KeyWorld: A goal inference game for evaluating resource-rational models

Amani Maina-Kilaas (amanirmk@mit.edu)
Department of Brain & Cognitive Sciences
Massachusetts Institute of Technology

Ingrid Wu (iswu@mit.edu)
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

Abstract

Inference plays a key role in everyday life. It can be as simple as guessing why your phone isn't turning on or deciphering a look on someone's face. Yet making these inferences is nontrivial and can be computationally-intensive. Accordingly, one might expect that the human mind is somehow optimized to minimize computations while maximizing the expected utility of the inference. This principle is called resource rationality and models capturing this often do well to predict both the successes and errors of human judgments. However, a utility-cost tradeoff is assumed for these models, and the specific method of collecting human judgments may shift the relative considerations for utility and cost. In this work we propose KEYWORLD, a goal inference game which collects rapid, low-stakes decisions from humans to implicitly test inferences. We conduct a pilot experiment, demonstrating that this framework can be used to evaluate resource-rational models.

Keywords: Bayesian inference; goal inference; intention inference; inverse planning; resource-rationality

Introduction

Imagine driving comfortably down a stretch of road with no stop lights. You pay no mind to the pedestrians on the sidewalk since there is no crossing for some time. Yet suddenly, one pedestrian turns to face the street and, before they even move, you start slowing down. This everyday scenario, driven solely by observation and preemptive action, raises questions about the underlying computations. How certain were you that they were about to cross before you acted? What made you re-evaluate the initial belief that they would continue on the sidewalk? How many possibilities did you imagine?

In quick decisions like this example, one must trade off the value of taking the best action with the cognitive cost of determining the best action. This notion of resource rationality (strategically allocating computational resources), is of great importance to both artificial intelligence (AI) and cognitive science. In cognitive science, resource rationality is a principle that can often lead to better behavioral predictions, explaining both cases of success and error (Hahn, Futrell, Levy, & Gibson, 2022). While traditionally in AI, only cases of success are valued, it is still desirable to achieve the most success for the least computation.

Scenarios like the one described also raise another essential question: in the context of limited communication, how do humans coordinate with other agents? When two agents in a collaborative setting are unable to communicate directly, they must form beliefs regarding the other's plans and goals based primarily on observing their actions, and be robust to situations where behavior may be ambiguous and misinterpreted. This task is often referred to as goal inference, and is particularly relevant to the field of human-robot interaction.

In this work, we present KEYWORLD, a game for investigating principles of resource-rationality in the domain of goal inference. In KEYWORLD, two agents collaborate in a setting with minimal communication, where inferring the correct goal is essential for success. We use the game to evaluate an inference model over a range of hyperparameters corresponding to different resource-rational assumptions and compare the model to human behavioral data. We find that humans are best predicted by hyperparameters favoring rationality over optimality, supporting the hypothesis that people make choices in alignment with minimizing computational resources.

Related Work

In this section, we outline several areas of research relevant to this work.

Goal Inference

Baker, Tenenbaum, and Saxe (2007) propose inverse planning as a framework for understanding goal inference. Since planning is a model from goals to actions, it can be inverted to obtain a model from actions to goals. Baker et al. showed that this framework led to good correlation with human inferences. In that work, they collected human judgments by showing a partial path of an agent and asking where the agent was likely headed. Although this approach has the advantage of directly collecting human inferences, it may impact an investigation of resource rationality. That is because pausing and asking could encourage participants to think more carefully about their answer, raising the amount of computation that people dedicate to the task. In our work, we measure inferences implicitly through the way it influences their actions. This allows us to investigate rapid, low-stakes choices, which more accurately represent the inferences people make in everyday life.

Zhi-Xuan, Mann, Silver, Tenenbaum, and Mansinghka (2020) study goal inference using a gridworld environment in which a single agent picks up keys to open doors and reach a desired gem. We took inspiration from this and sim-

ilar "key" games when creating our two-player, collaborative KEYWORLD.

Resource Rationality

Exact Bayesian inference provides an elegant framework for understanding human cognition, but fails to scale for many of the complex problems that people face, leading to the adoption of sample-based approximations. Even then, one concern with viewing cognition as Bayesian inference is that approximating a posterior distribution would require a large number of samples, far fewer than humans could realistically do. Vul, Goodman, Griffiths, and Tenenbaum (2014) analyzed the cost-benefit trade-off of how much to think and demonstrated that for a wide range of decision problems where there is one correct action, it is actually optimal to only take on the order of one sample. With our work, we can explore this idea within a more nuanced goal inference context, provide further evidence that few samples are sufficient to obtain high-quality performance. Although we do not explore the number of samples in our pilot experiment, and instead focus on other resource-rational choices, this is a future avenue of this work.

While reducing the computation involved in updating the posterior is a key aspect of resource-rationality, so is the question of when to update the posterior. In an autonomous vehicle intention inference context, Amatya, Ghimire, Ren, Xu, and Zhang (2022) use a reinforcement learning-based controller to determine when intention inference should be computed, outperforming several baselines. In similar spirit, Callaway, Gul, Krueger, Griffiths, and Lieder (2018) create a Bayesian metareasoning model that can decide how to allocate and terminate computation. Our work takes inspiration and explores several criteria, albeit simpler, for deciding when to update the posterior.

Collaboration with Limited Communication

When only limited communication between agents is allowed, collaboration can become quite challenging. The lack of direct and effective information potentially leads to misunderstanding an agent's behavior and goal. From the field of human-robot interaction, Saulnier, Sharlin, and Greenberg (2011) explore how minimal nonverbal robot cues can be used to initiate "interuption", signaling the importance and urgency of a situation to a human. Dragan, Bauman, Forlizzi, and Srinivasa (2015) study the effects of robot motion on collaboration with a human. They define functional, predictable, and legible motion, and find that legible motion—motion that enables the collaborator to infer the goal—leads to the most fluent collaborations. KEYWORLD provides a framework in which the effects of different types of motion on human-robot collaboration could be easily studied. In an interesting contrast with many prior work, our game presents a situation where the robot is leading the collaboration rather than supporting.

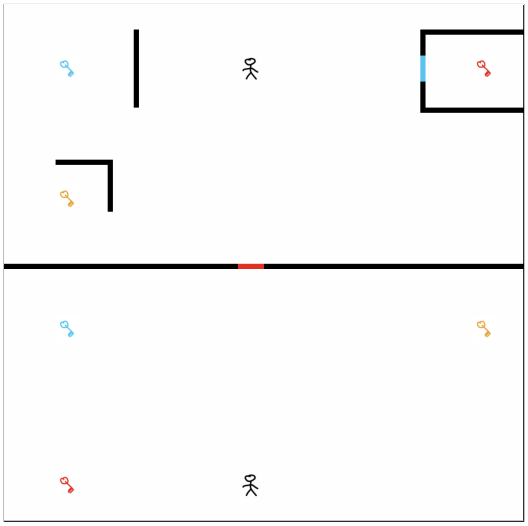


Figure 1: Example world (variant 11). The agents are represented as stick figures (top KNOWER, bottom WATCHER), walls as black lines, and doors as colored lines that can be opened by the key of the same color. Note that, when playing the game, the doors are depicted as gray since the WATCHER does not know which key corresponds to each door.

KeyWorld

In KEYWORLD, two agents collaborate to open a door that separates them and divides the world in half. To do so, both agents must bring the correct key to the door as quickly as they can. However, only the agent on the top half, the KNOWER, can see which key opens which door. The agent on the bottom half, the WATCHER, must observe the actions of the KNOWER in order to infer which key they should bring to the door. The KNOWER may have to navigate obstacles, including other locked doors, which complicate the WATCHER's inference process. Each agent takes turns in alternation, requiring decisive actions from the WATCHER before there may be sufficient information to act on. See Figure 1 for an example world.

Potential Investigations

We focus on a version of this game where the KNOWER is controlled by an artificial agent following a shortest path algorithm and the WATCHER is played by a human participant. We use this framework to gain insight into a human player's allocation of computational resources and how they may differ from what is optimal for the task.

However, KEYWORLD offers a platform for other investigations as well. Relating to the limited communication work (Dragan et al., 2015; Saulnier et al., 2011), one possibility we thought particularly interesting, but outside of the current scope of this work, is to investigate non-verbal pragmatics. In principal, the KNOWER could choose its movements in order to reduce the ambiguity of the correct key by applying Theory of Mind. A future extension of this work could model the collaboration between two human players with a pragmatic

framework such as Rational Speech Acts (Frank & Goodman, 2012).

Model

We aim to model the quick decisions a human participant must make when playing as the WATCHER. Our Bayesian model observes the KNOWER move, updates the inference for their goal (the correct key), and then selects the best move given the inference.

Planning Model

We follow the framework of goal inference as inverse planning (Baker et al., 2007), applying Bayes' rule to invert a model of planning $P(\text{action} = a \,|\, \text{goal} = g)$ into $P(g \,|\, a)$. However, this requires a model of the KNOWER's planning algorithm, which is not necessarily the actual algorithm used.

The likelihood of an action given a goal $P(a \,|\, g)$ is an increasing function of the utility of the action under the goal $U_g(a)$. In a reinforcement learning style view, the utility of an action would be equal to the expected utility of all paths stemming from said action. For problems of this type where computation grows exponentially, forms of Markov Chain Monte Carlo (MCMC) are often used to sample from possible paths and form an estimate of the expected value. We are interested in resource-rationality, which raises the question of how many samples N to take before acting. However, this is not the only way to reduce computation—for one, people may rely on cheaper heuristics. Since in the context of our game, utility is a decreasing function of the distance to the goal[1], we can use MCMC to explore the first K steps in each path and then estimate the remainder of the path using a shortest path heuristic[2]. At $K = \infty$, this becomes normal MCMC with N samples. At $K = 0$, only the action corresponding to the shortest path has any utility. In this way, both N and K become hyperparameters controlling the approximation of the true utility.

Once we have the expected distances of each action a_i to the goal, $d_g(a_i)$, we compute the utility as

$$U_g(a_i) = 1 - \frac{d_g(a_i)}{\sum_{a \in A} d_g(a)},$$

which values having a shorter distance relative to the other actions. However, since all actions are similar in distance (at most 2 apart), we have a "rationality" parameter α which controls how much preference the KNOWER agent should have for the optimal action. The final likelihood is calculated as

$$p(a_i \,|\, g) = \frac{U_g(a_i)^\alpha}{\sum_{a \in A} U_g(a)^\alpha}.$$

[1]To be accurate, the goal is not the key itself but to be at the main door with the correct key. Thus the distance d_g includes both picking up and delivering the key. This is a subtle nuance and in most cases it is safe to think of the key as the goal.

[2]Note that when imagining the possible paths of the KNOWER, our model pretends that a key can open any door since the WATCHER is not aware of the correct keys.

Belief Updates

Under Bayesian inference, the belief update after observing the KNOWER's action is

$$p(g \,|\, a_i) = \frac{p(a_i \,|\, g) p(g)}{p(a_i)}.$$

We initialize $p(g)$ according to the best action utility,

$$p(g_i) = \frac{\min_{a \in A} U_{g_i}(a)^\alpha}{\sum_{g \in G} \min_{a \in A} U_{g_i}(a)^\alpha},$$

although this results in a uniform prior for the worlds we explore, and then use the most recent beliefs in each subsequent update.

As another aspect of resource rationality, we explore several criteria for deciding when to update the beliefs: *turn-based*, *action-based*, and *goal-based*. In the turn-based criteria, the model updates its beliefs every nth turn. For action-based, the model updates its beliefs any time the estimated probability of the KNOWER's action is below a given threshold, $p(a) < p$. Similar to action-based, the goal-based criteria says to update whenever the KNOWER's action is unlikely, but specifically with respect to the currently-understood goal, $p(a \,|\, \text{argmax}_{g \in G} p(g))$.

Action Selection

Lastly, our model selects an action for the WATCHER according to the same planning model used for the KNOWER. That is, it selects $\text{argmax}_{a \in A} p(a)$, where

$$p(a_i) = \sum_{g \in G} p(a_i|g) p(g).$$

Pilot Experiment

To test the KEYWORLD domain and explore some of the resource-rational hyperparameters previously-described, we launched a pilot experiment with reduced dimensionality.

World Variants

We created four base worlds that all look very similar to each other: (a) only blue and orange keys with no obstacles, (b) only blue and orange keys with obstacles, (c) all three keys with no obstacles, and (d) all three keys with obstacles. For all, the key layouts were identical. With these four worlds, we permuted all combinations of keys that the doors could correspond to, resulting in 16 total world variants. Figure 1, variant 11, is an instance of (d).

Human Data

Due to various constraints, for this initial experiment, the two authors played the role of the participants. Each participant played all 16 world variants in a random order with short breaks in between games. We recorded both the sequence of moves and the amount of time used to make each move.

Hyperparameter Ranges

To reduce dimensionality, we only explore $K = 1$ (see model description for hyperparameter definitions). This means that each action's utility corresponds to the shortest path heuristic after having taken that action. This eliminates the MCMC sampling hyperparameter N, as the applicable paths are fully explored.

For the remaining hyperparameters we do a coarse grid search. We explore $\alpha \in [1, 4, 16, 256, 1024]$, which ranges from barely caring which move the KNOWER makes to completely requiring it be optimal. For the action-based and goal-based criteria, we explore $p \in [0.01, 0.17, 0.33, 0.49, 0.65]$, and for turn-based, we explore $n \in [1, 2, 3, 4, 5]$.

Results

For all of the following results, games were run for a maximum of 100 moves and were marked as unsolved if they reached the limit.

Performance

First, we take a look at the overall performance from both humans and models. We measure performance with the *speed ratio*, which is the number of moves it took the KNOWER to reach the goal over the number of moves for the WATCHER. When the world is unsolved, the WATCHER is considered as having taken 100 moves. The speed ratio is greater than 1 when the WATCHER reaches the door before the KNOWER, and less than 1 if after. Figure 2 shows our performance results.

In Figure 2a-b, we show the average performance on each of our world variants. There is notably-well alignment between the world variants that humans and models found difficult. The models often did better or equal to the humans where humans did well, but worse where the humans had difficulty.

Figure 2c shows the percent of worlds solved by each set of hyperparameters explored. We see that infrequent updates often leads to unsolved worlds ($p \leq 0.17$ for goal and action, $n \geq 3$ for turn), but that most hyperparameters consistently obtain the correct key. Figure 2d shows the same breakdown but for the speed ratio, which shows a clear preference for higher rationality parameters (α). We see that less frequent goal-based updates with high rationality achieves the best performance out of all combinations explored, averaging a speed faster than the KNOWER.

Human-Model Comparison

Next, we compare our hyperparameters with the human data to see which are the best fit. To do so, we have the model "watch" a replay of the human game, following along and imagining which move it would make in their position. We use this strategy to obtain the log-likelihood of a game under the model, and then compute the average log-likelihood of the human's move. These average log-likelihoods are shown in Figure 3a. Across the board, the best fits are found with $\alpha = 256$. Within that, for action-based and turn-based update criteria, there is a preference for mid-frequency updates.

As a second measure of human fit, we compare the surprisal of the model with human reaction time. Past studies in computational psycholinguistics have shown a linear linking function between word surprisal and reading times (Shain, Meister, Pimentel, Cotterell, & Levy, 2023; Smith & Levy, 2013). Given this, we may expect the reaction time of a human playing the game to correspond with the surprisal of the KNOWER's action. That is, a human would take longer to choose a move if the last action they observed from the KNOWER was surprising given their current beliefs. Figure 3b-c show the correlations between human reaction time and model surprisal, as measured by action-based surprisal, $-\log p(a)$, and goal-based surprisal, $-\log p(a \,|\, \text{argmax}_{g \in G} p(g))$, respectively. Wee see that surprisal generally correlates well with human reaction time, with the best fits at $\alpha = 64$.

Belief Trajectories

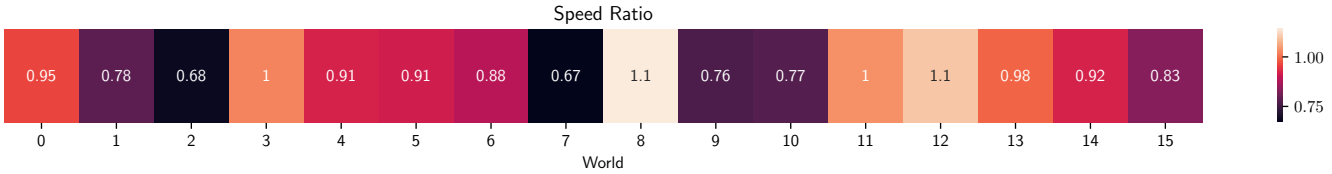
Lastly, we look qualitatively at the trajectory of model beliefs over the progression of a game, observing the effect of our parameters. In Figure 4, we show the belief trajectories for all hyperparameters playing the world shown in Figure 1. In this world variant, the KNOWER retrieves the blue key (which is near the orange key), opens the door to the red key, and then makes its way to the main door. We selected this variant for investigation because of the interesting belief change from blue to red. We see that increasing α leads to larger changes in beliefs and that decreasing the frequency of updates often results in a single change point from uniform beliefs to favoring the correct key. Although we do not have equivalent human data to compare to, many trajectories in the middle α values match intuitions.

Discussion

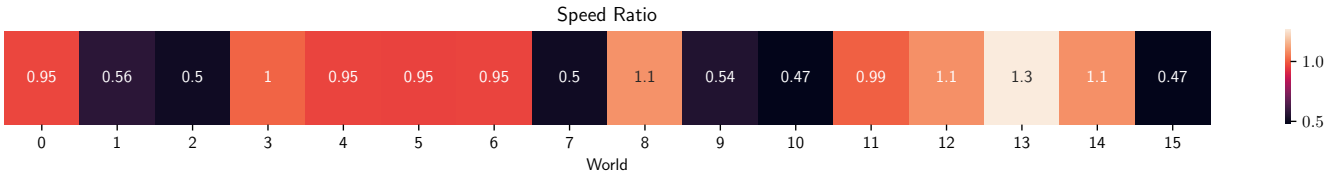
In this work, we explored hyperparameters and evaluated them based on both performance and fit to human data. We found that performance was best at high levels of rationality (α), which is to be expected since our KNOWER implementation only ever acts optimally. We also generally see that more frequent updates is better for performance, which is also unsurprising given that it results in more up-to-date beliefs. Perhaps the most interesting result is that the model (averaged over hyperparameters) tends to do better than humans in easy worlds and worse than humans in difficult worlds. We believe this may reflect (a) the non-optimality of humans, and (b) the robustness of humans. In cases where an irrational action from the KNOWER clearly signals the correct key, models do better; in cases where the initial inference may be deceiving, humans are better able to adapt.

The most difficult world variants for the human participants were 2 and 7. These correspond to the three-key world with no obstacles, where the orange key opens the main door. It seems that the presence of a locked-away, but unused key is misleading. Models also do very poorly on world variants

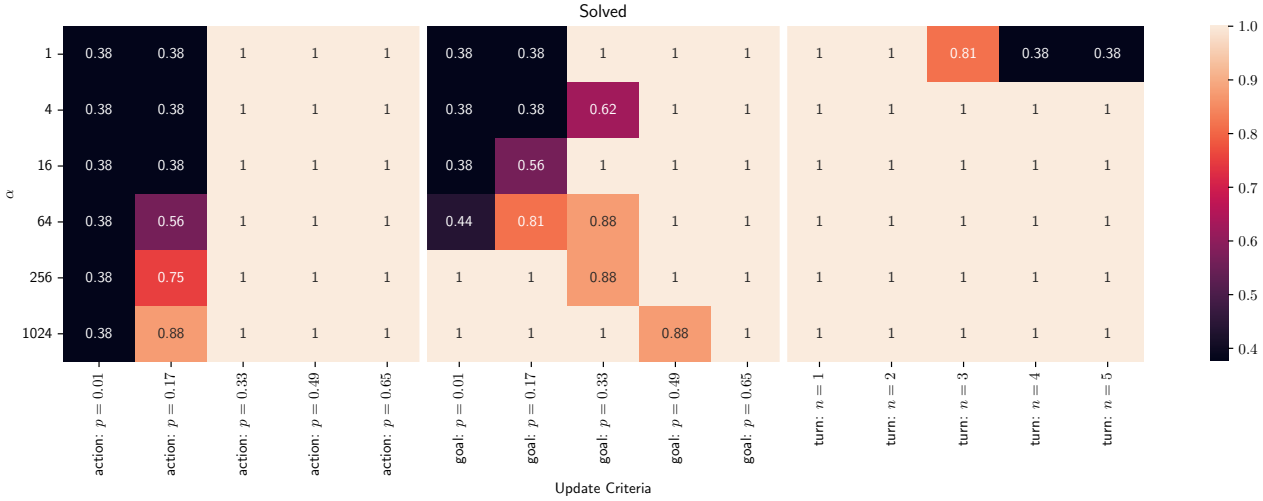
(a) Average human performance by world.



(b) Average model performance by world.



(c) Percent of worlds solved by hyperparameters.



(d) Average model performance by hyperparameters.

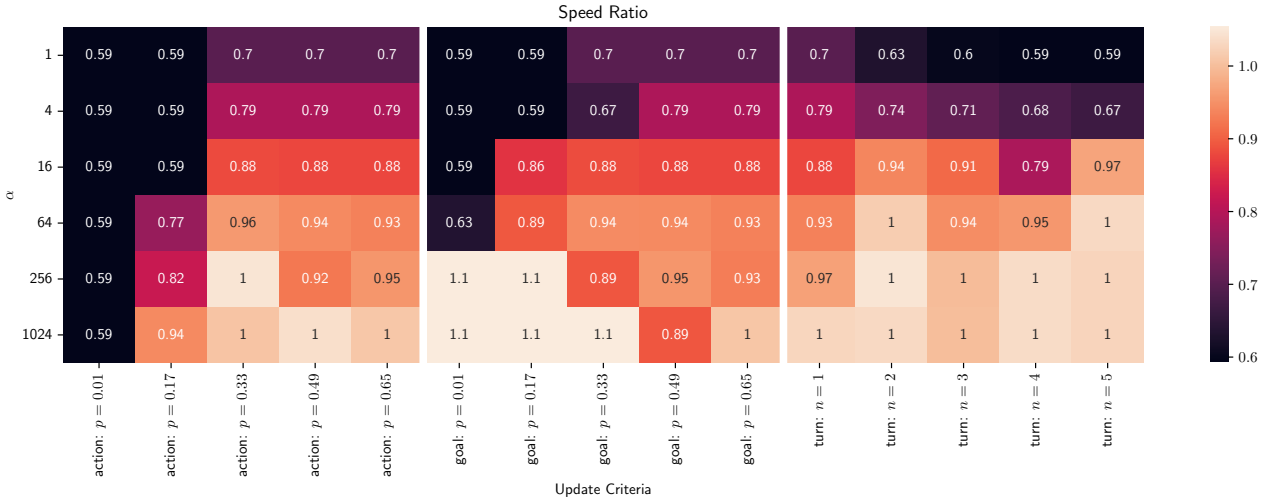
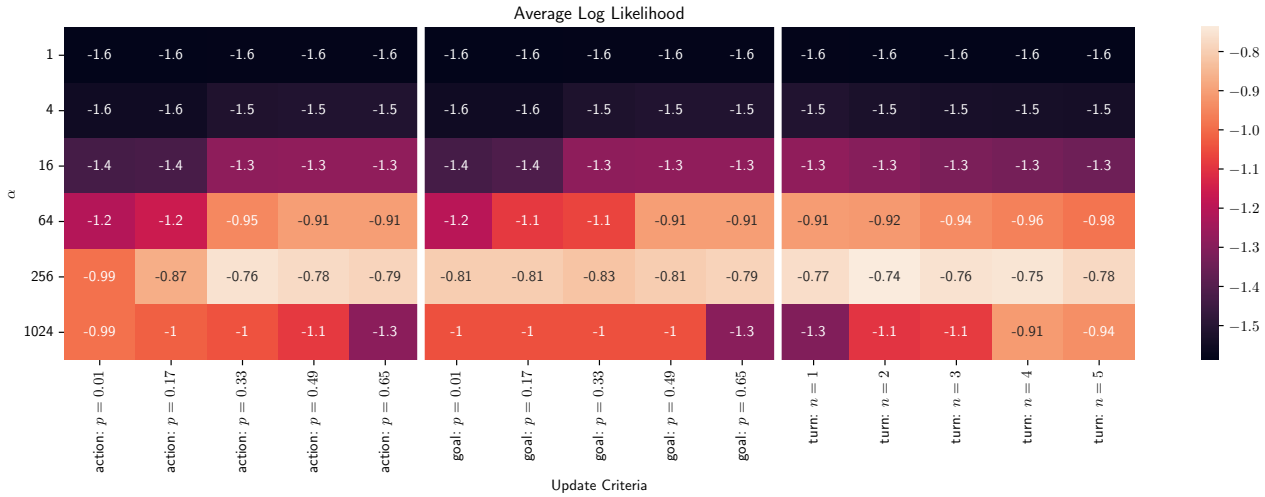
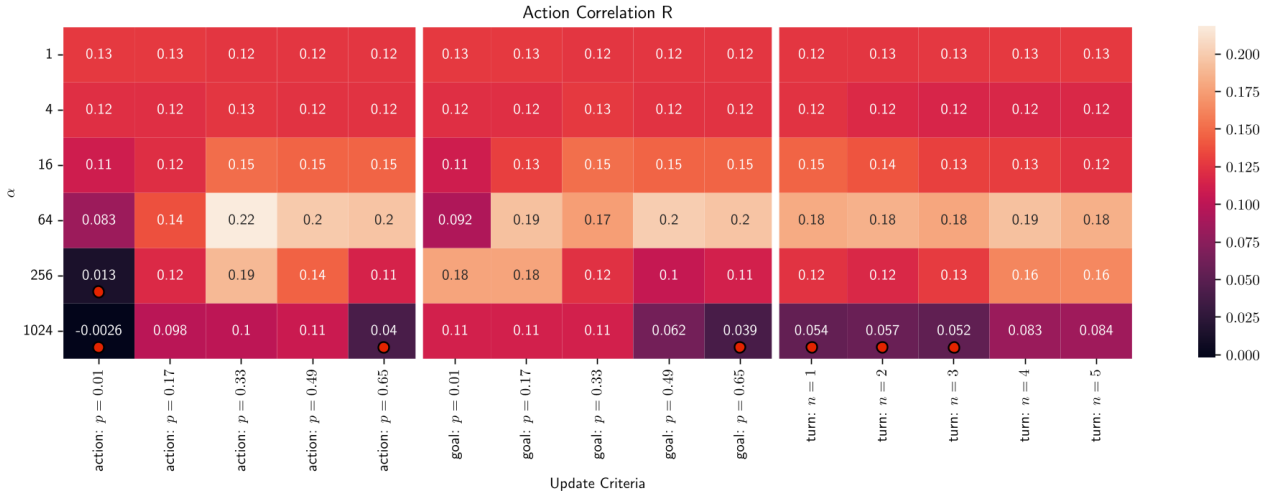


Figure 2: Performance. Here we present the performance, as measured by the speed ratio, on our world variants for **(a)** humans, averaged by subject, and **(b)** models, averaged by model hyperparameters. We also present **(c)** the percent of world variants solved (completed within 100 moves) and **(d)** average performance for each of the model hyperparameters tested.

(a) Average log-likelihood of human move according to model.



(b) Correlation of action-based surprisal to human reaction time.



(c) Correlation of goal-based surprisal to human reaction time.

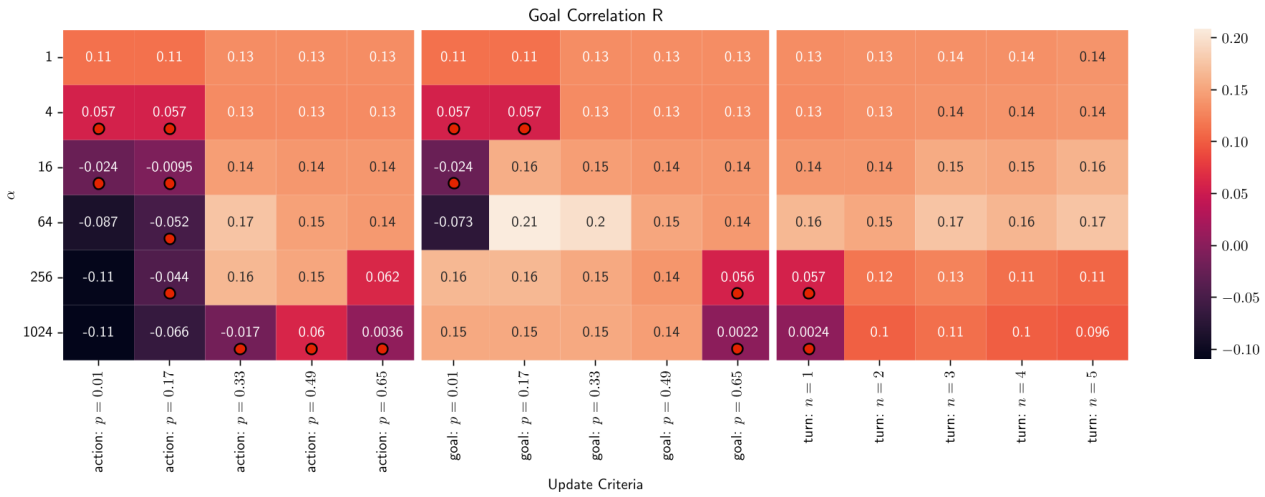


Figure 3: Human-Model Comparison. **(a)** shows the fit of model hyperparameters to human data, measured by the average log-likelihood of the human's move as predicted by the model. **(b)** and **(c)** show the Pearson correlation between the surprisal of the KNOWER's move and the subsequent reaction time; values that are *not* significant ($p > 0.05$) are indicated with a red dot.

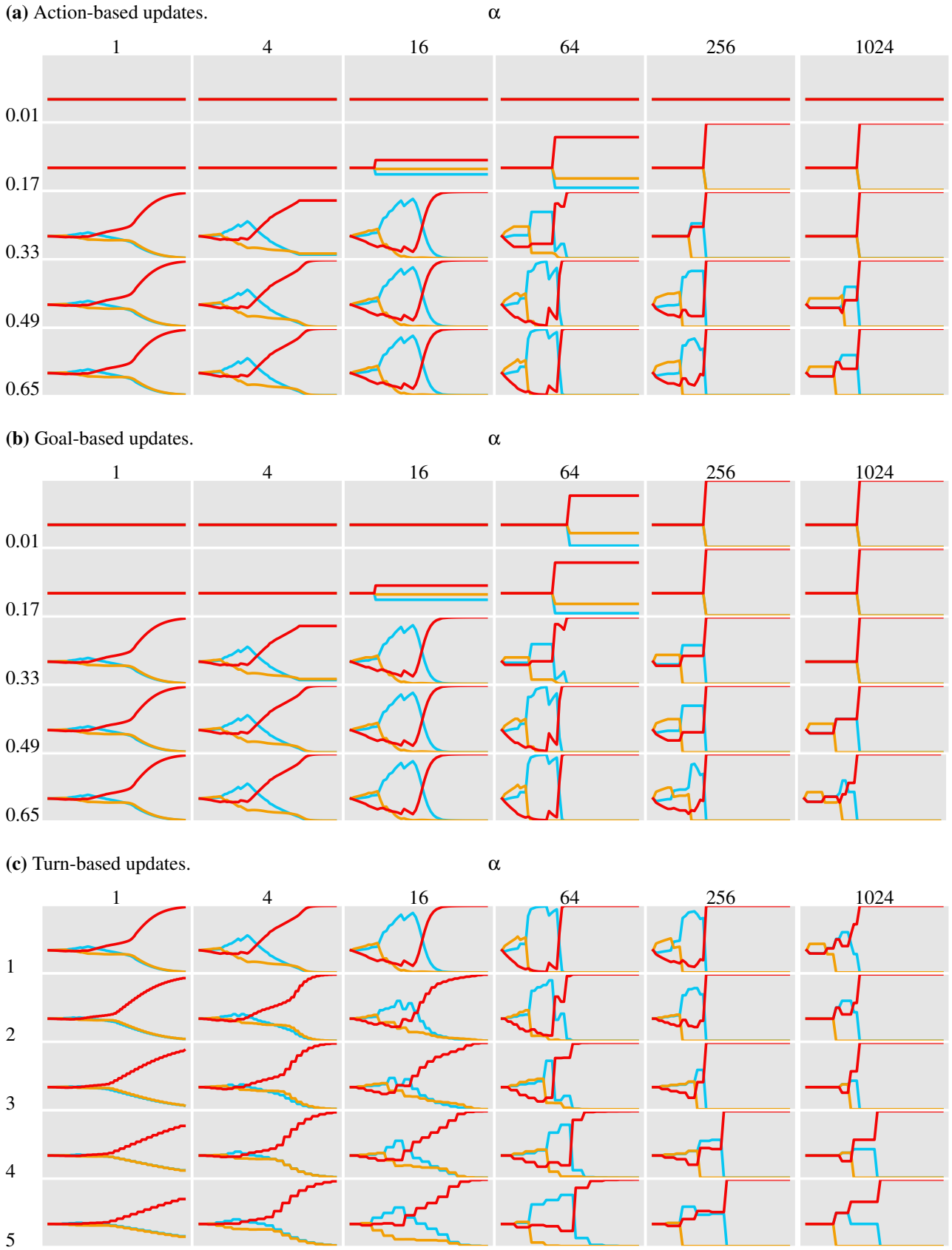


Figure 4: Belief trajectories for world variant 11 (see Figure 1). For each trajectory plot, the colors correspond to keys, the vertical axis is the level of belief a key is the goal, and the horizontal axis is the progression of the game (note that some games last much longer than others).

10 and 15, which are identical to 2 and 7 except with the presence of obstacles. It is possible that the obstacles helped disambiguate the task for humans, but given that the models did even worse with the obstacles, we hypothesize that the presence of obstacles reduced the expectation that the extra room should be involved.

We then found that the best fits to the human data were not the most optimal hyperparameters for performance. Most clearly, the fact that the best α was not the maximum means that humans do not expect completely rational behavior from the KNOWER. The best fits also indicated that it is good to update less frequently: the best turn-based criteria was updating every other step, while the best action-based criteria was updating when $p < 0.33$. The goal-based criteria did not have as clear of a preference, but was best fit by updating the most frequently ($p < 0.65$). This might be because the goal-based criteria does not take into account how confident the model is of that goal. It is also worth mentioning that our best average log-likelihood was a value of -0.74, which corresponds to a probability of 0.48. Roughly speaking, this is near the best that we could reasonably expect from a model, as with a clear goal there are typically two optimal moves in a gridworld.

We also saw that surprisal correlated, although not especially strongly, with human reaction time. Furthermore, this correlation was statistically significant in the majority of settings. Interestingly, there is a discrepancy between the log-likelihood fits and correlations: $\alpha = 256$ best predicts human moves, while $\alpha = 64$ best correlates with human reactions. We are not sure how to best explain this difference.

Limitations

This pilot study, while providing initial insights into human goal inference process, comes with some limitations. Firstly, the data collection was confined to a very small sample size, restricted to the two authors due to various constraints. The influence of our prior knowledge and experience with KEYWORLD, despite our attempts to mitigate this by shuffling variants and taking short breaks between each KEYWORLD game session, may have introduced a bias in our performance. Specifically, we might expect knowledge of the KNOWER algorithm to raise our expectations for rationality. We also might have shifted priors for the correct key. Recognizing this, we plan for future work within this framework to collect a more diverse and larger group of participants.

Another limitation comes with the concern of our choice of the rationality parameter α, which is designed to reflect WATCHER model's belief preference of KNOWER's ability in optimal planning. In our current framework, α is a static predetermined hyperparameter for each world setting, not adaptive to the changing states within that world. This approach while effective in modeling human preference and environment noise, potentially oversimplifies the process. We believe that a more refined model, incorporating an adaptive rationality parameter that adjusts according to the world's state, could offer a more nuanced understanding and yield results that better model human inference process. In accordance with this, we would also like to explore variants of the KNOWER algorithm that introduce degrees of irrationality. We acknowledge the importance of these aspects and will address them in future works.

Considering the dimensionality, we only explore $K = 1$, which eliminates the exploration of the MCMC sampling hyperparameter N from our study. Other important hyperparameters, including the threshold probability and the number of turns for belief updates, could be studied in a more extensive setting for deeper insights.

Furthermore, the data collection for the KNOWER model is limited to only one run in each KEYWORLD setting. This is mainly due to the constraints in available computational resource and time we have. Considering the amount of calculation required for our Bayesian KNOWER model and the extensive number of KEYWORLD scenarios (1440 in total), playing and collecting data for each game becomes a time-intensive process. We aim to collect more extensive data on KNOWER model to enhance our understanding across various KEYWORLD setting.

Conclusions

In this work, we presented KEYWORLD, a goal inference game that can be used to evaluate resource-rational models on rapid, low-stakes decisions. We connected our game with research in goal inference, resource rationality, and human-robot collaboration, illustrating how it could be used to investigate these areas. We then presented a pilot study, exploring an array of resource-rational model variants and comparing them to human behavioral data. Preliminary findings provide support in favor of models that more carefully allocate their computation, such as only updating their beliefs after observing a substantially incongruous action. While results are limited by both the human participants and the array of models explored, we believe KEYWORLD provides an interesting method for studying computational cognitive science.

Code Availability

Code for the game and model, along with all collected data, is available at https://github.com/amanirmk/key_world.

Author Contributions

Both authors contributed substantially to the codebase: the development of KEYWORLD, the models, the evaluation pipelines, and data analysis. Maina-Kilaas led the development of the game and model evaluation, while Wu led the development of the inference algorithm. Maina-Kilaas came up with the initial idea for the project and both authors contributed to the final paper.

Acknowledgments

We thank Josh Tenenbaum for early conversations and Tony Chen for feedback on the project proposal.

References

Amatya, S., Ghimire, M., Ren, Y., Xu, Z., & Zhang, W. (2022). When shall i estimate your intent? costs and benefits of intent inference in multi-agent interactions. In *2022 american control conference (acc)* (p. 586-592). doi: 10.23919/ACC53348.2022.9867155

Baker, C., Tenenbaum, J., & Saxe, R. (2007, 01). Goal inference as inverse planning. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*.

Callaway, F., Gul, S., Krueger, P., Griffiths, T., & Lieder, F. (2018). Learning to select computations. In R. Silva, A. Globerson, & A. Globerson (Eds.), *34th conference on uncertainty in artificial intelligence 2018, uai 2018* (pp. 776–785). Association For Uncertainty in Artificial Intelligence (AUAI). (Publisher Copyright: © 34th Conference on Uncertainty in Artificial Intelligence 2018. All rights reserved.; 34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018 ; Conference date: 06-08-2018 Through 10-08-2018)

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015). Effects of robot motion on human-robot collaboration. In *2015 10th acm/ieee international conference on human-robot interaction (hri)* (p. 51-58).

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998–998.

Hahn, M., Futrell, R., Levy, R., & Gibson, E. (2022). A resource-rational model of human processing of recursive linguistic structure. *Proceedings of the National Academy of Sciences*, *119*(43), e2122602119. Retrieved from https://www.pnas.org/doi/abs/10.1073/pnas.2122602119 doi: 10.1073/pnas.2122602119

Saulnier, P., Sharlin, E., & Greenberg, S. (2011). Exploring minimal nonverbal interruption in hri. In *2011 ro-man* (p. 79-86). doi: 10.1109/ROMAN.2011.6005257

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. (2023). Large-scale evidence for logarithmic effects of word predictability on reading time. In *36th annual conference on human sentence processing*.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302-319. Retrieved from https://www.sciencedirect.com/science/article/pii/S0010027713000413 doi: https://doi.org/10.1016/j.cognition.2013.02.013

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599-637. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12101 doi: https://doi.org/10.1111/cogs.12101

Zhi-Xuan, T., Mann, J. L., Silver, T., Tenenbaum, J. B., & Mansinghka, V. K. (2020). Online bayesian goal inference for boundedly-rational planning agents. In *Proceedings of the 34th international conference on neural information processing systems*. Red Hook, NY, USA: Curran Associates Inc.