Experimental designs and facets of evidence for computational theory of mind

Joel Michelson[1], Deepayan Sanyal[1], James Ainooson[1], Yuan Yang[1] and Maithilee Kunda[1]

[1]*Vanderbilt University Department of Computer Science, Nashville TN, USA*

Abstract

The competitive feeding paradigm is one of several experimental setups intended to test whether non-verbal subjects possess skills related to Theory of Mind. Competitive feeding focuses on the relationship between seeing and knowing. In this paper, we describe a highly-customizeable implementation of the competitive feeding paradigm for computational agents in a gridworld environment. We explore various modifications to the setup including shared rewards, alternate sequences of timed events, and asymmetrical values, that allow us to replicate a wide breadth of tests designed to study the social cognition skills of humans and animals. Finally, we describe how this paradigm can be expanded upon and used as a benchmark test to investigate social reasoning in artificially intelligent models.

Keywords

Theory of mind, machine learning, social cognition,

1. Introduction

One critical element of social cognition research is Theory of Mind (ToM), described originally by Premack and Woodruff in 1978 as a “system of inferences" regarding the mental states of others [1]. Specifically, mental states, which are unobservable, may only be *inferred* to both exist and relate to observable data. Because of their subjective nature, ToM skills and the mechanisms that produce them—in humans and other animals—are not thoroughly understood. Their detection and measurement has been and remains the subject of a lengthy ongoing debate.

A well-studied example of potential ToM reasoning in the animal kingdom is that of Western scrub-jays, who instinctively cache their food to save it for later. They tend to re-cache their food if they believe their behavior was observed by a competitor, who might try to pilfer the hidden prize. In doing so, they keep track of which individual witnesses are privy to information about different cache sites [2]. At first glance, such sophisticated behavior seems to imply that the jays are capable of inferring other competitors' mental states. By careful observation, however, it becomes apparent that directly observable information, e.g. “Polly's head was oriented towards this particular cache in the past” is sufficient for a successful re-caching strategy, without need for any mentalization, e.g. “Polly *knows* there is food here". Jays also display some degree of successful transfer between the roles of hiding and seeking: birds which have been thieves are more likely to re-cache their food when observed by competitors [3]. Could this pattern be evidence for experience projection? Although it is interesting behavior, this observation also

8th International Workshop on Artificial Intelligence and Cognition, June 15-17, 2022, Örebro, Sweden

fails to provide strong evidence for any reasoning about competitors' internal states, as it can be explained by ToM-free models [4].

While most literature on ToM focuses on humans and non-human animals, there exists a wealth of knowledge to be questioned, tested, and discovered in the realm of artificial intelligence. Michelson et al. [5] highlight the need for a standardized battery of tests that can be used by many to evaluate AI models' theory of mind skills. They describe several criteria and desiderata that make social cognition benchmark tests amiable to artificial intelligence researchers. Numerous tests of animal cognition examine ToM and related skills, including the popular Sally Anne test [6], knower guesser paradigm [7] [8], and competitive feeding paradigm [9]. The text of this paper covers the design, implementation, and use-cases of one such test environment—inspired by the competitive feeding paradigm—that serves as a foundation for such a test battery. The specific contributions of this paper include:

- A brief overview of the competitive feeding paradigm, a test framework designed to study whether non-verbal animals understand concepts of seeing and knowing, as well as its criticisms.
- A detailed description of the Standoff environment, a gridworld framework for running social cognition tests on computational agents.[1]
- Descriptions of how various specific modifications of competitive feeding under the Standoff framework allow for the measurement of a breadth of skills beyond those captured by competitive feeding.

2. Background

Povinelli and Vonk [10] point out the failure of existing paradigms for testing social cognition in that these tests generally do not distinguish reasoning about observable behavior from reasoning about unobservable mental states. Later, Penn and Povinelli provide a formalized definition of ToM so that its presence in a subject can be more systematically measured and falsified [11]. They describe ToM as the presence of a function, f_{ToM}, which a cognitive agent (the subject) may use to infer the mental state of another cognitive agent. As f_{ToM} is an inference, its output must be based solely on the perceptual inputs available to the subject. This definition avoids any specific interpretations of how f_{ToM} might be implemented or used. Compelling evidence of f_{ToM} must be in the form of behavior that demonstrates "the necessity of an f_{ToM} *in addition to* and *distinct from* the cognitive work that could have been performed without such a function."

The competitive feeding paradigm, which we describe in 2.1, is used by Penn and Povinelli as a case study for its *inability* to detect ToM [11]. With a few modifications, however, a new paradigm can be built that satisfies the requirements for proving and falsifying ToM hypotheses.

2.1. The Competitive Feeding Paradigm

The competitive feeding paradigm is a test setup designed to distinguish whether a non-verbal subject will change its behavior to account for what it believes a conspecific *knows*, based on

[1]The Standoff environment, along with instructions for generating the tests described in this paper, can be accessed at http://github.com/aivaslab/standoff

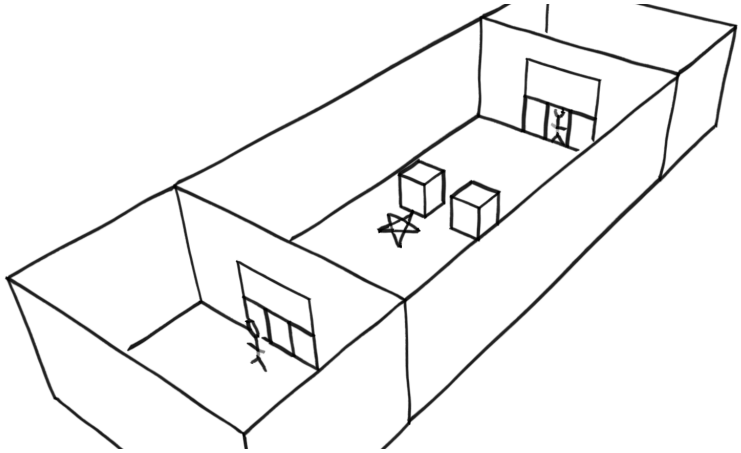

Figure 1: A diagram of a basic competitive feeding setup. The dominant and subordinate participates (stick figures, left and right) begin in cages on either side of the room. Here, a food treat (star) has been baited on the subordinate's side of a box, which occludes it from the dominant's view. During baiting events, the dominant's door might be closed, obscuring its vision of the food's placement.

evidence relating to what the conspecific *sees* [9]. The subject and one other participant must have an established social hierarchy, with the subject being 'subordinate' to the other 'dominant' participant.

2.1.1. Setup

The general setup of a competitive feeding test is as follows: The animals are kept in cages on either side of a central room, the subject's cage always opposite its one or more conspecifics' cages. During "baiting" events, large and small food rewards, or treats, are placed or moved in the central room. Although the placement of the treat is sometimes visible to the dominant, after one or both baiting events occur, the dominant is no longer able to see the treat. Eventually, both animals are released. Due to the nature of the social hierarchy, the subordinate will not challenge the dominant if the two would attempt to reach the same treat. So, if the subject believes the dominant will look for food in a particular location, we assume the subject will avoid that location. The subject's initial challenge, then, is determining where the dominant will decide to go. Once released, the subject's orientation or movement towards a treat is recorded.

2.1.2. Baiting events

During the baiting events, the dominant's door might be partially open, allowing it to see the baiting, or closed. By closing the dominant's door at specified times, researchers create scenarios in which it knows where the food is, it does not know, or it has a false belief about the food's presence or location (i.e. it knows where the food is initially, but is then unaware that the food has been moved).

By carefully observing what the dominant can and cannot see and then reasoning about what the dominant knows, the subject might choose to alter its behavior to secure more food

for itself. For example, if the subject believes that the dominant does not know the larger food pile's location, the subject might try going there for a greater reward, when it would otherwise leave the pile to the dominant.

2.1.3. Variants

Since its first use testing chimpanzees [9], multiple variants of competitive feeding have been proposed, implemented, and run on various animal species. Hare et al. published a compelling version of the test in 2001 featuring three experiments: "did", "who", and "which", referring to the subjects' beliefs about whether conspecifics witness different baiting events [12]. "Did" refers to the ability to distinguish whether an opponent did or did not observe an event, "who" involves understanding who of multiple opponents observed an event, and "which" involves understanding which of multiple baiting events an opponent observed. That test, and most following it, compare the subject's performance across at least four conditions: Informed, Uninformed, Control Misinformed, and Misinformed. The names of these variants refer to the dominant's awareness of baiting events. In the former two setups, one baiting event takes place, and the dominant is either aware or unaware of the food's location. In the latter two, the dominant is aware during one baiting event, but then is either aware or unaware of a second in which the foods' locations are swapped. The misinformed and control misinformed cases can be likened to the Sally Anne test, as the subject is tasked with identifying the presence of a change-of-location false belief.

2.2. Criticism

Because our subject has access to its own mental state, it is of critical importance to falsify the null hypothesis that it makes use of *only* its own mental state to determine its behavior. Behaviors that could be explained as a learned response to superficial perceptual input, e.g 'her eyes being pointed toward the food indicates that I should go somewhere else', do not suffice.

In general, ToM allows an agent to behave as though some other portion of the environment (read: another embodied agent) is expected to behave in accordance with a false belief. To be convinced that the agent has ToM, its behaviors under *all* alternate assumptions of truth values (and beliefs about truth values) must be known and compared. Penn and Povinelli describe two different alternatives to the competitive feeding paradigm that might aid in making such a comparison.

The first, called the opaque visor experiment, is a modification of a task described in [13]. The opaque visor experiment involves explicit generalization from novel first-person experience to third-person reasoning: the subject is given time to experiment with multiple visors, the opacity of which is only visible with physical proximity, before being evaluated about the visor's effect on an experimenter at a distance. Due to its emphasis on few-shot learning, the opaque visor experiment lies beyond the scope of this paper. The second, which motivates this work, adds a handful of modifications and variants meant to control for alternate explanations in animals' 'passing' behavior to Hare et al.'s competitive feeding paradigm [12].

2.2.1. Systematic Competitive Feeding

The improvements Penn and Povinelli suggest for a systematic competitive feeding paradigm (SCFP) are slightly more complex, but provide much more satisfying answers to questions of what, exactly, the subjects believe. To allow for satisfactory presumption of agents' behavior, they describe a specific training regime featuring steps that must be passed successfully, representing successful understanding of the test's fundamental components.

In Stage 1, subjects are trained in the absence of dominant competitors until they demonstrate proper goal-seeking behavior. Next, in Stage 2, they are trained to compete with a conspecific (as in all other competitive feeding tests) for food, and only those who successfully concede food to dominants are allowed to continue. If our subjects pass the first two stages, we can be certain that they understand the basics of how their reward can be maximized.

Finally, several variants are presented as test conditions in Stage 3. In this version, there are several buckets (food locations), and food is always placed in two of them during the baiting events. Because the number of buckets is usually greater than 2, the SCFP makes no cross-experiment distinction between the "did" and "which" cases of Hare et al. [12].

Instead of the four common test variants described above, the SCFP uses at least eight scenarios to comprehensively judge the subjects' understanding: Informed control, partially uninformed, removed informed, removed uninformed, moved, replaced, misinformed, and swapped. Like the four common competitive feeding variants, these scenarios differ from each other only by schedules of obscuring, baiting, hiding, and releasing events, performed by the experimenters. For full descriptions of each scenario, please refer to section 6b of Penn and Povinelli 2007 [11].

3. The Standoff Environment: A Gridworld Platform for Computational Theory of Mind Experiments

The Standoff Environment is a multiagent gridworld environment implemented as a partially-observable Markov decision process using the PettingZoo API [14]. SuperSuit [15] wrappers convert the environment's inputs and outputs into formats which can interface directly with off-the-shelf reinforcement learning paradigms in Stable-baselines3 [16] and RLlib [17]. Standoff replicates all SCFP variants as described in [11], and, as we will see, is capable of testing for ToM skills in a wide variety of settings.

3.1. Agents

Agents' views are bird's-eye representations of their surroundings. These views are either egocentric, in which the agent's body always appears in the same relative location, and orientation is aligned with the agent's current direction, or allocentric, in which the entire world is displayed with a uniform coordinate system, but areas outside the agent's perception are masked. In both cases, our agents' bird's-eye perceptions are notably different from real animals' first-person views, but we opt for the simpler and possibly easier perspective for the sake of programmer friendliness. Agents' action sets include movement of either of two kinds: directed (forward, backward, rotate left, and rotate right) and cardinal (North, South, East, and West).

3.2. Puppets

The Standoff environment supports multi-agent reinforcement learning, but its initial intent is studying the behavior of a single subject. As a starting point, the subject's conspecifics, be they collaborators or competitors, are implemented as hard-coded *puppets*. These puppets behave according to simple rulesets applied to their perceptions. Puppets appear identically to any agent—subject included—other than optional visual features that distinguish their values (see 4.6). Puppets have an explicit memory of relevant information that they witness (namely, treat locations), as well as basic navigation skills. Through this dynamic implementation, changing the sequence of information presented to a puppet causes predictable changes in its behavior. Various independent variables and environmental parameters can be edited to create different experimental conditions, to which puppets respond automatically. Puppets' behavior can be otherwise specified by the user to any degree of granularity, and they can even be controlled by custom artificially intelligent models. Note that the puppets' hard-coded behavior is intended as a starting point in absence of rational actors, though irrational behavior also warrants study.

3.3. Tutorial Stages

All tests of social cognition are based on a number of assumptions about their subjects' goals, knowledge, and abilities. Animal subjects' preferences for food are well-understood, and fundamental knowledge—like that doors open and close, or how to navigate simple environments—can generally be assumed without question or otherwise taught with repeated exposure.

The Standoff environment makes use of numerous 'tiles' with various behaviors and affordances that, at evaluation time, the subjects are assumed to understand. Curtains and boxes conceal their contents, treats grant rewards, gates (both transparent and opaque) open and close without warning, and other agents move about of their own volition. These 'commonsense' facts (along with many others) are established in the environment's provided tutorial stages, which expose a subject to various hardcoded and randomized settings so that it can explore the rules of the world, which imitate those of the other Standoff conditions.

3.4. Competitive Feeding

As a starting point, we shall introduce the Standoff implementation of competitive feeding, including all the systematic variants proposed by Penn and Povinelli [11]. In our computational version of the test paradigm, we closely imitate many aspects of the competitive feeding design: walls are opaque (occluded areas are masked by a special shadow color), gates may be opened or closed (and opaque or transparent), and treats are baited according to the same schedules.

Because the environment is a gridworld, many of these details are abstracted by a large degree. Treats are objects that provide reward to reinforcement learning agents (and often terminate the episode) when reached. The rewards granted by treats are dynamically determined following certain rules to ensure predictable optimal behavior. For example, if there are n boxes, the ratio of (positive) rewards between the larger and smaller treat must be greater than $n : 1$, otherwise strategies like "always approach the smaller treat" become valid strategies for achieving maximal total reward under undesirable circumstances. Likewise, a small negative reward placed on empty boxes reduces the expected value of random guesses. Treats may be 'hidden' in boxes,

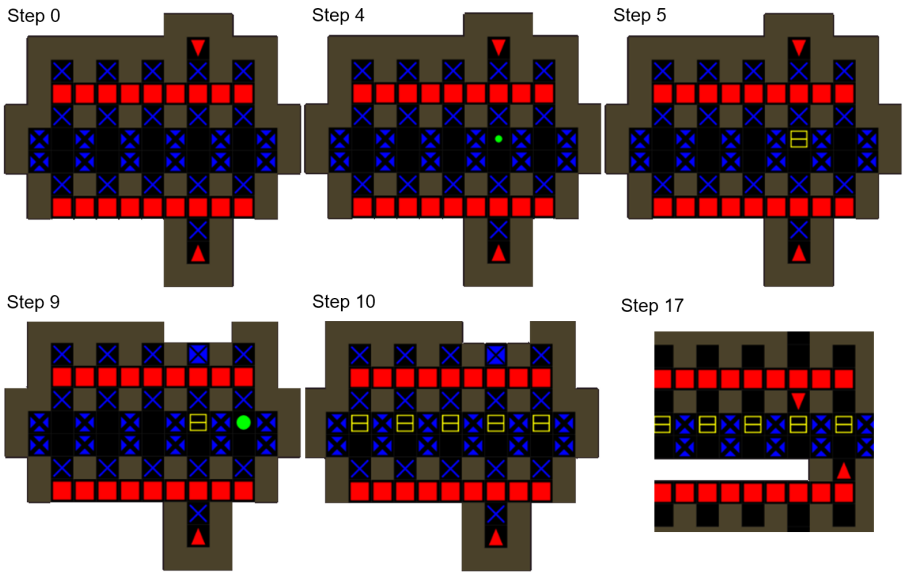

Figure 2: A *partially uninformed* scenario of the systematic competitive feeding paradigm showcased in the Standoff environment, pictured at selected timesteps. Both dominant puppet (top red triangle) and subordinate subject (bottom red triangle) begin with full views of the environment, and have movement impeded by transparent blocks (step 0). In view of both agents, the first treat, a small green circle, is baited (step 4) and hidden in a box (step 5). The dominant's view is obscured with an opaque block before the second, larger, treat is baited (step 9) and hidden alongside the decoy boxes at each treat location (step 10). Note that the dominant is occluded from the subject's view while the opaque block is present in steps 9 and 10. Both agents are released to the curtained area (red squares) to make decisions in private. Upon the second release event (step 17), the dominant progresses towards the small treat's location since it is unaware of the large treat's location, so, to the best of its knowledge, obtaining the small treat maximizes its expected value.

which obscure vision of the treat from both the dominant and (conditionally) the subordinate. Note that in the original competitive feeding paradigm design, the subordinate is always able to observe the treats' locations.

4. Independent Variables

By modifying various small sets of environmental parameters, we may computationally imitate many other social cognition experiments that have been performed on animals. Models' skills may be tested with a variety of lenses to gain insight into their fundamental capabilities and weaknesses.

All of the following variables can (and should) be investigated for transfer learning in a standardized manner. Then, models of ToM may be evaluated for their generalization capacity along various notable axes. As with scrub-jays, does experience with one role help an agent understand another? The test setups may be studied as well to find the extent to which success at a specific set of tests tends to predict other abilities.

4.1. Agent priority

In the standard competitive feeding tests, the subject is subordinate to its opponent, i.e. all else equal, the subject is at a competitive disadvantage. In the Standoff environment, this effect is achieved and signaled to the agents via treat locations; they are slightly closer to the dominant agent. On its own, allowing the subject to take on the role of the dominant invokes a trivially easy task, having an identical solution to Stage 1 of SCFP. In conjunction with other changes that we shall discuss below—especially visible decisions (see 4.3)—a dominant subject proves quite useful, as its decision can alter the subordinate puppet's behavior.

A transfer learning experiment of differing agent priorities is similar to the role-reversal experiment in [18]. Their experiment is of a collaborative nature, so note that the same experiment could be run under different conditions for anticipation valence (see 4.2) and reward sharing.

4.2. Anticipation valence

In the competitive feeding tests, the subject is expected to look for a treat in an area where it believes the dominant will *not* visit. Leslie and Polizzi find a significant difference between positive and negative desires, that is, looking where something is versus is not located, in the context of Sally Anne tests in human children [19]. A minor change in the rules of the Standoff environment inverts the negative valence in the competitive feeding paradigm: when the dominant reaches a treat, the treat shall remain and its value for the subordinate shall be *increased* to be maximal. Now, the task presented to the subject is arguably simpler: infer the dominant's goal, and adopt that goal as your own. There is no longer a need for extraneous decision-making regarding selecting the best goal alternative once the dominant's decision has been identified. Many other social cognition experiments, including most that involve collaboration, make use of positive anticipation.

We signal valence using treats' color for RGB inputs, and treats' identity for rich inputs. If the subject has the dominant priority (see 4.1), positive valence is achieved via reward sharing, that is, the subject is rewarded for a subordinate puppet's successful completion of the task.

4.3. Decision visibility

While evaluating all competitive feeding tests, it is of critical importance that the dominant (and, if the dominant has ToM, the subordinate) be given privacy while it decides which route to take. Otherwise, one agent could use the behavior of the other to inform its decision—a strategy that is clearly relevant to social cognition but interferes with our tests for attribution of already-established beliefs.

Allowing subordinate agents to make decisions while informed of other agents' decisions opens the possibility of testing imitation and emulation. By allowing the subordinate to view the dominant's decision before the decision is finalized, we can study the subordinate's ability to imitate (or avoid imitating, in the negative anticipation valence case). When the subordinate and dominant have differences in their abilities (be they perception, mental, or action), imitation may be directly compared with emulation. For example, a subject (occupying an empty room)

emulating a teacher (slowed by clutter) could navigate the room more efficiently than the teacher, as opposed to an imitating subject who would inefficiently copy the teacher's behavior.

When a dominant subject' decisions are visible, it might behave in a manner that strategically influences the subordinate puppet's decision. In the shared reward, positive anticipation version of this test, the subject's goal is to lead its conspecific to the treat. This altruistic variant, especially in conjunction with multiple value alignments, roughly evokes the Yummy-Yucky test described by [20], in which a subject is tasked with using knowledge of preferences to assist an experimenter.

4.4. Population size

An agent might solve the SCFP as defined by labeling events as 'seen by opponent' or 'unseen by opponent'. In this case, although an opponent's perception must be correctly inferred, it is unclear whether an f_{ToM} compartmentalizes the knowledge of a single opponent. In other words, we might pass all SCFP tests while operating under the assumption that all embodied opponents have a shared mental state. Note that this assumption could be correct in cases where opponents communicate with each other. In order to rule out this hypothesis, we must test for the "who" ability.

By increasing the population of puppets (each having individual vision-obscuring events), the subject may only find success by keeping track of *who* sees each baiting event. To accomplish this effect, multiple puppets are initialized, each in a separate starting room. Any number of puppets might be informed during the baiting events. During the release event, only one of the puppets is able to leave its cage. To pass these scenarios, the subject must determine whether or not the released puppet specifically was made privy to the pertinent information. In scenarios with more than one baiting event, the "informed" agent may or may not be informed of the irrelevant event(s). In conjunction with positive anticipation and visible decisions, the Standoff task becomes similar to the knower guesser paradigm [7], another popular test of social reasoning in animals.

4.5. Obscuring source

In the competitive feeding paradigm (real-life and Standoff), participants' vision is obscured using opaque doors that occlude baiting events. These doors may be replaced by one of any existing objects that have been established to be opaque (or not) during the agent's training. Numerous other methods may be devised for causing (and signaling) unawareness. Gaze, for one, has been extensively studied in humans and animals. By instructing puppets (with directional vision) to face away from the food during baitings, we can evoke a rudimentary replication of experiments involving gaze-originated unawareness.

4.6. Value alignment

A core assumption of previously described experiments is that all agents value treats similarly, yet a fundamental ToM skill involves empathizing with individuals with different preferences. We provide two alternative sets of preferences inspired by Leslie and Polizzi 1998 [19]: A negative-value agent prefers smaller treats to larger ones, and an avoidant agent prefers to

search boxes that contain no treats at all. Like anticipation valence, value alignment is signaled by agents with alternate color or numeric identity schemes.

4.7. Scenario complexity

Just as we would like to investigate our subjects' ability to compartmentalize their f_{ToM} functions to multiple different embodied agents (or distinguish between multiple inferred mental states), we might also test the complexity of f_{ToM} itself. Under what conditions, and to what extent, is it able to represent and distinguish between multiple goal states? In the multiple desires test [21] children are tested to study their comprehension of three different aspects of multiple desires. We can imitate this test by releasing the subject only after the puppet reaches its *first* goal, giving the puppet a chance to also reach a *second* goal before the subject.

Memory robustness is a closely related, fundamental skill for successful attribution. By increasing the complexity of the environment, an agent's memory will need improvement to succeed. The environment's scale or the number of potential treat locations can be trivially increased to achieve this effect. We may also increase the amount of time between baiting and releasing, as well as the number of relevant and irrelevant events, to stymie our agents' efforts to retain relevant information.

5. Future work

Although the Standoff environment can be used to systematically investigate a wide variety of skills, there are many aspects of social cognition that lie beyond its grasp. As mentioned previously, this environment is one of a set called for in [5]. Much additional work remains to be done in the task that is building models that solve our social reasoning tests.

5.1. Independent variables not covered by Standoff

Further environments, likely with different fundamental setups, will be required to replicate the design of social reasoning tests from the comparative cognition and developmental psychology literature.

Several classes of social cognition tests are not easily represented in the Standoff environment. One notable example is the goggles test (see opaque visor test described in 2.2), which demonstrates projection from first-person experience [11][22]. An environment capable of replicating this task would need to support both first-person viewpoints and memory sustained across repeated sessions to allow for testing one- and few-shot learning.

With significant modification, we hope to eventually cover a diverse set of tests which differentiate imitation and emulation. Just as the competitive feeding paradigm implementations make use of multiple vision-obscuring sources, tests of emulation include several sources of inefficient or unexpected conspecific behavior. These include irrationality or temporary inability [23] [24], accidents [25], and even moral transgressions [26].

Of particular note are tests involving deception beyond that which is allowed in 'decision viewing' scenarios. Despite having the label of deception, these tests involve hiding and communicating treats' locations in both collaborative and competitive settings. The box-locking

task, for example, asks its participant to aid or thwart a puppet by misinforming them or by physically preventing them from reaching their goal [27]. Other tasks involving deceptive behaviors tend to require repeated sessions, including penny hiding [28] and, as mentioned in 1, hiding belongings from onlooking competitors.

Similarly, we would like to point out that most of the inference the Standoff environment tests for is deductive in nature, although it is theoretically possible to test for abductive ToM reasoning. An accurate model of another agent's mental state should not only answer questions of what the agent will do, but should answer questions of *how* and *why* the agent displayed existing behavior. In the Standoff environment, *how* and *why* are generally answered by visible attributes of the environment, e.g. the opponent pursued the smaller goal *because* its body is colored blue and therefore experience dictates that it must have negative-value nature. This type of reasoning will likely prove necessary for successful one- and few-shot learning in ToM scenarios, a powerful but difficult skill to master.

5.2. From generating baselines to solving the ToM riddle

The overall difficulty of the various Standoff tasks is an important question whose answer lies beyond the scope of this paper. Several environmental parameters are included for practical ease of implementation, e.g. allowing for allocentric perception and cardinal movement actions might help with agents' spatial memory, which is a complex skill in its own right.

Many researchers have already made substantial headway towards artificial ToM, including those with their own versions of social cognition tests mentioned above, for example Rabinowitz et al., who test their models on a gridworld implementation of the Sally Anne test [29]. A wide variety of models and strategies have been employed, including deep reinforcement learning, Bayesian inference [30], and cognitive models [31]. A review of algorithms designed for ToM reasoning can be found in Hernandez-leal et al. 2019 [32].

Competitive feeding subjects might lack proper understanding of their rivals' mental states, but we, as scientists, must empathize with their struggle. We, too, have a long journey ahead of us as we attempt to overcome our own lack of understanding, not just about mental states, but about how mental states are understood. By continuing along this path of tests, with foundations in comparative literature, we hope to help uncover the mysteries that allow us to understand.

Acknowledgments

This work was supported in part by the Neurodiversity Inspired Science and Engineering (NISE) NSF program grant DGE 19-22697 (K. Stassun, PI). We also extend thanks to our anonymous reviewers for their helpful criticisms, comments, and suggestions.

References

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

[2] N. S. Clayton, J. M. Dally, N. J. Emery, Social cognition by food-caching corvids. the western scrub-jay as a natural psychologist, Philosophical Transactions of the Royal Society B: Biological Sciences 362 (2007) 507–522.
[3] N. J. Emery, N. S. Clayton, Effects of experience and social context on prospective caching strategies by scrub jays, Nature 414 (2001) 443–446.
[4] E. Van der Vaart, R. Verbrugge, C. K. Hemelrijk, Corvid re-caching without 'theory of mind': A model, PloS one 7 (2012) e32904.
[5] J. Michelson, D. Sanyal, J. Ainooson, Y. Yang, M. Kunda, Social cognition paradigms ex machinas (2021).
[6] H. Wimmer, J. Perner, Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception, Cognition 13 (1983) 103–128.
[7] M. A. Udell, N. R. Dorey, C. D. Wynne, Can your dog read your mind? understanding the causes of canine perspective taking, Learning & Behavior 39 (2011) 289–302.
[8] T. Bugnyar, Knower–guesser differentiation in ravens: others' viewpoints matter, Proceedings of the Royal Society B: Biological Sciences 278 (2011) 634–640.
[9] B. Hare, J. Call, B. Agnetta, M. Tomasello, Chimpanzees know what conspecifics do and do not see, Animal Behaviour 59 (2000) 771–785.
[10] D. J. Povinelli, J. Vonk, We don't need a microscope to explore the chimpanzee's mind, Mind & Language 19 (2004) 1–28.
[11] D. C. Penn, D. J. Povinelli, On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind', Philosophical Transactions of the Royal Society B: Biological Sciences 362 (2007) 731–744.
[12] B. Hare, J. Call, M. Tomasello, Do chimpanzees know what conspecifics know?, Animal behaviour 61 (2001) 139–151.
[13] C. M. Heyes, Theory of mind in nonhuman primates, Behavioral and brain sciences 21 (1998) 101–114.
[14] J. K. Terry, B. Black, N. Grammel, M. Jayakumar, A. Hari, R. Sullivan, L. Santos, C. Dieffendahl, C. Horsch, R. Perez-Vicente, et al., Pettingzoo: Gym for multi-agent reinforcement learning, Advances in Neural Information Processing Systems 34 (2021).
[15] J. K. Terry, B. Black, A. Hari, Supersuit: Simple microwrappers for reinforcement learning environments, arXiv preprint arXiv:2008.08932 (2020).
[16] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, N. Dormann, Stable-baselines3: Reliable reinforcement learning implementations, Journal of Machine Learning Research 22 (2021) 1–8. URL: http://jmlr.org/papers/v22/20-1364.html.
[17] E. Liang, R. Liaw, R. Nishihara, P. Moritz, R. Fox, K. Goldberg, J. Gonzalez, M. Jordan, I. Stoica, Rllib: Abstractions for distributed reinforcement learning, in: International Conference on Machine Learning, PMLR, 2018, pp. 3053–3062.
[18] D. J. Povinelli, K. A. Parks, M. A. Novak, Role reversal by rhesus monkeys, but no evidence of empathy, Animal Behaviour 44 (1992) 269–281.
[19] A. M. Leslie, P. Polizzi, Inhibitory processing in the false belief task: Two conjectures, Developmental science 1 (1998) 247–253.
[20] B. M. Repacholi, A. Gopnik, Early reasoning about desires: evidence from 14-and 18-month-olds., Developmental psychology 33 (1997) 12.

[21] M. Bennett, L. Galpert, Children's understanding of multiple desires, International Journal of Behavioral Development 16 (1993) 15–33.
[22] K. Karg, M. Schmelz, J. Call, M. Tomasello, The goggles experiment: Can chimpanzees use self-experience to infer what a competitor can see?, Animal Behaviour 105 (2015) 211–221.
[23] A. N. Meltzoff, Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli., Developmental psychology 24 (1988) 470.
[24] G. Gergely, H. Bekkering, I. Király, Rational imitation in preverbal infants, Nature 415 (2002) 755–755.
[25] J. Call, M. Tomasello, Distinguishing intentional from accidental actions in orangutans (pongo pygmaeus), chimpanzees (pan troglodytes) and human children (homo sapiens)., Journal of Comparative Psychology 112 (1998) 192.
[26] M. Killen, K. L. Mulvey, C. Richardson, N. Jampol, A. Woodward, The accidental transgressor: Morally-relevant theory of mind, Cognition 119 (2011) 197–215.
[27] B. Sodian, U. Frith, Deception and sabotage in autistic, retarded and normal children, Journal of child psychology and psychiatry 33 (1992) 591–605.
[28] G. Gratch, Response alternation in children: A developmental study of orientations to uncertainty, Vita humana (1964) 49–60.
[29] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. A. Eslami, M. Botvinick, Machine theory of mind, in: International conference on machine learning, PMLR, 2018, pp. 4218–4227.
[30] C. L. Baker, J. Jara-Ettinger, R. Saxe, J. B. Tenenbaum, Rational quantitative attribution of beliefs, desires and percepts in human mentalizing, Nature Human Behaviour 1 (2017) 1–10.
[31] T. N. Nguyen, C. Gonzalez, Theory of mind from observation in cognitive models and humans, Topics in Cognitive Science (2021).
[32] P. Hernandez-Leal, B. Kartal, M. E. Taylor, A survey and critique of multiagent deep reinforcement learning, Autonomous Agents and Multi-Agent Systems 33 (2019) 750–797.