E. Schulz. 2017. Rational inference of beliefs and desires from emotional expressions. *Cognitive Science* (2017).