Goal Recognition with Timing Information

Chenyuan Zhang,[1] Charles Kemp, [1] Nir Lipovetzky [1]

[1] The University of Melbourne

chenyuanz@student.unimelb.edu.au, c.kemp@unimelb.edu.au, nirlipo@gmail.com

Abstract

Goal recognition has been extensively studied by AI researchers, but most algorithms take only observed actions as input. Here we argue that the time taken to carry out these actions provides an additional signal that supports goal recognition. We present a behavioral experiment confirming that people use timing information in this way, and develop and evaluate a goal recognition algorithm that is sensitive to both actions and timing information. Our results suggest that existing goal recognition algorithms can be improved by incorporating a model of planning time on both synthetic data and human data, and that these improvements can be substantial in scenarios in which relatively few actions have been observed.

Introduction

Imagine that you and a friend are playing a strategic boardgame. Your friend just made a move that is compatible with only two possible goals: a low-reward goal that is easy to reach or a high-reward goal that is difficult to reach. If your friend made her move very rapidly, you might infer that she is aiming for the easily-reached goal, but if she thought for a long time you might conclude that she is aiming for the more ambitious goal.

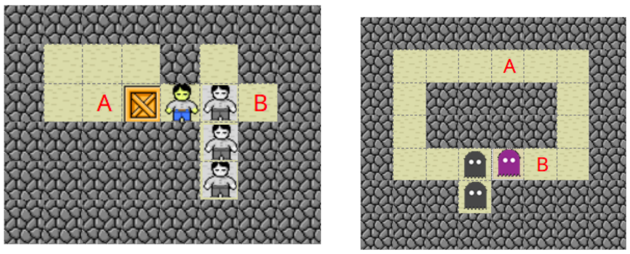
As this example suggests, people's inferences about the intentions of others are sensitive to information that goes beyond observed actions alone (Singh et al. 2018; Gates et al. 2021). Real-world interactions are embedded in time and timing information is almost always available. Current goal recognition algorithms, however, mostly focus on actions only and rarely take auxiliary information such as timing into consideration (Zhi-Xuan et al. 2020; Ramírez and Geffner 2010; Keren, Gal, and Karpas 2015; Pereira, Oren, and Meneguzzi 2017; Masters and Sardina 2019). In this paper we propose a new goal recognition framework that can exploit observed planning times, and evaluate it using both synthetic and human data.

The problem of goal recognition is the task of inferring an actor's real goal given a sequence of observations and a set of possible goals. Early approaches to this problem often used a plan library to perform goal inference and matched the sequence of observations with a library of historical observations associated with each goal candidate (Blaylock, Allen et al. 2003; Vered, Kaminka, and Biham 2016). Later, Ramirez and Geffner proposed a generative approach that uses planning algorithms over planning models and is known as plan recognition as planning (PRP) (Ramírez and Geffner 2010; Sohrabi, Riabov, and Udrea 2016).

A small amount of work in AI and cognitive science has explored how auxiliary information can be used to infer the mental states of others. Singh et al. (2018) used gaze information for intention recognition and found that gaze can help to reveal the hidden goals of players in a boardgame. Gates et al. (2021) developed a Bayesian model that aims to capture how people use response times when inferring the preferences of an actor who is observed to make a single decision. Our work generalizes the same underlying idea by exploring how timing information can be used in situations where actors generate rich sequences of actions, not just one-shot decisions. Perhaps closest to our own approach is the work of Avrahami-Zilberbrand, Kaminka, and Zarosim (2005), who developed a plan-recognition algorithm that incorporates constraints on action durations. Our work also highlights the role of time but focuses specifically on planning times that reflect the effort exerted by the actor when selecting actions.

Figure 1 illustrates two cases in which planning times are useful for goal recognition. In the Sokoban example (Figure 1a), the current position of the worker is shown in color and the grey workers show the trajectory the worker followed to reach this position. The actor is a real-time planner that performs a look-ahead search using Manhattan distance as a heuristic, and because the computational resources of the actor are limited it is not guaranteed to choose the optimal trajectory. Given the information in Figure 1a, goals A and B may seem equally likely because the observed trajectory is consistent with optimal paths to both goals. But if we observe in addition that the actor spent a relatively long time at the position shown, B now seems the more likely goal because A is easily achieved with a single push to the left, whereas the actor has to push the box away and then back to achieve B.

In Figure 1a, timing information breaks a tie between two goals that seem equally likely based on actions alone, but there may also be cases where timing information reverses the conclusion that would follow from actions alone. Figure



(a) (b)

Figure 1: Examples from two domains showing two ways in which timing information can influence goal recognition. (a) Timing information can break a tie between two goals. In this Sokoban example, observing the actor stop and think at the position shown with blue jeans suggests that the actor's goal is to push the box to B rather than A. (b) Timing information can reverse the inference that would follow from actions alone. In this navigation example, a protracted pause at the position shown in purple suggests that the goal may be A rather than B.

1b shows an example based on a navigation task. Here the observed action sequence suggests that B is the likely goal because this sequence is consistent with an optimal path to B but not A. But if we see that the actor spends a long time at the location shown, we might conclude that A is the actual goal because there would be no reason for the actor to stop and think if the goal were B rather than A.

Because timing information has received little attention in the literature, standard goal recognition benchmarks do not include this information. Most existing agent models do not produce useful planning times because they either allocate a constant amount of planning time for each step or do not consider this factor at all (Zhi-Xuan et al. 2020; Masters and Sardina 2019). We, therefore, develop an agent model that is inspired by human behaviour, and use this model to generate human-like timing information along with action sequences for standard goal recognition benchmarks. We also use the same agent model to develop our timing-sensitive goal recognition algorithms.

To preview some of our results, we find that the extent to which timing information helps in goal recognition depends on how closely the agent model assumed matches the agent actually generating the observations. Our agent model draws on the extensive response-time modelling literature in cognitive science (Ratcliff et al. 2016; Tavares, Perona, and Rangel 2017), but is not intended to capture all of the reasoning strategies that humans may use. Instead, we rely on two simple assumptions about human planners: (a) people carry out a forward search to make a decision (i.e. they are not reflex agents), and (b) planning time depends only on the current state and the true goal. Assumption (b) does not assume that people only consider the current state, as humans typically anticipate the consequence of future moves. Instead, the assumption is that the planning time for one move does not depend on the planning times for previous moves.

This paper makes a sequence of four contributions. First, we formally introduce a goal recognition framework that incorporates timing information and a novel goal recognition algorithm that can exploit this information. Second, inspired by the cognitive science literature, we develop a real-time agent model, and use it to generate observations with timing information for standard goal recognition benchmarks. Third, we use existing goal recognition datasets to show that timing information can be helpful when our framework can exploit an accurate model of timing. Finally, we show that the proposed goal recognition algorithm can exploit timing information in sequences generated by humans, and also accounts for human inferences in a behavioral study of goal recognition.

The next section reviews the relevant literature on goal recognition, online planning agents and response time modelling. We then present our agent model, formulate the problem of goal recognition with timing information, and present an algorithm that addresses this problem. Finally, we present a set of synthetic and behavioural experiment results and discuss prospects for future work.

Background

Theory of Mind and Goal Recognition

People's ability to infer the mental states of others is known as *Theory of Mind* (Leslie 1987), which is a classic topic in cognitive science. Many behavioural and neural studies have been done in this area while its computational basis has been extensively explored in the last two decades (Rescorla 2015; Baker et al. 2017). In recent years, industry labs have paid increasing attention to this field (Rabinowitz et al. 2018; Raileanu et al. 2018) because AI systems that interact with humans (e.g. self-driving cars) must be able to figure out the goals and intentions of human users.

In the literature on computational cognitive science, Baker *et. al.* developed a Bayesian model of theory of mind and showed that it makes human-like judgements when inferring people's goals and beliefs (Baker et al. 2017). Jara-Ettinger (2019) further suggested that theory of mind can be formalised as inverse reinforcement learning and involves inferring people's internal model of the world and their reward functions given some observed actions.

The automated planning community has proposed a variety of models for efficiently solving the goal recognition problem (Ramírez and Geffner 2010; Pereira, Oren, and Meneguzzi 2017). It is still unknown whether these models are able to account for human goal-recognition abilities but these models can inspire hypotheses about how people carry out goal recognition. Integrating ideas from cognitive science and automated planning literature is therefore a promising way to develop computational models of human goal recognition.

Online Agent Models and Suboptimal Behavior

When dealing with complex tasks with limited reasoning time, it is often impractical for both human and agent models to find a full plan from the current state to a goal state. Unlike classical planning algorithms, online agent models do not aim to find a full plan but rather focus on choosing which single action should be executed at the current state. A prominent approach used to develop online agent models is Monte Carlo Tree Search (MCTS), which has achieved striking success at playing Go (Chen 2016). MCTS has also been explored as a model of human problem solving (Kuperwajs, Van Opheusden, and Ma 2019; Krusche et al. 2018).

Current algorithms in the field of goal recognition usually assume full rationality, i.e. optimal behaviour for both the agent model and the observer model (Ramírez and Geffner 2010). In contrast, the cognitive literature suggests that people often depart from optimality (Tversky and Kahneman 1989; Baker et al. 2017; Gates et al. 2021). Masters and Sardina (2019) explore how goal-recognition systems can reason about irrational agents, but their approach has not yet been directly connected with research in cognitive science.

Response-time modeling

An extensive literature in psychology treats response times as a sign of underlying cognitive mechanisms. A prominent approach in this area focuses on one-shot decision making, and assumes that the decision-maker continually samples evidence about the available response options until some decision criterion is reached. This "evidence accumulation" framework is widely used to account for both reaction times and choice probabilities (Ratcliff et al. 2016).

There are a variety of evidence accumulation models that make different assumptions, and in recent years psychologists have explored which model gives the best account of behavioural data from perceptual decision-making tasks (Ratcliff et al. 2016; Tavares, Perona, and Rangel 2017). Little work has been done, however, in applying the evidence accumulation task to sequential decision-making problems. Solway and Botvinick (2015) take a step in this direction by showing how an evidence integration mechanism can be combined with a model-based tree search. Their work, however, focuses on simple two-step plans that are significantly simpler than those used in standard AI planning benchmarks. Ho et al. (2020) also consider sequential decision-making problems, and use value iteration to account for human reaction times.

Framework

In this section, we first describe an agent model that aims to produce human-like planning times by incorporating concepts from the evidence accumulation literature. We then propose a formal framework for modelling and solving the problem of goal recognition with timing information.

Agent Model

The goal recognition algorithms proposed later require models of planning times, and the datasets used to evaluate these algorithms must include planning times in addition to actions. Standard AI planning algorithms do not generate human-like response times, and we therefore developed a new agent model inspired by ideas from the evidence accumulation literature (Ratcliff et al. 2016; Tavares, Perona, and Rangel 2017).

Given a problem with a goal hypothesis g and start state s_0, we carry out the tree search described below until the goal state is found or the stop signal is triggered. The search tree starts with the current state s_0, or a subtree with root s_0 from the last planning step if a memory mechanism is included, and the algorithm traverses the tree using the UCB policy (Kocsis and Szepesvári 2006) until the leaf node is reached. If the leaf node is the goal state g, the tree search process stops. Otherwise, the node is expanded by generating all possible successor states except those states visited previously to avoid generating repeated states. Each successor state is initialized with the estimated cost-to-go and values of all ancestor nodes are then updated by averaging the obtained values of all visits passing through the node.

After each iteration (expansion), the stop trigger is executed to check if enough information has been collected to make the decision. The probability of triggering the stop signal is calculated as

$$Prob_{stop} = \frac{n}{n + I(s_0)\gamma \exp(-n/I(s_0))}, \tag{1}$$

where n is the number of iterations so far and γ is a parameter that controls the depth of the trajectories considered. The state importance $I(s)$ is defined as:

$$I(s) = \frac{v_{s,a}}{(1+\beta)v_{s,a'} - v_{s,a}}. \tag{2}$$

Here $v_{s,a}$ and $v_{s,a'}$ denote the cost estimates that result from choosing the best applicable action a and second-best applicable action a' towards a given goal from state s. The denominator of Equation 2 is therefore based on the estimated difference in cost between the top two applicable actions, and a small constant parameter β is included in order to avoid zero denominators when the top two applicable actions have the same costs. When the tree search stops, the agent model returns the number of iterations the planning time for the current state.

The stop probability in Equation 1 captures the idea that the actor will spend more planning time on states that have two or more applicable actions that seem equally good (or nearly so) while acting relatively fast in states with a dominating action. This approach is broadly consistent with the evidence accumulation literature, which suggests that people tend to keep gathering evidence until one option emerges as the winner (Ratcliff et al. 2016; Tavares, Perona, and Rangel 2017). Moreover, similar ideas of state importance have been used to summarise state trajectories over Pacman games (Amir and Amir 2018).

Problem Formulation

We now formalize the problem of Goal Recognition with Timing information (GRT). For simplicity, we assume a fully-observable deterministic environment, but the framework and goal recognition algorithms introduced later can be extended to partial observability and/or probabilistic settings by choosing appropriate cost-to-go estimators.

The planning domain is a planning problem without a goal, which can be defined as follows.

Definition 1 *A planning domain $D = \langle S, s_0, A, f, c\rangle$ consists of a finite set of discrete states S, an initial state $s_0 \in S$, a finite set of actions A, a state transition function $f : S \times A \rightarrow S$ that maps a state-action pair (s, a) into another state s' and a cost function $c : S \times A \rightarrow \mathbb{R}$ which specifies the cost $c(s, a)$ incurred when applying action $a \in A$ on state $s \in S$.*

A planning problem $D[g]$ is instantiated by adding a goal g to the planning domain D. For a goal recognition problem, we have a set of possible goals along with a sequence of observations in a planning domain.

Definition 2 *A goal recognition problem with timing information (GRT) is a tuple $\langle D, G, Prior, O\rangle$, where $D = \langle S, s_0, A, f, c\rangle$ is the planning domain, $G = \{g_1, g_2, ..., g_n\}$ is a set of possible goals for the planning domain, $Prior$ is the prior probability over G, and O is a sequence of observations $\langle a_0, t_0\rangle, ..., \langle a_m, t_m\rangle$, where $a_i \in A$ is an action, and t_i is a non-negative real number denoting the planning time used to select a_i for execution.*

The key difference compared to a classical goal recognition setup is that we include planning times in the observation sequence.

Timing-sensitive Goal Recognition Algorithm

We assume that actions and planning times only depend on the current state and the true goal (Markovian), and that planning time and action are conditionally independent on states and goals. Using a uniform prior, we can decompose the likelihood $P(O|g)$ as :

$$\begin{aligned}
P(O|g) =& P(\langle a_0, t_0\rangle, ..., \langle a_m, t_m\rangle|g) \\
=& \prod_{j=0}^{m} P(t_j|g, \langle a_0, t_0\rangle, ..., \langle a_{j-1}, t_{j-1}\rangle) \\
& P(a_j|g, \langle a_0, t_0\rangle, ..., \langle a_{j-1}, t_{j-1}\rangle, t_j) \\
=& \prod_{j=0}^{m} P(a_j|g, s_j) \prod_{j=0}^{m} P(t_j|g, s_j)
\end{aligned}$$

We call the product $\prod_{j=0}^{m} P(a_j|g, s_j)$ the *action component* and $\prod_{j=0}^{m} P(t_j|g, s_j)$ the *timing component*. The next section explains how we estimate both components, and we then discuss how these components are combined to produce a GRT solution.

Action Component We follow the PRP approach proposed by Ramírez and Geffner (2010) to estimate $\prod_j P(a_j|g, s_j)$. Rather than estimating the probability for each step of the observation sequence, their approach approximates the full sequence directly as $\prod_j P(a_j|g, s_j) \propto \exp(v^*_{s_0}(g) - v^*_{s_0}(g, O))$, where $v^*_{s_0}(g)$ denotes the optimal (thus smallest) cost-to-go from start state s_0 while $v^*_{s_0}(g, O)$ represents the cost of the best path consistent with current observations O. Their approach uses the full observation trajectory and is computationally expensive, as optimal planning is hard unless approximated with suboptimal planners or suitable relaxations (Bylander 1994). Thus, we propose a novel method, namely **real-time PRP** via simulation through the agent model we proposed. Compared to the original PRP, real-time PRP assumes the problem to be Markovian and considers each step independently:

$$\prod_j P(a_j|g, s_j) \propto \prod_j \exp(v_{s_j}(g) - v_{s_j}(g, a_j)),$$

where $v_{s_j}(g)$ denotes the approximation of cost-to-go from the state s_j and $v_{s_j}(g, a)$ denotes the approximation of cost-to-go if action a is taken on s_j. This allows for real-time performance instead of computing a full plan as in PRP.

Timing Component We define expected planning time for state s_j given goal g as $t^*(s_j, g)$, and decision cost (which captures the total effort needed by an actor to choose the move at state s_j when pursuing goal g) as $t(s_j, g)$. In this paper, we assume for simplicity that the decision cost $t(s_j, g)$ is identical to t_j, the time recorded in the observation sequence.

We use $\exp(-|t^*(s_j, g) - t(s_j, g)|) = \exp(-|t^*(s_j, g) - t_j|)$ to estimate $P(t_j|g, s_j)$. We propose two approaches to approximate the expected planning time $t^*(s_j, g)$:

- **Agent-based**. Given goal g, $t^*(s_j, g)$ is considered as the number of iterations to make the decision at state s_j via simulation by the agent model described above.
- **Importance-based**. In this approach, $t^*(s_j, g)$ is estimated directly by the state importance $I(s_j)$ defined in the agent model shown in Equation 2.

Note that $t^*(s_j, g)$ and $t(s_j, g)$ may be measured on different scales. $t(s_j, g)$ is typically measured in seconds, whereas $t^*(s_j, g)$ is generated by the timing component of the model and has iterations or importance as units. To map between these different scales, we normalize $t^*(s_j, g)$ by scaling its sum over s_j to match the sum of $t(s_j, g)$.

Combining Components We use two approaches to combine the action and timing components. The first one adds evidence from the two components and uses the resulting sum to rank the goals. Let $p_t(g)$ be the $\log$ probability of $\prod_j P(t_j|g, s_j)$ and $p_a(g)$ be the $\log$ probability of $\prod_j P(a_j|g, s_j)$ for the potential goal g. Then the combined probability of goal g is $\frac{p_t(g)}{\sum_j p_t(g_j)} + w\frac{p_a(g)}{\sum_j p_a(g_j)}$ where w is an adjustable balance factor.

The second approach uses the evidence from the action component to rank the goals, and relies on the timing component only to break ties. In this approach, the timing component cannot reverse the inference suggested by the action component, and can contribute only when the action component does not provide enough information to infer a single most likely goal.

Synthetic Experiment

This section describes an experiment that uses standard goal-recognition data sets to evaluate whether timing information can improve the performance of goal-recognition algorithms. Existing goal recognition algorithms return the set of most likely goals, and accuracy is typically used as an evaluation metric. Here we use fractional ranking to evaluate the extent to which timing information helps distinguish between equally likely goals. Fractional ranking generates the same mean rank as ordinal ranking but allows for ties. For example, if the likelihoods of 4 potential goals were 0.8, 0.5, 0.5, 0.2, the ordinal ranks would be 1,2,3,4 and the fractional ranks would be 1,2.5,2.5, 4.

For each instance, the performance of an algorithm is measured by the fractional rank assigned by the algorithm to the true goal. The performance on an entire domain is measured by the average performance across all instances of that domain. Given the average fractional rank r_D for an algorithm on each domain D, the overall normalized score for that algorithm is $\sum_D \frac{2r_D}{k_D+1}$, where k_D is the number of potential goals in domain D. Note that $\frac{k_D+1}{2}$ is the expected fractional rank achieved by a random algorithm in the domain D. Overall, lower fractional ranks or normalized scores indicate better performance.

Experiment Configuration

We evaluated goal recognition algorithms on 10 domains from the goal recognition dataset of Pereira, Oren, and Meneguzzi (2017). Because this dataset does not include timing information, we used our agent model to supplement the trajectories with times: for each state s, we ran the agent model (without memory mechanism) given the real goal g and took the average number of iterations over 100 runs as the planning time for that state. Although we record the planning time, we disregard the action chosen by the agent to ensure that the trajectories remain consistent with the original dataset.

We use the satisfying planner DUAL-BFWS (Lipovetzky and Geffner 2017) to approximate the optimal cost-to-go in the goal recognition algorithm PRP (Ramírez and Geffner 2010). For initializing node values in the agent model and computing the importance-based timing component, we use the heuristic function h_{ff} (Hoffmann and Nebel 2001). All experiments were conducted on 4 servers each running Intel®Xeon®Gold 6138 CPU @ 2.00GHz with 4 CPUs, and 8GB of RAM each.

All action costs were set to 1. Constants in the agent model ($\gamma = 10000, \beta = 0.2$) were chosen manually so that the model generated human-like response times in navigation tasks like Figure 1b. Except when mentioned otherwise, the observation ratio is set to 0.25, which means that we use the first quarter of observations in a trajectory as the input to the goal recognition algorithms. The adjustable weight w is set to 1, which means that we weigh the action and timing components equally.

Experiment Results

Table 1 shows the performance of 8 goal-recognition algorithms along with a random baseline.

Action-only Algorithms Columns rtPRP and PRP in Table 1 show the results of real-time PRP and standard PRP (both without a timing component). In DEPOTS, DWR, MICONIC, DRIVERLOG, FERRY, BLOCKSWORLD and LOGISTICS, real-time PRP outperforms PRP. In SOKOBAN and EASYIPCGRID, PRP performs better while in INTRUSION-DETECTION, both approaches have the same performance. Overall, the normalized score for rtPRP is 6.06, which is slightly better than the score of 6.22 achieved by PRP.

These results suggest that real-time PRP performs similarly to PRP, which implies that computing a full solution might not be necessary for goal recognition even when considering the action component alone.

Effect of Timing Components When supplied with the agent-based timing component, rtPRP-a and PRP-a receive overall scores of 3.76 and 3.82 respectively, while importance-based timing components increase these scores to 7.48 and 7.10. The scores for importance-based timing components are worse than those for the corresponding algorithms without timing components (6.06 and 6.22). These results indicate that an accurate timing component can substantially increase the performance of both PRP and rtPRP, but that incorporating evidence from an inconsistent timing component using a sum can be harmful.

Using the timing component as a tiebreaker (AF-a) performs worse (4.43) than the sum of evidence algorithms. On the other hand, AF-i (6.16) is slightly better than the action-only algorithm PRP (6.22). These findings imply that the agent-based timing component can sometimes reverse incorrect inferences made by the action component alone, while even importance-based timing components can be helpful for breaking ties between goals. They also suggest that non-linear evidence combination strategies are likely to be superior to the simple sum used by rtPRP-i and PRP-i.

Observation ratio To explore whether timing information is especially valuable in scenarios with relatively few observed actions, we ran rtPRP / rtPRP-a on BLOCKSWORLD with observation ratios set to 0.25, 0.5, 0.75 and 1.

Table 2 shows that given the timing goal recognition dataset, rtPRP-a has the largest performance boost when fewest observations are available. As expected, timing information appears to be especially valuable when the information conveyed by the action trajectory is relatively minimal.

Discussion

The result suggests that if we want to take advantage of timing information, then we have to access an accurate timing model or at least a good approximation. Our experimental results are in line with the findings in theory of mind (Leslie 1987): if you can construct an accurate model of an actor's mind, then you stand a good chance of correctly inferring their intentions. On the other hand, an inaccurate model is likely to lead to faulty inferences about others.

One possible criticism of our synthetic experimental setup is that the algorithms with the agent-based model timing component rely on the same mechanism used to generate the timing data, and it is therefore not surprising that timing information turns out to be useful to infer the real goal. The next section addresses this concern by demonstrating that the agent-based timing component is still useful when goal inference is performed on human data. Our results for synthetic data, however, still make a useful point: they demonstrate that timing information can be used to distinguish between candidate goals that are not distinguishable based on

Algorithm	rtPRP	rtPRP-a	rtPRP-i	PRP	PRP-a	PRP-i	AF-a	AF-i	Random
DEPOTS	4.61	**2.46**	5.14	4.96	3.36	5.32	3.71	5.09	5.5
MICONIC	1.48	**1.15**	1.38	2.18	1.45	1.55	1.90	1.85	3.5
DWR	2.98	**1.57**	3.70	3.38	2.04	3.52	2.50	3.38	3.5
SOKOBAN	3.52	2.05	3.55	2.02	**1.23**	2.50	1.41	2.09	3.5
EASYIPCGRID	3.58	1.74	3.45	3.07	1.44	3.73	**1.31**	3.21	5.5
DRIVERLOG	2.93	1.71	2.98	3.14	**1.54**	3.21	2.18	3.04	3.5
INTRUSIONDETECTION	1.99	3.16	2.13	1.99	3.16	2.13	**1.93**	2.13	8.83
FERRY	1.71	**1.14**	2.64	2.04	1.39	2.52	1.82	2.02	4
BLOCKSWORLD	4.43	2.83	10.08	5.84	**2.54**	9.51	2.88	5.76	11
LOGISTICS	2.18	1.33	4.33	2.36	**1.23**	3.51	1.67	2.33	5.5
Normalized Score	6.06	**3.76**	7.48	6.22	3.82	7.10	4.43	6.16	10

Table 1: Performance of eight goal recognition algorithms on the timing goal recognition dataset: real-time PRP (rtPRP), real-time PRP with agent-based timing component (rtPRP-a), real-time PRP with importance-based timing component (rtPRP-i), PRP, PRP with agent-based timing component (PRP-a), PRP with importance-based timing component (PRP-i), action first with agent-based timing component (AF-a) and action first with importance-based timing component (AF-i). The best algorithm for each domain is shown in bold. Both AF-a and AF-i use PRP as the action component.

Ratio	Quarter	Half	Three-quarter	Full
rtPRP	2.53	1.5	1.17	1.03
rtPRP-a	1.67	1.17	1.13	1
Difference	0.86	0.33	0.04	0.03

Table 2: Performance of rtPRP-a and rtPRP on BLOCKSWORLD with different observation ratios.

action sequences alone (as in Figure 1a), and can even reverse weak inferences based on action sequences alone (as in Figure 1b).

Over the past decade, several goal recognition algorithms have been developed based on PRP that outperform the original PRP in certain conditions (Pereira, Oren, and Meneguzzi 2017; Santos et al. 2021). These alternatives may perform slightly better than PRP in Tables 1 and 2, but this would not affect our main conclusions. For the behavioral experiments described in the next section, these alternatives would yield the same goal inference as PRP because the action trajectories provide no information about the goal.

Behavioral Experiments

The major question left open by our synthetic experiment is whether timing information can still be exploited when the process generating planning times is not fully known. In real-world settings, for example, we might aspire to make inferences about the goals of human actors even in the absence of a veridical model of human planning. We therefore developed the behavioral experiments to explore whether our current agent model matches humans closely enough to allow rtPRP-a to exploit timing information when inferring the goals of humans.

Problem-solving experiment

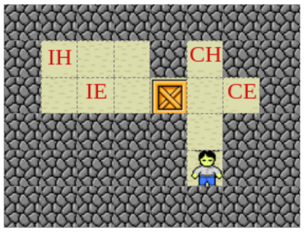
Our first experiment collected human actions and planning times on a series of Sokoban problems. We used these data to ask whether the agent model proposed earlier can generate human-like planning times, and whether timing information can be exploited when inferring the goals of humans. Our experiment was carried out with approval from Human Ethics Advisory Group at the University of Melbourne.

Experiment configuration 50 participants (21 females and 29 males with a median age of 27) were recruited using Prolific and asked to complete 24 Sokoban instances each.

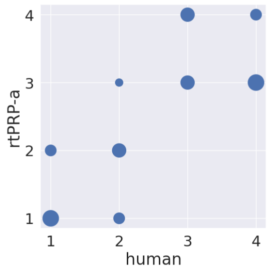
Sokoban is a classic puzzle game where the player must push boxes to designated locations while navigating a maze-like environment. The goal is to successfully move all the boxes to their targets without getting stuck or blocking the path. For simplicity, all off our instances included a single box only. The 24 instances were designed to fall into 6 sets, where each set includes 4 different goal positions located on the same map configuration. One such set is shown in Figure 2a. The presentation order of all 24 instances was randomized for each participant.

Within each set, the 4 goal positions were chosen as follows. A goal is deemed *intuitive* if the first box push on an optimal path to the goal reduces the distance between the box and the goal, and *counter-intuitive* otherwise. In Figure 2a, goals IE and IH are intuitive but goals CH and CE are not. Of the two intuitive goals, IE denotes the "easier" goal and IH the "harder" goal, which in some instances can be unreachable. The difficulty is formalized based on the number of nodes expanded by the A* algorithm. Similarly, CH and CE denote the harder and easier of the two counter-intuitive goals. Choosing goals in this way was inspired by results from the psychological literature suggesting that people tend to spend more time planning when the solution length is long and when the solution involves counter-intuitive moves (MacGregor and Chu 2011; Newell, Simon et al. 1972).

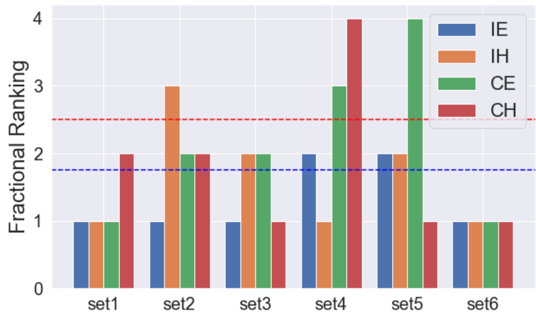
Human planning times are typically highly variable, and to minimize the variance we use only the initial action and initial planning time to generate goal recognition instances with timing information (GRT). We generate 4 separate GRT instances for each set, and each GRT instance includes all four goal positions (IE, IH, CE, and CH) as candidates. All instances were designed so that the first move is forced: in



(a)



(b)



(c)

Figure 2: (a) GRT set 2. One of the 4 potential goal positions is shown in the problem-solving experiment, and the resulting timing is used for goal recognition. (b) Comparison between human initial planning time rank and agent model prediction rank within each group. All instances have similar ranks, with a maximum difference of 1. (c) Performance of rtPRP-a on all GRT instances. The red dotted line denotes the performance of the rtPRP algorithm (2.5) and the blue dotted line denotes the average performance of rtPRP-a (1.75).

Figure 2a, for example, the agent has no option except to move up on the first move. As a result the initial action provides no information about the goal position, but the time taken before this action is potentially informative.

Results and Discussion Some of the "hard" goals in the task are actually unachievable, including goals IH and CH in Figure 2a. We used rtPRP-a as a goal recognition algorithm with both timing and action components and rtPRP as an action-only algorithm.

First, we asked whether rtPRP-a generates human-like planning times. Within each set, we ranked the 4 instances separately by human planning time and by the prediction of rtPRP-a. Figure 2b compares these ranks. Most instances lie along the diagonal, which means that rtPRP-a and humans both give the same rank to those instances. When the rtPRP-a ranking departs from the human ranking, the rank difference for any instance is no more than 1.

We then applied rtPRP-a to the goal recognition task. Because the observation sequences include a single action only, our previous method for aligning t (measured in seconds) with t^* (measured in iterations) no longer applies. We therefore align the two by using ranks relative to the entire set

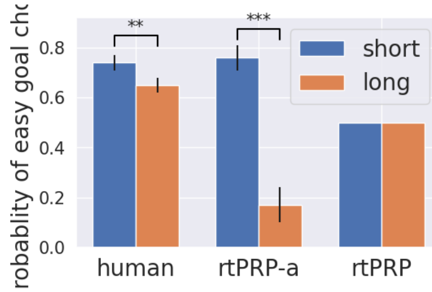


Figure 3: Human and algorithm responses on the GRT instances. The effect of thinking time is significant for humans ($p < 0.05$) and for rtPRP-a ($p < 0.005$) but there is no effect for rtPRP. Error bars show the standard deviation of the mean.

of 24 instances. For example, the median t across this set is mapped to the median t^*.

When choosing among 4 possible goals, rtPRP achieves an average fractional rank of 2.5 (the same as random choice) because by design the initial action is uninformative about the goal. rtPRP-a achieves an average ranking of 1.75, and a paired t-test suggests that the improvement with respect to rtPRP is statistically significant ($t(23) = -3.89, p < 0.001$).

Even though our agent model is at best a coarse approximation of the strategies used by our experimental participants, our results suggest that this approximation is good enough to be usefully incorporated into our goal recognition framework.

Goal recognition experiment

Our work is motivated in part by the idea that humans take timing information into account when faced with goal recognition problems such as those in Figure 1a. To our knowledge, this idea has not been previously tested, and we therefore designed a second behavioral experiment to verify that timing information can influence human goal recognition.

Experiment configuration We designed 13 pairs of goal recognition instances based on the Sokoban domain. One pair used the configuration in Figure 1a. The instances in each pair included the same map and the same two potential goals. One goal (e.g. A in Figure 1a) was easy and the other (e.g. B in Figure 1a) was hard, where "hard" and "easy" are defined as for the previous experiment using A*.

For each member of a pair, participants saw the same sequence of three actions, and the only difference within a pair was the time observed for the third action. For "long" instances, the time associated with the third action was 3 seconds, and for "short" instances the time was only 0.5 seconds. The first two actions were always forced (e.g. in Figure 1a an actor who does not backtrack has no option but to move up twice), and the time for both actions was always set to 0.1 seconds.

For each instance, participants observed the sequence of three actions and then indicated whether A or B was more likely to be the goal pursued by the actor. For each pair of instances, we anticipated that participants would be more likely to choose the hard goal in the long version than the short version.

The same 50 participants who completed the problem-solving experiment also completed the goal-recognition experiment, and the goal-recognition experiment was always completed second. As a result participants were familiar with the Sokoban domain by the time they started the goal-recognition task. The presentation order of the 26 goal recognition instances was randomized within participants.

Results and Discussion Figure 3 shows the average probability of choosing the easy goal across all 13 pairs of instances. As predicted, humans are more likely to choose the easy goal given a short instance than when given the corresponding long instance. A paired t-test reveals that this difference between long and short instances is statistically significant, and confirms that human goal inference is sensitive to timing information. rtPRP-a ($t(12) = 5.48, p < 0.001$) shows the same pattern as humans ($t(12) = 4.26, p = 0.001$) but the action-only algorithm rtPRP does not consider timing information and therefore generates identical responses to short and long instances.

Although humans and rtPRP-a are both sensitive to timing information, they respond differently to long instances. Humans prefer to choose easy goals even for long instances, but rtPRP-a is more likely to choose hard goals than easy goals across the set of long instances. This difference may reveal a lack of calibration between rtPRP-a and humans. For example, if the true goal were easy, spending 3 seconds on a single move would be highly anomalous according to rtPRP-a, but is apparently less anomalous according to people. Future work can attempt to better calibrate the predictions of rtPRP-a by aligning human and model planning times across responses to a large set of planning problems.

Future Directions

Our framework opens up a number of additional directions for future work, and here we consider four that seem especially important. First, as mentioned in our discussion of our synthetic experiment, a generative approach that makes accurate inferences based on human planning times will need to incorporate an accurate generative model of human planning times. Our behavioral experiments suggest that our current agent model is accurate enough to support useful inferences about human planners. This model, however, is far from a comprehensive account of human planning and future agent models can incorporate additional factors that influence human planning times. For example, future models may be able to capture the notion of action commitment by incorporating a meta-reasoning process about when to stop searching and add the current best action to the execution queue (Gu et al. 2022). Future versions of the model can also take a bounded-rationality approach and explicitly incorporate human memory limitations (Simon 1990).

A second direction is to develop agent models that allow for individual differences. Our behavioral data suggested that planning strategies are highly variable across individuals: some participants seem to compute a complete path to the goal, while others seem to focus only on the next few steps. Future versions of our agent model could therefore include adjustable parameters that reflect individual differences in planning strategies, and the values for these parameters could be inferred on a per-participant basis.

Third, our current analyses assume that decision cost for a move (i.e. the total effort required to select the move) is proportional to the observed time for that move. This assumption holds if an agent is memoryless, and must carry out a fresh search on each move without using any information computed on previous moves. In reality, however, decision costs may be amortized over multiple moves, because humans and other memory-based agents may reuse information (such as search trees) computed on previous moves (Krusche et al. 2018; Van Opheusden et al. 2017). Future models can therefore consider ways to use observed planning times to infer the total decision cost associated with each move. One possible approach is to model the total decision cost for a given move as a discounted accumulated sum that incorporates some fraction of the observed times recorded for previous moves.

Finally, although timing information is often informative about the goals of an actor, this relationship may not hold in contexts in which actors use strategies other than forward search to make decisions. In some scenarios, especially when people are dealing with familiar situations, they might act immediately in a reflex way without thinking or reasoning (Kuperwajs, Van Opheusden, and Ma 2019). Whether or not actors carry out forward search could potentially be inferred on the basis of timing information. Future extensions of our model could therefore adopt a hierarchical approach that supports two inferential phases: the first phase aims to identify moves for which an actor has relied on forward search, and the second phase uses only these moves to infer the goal pursued by the actor.

Conclusion

Goal recognition is an important problem for both AI and cognitive science researchers. Most work in this area considers action sequences only, but we showed that humans are sensitive to timing information and introduced a goal recognition framework that can take timing information into account. To develop and evaluate this framework we introduced an agent model with a response-time mechanism inspired by the evidence accumulation literature in cognitive science. Our results suggest that incorporating an accurate model of timing is a promising way to improve the performance of goal recognition algorithms, and that our agent model captures human planning closely enough to support useful inferences about the goals pursued by human actors. Because timing information is easy to acquire and generally observable, exploiting this information can potentially provide payoffs across many different settings.

References

Amir, D.; and Amir, O. 2018. Highlights: Summarizing agent behavior to people. In *Proceedings of the 17th Inter-*

national Conference on Autonomous Agents and MultiAgent Systems, 1168–1176.

Avrahami-Zilberbrand, D.; Kaminka, G.; and Zarosim, H. 2005. Fast and complete symbolic plan recognition: Allowing for duration, interleaved execution, and lossy observations. In *Proc. of the AAAI Workshop on Modeling Others from Observations, MOO*, 25–19. Citeseer.

Baker, C. L.; Jara-Ettinger, J.; Saxe, R.; and Tenenbaum, J. B. 2017. Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, 1(4): 0064.

Blaylock, N.; Allen, J.; et al. 2003. Corpus-based, statistical goal recognition. In *IJCAI*, volume 3, 1303–1308.

Bylander, T. 1994. The computational complexity of propositional STRIPS planning. *Artificial Intelligence*, 69(1-2): 165–204.

Chen, J. X. 2016. The evolution of computing: AlphaGo. *Computing in Science & Engineering*, 18(4): 4–7.

Gates, V.; Callaway, F.; Ho, M. K.; and Griffiths, T. 2021. A rational model of people's inferences about others' preferences based on response times.

Gu, T.; Ruml, W.; Shperberg, S. S.; Shimony, E. S.; and Karpas, E. 2022. When to Commit to an Action in Online Planning and Search. In *Proceedings of the International Symposium on Combinatorial Search*, volume 15, 83–90.

Ho, M. K.; Abel, D.; Cohen, J. D.; Littman, M. L.; and Griffiths, T. L. 2020. The efficiency of human cognition reflects planned information processing. *arXiv preprint arXiv:2002.05769*.

Hoffmann, J.; and Nebel, B. 2001. The FF planning system: Fast plan generation through heuristic search. *Journal of Artificial Intelligence Research*, 14: 253–302.

Jara-Ettinger, J. 2019. Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, 29: 105–110.

Keren, S.; Gal, A.; and Karpas, E. 2015. Goal recognition design for non-optimal agents. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Kocsis, L.; and Szepesvári, C. 2006. Bandit based monte-carlo planning. In *European conference on machine learning*, 282–293. Springer.

Krusche, M. J.; Schulz, E.; Guez, A.; and Speekenbrink, M. 2018. Adaptive planning in human search. *BioRxiv*, 268938.

Kuperwajs, I.; Van Opheusden, B.; and Ma, W. J. 2019. Prospective planning and retrospective learning in a large-scale combinatorial game. In *2019 Conference on Cognitive Computational Neuroscience*, 13–16.

Leslie, A. M. 1987. Pretense and representation: The origins of" theory of mind.". *Psychological review*, 94(4): 412.

Lipovetzky, N.; and Geffner, H. 2017. Best-first width search: Exploration and exploitation in classical planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

MacGregor, J. N.; and Chu, Y. 2011. Human performance on the traveling salesman and related problems: A review. *The Journal of Problem Solving*, 3(2): 2.

Masters, P.; and Sardina, S. 2019. Goal recognition for rational and irrational agents. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 440–448.

Newell, A.; Simon, H. A.; et al. 1972. *Human problem solving*, volume 104. Prentice-hall Englewood Cliffs, NJ.

Pereira, R.; Oren, N.; and Meneguzzi, F. 2017. Landmark-based heuristics for goal recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Rabinowitz, N. C.; Perbet, F.; Song, H. F.; Zhang, C.; Eslami, S.; and Botvinick, M. 2018. Machine theory of mind. *arXiv preprint arXiv:1802.07740*.

Raileanu, R.; Denton, E.; Szlam, A.; and Fergus, R. 2018. Modeling others using oneself in multi-agent reinforcement learning. *arXiv preprint arXiv:1802.09640*.

Ramírez, M.; and Geffner, H. 2010. Probabilistic plan recognition using off-the-shelf classical planners. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

Ratcliff, R.; Smith, P. L.; Brown, S. D.; and McKoon, G. 2016. Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, 20(4): 260–281.

Rescorla, M. 2015. The computational theory of mind.

Santos, L. R.; Meneguzzi, F.; Pereira, R. F.; and Pereira, A. G. 2021. An LP-based approach for goal recognition as planning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 11939–11946.

Simon, H. A. 1990. Bounded rationality. In *Utility and probability*, 15–18. Springer.

Singh, R.; Miller, T.; Newn, J.; Sonenberg, L.; Velloso, E.; and Vetere, F. 2018. Combining planning with gaze for online human intention recognition. In *Proceedings of the 17th international conference on autonomous agents and multiagent systems*, 488–496.

Sohrabi, S.; Riabov, A. V.; and Udrea, O. 2016. Plan Recognition as Planning Revisited. In *IJCAI*, 3258–3264.

Solway, A.; and Botvinick, M. M. 2015. Evidence integration in model-based tree search. *Proceedings of the National Academy of Sciences*, 112(37): 11708–11713.

Tavares, G.; Perona, P.; and Rangel, A. 2017. The attentional drift diffusion model of simple perceptual decision-making. *Frontiers in neuroscience*, 11: 468.

Tversky, A.; and Kahneman, D. 1989. Rational choice and the framing of decisions. In *Multiple criteria decision making and risk analysis using microcomputers*, 81–126. Springer.

Van Opheusden, B.; Galbiati, G.; Bnaya, Z.; Li, Y.; and Ma, W. J. 2017. A computational model for decision tree search. In *CogSci*.

Vered, M.; Kaminka, G. A.; and Biham, S. 2016. Online goal recognition through mirroring: Humans and agents. In *The Fourth Annual Conference on Advances in Cognitive Systems*.

Zhi-Xuan, T.; Mann, J. L.; Silver, T.; Tenenbaum, J. B.; and Mansinghka, V. K. 2020. Online bayesian goal inference for boundedly-rational planning agents. *arXiv preprint arXiv:2006.07532*.