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

Kanishk Gandhi[a] Gala Stojnic[a] Brenden M. Lake[a] Moira R. Dillon[a]

Abstract— **Human infants intuitively make rich inferences about the goals, preferences, and actions of other agents in the environment. To achieve human-like common sense about everyday life, artificial intelligence and machine learning systems must also understand and reason about such intentionality in other agents. Directly informed by research on infant cognition, we present the Baby Intuitions Benchmark (BIB), which challenges machines to achieve generalizable, common-sense reasoning about other agents like human infants do. We show that while deep-learning-based agency reasoning models fail on BIB, 11-month old infants tested in a pilot validation succeed. This interdisciplinary investigation both suggests a critical gap in machines' common-sense reasoning and also provides new insight into the abstractness and generalizability of infants' knowledge.**

I. 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 & Simmel's (1944) [21] seminal study (see video[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 make rich inferences about the intentions underlying other agents' actions. For example, infants expect agents: to have object-based goals [19, 30, 48, 51, 52, 53]; to have goals that reflect preferences [37, 26, 10]; to engage in instrumental actions to bring about goals [11, 14, 23, 20, 40, 53]; and to act efficiently towards goals [19, 17, 18, 28, 29, 12].

Artificial Intelligence (AI) systems, in contrast, are much more limited compared even to human infants in their understanding of other agents. AI systems typically aim to predict actions of interest (e.g., churn, clicks, likes, etc.) rather than to learn about the goals and preferences that underlie those actions. Addressing this difference is crucial if research in machine learning and artificial intelligence aims to approximate the flexibility of human common-sense reasoning about agents [27].

Common-sense reasoning about agents has thus recently been the focus of AI research relying on diverse approaches: inverse reinforcement learning [32, 1, 56]; Bayesian models [50, 8, 6, 7, 24], game theoretic models [see survey: 2]; and learning-based neural network models [34, 35]. Despite the increasing interest in this research area, such models have not been evaluated or compared using a comprehensive benchmark that captures early emerging human knowledge about agents.[2]. In this paper, we thus present the Baby Intuitions Benchmark (BIB) to evaluate machines' reasoning about agents. By presenting a canonical split between training and evaluation sets, BIB tests for flexible, generalizable common-sense reasoning. BIB adapts experimental stimuli from research in developmental cognitive science that has captured the abstract nature of infants' knowledge [3, 9]. Moreover, BIB adopts a "violation of expectation" (VOE) paradigm (similar to Riochet et al. (2018) [38] and Smith et al. (2019) [46]), 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 provide pilot validation data on BIB with human infants. BIB thus serves as a key step in bridging machines' impoverished understanding of the intentions that drive other agents' actions with humans' rich understanding.

II. Baby Intuitions Benchmark

BIB presents a set of agency-reasoning tasks for AI based on findings from developmental cognitive science and adopting its VOE paradigm. 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 may 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? and 5) can it learn that an agent acts efficiently towards a goal object?

Following the rationale of the VOE paradigm, each of the BIB tasks consists of a familiarization phase and a test phase. The familiarization phase includes a succession of eight trials that introduce the main elements of the visual displays used in the test phase. Familiarization allows the observer to form expectations about the future behavior of those elements. The test phase includes an expected and unexpected outcome based on what was observed during familiarization. The expected outcome is typically perceptually dissimilar to the events in the familiarization while the unexpected outcome is

[a] New York University

[1] https://www.youtube.com/watch?v=VTNmLt7QX8E

[2] AGENT [43], 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 merely their actions. We see AGENT as complementary to our efforts

typically perceptually similar. So, in order for the outcome to be unexpected, it must be so at the conceptual, rather than perceptual, level. When the VOE paradigm is used with infants, their looking time to each event is measured; longer looking time is interpreted as reflecting an infant's surprise by an outcome that ”violates her expectations” [5, 49, 33][3].

BIB's stimuli include a set of animated videos of a ”grid-world” environment, which is presented from an overhead perspective and is populated with simple shapes that take on different roles (e.g. “agents,” “objects,” “tools”). We assume the the environment is fully observable to the agent (i.e., the agent can see over the walls) and the viewer. We chose this type of environment as particularly suitable for testing AIs [e.g., 7, 34] because it allows for procedural generation of a large number of episodes, and the simple visuals focus the problem on reasoning about agents.

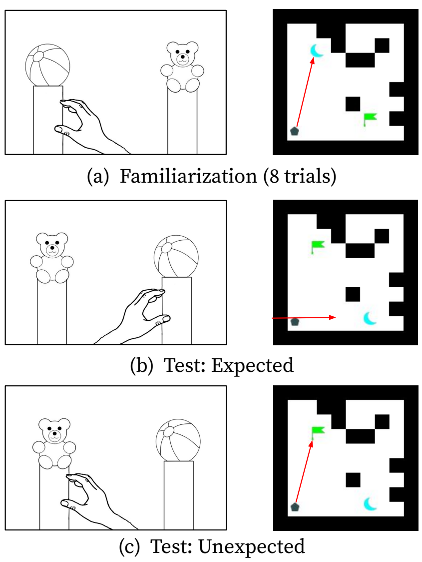


Fig. 1: BIB's Preference Task. Inspired by the Woodward et al.'s [51] original study with infants (left), our version of the task (right) displays an agent moving to the same object in approximately the same location in a grid world across eight familiarization trials. In the expected test trial, the agent moves to the preferred object in a new location, and in the unexpected test trial, the agent moves to the nonpreferred object in a familiar location.

A. Preference Task.

Infants attribute object-based (rather than location-based) preferences and goals to agents [19, 30, 48, 51, 52, 53]. As illustrated in Figure 1 (left), Woodward's [51] 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. For BIB, the familiarization presents an agent repeatedly moving towards a specific object in a world with two objects. 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. 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. There are two possible outcomes: the agent moves to the object that had been their goal during the familiarization (expected); or, the agent moves to the nonpreferred object (unexpected). The model is successful if it expects the agent to go to the preferred object in a different location (see Figure 1 (right)).

[3]It should be noted, however, that interpreting infants' looking behavior as measured by VOE has been a matter of debates among developmental scientists ([13, 31]) as infants might sometime look longer to the stimulus both because they detect a conceptual violation in it, but also because they prefer to look at perceptually more familiar events (e.g. [41])

B. Multi-Agent Task.

Infants attribute specific preferences to specific agents [10, 22, 26, 37]. The familiarization presents an agent consistently choosing one object over the other, but objects appear at widely varying locations in the grid world . The test includes four possible outcomes: the same agent approaches the preferred object (expected); a new agent approaches the object preferred by the first agent (no expectation); the same agent approaches the nonpreferred object (unexpected); or the new agent approaches the nonpreferred object (no expectation). The model is successful if it has weak or no expectations about the preferences of the new agent (see Figure 2).

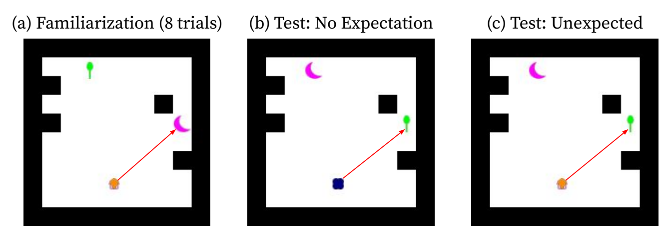


Fig. 2: Evaluation of whether machines can bind specific goals to specific agents. One agent shoes a consistent preference for one of the two objects during familiarization. This should lead to a strong expectation about that agent's preference at test but should lead to no or weak expectations about another agent.

C. Inaccessible Goal Task

. Infants' expectations about what object an agent is likely to approach may depend on the accessibility of that object within a particular environment, e.g., if it is blocked by a physical barrier [42]. The familiarization presents an agent consistently choosing one object over the other, and objects appear at widely varying locations in the grid world. The test presents two new object locations and two possible outcomes: the preferred object is now inaccessible, blocked on all sides by fixed, black barriers, and the agent moves to the nonpreferred object (expected); or, both of the objects remain accessible, and the agent moves to the nonpreferred object (unexpected). The model is successful if it expects the agent to move to the nonpreferred object only when the preferred object is inaccessible (see Figure 3).

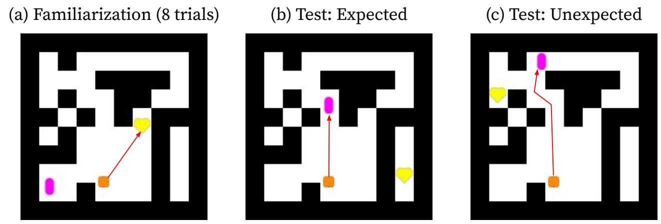


Fig. 3: Evaluation of whether machines can understand that obstacles restrict actions and that an agent might move to a nonpreferred object when their preferred object is inaccessible.

D. Instrumental Action Task.

Infants represent an agent's sequence of actions as instrumental to achieving a higher-order goal [11, 14, 23, 20, 40, 47, 53]. The familiarization includes five main elements: an agent; a goal object; a key; a lock; and a green removable barrier. The green barrier initially restricts the agent's access to the object. The agent removes the barrier by collecting and then inserting the key into the lock. The agent then moves to the object. The test phase presents three different scenarios for a total of six different outcomes. In the scenario with no green barrier: the agent moves directly to the object (expected); or to the key (unexpected). In the scenario with an inconsequential green barrier: the agent moves directly to the object (expected); or to the key (unexpected). In the scenario with variability in the presence/absence of the green barrier: the barrier blocks the agent's access to the object, and the agent moves to the key (expected); or, the barrier does not block the object and the agent goes to the key (unexpected). The model is successful if it expects the agent to go to the key only when the green removable barrier is blocking that object (see Figure 4).

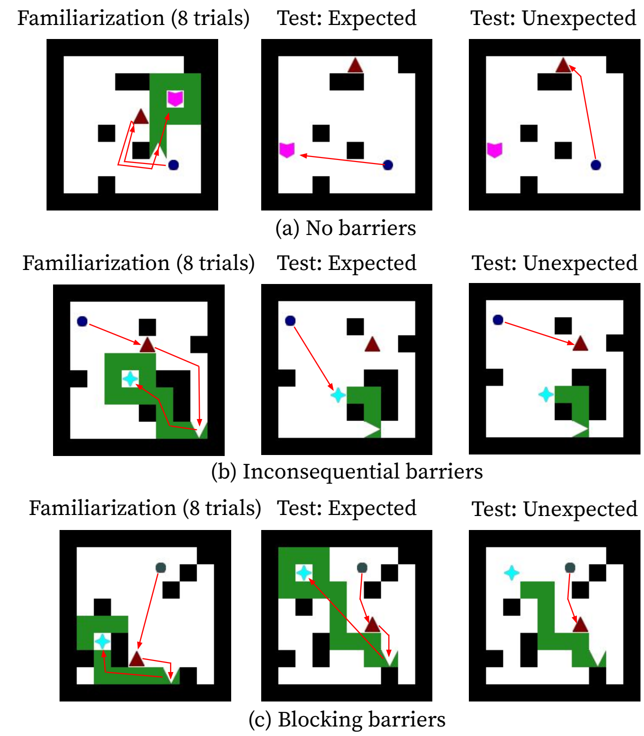


Fig. 4: The three types of test trials evaluate machines' understanding of an agent's instrumental actions towards a higher-order goal. The goal is initially inaccessible (blocked by a green removable barrier). During familiarization, the agent removes the barrier by retrieving the key (triangle) and inserting it into the lock. At test, it is expected that the agent should move directly to the goal when it is accessible.

E. Efficiency Task.

Infants expect agents to move efficiently to their goals and to modify their paths to goals based on the presence or absence of obstacles [4, 19, 17, 18, 29, 28, 12].In a seminal study by Gergely et al. [19], for example, 12-month-old infants repeatedly saw a small circle jumping over an obstacle to get to a big circle (see Figure 5 left). For BIB, the familiarization includes two different scenarios: a rational agent consistently moves along an efficient path to its goal object around a fixed black barrier in the gird world; or, an irrational agent moves along these same paths as the rational agent, but there is no barrier in the way. The test includes two possible scenarios. One scenario shows only the rational agent, and it presents one of the familiarization trials but with the barrier between the agent and the goal object removed. The agent either moves along an efficient path to its goal (expected) or the agent moves along the exact same, but now inefficient, path that it had during familiarization (path control) 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 starts off closer to the agent). The second scenario shows either the rational or irrational agent taking an inefficient path towards its goal. 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 (5 right).

III. Baseline Models' Performance on BIB

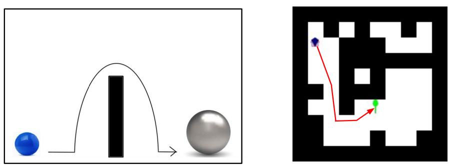
A. Background Training

We provide a set of background training tasks for the models to learn about the grid worlds, their elements, and the structure of the trials. Importantly, when participating in VOE 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.

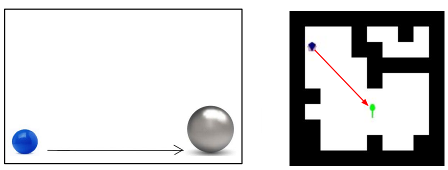
The episodes in the background training are structured similarly to those in the evaluation, although the familiarization and test trials in the background training are drawn from the same distribution within each episode. Similar to IntPhys [38] and ADEPT [46], we only provide the expected outcomes during training. There are four training tasks (see Figure 6): Single Object Task; No-Navigation Preference Task; No-Preference No-Navigation Multi-Agent Task; and Agent-Blocked Instrumental Action Task.

TABLE I: Performance of the baseline models on BIB. The scores quantify pairwise VOE judgements.

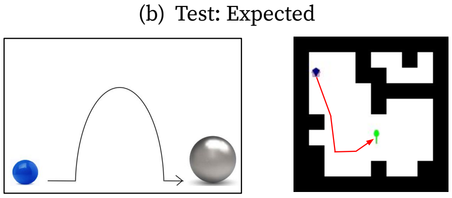
BIB Agency Task	BC-MLP	BC-RNN	Video-RNN
Preference	26.3	48.3	47.6
Multi-Agent	48.7	48.2	50.3
Inaccessible Goal	53.1	46.6	66.0
Efficiency: Path control	96.0	95.8	99.8
Efficiency: Time control	99.1	99.1	99.9
Efficiency: Irrational agent	73.4	48.8	50.0
Efficient Action Average	85.5	73.1	74.9
Instrumental: No barrier	98.8	98.8	99.7
Instrumental: Inconsequential barrier	56.7	78.2	76.7
Instrumental: Blocking barrier	48.2	55.9	58.2
Instrumental Action Average	67.9	77.6	78.2



(a) Familiarization (8 trials)

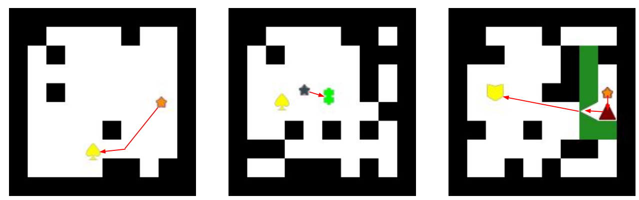


(b) Test: Expected

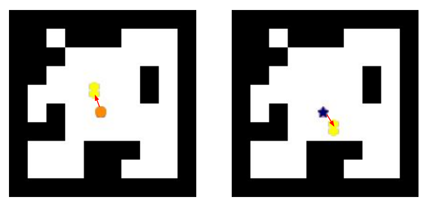


(c) Test: Unexpected

Fig. 5: Inspired by Gergely et al. (1995) [19] (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.



(a) Single object (b) Preference (c) Instrumental action



(d) Multi-agent

Fig. 6: The four tasks from the background training set. Only the test trials are shown here.

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 Preference Task, which evaluates the underlying preferences guiding agents' goal-directed navigation.

B. Baseline Models

When being evaluated on BIB, a model cannot actively sample from the environment; it can only use the samples provided in the dataset. We therefore did not test baseline models using traditional approaches in imitation learning (IL), inverse RL (IRL), and RL [32, 1, 56] because they require substantial privileged information such as access to the environment to actively sample trajectories using the modelled policy and an observable reward for RL algorithms. Moreover, these approaches often model one agent at a time, and BIB requires the same model to infer the behavior of different agents across different episodes (although recent approaches in deep RL and IRL try to mitigate this latter issue with work in meta-RL and meta-IRL [54, 55, 36] that allows for similar cross-episode adaptation). Although this feature of BIB makes it less suitable for testing RL models, it is essential to BIB's design because it reflects infants' understanding and reasoning. Infants rely on little to no active interaction with a particular environment to make meaningful inferences and predictions about the agents in that environment, and infants' inferences are far more abstract than their particular observed or active experience [45, 16, 28, 57].

We thus tested two baseline models including video modeling and behavior cloning (BC). Models were trained passively and through observation only. Our baseline models either predict the next frame in the video (see Figure 7 for architecture) or the actions taken by the agent. To encode the context in the form of the familiarization trials, we use a sequence of frames (for the video model) and frame-action pairs (for the BC models). In terms of the architecture, the baseline models take inspiration from a state-of-the-art, neural-network-based approach to encode the characteristic of an agent: the theory of mind net (ToMnet) model in Rabinowitz et al. (2018) [34]. We encode the familiarization trials as context either using a bidirectional LSTM or an MLP. In addition to video modeling and BC, we also try an offline-RL baseline [44].

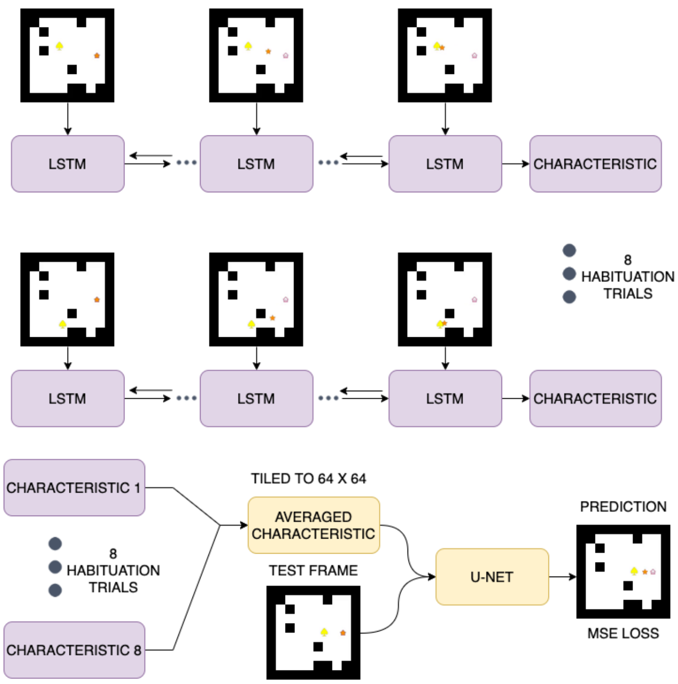
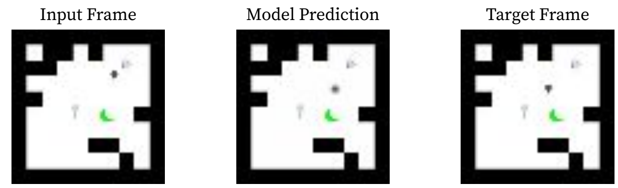


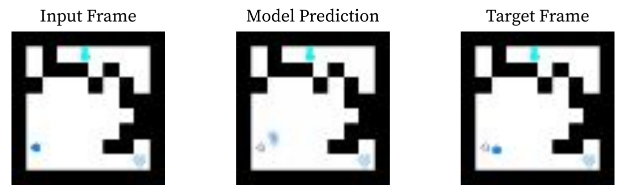
Fig. 7: Architecture of the video baseline model inspired by Rabinowitz et al. (2018) [34]. An agent-characteristic embedding is inferred from the familiarization trials using a recurrent net. This embedding, with the state at test time, is used to predict the next frame of the video using a U-Net [39].

C. Results

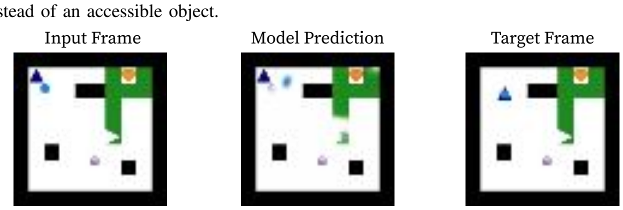
The models with an RNN perform at chance on the Preference Task (see Figure 8a for predictions made by the video model); they tend to predict that an agent will go to the closer object (this prediction is made in about 70% of trials). The model thus neglects the agent's preference established during familiarization. This is striking because the model does take into account the familiarization phase when succeeding in the No-Navigation Preference Task in the background training. This difference could result from differences in the distance at which the objects are placed in the scene. In the background training, the objects are close to the agent, and familiarization trial lengths are short. The characteristic encoder RNN might find it difficult to generalize to the longer sequences seen in the evaluation tasks. The BC-MLP model is confused by how the object locations correlate with their identity, encoding an agent's preference for location instead of objects. This is surprising



(a) The model predicts that the brown agent will go to the green object instead of the grey object, its preferred object goal during familiarization.



(b) The model predicts that the blue agent will move to the inaccessible cyan object instead of an accessible object.



(c) The model predicts that the blue agent will directly go to the inaccessible orange object instead of performing the instrumental action to first collect the triangular key.

Fig. 8: 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.

as the background training provides evidence that agents prefer object identities, not specific locations.

The models also fail 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 models also fail to map specific preferences to specific agents. The models do slightly better than chance on the Inaccessible Goal Task. As seen in Figure 8b, the video model still, nevertheless, frequently predicts that the agent will go to the inaccessible goal. The models are proficient at finding the shortest path to the goal in the Efficiency Task (appendix Figure 16a), leading to high accuracy on both sub-evaluations that test for efficient action: Path Control and Time Control (Table I). However, the RNN-based-models fail to modulate their predictions based on whether the agent was rational or irrational during familiarization (Table I). In contrast, BC-MLP has a weaker expectation of rationality from an irrational agent, scoring 73.4% on this task. Finally, the models perform above chance on the Instrumental Action Task, but performance on the sub-evaluations (Table I) indicate that they rely on the simple heuristic of directly going to the goal object rather than understanding the nature of the instrumental action (Figure 8c). This leads to higher scores on sub-evaluations with no barrier and an inconsequential barrier (Table I) 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 6c) and evaluation (where the object is confined; Figure 4).

IV. Pilot Validation of BIB with 11-month-old Infants

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 in past research. In addition to allowing for allowing for the direct comparison of machine and human performance, validating BIB with infants addresses new questions about agency-reasoning in infancy. For example: 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 as agents with intentionality; can infants predict an agent's goal-directed navigation when the location of a preferred object varies greatly during a familiarization phase; and can infants expect agents to move towards a nonpreferred object, versus not move at all, when a preferred object is inaccessible?

In our pilot validation, we focused on two of BIB's subtasks: the Preference Task and the Efficiency Task. We focused on these tasks both because there is a rich literature suggesting infants' success on similar tasks [3, 15] and also because state-of-the-art computational models do not show rich, infant-like agency reasoning on these tasks. This pilot is a first step towards a a comprehensive evaluation of all the BIB tasks with infants.

A. Participants and Methods.

Twenty-six 11-month-old infants (12 = female, *M*age = 11.12 months, SD = 0.38) were tested on both the Preference and the Efficiency Tasks with half of the infants receiving each task first. Two infants completed only the Preference Task, and two infants completed only the Efficiency Task, leading to the total of 24 infants per task. Infants were tested online using the Zoom teleconferencing system, and they sat on their parents' laps or in highchairs facing the computer screen. An experimenter, naïve to when the test trials were presented and to the order of the test trial outcomes, live-coded infants' looking times using the PyHab-online software [25]. PyHab-online was set up to control the presentation directly from a parent's browser via slides.com. The stimuli were the videos chosen directly from the large set of videos used to evaluate the baseline models, but they were presented at a slower pace, more appropriate for infant viewing. Expected and unexpected outcomes were also presented sequentially after the eight familiarization trials. Each trial lasted for a maximum of 60 seconds and was introduced by a five-second "attention grabber" (a swirling colorful blob accompanied by a chiming sound) to focus infants' looking to the screen. The video froze after the agent reached an object, and the last frame remained on the screen until infants looked away for two consecutive seconds or for the total duration of the trial (i.e., 60 seconds). Which of the two objects the agent preferred in the Preference Task as well as the order of the test trial outcomes in both tasks was counterbalanced across infants. Figure 9 shows stills from an infant's testing session.

B. Results.

TABLE II: Infants' mean looking times on the Preference and the Efficiency Tasks

	Preference Task		**Efficiency Task**	
	expected	unexpected	expected	unexpected
Valid	24	24	24	24
Missing	2	2	2	2
Mean	5.260	8.503	7.964	12.470
Std. Error of Mean	0.677	1.402	1.083	1.709

Infants' raw looking times to the test trials on the Preference and Efficiency Tasks are summarized in Appendix, Table II. The main analysis used one linear mixed effects model for each task with Expectancy (expected vs. unexpected) as a fixed effect factor and Participant as a random effects intercept. Looking time was counted from the start of the video.[4] Our analysis revealed that infants looked longer to the unexpected test trials in both the Preference Task ($\beta = 3.24, p = .040$) and the Efficiency Task ($\beta = 4.50, p = .016$)(see Figure 10). We also analyzed infants' raw looking times using a linear mixed effects model with Task (Preference vs. Efficiency) and Expectancy (expected vs. unexpected) as fixed effect factors and Participant as a random effects intercept. This analysis revealed longer looking at the unexpected vs. expected outcomes ($\beta = 4.506, p = .007$), longer looking to the Efficiency vs. Preference Task ($\beta = -2.934, p = .07$; likely because the action sequences were longer in the Efficiency Task), and critically, no interaction between Task and Expectancy ($\beta = -1.263, p = .585$). These results suggest that infants succeed on the very same highly abstract agency-reasoning tasks that state-of-the-art computational models fail on.

V. 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.

[4]Because we included only the path-matched version of the Efficiency Task, where the length of the action sequence in the unexpected outcome was longer than in the expected outcome, and because there was some variation in the length of the action sequence in the two outcomes of the Preference Task, we also analyzed infants' looking times from the end of the video. This analysis revealed the same pattern of looking as in the main analysis, but the effect was weaker: Preference Task ($\beta = 2.19, p = .090$, *M*expected = 2.91, SE = 0.65, *M*unexpected = 5.10, SE = 1.08); Efficiency Task ($\beta = 2.30, p = .159$, *M*expected = 4.43, SE = 1.07, *M*unexpected = 6.73, SE = 1.42). That said, looking time using this metric was much lower overall, and this metric could not capture infants' surprise at an agent's unexpected behavior occurring during the video (e.g., the agent's movement indicated a path towards an nonpreferred object very soon after that movement started). In the final version of this task, we will thus equate the length of all test trials within the same task.

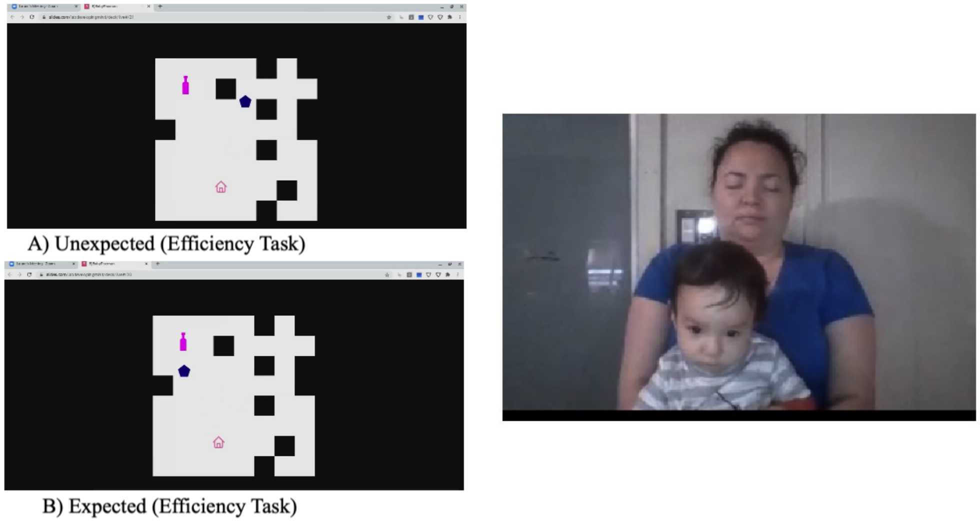
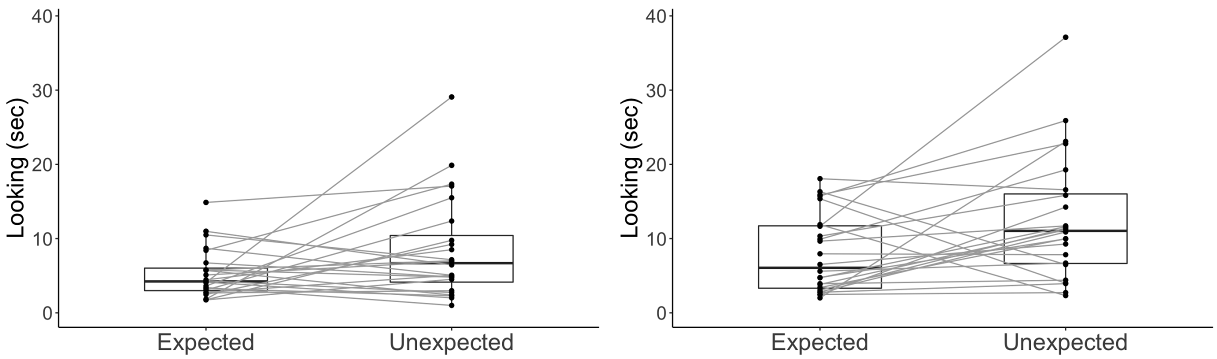


Fig. 9: The right panel shows a participant looking at the stimuli during the testing session. The left panel shows how the stimuli displayed on the participant's screen ((A) Unexpected outcome, Efficiency task and (B) Expected outcome, Efficiency task.



(a) Preference (b) Efficiency

Fig. 10: Box-plots of infants' looking time (in seconds) towards the expected and the unexpected outcome in (a) the Preference task and b) the Efficiency task. The horizontal lines indicate medians. Black points, connected by gray lines across boxes indicate looking time for individual participants.

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 [34]. 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.

Our pilot validation of BIB with 11-month-old infants both provides additional support for this suggestion and also provides new insight into the abstractness and generalizability of infants' knowledge. While studies with infants have relied on both simple and complex visual displays, they have not relied on one comprehensive battery with agents' intentionality conveyed at a high level of abstraction, as in the seminal Heider & Simmel (1944) [21] displays, which adults find very compelling. Moreover, infants' preliminary success on BIB informs our understanding of the generlizability of infant's knowledge. For example, infants in BIB's Preference Task successfully predicted an agent's goal-directed navigation when the location of the agent's preferred object varied greatly during familiarization.

The pilot validation with infants presented here converges with the main findings from conceptually and methodologically similar experiments from developmental psychology [19, 51, 52]. In future work, we plan to test infants on all of the BIB tasks, which would provide a comprehensive validation of the benchmark and a stronger case for the conclusion that state-of-the-art machine learning models are fundamentally different from human infants in their reasoning about agents.

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 commonsense 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.

Acknowledgements

This research was supported by the DARPA Machine Common Sense program (HR001119S0005), as well as NSF grant (DRL-1845924). We thank Victoria Romero, Koleen McKrink, David Moore, Lisa Oakes, Clark Dorman, and Thomas Schellenberg, Dean Wetherby, and Brian Pippin for their generous feedback.

References

[1] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the 21st International Conference on Machine learning*, 2004, p. 1.

[2] S. V. Albrecht and P. Stone, "Autonomous agents modelling other agents: A comprehensive survey and open problems," *Artificial Intelligence*, vol. 258, pp. 66–95, 2018.

[3] R. Baillargeon, R. M. Scott, and L. Bian, "Psychological reasoning in infancy," *Annual review of psychology*, vol. 67, pp. 159–186, 2016.

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

[5] R. Baillargeon, E. S. Spelke, and S. Wasserman, "Object permanence in five-month-old infants," *Cognition*, vol. 20, no. 3, pp. 191–208, 1985.

[6] C. Baker, R. Saxe, and J. Tenenbaum, "Bayesian theory of mind: Modeling joint belief-desire attribution," in *Proceedings of the annual meeting of the cognitive science society*, vol. 33, 2011.

[7] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, "Rational quantitative attribution of beliefs, desires and percepts in human mentalizing," *Nature Human Behaviour*, vol. 1, no. 4, pp. 1–10, 2017.

[8] C. L. Baker, R. Saxe, and J. B. Tenenbaum, "Action understanding as inverse planning," *Cognition*, vol. 113, no. 3, pp. 329–349, 2009.

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

[10] J. S. Buresh and A. L. Woodward, "Infants track action goals within and across agents," *Cognition*, vol. 104, no. 2, pp. 287–314, 2007.

[11] M. Carpenter, J. Call, and M. Tomasello, "Twelve-and 18-month-olds copy actions in terms of goals," *Developmental science*, vol. 8, no. 1, pp. F13–F20, 2005.

[12] M. Colomer, J. Bas, and N. Sebastian-Galles, "Efficiency as a principle for social preferences in infancy," *Journal of Experimental Child Psychology*, vol. 194, p. 104823, 2020.

[13] K. Dunn and J. G. Bremner, "Investigating looking and social looking measures as an index of infant violation of expectation," *Developmental Science*, vol. 20, no. 6, p. e12452, 2017.

[14] B. Elsner, P. Hauf, and G. Aschersleben, "Imitating step by step: A detailed analysis of 9-to 15-month-olds' reproduction of a three-step action sequence," *Infant Behavior and Development*, vol. 30, no. 2, pp. 325–335, 2007.

[15] G. Gergely and G. Csibra, "Navigating the social world—what infants, children, and other species can teach us," 2013.

[16] G. Gergely, H. Bekkering, and I. Király, "Rational imitation in preverbal infants," *Nature*, vol. 415, no. 6873, pp. 755–755, 2002.

[17] G. Gergely and G. Csibra, "Teleological reasoning in infancy: The infant's naive theory of rational action: A reply to premack and premack," *Cognition*, vol. 63, no. 2, pp. 227–233, 1997.

[18] ——, "Teleological reasoning in infancy: The naıve theory of rational action," *Trends in cognitive sciences*, vol. 7, no. 7, pp. 287–292, 2003.

[19] G. Gergely, Z. Nádasdy, G. Csibra, and S. Bíró, "Taking the intentional stance at 12 months of age," *Cognition*, vol. 56, no. 2, pp. 165–193, 1995.

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

[21] F. Heider and M. Simmel, "An experimental study of apparent behavior," *The American journal of psychology*, vol. 57, no. 2, pp. 243–259, 1944.

[22] A. M. Henderson and A. L. Woodward, "Nine-month-old infants generalize object labels, but not object preferences across individuals," *Developmental science*, vol. 15, no. 5, pp. 641–652, 2012.

[23] M. Hernik and G. Csibra, "Infants learn enduring functions of novel tools from action demonstrations," *Journal of experimental child psychology*, vol. 130, pp. 176–192, 2015.

[24] J. Jara-Ettinger, "Theory of mind as inverse reinforcement learning," *Current Opinion in Behavioral Sciences*, vol. 29, pp. 105–110, 2019.

[25] J. F. Kominsky, "Pyhab: Open-source real time infant gaze coding and stimulus presentation software," *Infant Behavior and Development*, vol. 54, pp. 114–119, 2019.

[26] V. Kuhlmeier, K. Wynn, and P. Bloom, "Attribution of dispositional states by 12-month-olds," *Psychological science*, vol. 14, no. 5, pp. 402–408, 2003.

[27] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, "Building machines that learn and think like people," *Behavioral and brain sciences*, vol. 40, 2017.

[28] S. Liu, N. B. Brooks, and E. S. Spelke, "Origins of the concepts cause, cost, and goal in prereaching infants," *Proceedings of the National Academy of Sciences*, vol. 116, no. 36, pp. 17 747–17 752, 2019.

[29] S. Liu, T. D. Ullman, J. B. Tenenbaum, and E. S. Spelke, "Ten-month-old infants infer the value of goals from the costs of actions," *Science*, vol. 358, no. 6366, pp. 1038–1041, 2017.

[30] Y. Luo, "Three-month-old infants attribute goals to a non-human agent," *Developmental science*, vol. 14, no. 2, pp. 453–460, 2011.

[31] Y. Munakata, "Challenges to the violation-of-expectation paradigm: Throwing the conceptual baby out with the perceptual processing bathwater?" *Infancy*, vol. 1, no. 4, pp. 471–477, 2000.

[32] A. Y. Ng, S. J. Russell *et al.*, "Algorithms for inverse reinforcement learning." in *Proceedings of the 17th International Conference on Machine learning*, vol. 1, 2000, p. 2.

[33] L. M. Oakes, "Using habituation of looking time to assess mental processes in infancy," *Journal of Cognition and Development*, vol. 11, no. 3, pp. 255–268, 2010.

[34] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. M. A. Eslami, and M. Botvinick, "Machine theory of mind," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. Stockholmsmässan, Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 4218–4227. [Online]. Available: http://proceedings.mlr.press/v80/rabinowitz18a.html

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

[36] K. Rakelly, A. Zhou, C. Finn, S. Levine, and D. Quillen, "Efficient off-policy meta-reinforcement learning via probabilistic context variables," in *International conference on machine learning*. PMLR, 2019, pp. 5331–5340.

[37] B. M. Repacholi and A. Gopnik, "Early reasoning about desires: evidence from 14-and 18-month-olds." *Developmental psychology*, vol. 33, no. 1, p. 12, 1997.

[38] R. Riochet, M. Y. Castro, M. Bernard, A. Lerer, R. Fergus, V. Izard, and E. Dupoux, "Intphys: A framework and benchmark for visual intuitive physics reasoning," *CoRR*, vol. abs/1803.07616, 2018. [Online]. Available: http://arxiv.org/abs/1803.07616

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[40] R. Saxe, T. Tzelnic, and S. Carey, "Knowing who dunnit: Infants identify the causal agent in an unseen causal interaction." *Developmental psychology*, vol. 43, no. 1, p. 149, 2007.

[41] G. Schöner and E. Thelen, "Using dynamic field theory to rethink

infant habituation." *Psychological review*, vol. 113, no. 2, p. 273, 2006.
[42] R. M. Scott and R. Baillargeon, "Do infants really expect agents to act efficiently? a critical test of the rationality principle," *Psychological science*, vol. 24, no. 4, pp. 466–474, 2013.
[43] T. Shu, A. Bhandwaldar, C. Gan, K. Smith, S. Liu, D. Gutfreund, E. Spelke, J. B. Tenenbaum, and T. D. Ullman, "AGENT: A Benchmark for Core Psychological Reasoning," *arXiv preprint arXiv:2102.12321*, 2021.
[44] N. Y. Siegel, J. T. Springenberg, F. Berkenkamp, A. Abdolmaleki, M. Neunert, T. Lampe, R. Hafner, N. Heess, and M. Riedmiller, "Keep doing what worked: Behavioral modelling priors for offline reinforcement learning," *arXiv preprint arXiv:2002.08396*, 2020.
[45] A. E. Skerry, S. E. Carey, and E. S. Spelke, "First-person action experience reveals sensitivity to action efficiency in prereaching infants," *Proceedings of the National Academy of Sciences*, vol. 110, no. 46, pp. 18 728–18 733, 2013.
[46] K. Smith, L. Mei, S. Yao, J. Wu, E. Spelke, J. Tenenbaum, and T. Ullman, "Modeling expectation violation in intuitive physics with coarse probabilistic object representations," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8985–8995.
[47] J. A. Sommerville and A. L. Woodward, "Pulling out the intentional structure of action: the relation between action processing and action production in infancy," *Cognition*, vol. 95, no. 1, pp. 1–30, 2005.
[48] H.-j. Song, R. Baillargeon, and C. Fisher, "Can infants attribute to an agent a disposition to perform a particular action?" *Cognition*, vol. 98, no. 2, pp. B45–B55, 2005.
[49] N. B. Turk-Browne, B. J. Scholl, and M. M. Chun, "Babies and brains: habituation in infant cognition and functional neuroimaging," *Frontiers in human neuroscience*, vol. 2, p. 16, 2008.
[50] T. Ullman, C. Baker, O. Macindoe, O. Evans, N. Goodman, and J. B. Tenenbaum, "Help or hinder: Bayesian models of social goal inference," in *Advances in neural information processing systems*, 2009, pp. 1874–1882.
[51] A. L. Woodward, "Infants selectively encode the goal object of an actor's reach," *Cognition*, vol. 69, no. 1, pp. 1–34, 1998.
[52] ——, "Infants' ability to distinguish between purposeful and non-purposeful behaviors," *Infant behavior and development*, vol. 22, no. 2, pp. 145–160, 1999.
[53] A. L. Woodward and J. A. Sommerville, "Twelve-month-old infants interpret action in context," *Psychological Science*, vol. 11, no. 1, pp. 73–77, 2000.
[54] K. Xu, E. Ratner, A. Dragan, S. Levine, and C. Finn, "Learning a prior over intent via meta-inverse reinforcement learning," in *International Conference on Machine Learning*. PMLR, 2019, pp. 6952–6962.
[55] L. Yu, T. Yu, C. Finn, and S. Ermon, "Meta-inverse reinforcement learning with probabilistic context variables," *arXiv preprint arXiv:1909.09314*, 2019.
[56] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning." in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.
[57] N. Zmyj, M. M. Daum, and G. Aschersleben, "The development of rational imitation in 9-and 12-month-old infants," *Infancy*, vol. 14, no. 1, pp. 131–141, 2009.