Baby Intuitions Benchmark (BIB): Discerning the goals, preferences, and actions of others

Kanishk Gandhi* **Gala Stojnic** **Brenden M. Lake** **Moira R. Dillon**
New York University

Abstract

To achieve human-like common sense about everyday life, machine learning systems must understand and reason about the goals, preferences, and actions of others. Human infants intuitively achieve such common sense by making inferences about the underlying causes of other agents' actions. Directly informed by research on infant cognition, our benchmark BIB challenges machines to achieve generalizable, common-sense reasoning about other agents like human infants do. As in studies on infant cognition, moreover, we use a violation of expectation paradigm in which machines must predict the plausibility of an agent's behavior given a video sequence, making this benchmark appropriate for direct validation with human infants in future studies. We show that recently proposed, deep-learning-based agency reasoning models fail to show infant-like reasoning, leaving BIB an open challenge.

1. Introduction

Humans have a rich capacity to infer the underlying intentions of others by observing their actions. For example, when we watch the simple animations from Heider and Simmel (1944)'s seminal study (see video[1] and Figure 1), we attribute goals and dispositions to simple 2D figures moving around a flat world. Using behavioral experiments presenting both simple and complex visual displays, developmental cognitive scientists have found that even young infants also infer intentionality in the actions of other agents. Infants expect other agents: to have object-based goals (Gergely et al., 1995; Luo, 2011; Song et al., 2005; Woodward, 1998, 1999; Woodward and Sommerville, 2000); to have goals that reflect preferences (Buresh and Woodward, 2007; Kuhlmeier et al., 2003; Repacholi and Gopnik, 1997); to engage in instrumental actions to bring about goals (Carpenter et al., 2005; Elsner et al., 2007; Gerson et al., 2015; Hernik and Csibra, 2015; Saxe et al., 2007; Woodward and Sommerville, 2000); and to act efficiently towards goals (Colomer et al., 2020; Gergely and Csibra, 1997, 2003; Gergely et al., 1995; Liu et al., 2019, 2017).

Machine-learning and AI systems, in contrast, are much more limited compared even to human infants in their understanding of other agents. One reason this might be the case is that machine learning and AI systems typically treat human behavioral data like any other type of data: Statistical models are trained to predict outcomes of interest (e.g., churn, clicks, likes, etc.) rather than to learn about the goals and preferences that underlie such outcomes. The resulting, impoverished "machine theory of mind"[2] may thus be a critical difference between human and machine intelligence more generally (Lake et al., 2017). Addressing this difference is crucial if machine learning aims to approximate the flexibility of human common sense and reasoning.

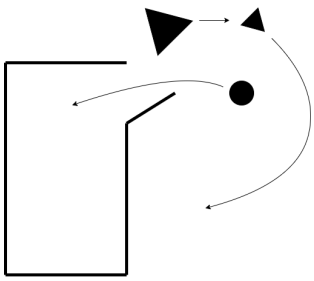


Figure 1: A still from Heider and Simmel (1944). In this animation, the large triangle chases the small triangle and the circle who cooperate to avoid it.

Understanding reasoning about agents has so far received substantially more attention from researchers in cognitive development than in AI. However, recent computational work has aimed to focus on such reasoning by adopting several approaches. Inverse reinforcement learning (Abbeel and Ng, 2004; Ng et al., 2000; Ziebart et al., 2008) and Bayesian approaches (Baker et al., 2011, 2017, 2009; Jara-Ettinger, 2019; Ullman et al., 2009) have modeled other agents as rational, yet noisy, planners. In these models, rationality serves as the tool by which to infer the underlying intentions that best explain an agent's

*Corresponding author: kanishk.gandhi@nyu.edu
Data and code available here: https://www.kanishkgandhi.com/bib
[1]https://www.youtube.com/watch?v=VTNmLt7QX8E

[2]Note that in the cognitive development literature, "theory of mind" typically refers to the attribution of mental states, such as phenomenological or epistemic states (e.g., perceptions or beliefs) to other intentional agents (Premack and Woodruff, 1978). In this paper, we address on only one potential component of theory of mind, present from early infancy, which focuses on reasoning about the intentional states, not the phenomenological or epistemic states, of others (Spelke, 2016)

observed behavior. Game theoretic models have aimed to capture an opponent's thought processes in multi-agent interactive scenarios (see survey: Albrecht and Stone, 2018), and learning-based, neural network approaches have focused on learning predictive models of other agents' latent mental states, either through structured architectures that encourage mental-state representations (Rabinowitz et al., 2018) or through the explicit modeling of other agents' mental states using a different agent's forward model (Raileanu et al., 2018).

Despite the increasing sophistication of these computational models, they have not been evaluated or compared using a comprehensive benchmark that captures early emerging human competencies about agents. For example, some existing evaluations have provided fewer than 100 sample episodes (Baker et al., 2011, 2017, 2009), making it infeasible to evaluate learning-based approaches that require substantial training. Other evaluations have used largely the same distribution for both training and test episodes (Rabinowitz et al., 2018), making it difficult to measure how abstract or flexible a model's performance might be. Moreover, existing evaluations have not used or been translatable to the behavioral paradigms that test infant cognition. They therefore cannot be validated with infants nor can their results be analyzed in terms of the representations and processes that support human performance. AGENT (Shu et al., 2021), a benchmark developed contemporaneously to the one presented here, is inspired by studies with infants and has been validated with behavioral data from adults. Moreover, it challenges machines to reason about the underlying intentions of agents as opposed to their actions. We see AGENT as largely complementary to our efforts, covering a distinct (yet overlapping) set of infant abilities. There are other differences, including the ease of evaluating new models: AGENT involves training on many different leave-out splits, where most splits have relatively minor differences between training and test. In contrast, BIB offers a single canonical split designed to evaluate the abstractness and flexibility of the underlying representations of other agents. Ultimately we hope that new models will be evaluated on both benchmarks, further probing their breadth and sensitivity to design choices.

In this paper, we present a comprehensive benchmark, the Baby Intuitions Benchmark (BIB), which is directly inspired by infant cognition. BIB adapts experimental stimuli from research in developmental cognitive science that has captured the abstract nature of infants' reasoning about agents (Baillargeon et al., 2016; Banaji and Gelman, 2013). Moreover, BIB adopts a "violation of expectation" (VOE) paradigm (similar to Riochet et al. (2018); Smith et al. (2019)), commonly used in behavioral research with infants, which both makes its direct validation with infants possible and also makes its results interpretable in terms of human performance. Finally, we design the BIB training and evaluation sets so that they test for flexible, generalizable common sense reasoning. BIB thus serves as a key step in bridging machines' impoverished understanding of intentionality with humans' rich one.

2. Baby Intuitions Benchmark (BIB)

BIB presents a battery of agency-reasoning tasks, based on findings from developmental cognitive science and adopting its VOE paradigm, to evaluate computational models. We focus on the following five questions: 1) can an AI system represent an agent as having a particular object-based goal? 2) can it bind specific preferences for goal objects to specific agents? 3) can it understand that there may be obstacles that restrict an agent's actions and that an agent will move to a previously nonpreferred object when their preferred object becomes inaccessible? 4) can it represent an agent's sequence of actions as instrumental, directed towards a higher-order goal object? 5) can it learn that an agent acts efficiently towards a goal object?

We also adopt the VOE paradigm, which involves presenting visual stimuli in two phases, a familiarization phase and a test phase. We refer to the two phases together as an "episode." The familiarization phase includes a succession of eight trials that introduce the main elements of the visual displays used in the test phase. This introduction also allows the observer to form expectations about the future behavior of those elements based on their prior knowledge or learning. The test phase includes an unexpected and expected outcome, based on what was observed during familiarization. The unexpected outcome is typically perceptually similar to the events in the familiarization while the expected outcome is typically more perceptually different. So, in order for the outcome to be unexpected, it must be so at the conceptual, rather than perceptual, level. When this paradigm is used with infants, their looking time to each event is measured, and infants tend to look longer at unexpected outcomes, i.e., outcomes that "violate their expectations" (Baillargeon et al., 1985; Oakes, 2010; Turk-Browne et al., 2008).

2.1. Can an AI system represent an agent as having a particular object-based goal?

Developmental Background. Infants attribute object-based—as opposed to location-based—goals to agents (Gergely et al., 1995; Luo, 2011; Song et al., 2005; Woodward, 1998, 1999; Woodward and Sommerville, 2000). As illustrated in Figure 2 (left), Woodward

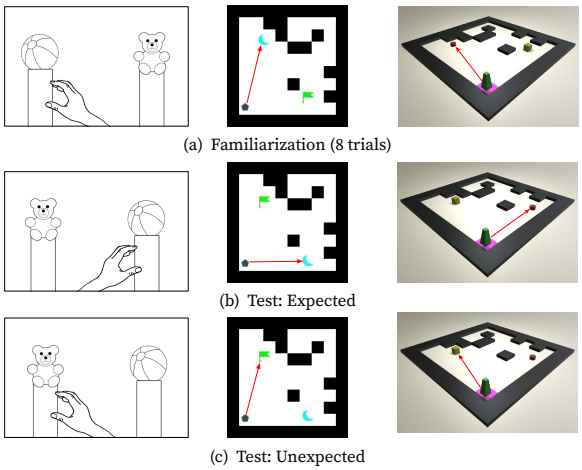


Figure 2: Evaluation of whether machines can represent preferences of agents. Inspired by the Woodward (1998)'s original study with infants (left), our version of the task is rendered in both 2D (middle) and 3D (right). The familiarization trials establish the preference of the agent.

(1998, 1999)'s seminal study showed that when 5- and 9-month-old infants saw a hand repeatedly reaching to a ball on the left over a bear on the right, they then looked longer when the hand reached to the left for the bear, even though the direction of the reach was more similar in that event to the events in the previous trials. These results suggest that the infants expected that the hand would reach consistently to a particular goal object as opposed to a particular goal location. Other studies have shown that infants' interpretations are not restricted to reaching events. For example, infants attribute an object-based goal to a 3D box during a live puppet show when that box seemingly exhibits self-propelled motion. (Luo, 2011; Luo and Baillargeon, 2005; Shimizu and Johnson, 2004). When shown an agent repeatedly moving to the same object at approximately the same location, do AIs, like infants, infer that the agent's goal is the object and not the location?

Familiarization Trials. The familiarization shows an agent repeatedly moving towards a specific object in a world with two objects (Figure 2a right). The agent's starting position is fixed across trials, and the locations of the objects are correlated with their identities such that the preferred object and nonpreferred object appear in generally the same location across trials (see appendix Figure 11 and 12).

Test Trials. The test uses two object locations that had been used during one familiarization trial, but the identity of the objects at those locations has been switched. In the expected outcome (Figure 2b right), the agent moves to the object that had been their goal during the familiarization, i.e., their preferred object, but the trajectory of their motion and the location of that object is different from familiarization. In contrast, in the unexpected outcome (Figure 2c), the agent moves to the nonpreferred object, but the trajectory of their motion and the location they move to is the same as familiarization. The model is successful if it expects the agent to go to the preferred object in a different location.

2.2. Can an AI system bind specific preferences for goal objects to specific agents?

Developmental Background. Infants are capable of attributing specific preferences to specific agents (Buresh and Woodward, 2007; Henderson and Woodward, 2012; Kuhlmeier et al., 2003; Repacholi and Gopnik, 1997). For example, while 9- and 13-month-old infants looked longer at test when an actor reached for a toy that they did not prefer during habituation, infants showed no expectations when the habituation and test trials featured different actors (Buresh and Woodward, 2007). When shown an one agent repeatedly moving to the same object, do AIs, like infants, expect that that object is preferred to that specific agent?

Familiarization Trials. The familiarization shows an agent consistently choosing one object over the other, as above, but objects appear at widely varying locations in the grid world.

Test Trials. The test includes two possible scenarios. One scenario presents an expected outcome, in which the familiar agent goes to the object it prefers, and another outcome, in which a new, unfamiliar agent goes to the object preferred by the familiar agent. While the latter outcome is not necessarily unexpected, the familiar agent going to the preferred object should be more expected given the familiarization (appendix Figure 14). The second scenario presents an unexpected outcome, in which the familiar agent goes to the nonpreferred object, and another outcome, in which the new agent goes to the object not preferred by the familiar agent. Here, the familiar agent going to the nonpreferred object should be more unexpected (Figure 3). The model is successful if it has weak or no expectations about the preferences of the new agent.

2.3. Can an AI system understand that there may be obstacles that restrict an agent's actions and that an agent will move to a previously nonpreferred object when their preferred object becomes inaccessible?

Developmental Background. Infants understand the principle of solidity (e.g., that solid objects cannot pass through one another), and they apply this principle to both inanimate entities (Baillargeon, 1987; Baillargeon

et al., 1992; Spelke et al., 1992) and also animate entities, such as human hands (Luo et al., 2009; Saxe et al., 2006). Infants' expectations about the objects agents might approach are also informed by object accessibility. Scott and Baillargeon (2013) demonstrate, for example, that 16-month-old infants expected an agent, facing two identical objects, to reach for the one in the container without a lid versus the one in the container with a lid.

Familiarization Trials. The familiarization shows an agent consistently choosing one object over the other, as above, and objects appear at widely varying locations in the grid world. (Figure 4).

Test Trials. The test presents two new object locations. In the expected outcome, the preferred object is now inaccessible, blocked on all sides by the fixed, black barriers, and the agent moves to the nonpreferred object. In the unexpected outcome, both of the objects remain accessible, and the agent moves to the nonpreferred object (Figure 4). The model is successful if it expects the agent to move to the nonpreferred object only when the preferred object is inaccessible.

2.4. Can an AI system represent an agent's sequence of actions as instrumental, directed towards a higher-order goal object?

Developmental Background. Infants represent an agent's sequence of actions as instrumental to achieving a higher-order goal (Carpenter et al., 2005; Elsner et al., 2007; Gerson et al., 2015; Hernik and Csibra, 2015; Saxe et al., 2007; Sommerville and Woodward, 2005; Woodward and Sommerville, 2000). For example, Sommerville and Woodward (2005) showed that 12-month-old infants understand an actor's pulling a cloth as a means to getting the otherwise out-of-reach object placed on it. When shown an agent repeatedly taking the same action to effect a change in the environment that enables them to

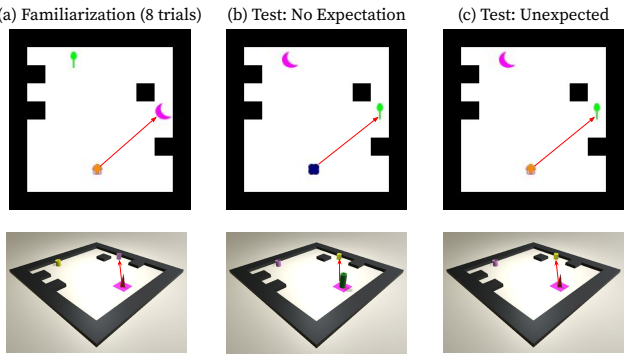


Figure 3: Evaluation of whether machines can bind specific goals to specific agents. The familiarization trials establish the preference of the agent.

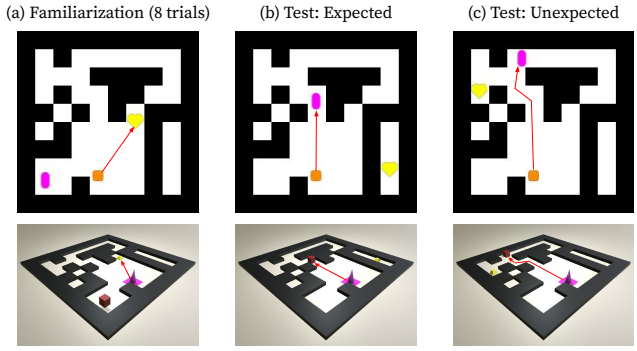


Figure 4: Evaluation of whether machines can understand that obstacles restrict actions. The familiarization trials establish the preference of the agent.

move towards an object, do AIs, like infants, expect that that object is the goal, as opposed to the sequence of actions?

Familiarization Trials. The familiarization includes five main elements: an agent; a goal object; a key; a lock; and a green removable barrier (see Figure 5). The green barrier initially restricts the agent's access to the object. And so, the agent removes the barrier by collecting and then inserting the key into the lock. The agent then moves to the object.

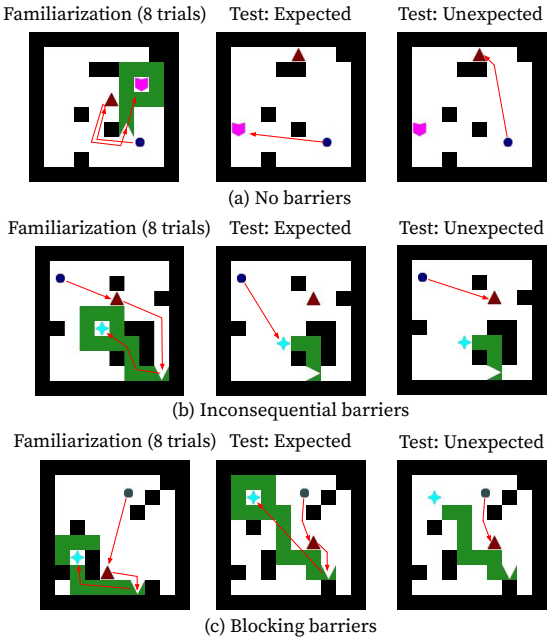


Figure 5: The three types of trials that test machines' understanding of an agent's actions towards a higher-order goal. The goal is initially inaccessible (blocked by a green removable barrier). During familiarization, the agent removes the barrier by retrieving the key (triangle) and inserting it into the lock.

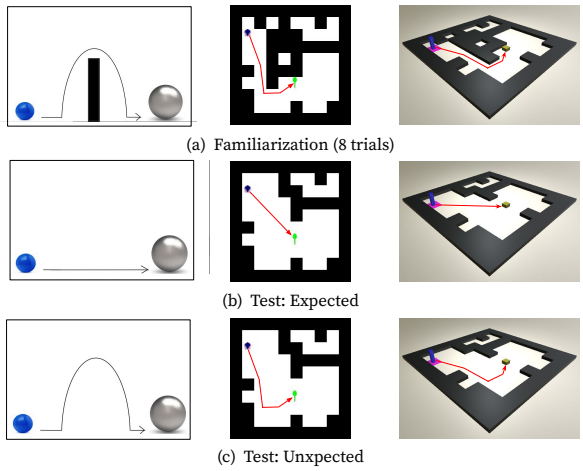


Figure 6: Inspired by Gergely et al. (1995) (left) we ask whether machines expect that agents move efficiently towards goal objects. At test, the agent moves along one of the same paths they moved along during familiarization, but unlike familiarization, there is no barrier between the agent and the object. So, this inefficient action is unexpected.

Test Trials. The test includes three possible scenarios. One scenario presents no green barrier. In the expected outcome, the agent moves directly to the object while in the unexpected outcome the agent moves to the key (Figure 5a). The second scenario presents a green barrier, but it does not restrict the agent's access to the object. In the expected outcome, the agent moves directly to the object while in the unexpected outcome the agent moves to the key (Figure 5b). The third scenario presents an expected outcome, in which the barrier restricts the agent's access to the object and the agent moves to the key. In the unexpected outcome, the barrier does not block the object and the agent goes to the key (Figure 5c). Including these three scenarios allows us to test for simple heuristics that models might use to solve these tasks. If the model uses the heuristic that the key should be visited first and then the object, it will fail on the no barrier and inconsequential barrier scenarios. If the model uses the heuristic that the key should be visited only when a removable barrier is present, then it will fail on the inconsequential barrier scenario. Finally, the heuristic of always going to the object directly will fail on the blocking barrier scenario. The model is successful if it expects the agent to go to the key only when the removable barrier is blocking that object.

2.5. Can an AI system understand that agents act efficiently towards a goal object?

Developmental Background. Infants expect agents to move efficiently towards their goals (Baillargeon et al., 2015; Colomer et al., 2020; Gergely and Csibra, 1997, 2003; Gergely et al., 1995; Liu et al., 2019, 2017). In a seminal study by Gergely et al. (1995), for example, 12-month-old infants repeatedly saw a small circle jumping over an obstacle to get to a big circle (see Figure 6 left). At test, the obstacle was removed, and the small circle either performed the same, now inefficient, action to get to the big circle or performed the straight, now efficient action. Infants were surprised when the agent performed the familiar but inefficient action. These findings have been replicated by instantiating the agent and object in different ways (as, e.g., humans, geometric shapes, or puppets) and by using different kinds of presentations (e.g., prerecorded or live) (Colomer et al., 2020; Liu et al., 2017; Phillips and Wellman, 2005; Sodian et al., 2004; Southgate et al., 2008). When infants see an irrational agent, i.e., one moving inefficiently to their goal from the start, however, they do not form any expectations about that agent's efficient action at test (Gergely et al., 1995; Liu and Spelke, 2017). When shown a rational agent repeatedly taking an efficient path around a barrier to its goal object, do AIs, like infants, expect that that agent will continue to take efficient paths as opposed to similar-looking paths, once that barrier is removed?

Familiarization Trials. The familiarization includes two different scenarios. In one scenario, a rational agent consistently moves along an efficient path to its goal object around a fixed, black barrier in the gird world (Figure 6a). In the other scenario, an irrational agent moves along these same paths, but there is no barrier in the way. So in this latter scenario, the irrational agent is acting inefficiently from the start (Figure 7).

Test Trials. The test includes two possible scenarios. One scenario shows only the rational, efficient agent during familiarization, and at test, it presents one of the familiarization trials but with the barrier between the agent and the goal object removed. In the expected outcome, the agent moves along a straight, efficient path to its goal. In the unexpected outcome, the agent either

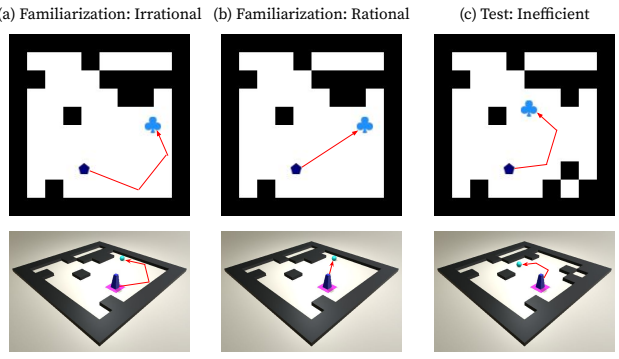


Figure 7: Inspired by Gergely et al. (1995), we ask whether machines expect either rational or irrational agents to move efficiently towards their goals.

moves along the exact same, but now inefficient, path that it had during familiarization (path control), Figure 6) or along a path that is inefficient but takes the same amount of time as the efficient path (in this latter case, the goal object is closer to the agent, appendix Figure 13). As in the original studies with infants Gergely et al. (1995), these path/timing variations focus the solution on efficiency as opposed to other variables that often correlate with efficiency.

The second scenario shows either the rational or irrational agent during familiarization, and at test, it presents that agent taking an inefficient path towards its goal (Figure 7). This outcome should be unexpected in the case of the rational agent, but should yield no expectation in the case of the irrational agent. The model is successful if it expects only a rational agent to modify its path based on the presence or absence of barriers and move efficiently to its goal. Moreover, a model that ignores the familiarization phase during which the rationality of the agent is established will fail.

2.6. Generating the Evaluations

Inspired by Heider and Simmel (1944), the primary set of visual stimuli present "grid-world" animations, shown from an overhead perspective and populated with simple shapes that take on different roles (e.g. "agents", "objects", "tools"), and we assume the environment is fully observable to the agent (i.e., the agent can see over the walls) and the observer. We chose this type of environment as particularly suitable for testing AIs (e.g., Baker et al., 2017; Rabinowitz et al., 2018) because it allows for procedural generation of a large number of episodes, and the simple visuals focus the problem on reasoning about agents.

For each of the five evaluation tasks, we generated 1000 episodes, each with one expected and one unexpected outcome (2000 videos), by sampling the locations of barriers, agents, and objects in the 10×10 grid. The locations are controlled to account for the distances and obstacles between the agent and the objects so that, e.g., preferred objects are not consistently closer or farther from agents. We provide two evaluation sets, one with objects and agents seen during background training and the other with new shapes for the objects and agents. Finally, as a means to vary the perceptual difficulty of the benchmark, we also include 3D versions of the stimuli rendered to match the 2D versions and presented at a three-quarters point of view (Figure 2).The 2D stimuli (except for the instrumental action tasks) are directly translated to 3D using the AI2THOR (Kolve et al., 2019) framework. For both 2D and 3D videos, we provide scene configuration files describing the objects and agents present in the scene.

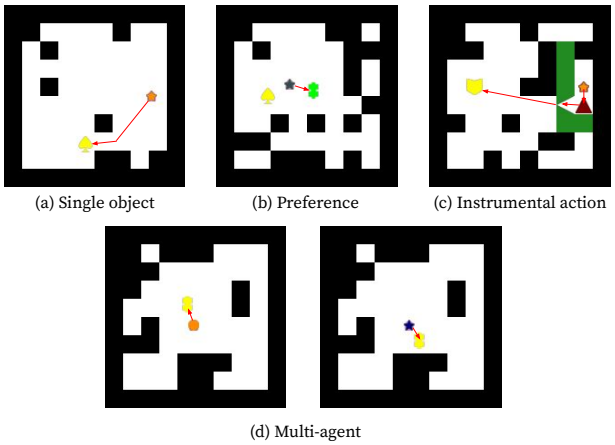


Figure 8: The four tasks from the background training set. Only the test trials are shown here.

3. Background Training

We provide a set of background training tasks for the models to learn about agents and objects in our grid worlds and the structure of the trials. Although we provide a training set, we do not intend to limit models to just these data prior to being tested. Additional out-of-distribution training data is allowed, just as infants get varied experience with agents in the real world. Importantly, when participating in a lab study, infants can make meaningful inferences about novel stimuli/environments with only a relatively brief familiarization phase. We include tens of thousands of background episodes as a generous stand-in for this type of in-lab familiarization so AI systems are not surprised merely by the various elements and dynamics used in the evaluation. Although learning-centric approaches will learn something about other agents if trained on the background set, we do not intend it to be sufficient for acquiring genuine, abstract agent representations. We intend that either supplemental pretraining or additional prior knowledge can be enriched by the background training to approach the benchmark successfully.

The episodes in the background training are structured similarly to those in the evaluation, although the familiarization and test trials are now drawn from the same distribution within each episode. Similar to IntPhys (Riochet et al., 2018) and ADEPT (Smith et al., 2019), we only provide the expected outcomes during training. There are four training tasks:

Single Object Task. The agent navigates to an object at some varied location in the scene (Figure 8a). This task is different from the evaluation task in that it presents only a single object. With this training, models can learn how agents start and end trials, how agents move, and how barriers influence agent motion. We provide 10,000 episodes of this type.

No-Navigation Preference Task. Two objects are located very close to the agent's starting location, and the agent approaches one object consistently across trials (Figure 8b). The task allows the model to learn that agents have preferences. Critically, the navigation in these trials is trivial compared to the evaluation trials, so navigation to goal objects is not trained. We provide 10,000 episodes of this type.

No-Preference, Multiple-Agent Task. One object is located very close to the agent's initial starting location (Figure 8d). At some point during the episode, a new agent takes the initial agent's place (for example, the initial agent could be replaced at the fourth trial and all subsequent trials would have the new agent). The task allows the model to learn that multiple agents can appear across trials, but this task differs from the evaluations, in which the new agent appears only in the test trials. We provide 4,000 episodes of this type.

Agent-Blocked Instrumental Action Task. The agent starts confined to a small region of the grid world, blocked by a removable green barrier (Figure 8c). The agent collects a key and inserts it into a lock to make the barrier disappear. The agent then navigates to the object. This task allows the model to learn that the green barrier obstructs navigation and how the key and lock remove that barrier. These trials differ from the evaluation in that the removable barriers are around the agent instead of the object. We provide 4,000 episodes trials of this type.

To be successful at the evaluations, models must acquire or enrich their representations of agents for flexible and systematic generalization. For example, models have to combine acquired knowledge of navigation (Single Object Task) and agent preferences (No-Navigation Preference Task) to be successful at the first evaluation testing the underlying preferences guiding agents' goal-directed actions (section 2.1).

4. Baseline Models

The baseline models are variants of a state-of-the-art, neural-network approach to reasoning about agents: the theory of mind net (ToMnet) model in Rabinowitz et al. (2018). These models are trained passively and through observation only. We use a self-supervised learning setup where the objective is to predict the future actions of the agent. During evaluation, the expectedness of a test trial, in the context of the previous familiarization trials, is defined by its error on the most 'unexpected' video frame (frame with the highest error).

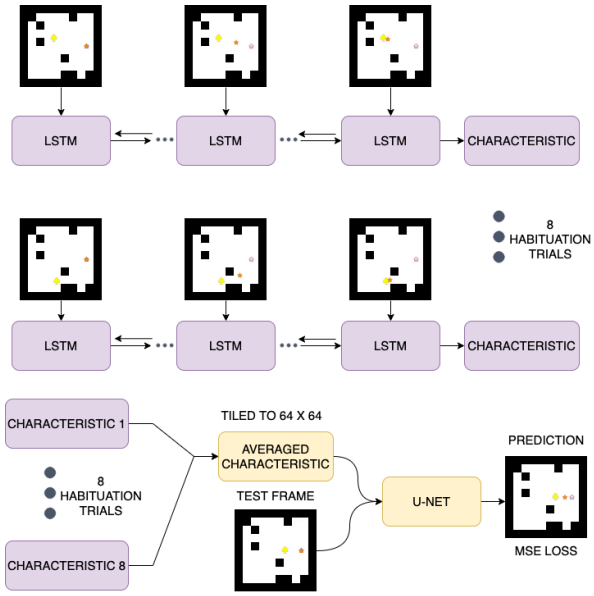


Figure 9: Architecture of the video baseline model inspired by Rabinowitz et al. (2018). An agent-characteristic embedding is inferred from the familiarization trials using a recurrent net. This embedding, with the state at test time, is used to predict the next frame of the video using a U-Net (Ronneberger et al., 2015).

We test two baseline models (see appendix B for full model specifications), one that operates directly on the videos and another that operates on the mask representations of the elements (i.e., individual elements – agents, objects, etc. — in a scene are split into different channels). The objective of the mask model (see appendix Figure 17) is to predict the trajectory of the agent in the test trial (see appendix B.1).

The video model (see Figure 9) operates on videos sampled at 3 fps and resized to 64×64. Each frame in each familiarization trial is encoded using a convolutional neural network. The frame embeddings in a trial are passed to a bidirectional LSTM. The last output embdedding of the LSTM represents the characteristic of the agent in the trial. These embeddings are averaged across familiarization to obtain a characteristic embedding for an agent. The characteristic embedding is tiled to a 64×64 spatial resolution, concatenated to a frame from the test trial, and passed through a U-net to predict the next frame in the trial. A mean squared error loss is used to train the network.

5. Results

The models were trained on 80% of the background training episodes (training set), and the rest of the episodes were used for validation (validation set). A comparison of the MSE loss on the training and validation sets and a qualitative evaluation of the video model's performance indicates that it learned the training tasks successfully

BIB Agency Task	Mask		Video		Video (New Shapes)		3D Video	
	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.	Rel.	Abs.
Preference	69.0	69.0	47.8	47.6	47.4	47.8	49.2	48.3
Multi-Agent	50.0	49.8	50.3	50.3	50.0	51.5	50.0	51.0
Inaccessible Goal	50.7	52.4	66.0	61.4	61.7	60.9	40.0	43.2
Efficiency: Path control	95.6	94.3	99.8	92.0	98.5	92.1	66.3	57.9
Efficiency: Time control	94.8	91.4	99.9	90.1	96.9	90.3	75.4	61.8
Efficiency: Irrational agent	50.0	50.0	50.0	50.0	47.8	49.5	50.0	50.0
Efficient Action Average	72.6	69.9	74.9	70.3	72.7	70.0	62.9	55.0
Instrumental: No barrier	98.2	98.4	99.7	94.0	93.0	88.1	-	-
Instrumental: Inconsequential barrier	89.5	83.0	76.7	57.8	66.0	56.0	-	-
Instrumental: Blocking barrier	77.3	56.2	58.2	57.5	59.7	58.0	-	-
Instrumental Action Average	85.6	71.8	73.0	56.9	69.6	55.8	-	-

Table 1: Performance of the baseline models on BIB. Scores for the mask model on 2D videos, the video model on the 2D videos, 2D videos with new elements, and 3D videos are shown above. Relative accuracy (Rel.) scores quantify pairwise VOE judgements. Absolute scores (Abs.) quantify VOE judgements on each video independently, requiring the prediction error to be lower on the expected videos. The absolute score is the Area Under the ROC Curve (AUC), where the true positive rate is plotted against the false positive rate for different threshold values.

(see appendix Figure 15).

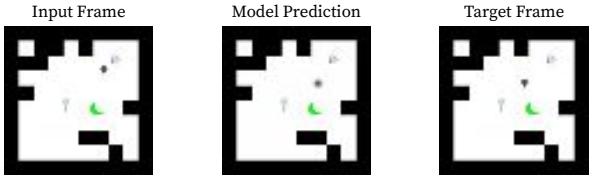
For each evaluation episode, we first calculated the model's relative accuracy, i.e., whether the model found the expected video in each pair more expected than the unexpected video (chance is 50%). We also calculated the model's absolute score, i.e., the model's prediction of each video's plausibility independent of the pairing. This is measured by the Area Under the ROC Curve (AUC), which plots true positive rates against the false positive rate for different threshold values.

The results of our baseline models are presented in Table 1. The video model performs at chance on the Preference Task (see Figure 10a for predictions made by the video model); it tends to predict that an agent will go to the closer object (this prediction is made in about 70% of trials). The model thus neglects the agent's preference, established during familiarization. This is particularly striking because the model does take into account the familiarization phase when succeeding in the No-Navigation Preference Task in the background training.

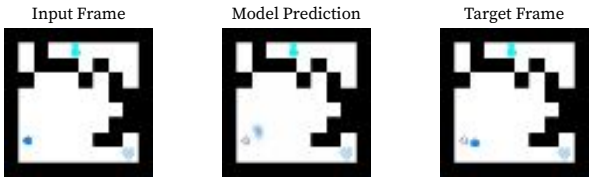
The video model also fails on the Multi-Agent Task, again tending to predict that an agent will go to the closer object regardless of any established preferences. Consistent with this failure, the model also fails to map specific preferences to specific agents.

This model does slightly better than chance on the Inaccessible Goal Task. As seen in Figure 10b, it still nevertheless, frequently predicts that the agent will go to the inaccessible goal.

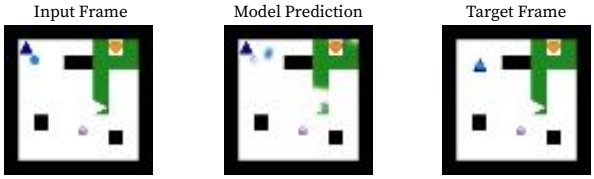
The video model is proficient at finding the shortest path to the goal in the Efficiency Task (appendix Figure 19a), leading to high accuracy on both sub-evaluations that test for efficient action: Path Control and Time Control (Table 1). However, the model fails



(a) Preference Task: The model predicts that the brown agent would go to the green object instead of the established preference of the grey object.



(b) Inaccessible goal task: The model predicts that the blue agent would head to the inaccessible cyan object.



(c) Instrumental action task C: The model predicts that the blue agent would directly go to the inaccessible orange object goal instead of performing the instrumental action by first collecting the triangular key.

Figure 10: The most surprising frame (the frame with the highest prediction error) from the test trial for the video model taken from the evaluation tasks. Failure cases are shown here.

to modulate its predictions based on whether the agent was rational or irrational during familiarization (Table 1).

Finally, the video model performs above chance on the Instrumental Action Task, but performance on the sub-evaluations (Table 1) indicates that it relies on the simple heuristic of directly going to the goal object rather than understanding the nature of the instrumental action (Figure 10c). This leads to higher scores on sub-evaluations with no barrier and an inconsequential barrier (Table 1) but lower ones on the sub-evaluation with a blocking barrier. This poor performance may be due to the difference between the agent and barrier conditions in the background training (where the agent is confined; Figure 8c) and evaluation (where the object is confined; Figure 5).

The mask model shows similar performance to the video model across the tasks (see appendix B) for a detailed analysis).

Moreover, when we replace the elements in the evaluation set with new ones, the video model scores fall slightly, but the trends remain the same (Table 1). Finally, the video model performs similarly on the 3D videos of the tasks, although performance is generally worse overall with 3D videos. This is likely because perceiving the trajectories of agents in 3D is more difficult for a predictive model in pixel space. The predictive networks trained with MSE find it challenging to model trajectories in depth.

6. General Discussion

In this paper we introduced the Baby Intuitions Benchmark (BIB), which tests machines on their ability to reason about the underlying intentionality of other agents by observing only agents' actions. BIB is directly inspired by the abstract reasoning about agents that emerges early in human development, as revealed by behavioral studies with infants. BIB's adoption of the VOE paradigm, moreover, means its results can be interpreted in terms of human performance and makes it appropriate for direct validation with human infants in future studies.

While baseline, deep-learning models successfully generalize to BIB's training tasks, they fail to systematically generalize to the evaluation tasks even though the models incorporate theory-of-mind-inspired architectures (Rabinowitz et al., 2018). In particular, the baseline models performed at about chance when required to reason that agents have preferred goal objects, that preferences are tied to specific agents, and that goal objects can be physically inaccessible. When presented with instrumental actions, moreover, the models succeeded only by relying on a simple heuristic of going directly to the goal object, rather than on a more sophisticated understanding of an agent's sequence of actions. Finally, the models failed to modulate their predictions about efficient action for irrational versus rational agents. These results suggest that state-of-the-art AI models do not have a common-sense understanding of agents the way human infants do.

BIB is rooted in the findings and methods of developmental cognitive science, but there are still critical differences between its stimuli and the stimuli used with infants, and its particular tasks have not yet been validated with infants. First, while the simplicity of the grid-world environment, for example, makes it ideal for procedural generation to test AIs, such displays may not be compelling enough to engage infants' intuitions about agents, and overhead, object-directed navigation events may not be the most intuitive context in which to engage infants' representations of other agents (in contrast to, e.g., perspectival reaching events). Can infants reason about agents' actions when viewing them from an overhead perspective? Can infants recognize simple shapes with simple movements and minimal cues to animacy (e.g., no eyes/gaze direction, no distinctive sounds, and no emotional expressions) as agents with intentionality? Most of the existing infant literature off of which BIB is based presents infants with richer cues to animacy and in the form of live-action or animated displays from a frontal or three-quarters points of view. Second, some of the variability introduced across the evaluation videos may make it difficult for infants to track and stably represent the different elements. For example, the location of the preferred object varies greatly during the familiarization phase in the evaluation that links specific agents to specific preferences. No study with infants, to our knowledge, has shown that infants succeed in predicting an agent's goal-directed actions under these conditions. Third, some inferences about agents included in this benchmark are yet to be tested with infants. For example, no study to our knowledge has examined whether infants expect agents to move towards a nonpreferred object, versus not move at all, when a preferred object is inaccessible. And, no study has examined whether infants expect a goal object in a two-alternative forced-choice scenario to generalize across agents when infants are familiarized to both agents both moving to the same object when there is only that one object present. Finally, the "extended familiarization" needed for training AI models (i.e., the background training), reveals a striking difference between how BIB might challenge minds versus machines. While both infants and AIs may have built-in knowledge and/or pretraining (e.g., from infants' everyday experience or from AIs' simulated experience), infants may need to

watch only eight, as opposed to thousands, of videos of shapes moving around grid worlds to successfully apply their reasoning about agents to new, test events presented in that medium.

The origins and development of human, intuitive understanding of agents and their intentional actions have been studied extensively in developmental cognitive science. The representations and computations underlying such understanding, however, are not yet understood. BIB serves as a test for computational models with different priors and learning-based approaches to achieve the common-sense reasoning about agents that human infants have. A computational description of how we reason about agents could ultimately help us build machines that better understand us and that we better understand.

Finally, BIB serves as a key step in bridging machines' impoverished understanding of intentionality with humans' rich one, since intentionality is one key component to understanding and reasoning about others in terms of their underlying mental states, including their beliefs and desires. A benchmark that focuses on reasoning about agents' intentional states, as well as their phenomenological and epistemic states, such as false-beliefs (a litmus test of human theory of mind (e.g. Baron-Cohen et al. (1985); Leslie (1987)), is thus a natural extension of BIB and could further advance our understanding of both human and artificial intelligence.

Acknowledgements

This worked was supported by the DARPA Machine Common Sense program (HR001119S0005). We thank Victoria Romero, Koleen McKrink, David Moore, Lisa Oakes, Clark Dorman, and Amir Tamrakar for their generous feedback. We are especially grateful to Thomas Schellenberg, Dean Wetherby, and Brian Pippin for their development effort in porting the benchmark to 3D.

References

Abbeel, P. and Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the 21st International Conference on Machine learning*, page 1.

Albrecht, S. V. and Stone, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, 258:66–95.

Baillargeon, R. (1987). Object permanence in 31/2-and 41/2-month-old infants. *Developmental psychology*, 23(5):655.

Baillargeon, R., Needham, A., and DeVos, J. (1992). The development of young infants' intuitions about support. *Early development and parenting*, 1(2):69–78.

Baillargeon, R., Scott, R. M., and Bian, L. (2016). Psychological reasoning in infancy. *Annual review of psychology*, 67:159–186.

Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K.-s., Wu, D., and Bian, L. (2015). *Psychological and sociomoral reasoning in infancy.* American Psychological Association.

Baillargeon, R., Spelke, E. S., and Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition*, 20(3):191–208.

Baker, C., Saxe, R., and Tenenbaum, J. (2011). Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society*, volume 33.

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):1–10.

Baker, C. L., Saxe, R., and Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113(3):329–349.

Banaji, M. R. and Gelman, S. A. (2013). *Navigating the social world: What infants, children, and other species can teach us.* Oxford University Press.

Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21(1):37–46.

Buresh, J. S. and Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, 104(2):287–314.

Carpenter, M., Call, J., and Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental science*, 8(1):F13–F20.

Colomer, M., Bas, J., and Sebastian-Galles, N. (2020). Efficiency as a principle for social preferences in infancy. *Journal of Experimental Child Psychology*, 194:104823.

Elsner, B., Hauf, P., and Aschersleben, G. (2007). Imitating step by step: A detailed analysis of 9-to 15-month-olds' reproduction of a three-step action sequence. *Infant Behavior and Development*, 30(2):325–335.

Gergely, G. and Csibra, G. (1997). Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to premack and premack. *Cognition*, 63(2):227–233.

Gergely, G. and Csibra, G. (2003). Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, 7(7):287–292.

Gergely, G., Nádasdy, Z., Csibra, G., and Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165–193.

Gerson, S. A., Mahajan, N., Sommerville, J. A., Matz, L., and Woodward, A. L. (2015). Shifting goals: Effects of active and observational experience on infants' understanding of higher order goals. *Frontiers in Psychology*, 6:310.

Heider, F. and Simmel, M. (1944). An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259.

Henderson, A. M. and Woodward, A. L. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental science*, 15(5):641–652.

Hernik, M. and Csibra, G. (2015). Infants learn enduring functions of novel tools from action demonstrations. *Journal of experimental child psychology*, 130:176–192.

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

Kolve, E., Mottaghi, R., Han, W., VanderBilt, E., Weihs, L., Herrasti, A., Gordon, D., Zhu, Y., Gupta, A., and Farhadi, A. (2019). Ai2-thor: An interactive 3d environment for visual ai.

Kuhlmeier, V., Wynn, K., and Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological science*, 14(5):402–408.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and brain sciences*, 40.

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

Liu, S., Brooks, N. B., and Spelke, E. S. (2019). Origins of the concepts cause, cost, and goal in prereaching infants. *Proceedings of the National Academy of Sciences*, 116(36):17747–17752.

Liu, S. and Spelke, E. S. (2017). Six-month-old infants expect agents to minimize the cost of their actions. *Cognition*, 160:35–42.

Liu, S., Ullman, T. D., Tenenbaum, J. B., and Spelke, E. S. (2017). Ten-month-old infants infer the value of goals from the costs of actions. *Science*, 358(6366):1038–1041.

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental science*, 14(2):453–460.

Luo, Y. and Baillargeon, R. (2005). Can a self-propelled box have a goal? psychological reasoning in 5-month-old infants. *Psychological Science*, 16(8):601–608.

Luo, Y., Kaufman, L., and Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive psychology*, 58(4):441–486.

Ng, A. Y., Russell, S. J., et al. (2000). Algorithms for inverse reinforcement learning. In *Proceedings of the 17th International Conference on Machine learning*, volume 1, page 2.

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, 11(3):255–268.

Phillips, A. T. and Wellman, H. M. (2005). Infants' understanding of object-directed action. *Cognition*, 98(2):137–155.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, 1(4):515–526.

Rabinowitz, N., Perbet, F., Song, F., Zhang, C., Eslami, S. M. A., and Botvinick, M. (2018). Machine theory of mind. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4218–4227, Stockholmsmässan, Stockholm Sweden. PMLR.

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

Repacholi, B. M. and Gopnik, A. (1997). Early reasoning about desires: evidence from 14-and 18-month-olds. *Developmental psychology*, 33(1):12.

Riochet, R., Castro, M. Y., Bernard, M., Lerer, A., Fergus, R., Izard, V., and Dupoux, E. (2018). Intphys: A framework and benchmark for visual intuitive physics reasoning. *CoRR*, abs/1803.07616.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Saxe, R., Tzelnic, T., and Carey, S. (2006). Five-month-old infants know humans are solid, like inanimate objects. *Cognition*, 101(1):B1–B8.

Saxe, R., Tzelnic, T., and Carey, S. (2007). Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction. *Developmental psychology*, 43(1):149.

Scott, R. M. and Baillargeon, R. (2013). Do infants really expect agents to act efficiently? a critical test of the rationality principle. *Psychological science*, 24(4):466–474.

Shimizu, Y. A. and Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental science*, 7(4):425–430.

Shu, T., Bhandwaldar, A., Gan, C., Smith, K., Liu, S., Gutfreund, D., Spelke, E., Tenenbaum, J. B., and Ullman, T. D. (2021). AGENT: A Benchmark for Core Psychological Reasoning. *arXiv preprint arXiv:2102*.

Smith, K., Mei, L., Yao, S., Wu, J., Spelke, E., Tenenbaum, J., and Ullman, T. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32*, pages 8985–8995. Curran Associates, Inc.

Sodian, B., Schoeppner, B., and Metz, U. (2004). Do infants apply the principle of rational action to human

agents? *Infant Behavior and Development*, 27(1):31–41.

Sommerville, J. A. and Woodward, A. L. (2005). Pulling out the intentional structure of action: the relation between action processing and action production in infancy. *Cognition*, 95(1):1–30.

Song, H.-j., Baillargeon, R., and Fisher, C. (2005). Can infants attribute to an agent a disposition to perform a particular action? *Cognition*, 98(2):B45–B55.

Southgate, V., Johnson, M., and Csibra, G. (2008). Infants attribute goals to biomechanically impossible actions. *Cognition*, 107(3):1059–1069.

Spelke, E. S. (2016). Core knowledge and conceptual change. *Core knowledge and conceptual change*, 279:279–300.

Spelke, E. S., Breinlinger, K., Macomber, J., and Jacobson, K. (1992). Origins of knowledge. *Psychological review*, 99(4):605.

Turk-Browne, N. B., Scholl, B. J., and Chun, M. M. (2008). Babies and brains: habituation in infant cognition and functional neuroimaging. *Frontiers in human neuroscience*, 2:16.

Ullman, T., Baker, C., Macindoe, O., Evans, O., Goodman, N., and Tenenbaum, J. B. (2009). Help or hinder: Bayesian models of social goal inference. In *Advances in neural information processing systems*, pages 1874–1882.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, 69(1):1–34.

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant behavior and development*, 22(2):145–160.

Woodward, A. L. and Sommerville, J. A. (2000). Twelve-month-old infants interpret action in context. *Psychological Science*, 11(1):73–77.

Ziebart, B. D., Maas, A. L., Bagnell, J. A., and Dey, A. K. (2008). Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA.

A. Data Specifications

Each video has a resolution of 200 x 200 at 25 fps (the videos can be converted to a higher resolution if required). In addition to the videos, we provide metadata in the form of json files describing every frame in the video. This description contains information about the layout of the scene and the objects present.

Each video has a json file associated with it. A video has 9 trials which correspond to the 9 items in the json file. These 9 trials have a variable number of frames. Each frame is described by the objects contained in it.

These include:

- The 'size' attribute specifies the resolution of the frame.
- The 'walls' attribute has a list of [bottomleft, extent] attributes describing the barriers. The bottomleft attribute is 2-dimensional and is defined by an x and y coordinate. Similarly, the extent for each wall is 2-dimensional and describes the width and height of the wall.
- The 'objects' attribute is defined as a list of attributes [bottomleft, size, image, color]. The bottomleft attribute is 2 dimensional and is defined by an x and y coordinate. The size is the half of the side of the square shape that the image of the object would be resized to. So, if the size is 10, an object image of size 100x100 would be resized to 20x20. The image attribute gives the path of the object image. The color attribute gives the color of the object in RGB format in the range [0, 255].
- The 'home', 'agents', 'key' and 'lock' attributes have a similar structure to the objects attribute.
- The 'fuse' attribute corresponds to the removable barrier and has a similar structure to the 'walls' attribute.

B. Baseline Details

B.1. Mask Model

Model Description. Each trial is represented in the form of its initial state and the trajectory taken by the agent (see Figure 17). The states and trajectories are approximated to a grid of size 10×10. The initial state is approximated from the frame in the form of a downsampled representation of size $10 \times 10 \times |O|$, where $|O|$ represents the number of possible elements in the scene. These include target objects (14), agents (5), walls (1), home (1), key (1), lock (1) and removable barriers (1) with a total of 24 possible objects in the environment. The trajectory of the agent for a trial is provided in the form of a flat 10×10 grid where the cells visited by the agent have a value of 1 while the rest are 0.

The objective of the model is to predict the trajectory of the agent in the test trial conditioned on the initial state of the trial and the eight familiarization trials, presented in the form of initial state and agent trajectory pairs. To encode a trial, the trajectory is concatenated with every channel of the state representation and passed through a two convolutional layers (3×3, 2 output channels, with batchnorm (BN) and residual connections). The outputs of this network are concatenated and passed through another convolutional neural network (1×1, 24 output channels, BN $\rightarrow$ 3×3, 24, BN $\rightarrow$ 3×3, 24, BN with residual connections), flattened and passed through a fully connected layer to get an agent characteristic embedding for the trial (1×8). The trial embeddings from the eight familiarization trials are averaged to get an agent characteristic embedding (1×8). This embedding is spatialised (tiled) to a 10×10 grid ($10 \times 10 \times 8$) and concatenated to the initial state representation of the test trial and passed through a fully convolutional network (Residual Net [3×3; 32; BN] $\times 4$ + Sigmoid) to predict the trajectory of the agent ($10 \times 10 \times 1$).

A binary cross entropy objective with a focal loss is used to train the agent. Although multiple plausible trajectories exist for a trial and no one trajectory is the 'right' one, we expect that a model can learn reasonable expectations about agent behavior. We train the model with an Adam optimizer with a learning rate of 1e-4 (betas=(0.9, 0.999)) for 21 epochs.

Training Tasks. The performance of the mask model on the background training tasks is shown in appendix Table 2 and appendix Figure 15. For the mask model, each grid cell is treated as a separate binary classification problem (if the agent will visit the cell or not). We compute the precision and recall for these binary classification problems. For the preference task, we also analyse if the model predicts that the agent will visit the cell of the object goal. The model predicts the cell of the preferred object 83.2%, the cell of the less preferred object 6.9%, of both objects 8.4% and no object 2.4% of the times. We see that the model successfully generalizes to the training tasks.

Evaluation Results. The performance of the mask model can be seen in Table 1 and appendix Figure 16. The mask model quickly learns to find the shortest path between the agent and the object. It fails on the multi-agent, inaccessible goal and the efficient action task with an irrational agent. The model does not have different expectations for the preferences of the new agent and

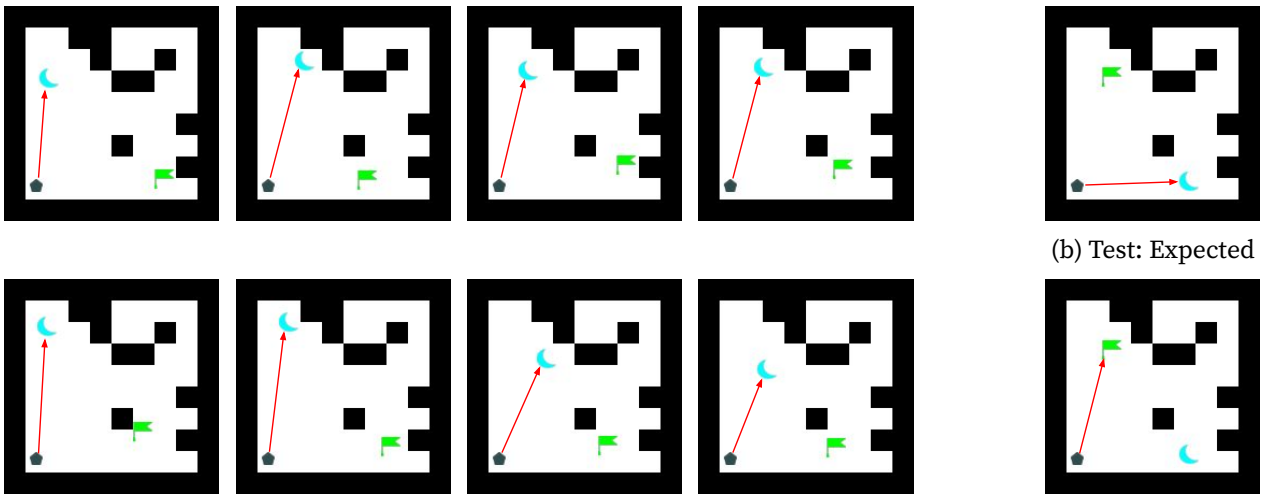


Figure 11: Evaluation task to test if machines can represent preferences of agents. 2D versions of the stimuli are shown here.

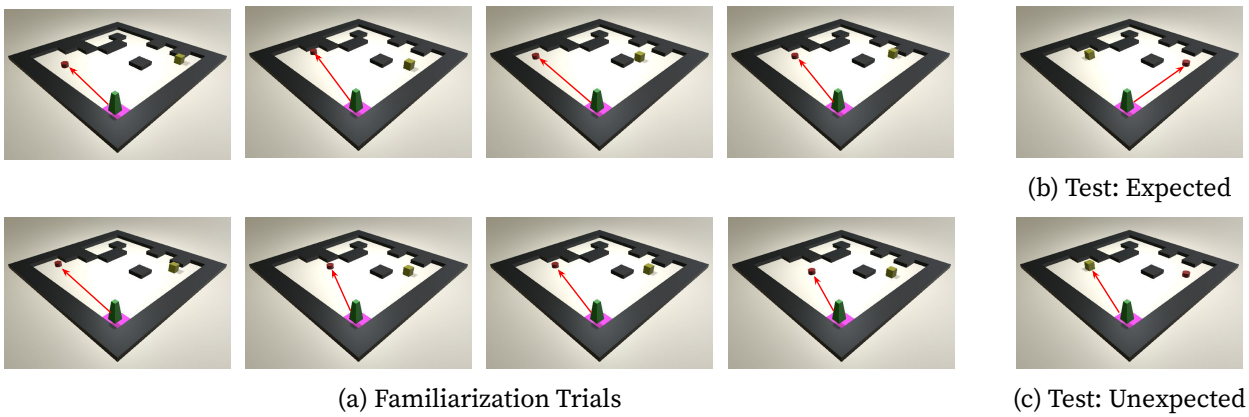


Figure 12: Evaluation task to test to test if machines can represent preferences of agents. 3D versions of the stimuli are shown here.

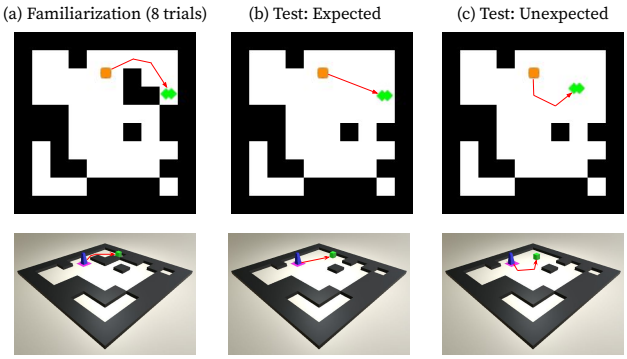


Figure 13: We draw inspiration from Gergely et al. (1995) to design an equivalent task to test if machines can understand if agents act efficiently towards their goals. In this task, the time taken by the agent to reach the goal in the expected and unexpected cases is the same.

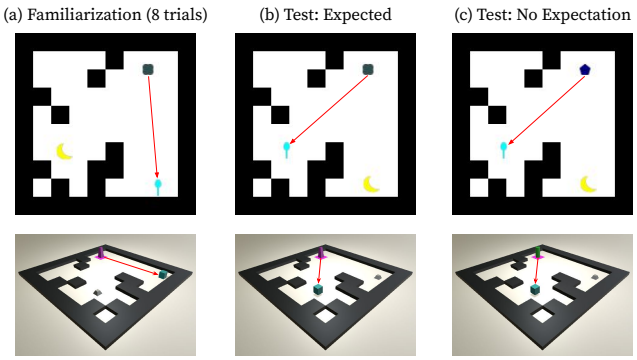
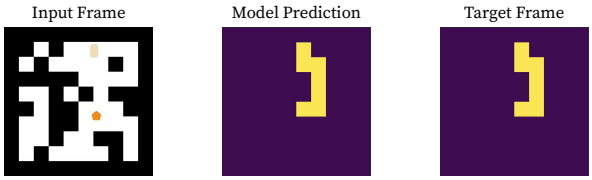
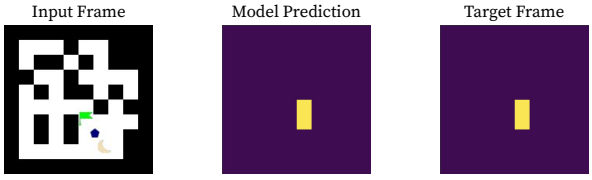


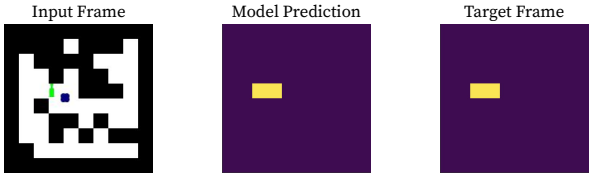
Figure 14: Evaluation of binding specific preferences to specific agents. The familiarization trials establish the preference of the agent.



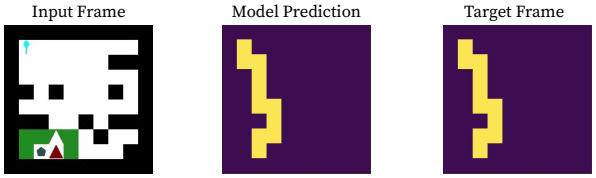
(a) Background Single Object Task: The model correctly predicts that the orange agent will go around the barriers to reach the beige object goal.



(b) Background No-Navigation Preference Task: The model correctly predicts that the blue agent will go to the preferred green object goal.



(c) Background No Preference Multi-Agent Task: The model predicts that the blue agent will go the object goal in the trial.



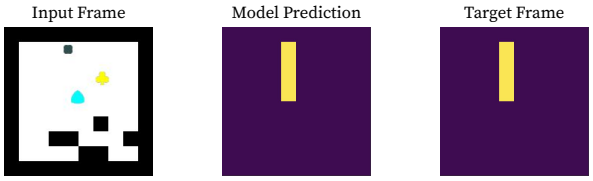
(d) Background Agent-Blocked Instrumental Action Task: The model correctly predicts the locations visited by the agent to perform the instrumental action and visit the object goal (with the caveat that the model does not have the capacity to understand the sequence in which the cells in the grid are visited).

Figure 15: Agent trajectory predictions on the background training set in the test trial made by the model working on abstract mask representations. Test trials are shown here.

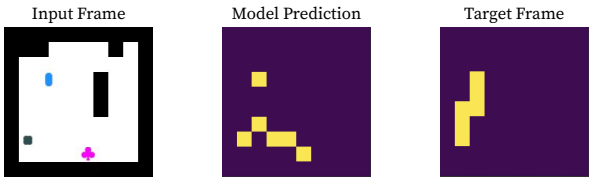
makes the same predictions as those for the familiar agent. For the inaccessible goal task, the model predicts that the agent will go to both objects in the test trial (with the trajectory blocked by the obstacle around the goal)(appendix Figure 16d). The model performs better than chance on the preference task but frequently predicts that the agent will go to both objects in the scene (see appendix Figure 16b). As the mask model tries to predict the complete trajectory of the agent in a trial (ignoring the sequence of the actions), it solves a weaker proxy of the instrumental action task, achieving a score higher than the video model.

B.2. Video Model

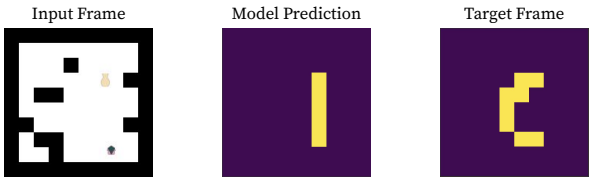
Model Description. In the video model, the frames of a familiarization trial are encoded using a residual convolutional network with 4 blocks, each with two



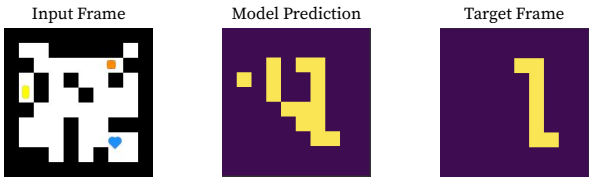
(a) Preference task: The model correctly predicts that the dark grey agent will go to the preferred cyan object (established in the familiarization)



(b) Preference task: The model predicts a trajectory going to the wrong magenta object but also highlights the blue preferred object. This shows a case of failure.



(c) Efficient action task: A successful case is shown here where the model predicts that the agent will take the shortest path to the beige object goal. The target frame here is from the unexpected episode.



(d) Inaccessible goal task: A failure case is shown here where the model predicts that the orange agent will go to the less preferred blue object and also to the preferred yellow object but the trajectory is blocked by the walls.

Figure 16: Agent trajectory predictions on the evaluation set in the test trial made by the model working on mask representations.

BIB Task	Precision	Recall
Single object	0.88	0.67
Preference	0.92	0.57
Multi-agent	0.97	0.56
Instrumental action	0.89	0.74

Table 2: The performance of the mask model on the background training tasks.

3×3 convolutional operations with 16 feature maps. This is followed by a 1×1 convolutional layer to map the 16 feature maps to one map. This representation is flattened and passed sequentially to a bi-directional LSTM. The output from the last timestep is used as the agent characteristic representation of size 1×16 for the trial (see Figure 9). The characteristic embedding across the 8 familiarization trials is averaged to get a final agent characteristic embeddding. This embedding is tiled to get a vector of size $64 \times 64 \times 16$ and concatenated to the current frame from the test trial. This vector of size $64 \times 64 \times 19$ is passed to a U-Net (Ronneberger et al., 2015) to predict the next frame. We train the model with an Adam optimizer with a learning rate of 1e-4 (betas=(0.9, 0.999)). We train the 2D video model for 11 epochs and the 3D model for 10 epochs.

Background Training. The errors on the validation set for the model are shown in appendix Table 3. Some of the predictions made by the model can be seen in Figure 18. Only the preference task requires the model to take the familiarization phase into consideration.

Evaluation Tasks. The model fails to reliably understand the preference of the agent. This could be a result of differences in the distance at which the objects are placed in the scene. In the background training, the objects are placed close (section 3) to the agent, making the length of the familiarization trials short. The characteristic encoder LSTM might find it difficult to extract characteristcs from longer sequences that are seen in the evaluation tasks.

The model learns the simple heuristic of always going to the object in the instrumental action task. This could be caused due to a difference in the distribution of the background training and evaluation tasks. In the background training task (Figure 8c), the agent is confined in a small space within green removable barriers with the key and the lock. The number of samples where the model has to predict that the agent goes to the key or the lock is relatively small compared to that of the barriers disappearing and the agent moving towards the object goal. In the evaluation tasks (Figure 5c), the number of steps to reach the key and the lock are significantly higher (as the object goal is confined in the removable barriers). The model thus has trouble generalizing to this case (Table 1 Instrumental: Blocking barriers task).

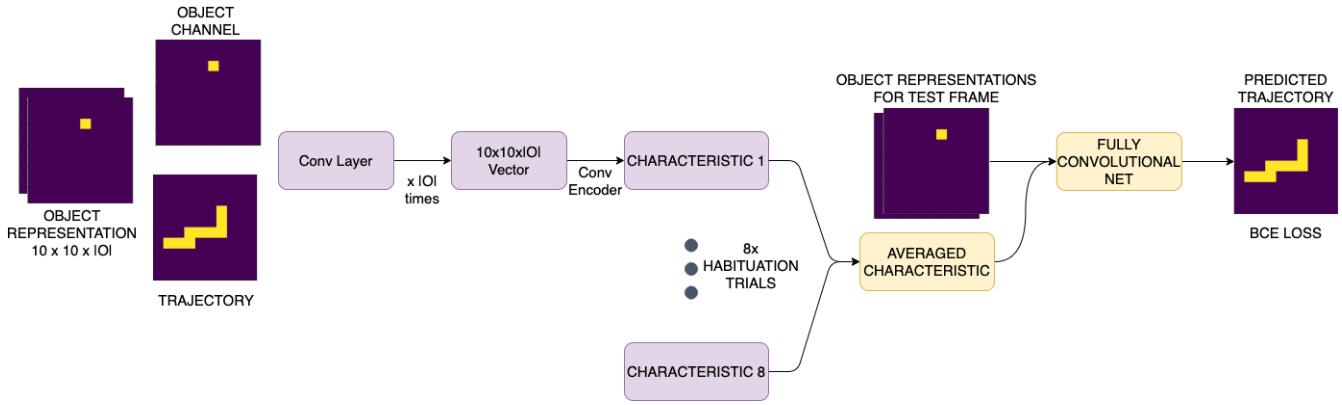
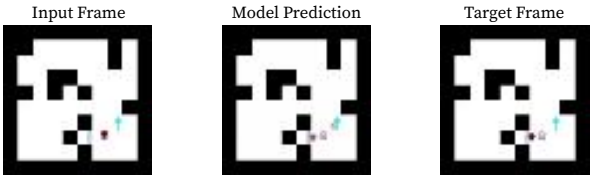


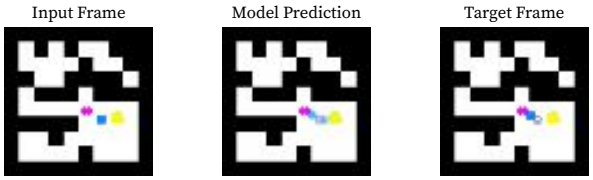
Figure 17: Architecture of our baseline model working on abstract mask representations inspired from Rabinowitz et al. (2018). The objective of the model is to predict the trajectory of the agent.

BIB Task	MSE
Single object	3.3×10^{-4}
Preference	5.4×10^{-4}
Multi-agent	2.4×10^{-4}
Instrumental action	9×10^{-4}

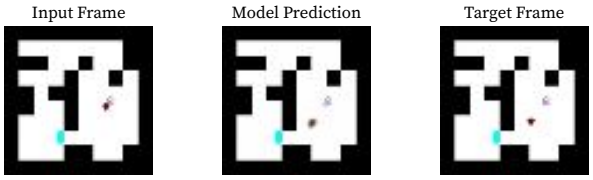
Table 3: The performance of the video model on the 2D background training tasks.



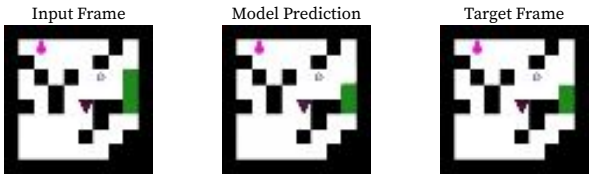
(a) A trial from the training set where the model predicts that thebrown agent will go to the preferred (established in the familiarization) grey object.



(b) A trial from the training set where the model predicts that the blue agent will go to the preferred magenta object (established in the familiarization). We see that there is blurred blue prediction close to the yellow object but the model thinks that it is more likely that the agent will go to the magenta one.

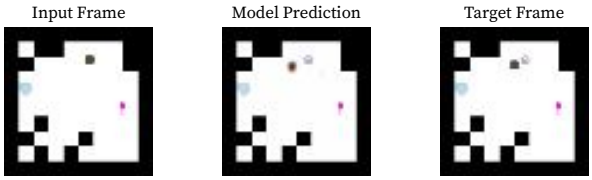


(c) The model correctly predicts that the agent will take the shortest path to go to the object goal.

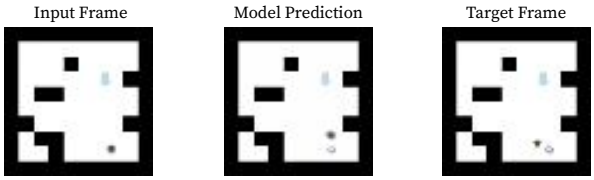


(d) The model correctly predicts that in the instrumental action task, when the key is inserted into the lock, the removable barriers will slowly disappear.

Figure 18: Predictions of the video model on the background training tasks. (a) and (b) show model predictions for two preference trials where the model splits its predictions between the two objects but thinks that going to the preferred object (established during the familiarization phase) is more likely. (c) shows model predictions for the single object task where the model predicts that the agent will take the shortest path to the object. (d) shows the instrumental action task where the model predicts the disappearance of the removable barriers. Test trials are shown here.



(a) Preference Task: The model correctly predicts that the brown agent will go to the preferred object that has been established during the familiarization (gray heart).



(b) Efficient action task: The model correctly predicts that the brown agent will take the shortest path to go towards the object goal. The target frame from the unexpected trial is shown above.

Figure 19: The most unexpected frame (the frame with the highest prediction error) from the test trial for the video model taken from the evaluation tasks. Successful examples shown here.