Learning in Social Environments with Curious Neural Agents

Megumi Sano (megsano@stanford.edu)
Department of Computer Science, Stanford University

Julian De Freitas (defreitas@g.harvard.edu)
Department of Psychology, Harvard University

Nick Haber* (nhaber@stanford.edu)
Graduate School of Education, Stanford University

Daniel L. K. Yamins* (yamins@stanford.edu)
Department of Psychology and Computer Science, Stanford University

Abstract

From an early age, humans are capable of learning about their social environment, making predictions of how other agents will operate and decisions about how they themselves will interact. In this work, we address the problem of formalizing the learning principles underlying these abilities. We construct a curious neural agent that can efficiently learn predictive models of social environments that are rich with external agents inspired by real-world animate behaviors such as peekaboo, chasing, and mimicry. Our curious neural agent consists of a controller driven by γ-*Progress*, a scalable and effective curiosity signal, and a *disentangled world model* that allocates separate networks for interdependent components of the world. We show that our disentangled curiosity-driven agent achieves higher learning efficiency and prediction performance than strong baselines. Crucially, we find that a preference for animate attention emerges naturally in our model, and is a key driver of performance. Finally we discuss future directions including applications of our framework to modeling human behavior and designing early indicators for developmental variability.

Keywords: world models; curiosity; social cognition

Introduction

Imagine a toddler at a busy playground, surrounded by a cornucopia of potentially interesting stimuli — from the leaves rustling along the ground, to the kickball lying in the sandbox, to the group of other children playing hide and seek, with their dynamic moves and complex interplay. Amongst this blooming, buzzing confusion, the child still manages to quickly learn about the world and its various dynamics. Underlying these abilities is the child's facility at building *world models*, predictive models of the environment that enable compact abstractions of high bandwidth sensory inputs and planning across long temporal horizons.

Crucially, as in the playground example, humans are effective world model learners even in complex social environments involving other agents. Such environments contain a diverse range of dynamics with varying levels of learnability. Inanimate stimuli such as a kickball display dynamics that are easy to learn. On the other end of the learnability spectrum, some stimuli, such as falling leaves, exhibit random noise-like dynamics. Lying in a "sweet spot" on this spectrum are animate agents that have interesting and complex yet learnable dynamics, e.g. children playing hide-and-seek. Balancing attention amidst a sea of diverse dynamics in a way that maximizes learning progress is a challenging problem. Particularly difficult is solving the "white noise" problem (Schmidhuber, 2010; Pathak, Agrawal, Efros, & Darrell, 2017; Burda, Edwards, Storkey, & Klimov, 2018), i.e distinguishing between unlearnable dynamics and learnable yet complex dynamics. Another key challenging property of social environments is that agents have complex interdependencies in their dynamics. Thus, understanding an agent-rich social environment entails identifying how agents are disentangled into their respective interdependent groups.

In this work, we address the problem of how to effectively learn world models in complex *social* environments — that is, in environments that are rich with both inanimate and animate stimuli, executing diverse dynamics with different levels of learnability. Specifically, we build a *curious neural agent* embedded in a custom-built 3D virtual world environment, filled with external agents displaying a wide spectrum of realistic animate behaviors such as peekaboo, chasing, and mimicry.

Our neural agent has two key components: first, a progress-driven curiosity signal, which we term γ-*Progress*, that rewards the neural agent proportionally to learning progress, estimated in a computationally scalable fashion. Intrinsically motivating the neural agent to maximize learning progress enables it to overcome the white noise problem as the agent ends up preferentially attending to stimuli with high learnability. Second, the neural agent has an agent-centric *disentangled* world model that allocates separate networks for interdependent agent groups. This allows the neural agent to ignore spurious correlations in the environment dynamics, thereby improving predictive performance.

We show that our disentangled curiosity-driven neural agent achieves higher learning efficiency and prediction performance than strong baselines. Our analysis shows that higher performance is in large part driven by the emergence of animate attention. Finally, we discuss future directions including applications of our framework to modeling human behavior and designing early indicators for developmental variability.

Related Works

Intuitive physics and object-based priors. Humans excel at intuitively predicting object dynamics (Battaglia et al., 2018).

*equal contribution

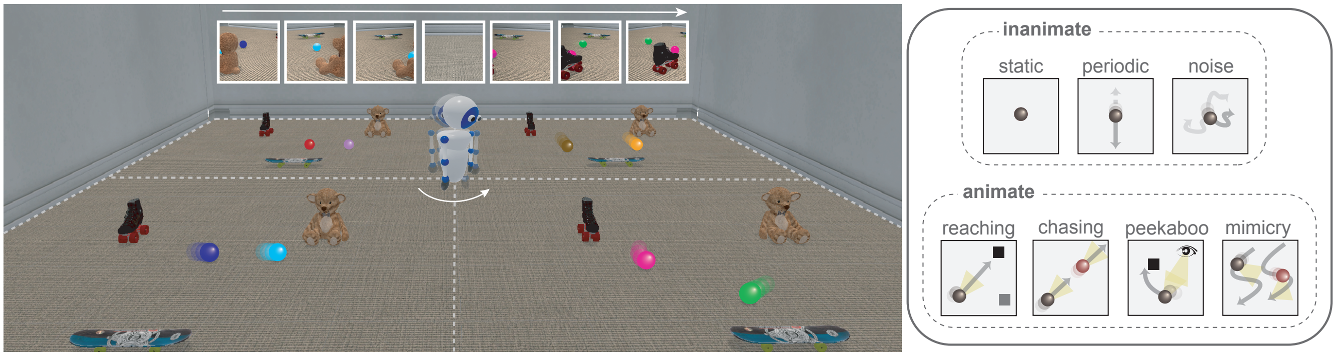


Figure 1: **Virtual Environment.** Our 3D virtual environment is a distillation of key aspects of real-world social scenes. The *curious neural agent* (white robot) is centered in a room, surrounded by various *external agents* (colored spheres) contained in different quadrants, each with dynamics that correspond to a realistic inanimate or animate behavior (right box). The curious neural agent can rotate to attend to different behaviors as shown by the first-person view images at the top. See https://bit.ly/2uf7lEY for videos.

A key framework underlying such abilities is object-centric attention allocation. Humans are able to keep track of objects over time, even as they become occluded or leave the visual frame (Piaget, 1952). In this work, we include object-based attention and object permanence as neural architectural biases.

Curiosity and active learning. Humans interact with the world to learn how it works. Infants actively gather information from their environment by attending to objects in a highly non-random manner (Smith et al., 2019), devoting more attention to objects that violate their expectations (Stahl & Feigenson, 2015). They also self-generate learning curricula, preferring stimuli that are complex enough to be interesting but still predictable (Kidd, Piantadosi, & Aslin, 2012). We study active learning by means of attention allocation.

Animate attention. From early infancy, humans effectively distinguish between inanimate and animate agents, preferentially paying attention to animate features like faces (Maurer, Le Grand, & Mondloch, 2002). Even in the absence of such visual features, infants preferentially attend to spatiotemporal kinematics indicative of animacy, such as efficient movement towards targets (Gergely, Nádasdy, Csibra, & Bíró, 1995) and contingent behavior between agents (Frankenhuis, House, Barrett, & Johnson, 2013). Such kinematic patterns give rise to an irresistable sense of animacy, even when the moving objects are simple shapes (Heider & Simmel, 1944). In this work, instead of injecting biases for animate attention, we test whether it emerges naturally, albeit with the right choice of curiosity.

Social prediction and Theory of Mind. A more sophisticated ability emerging later in development is understanding and predicting other agents' behaviors as consequences of their underlying mental states, aka *Theory of Mind* (Astington, Harris, & Olson, 1990). In this work, our model learns to predict what other agents will do next through the use of a disentangled architecture that leverages the idea of different agents having different internal states.

Artificial Intelligence. Our method is a form of *artificial curiosity* (Schmidhuber, 2010), a framework in which a reinforcement learner receives an intrinsic reward signal, often generated using the state of its world model, to encourage actions that maximize prediction gain. Prior works have explored prediction error (Pathak et al., 2017), novelty (Burda et al., 2018), and disagreement (Pathak, Gandhi, & Gupta, 2019). This work proposes a learning progress-based curiosity signal.

Virtual World Environment

To faithfully replicate the algorithmic challenges we face in the real world, we design our 3D virtual environment to preserve the following key properties of real-world environments: *diverse dynamics* consisting of various agent-specific programs, *partial observability*, which limits the agent's learning to what lies within view, and *interactivity*, allowing the agent's actions to influence the state of the world.

Our virtual environment consists of two main components, a *curious neural agent* and various *external agents*.

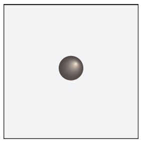
The **curious neural agent**, embodied by an avatar, is fixed at the center of a room (Figure 1). Just as a human toddler can control her gaze to visually explore her surroundings, the agent is able to partially observe the environment based on what lies in its field of view (see top of Figure 1). The agent can choose from 9 actions: rotate $12^\circ, 24^\circ, 48^\circ$, or 96°, to the left/right, or stay in its current orientation.

The **external agents** are spherical avatars that act under hard-coded policies inspired by real-world inanimate and animate stimuli. An *external agent behavior* consists of either one external agent, e.g reaching, or two interacting ones, e.g chasing. Since external agents are devoid of surface features, the curious agent must learn to attend to them based on spatiotemporal kinematics alone. We experiment with external agent behaviors (see Figure 1, right) including static, periodic, noise, reaching, chasing, peekaboo, and mimicry. The animate behaviors are inspired by stimuli used in the developmental psychology literature (Foster et al., 1973; Frankenhuis et al., 2013; Gergely et al., 1995; Johnson, 2003; Premack, 1990). We designed deterministic and stochastic variants of each animate behavior, where the stochastic variant preserves the core dynamics of the behavior, albeit with more randomness.

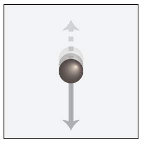
We divide the room into four quadrants, each of which contains three auxiliary objects (e.g teddy bear, roller skates, surfboard) and one external agent behavior. The room is designed such that the curious agent can see at most one external agent behavior at any given time.

Below, we describe all external agent behaviors in detail. See https://bit.ly/2uf7lEY for video descriptions of the environment and behaviors.

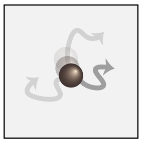
Inanimate behaviors



Static Inspired by stationary objects such as couches, lampposts, and fire hydrants, the *static agent* remains at its starting location and stays immobile.

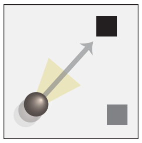


Periodic Inspired by objects exhibiting periodic motion such as fans, flashing lights, and clocks, the *periodic agent* regularly moves back and forth between two specified locations in its quadrant.

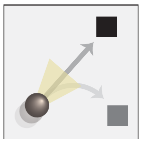


Noise Inspired by random motion in inanimate elements such as wind, the *noise agent* repeatedly moves with a fixed step size in some randomly sampled direction, while remaining within its quadrant.

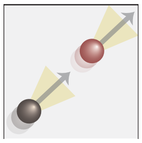
Animate Behaviors



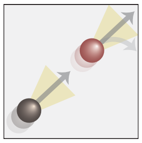
Reaching (deterministic) We often exhibit goal-oriented behavior by interacting with objects. The *reacher agent* approaches each auxiliary object in its quadrant sequentially, such that object positions fully determine its trajectory. Objects periodically shift locations so that predicting agent behavior requires knowing the current object positions.



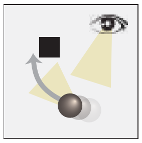
Reaching (stochastic) The order in which the reacher agent visits the objects is stochastic (uniform sampling from the three possible objects). However, once the reacher agent starts moving towards an object, its trajectory for the next few time steps, before it chooses a different object to move to, is predictable.



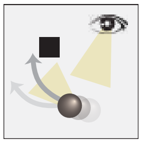
Chasing (deterministic) We often act contingently on the actions of other agents, which in turn depend on our own. In chasing, a *chaser agent* chases a *runner agent*. If the runner is too close to quadrant bounds, it then escapes to one of a few escape locations away from the chaser. Thus, the chaser's position affects the runner's trajectory and vice versa.



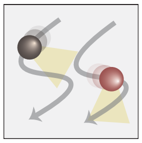
Chasing (stochastic) When the runner is too close to the bounds, it escapes by picking any random location away from the chaser, making the behavior harder to predict.



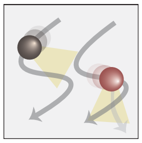
Peekaboo (deterministic) The *peekaboo agent* acts contingently on the curious agent. If the curious agent stares at it, it hides behind an auxiliary object. If the curious agent continues to stare, it starts *peeking* out by moving to a close fixed location. Once the curious agent looks away, it stops hiding, returning to an exposed location.



Peekaboo (stochastic) There are multiple peeking locations near the hiding object that the peekaboo agent can visit randomly during its peeking behavior.



Mimicry (deterministic) From an early age, we learn by imitating others. Mimicry consists of an *actor agent* and an *imitator agent*, each staying in one half of the quadrant. The actor acts identically to the noise agent, while the imitator mirrors the actor with a delay, such that the past trajectory of the actor determines the future trajectory of the imitator.



Mimicry (stochastic) The imitator agent is imperfect and produces a noisy reproduction of the actor agent's trajectory.

Methods

In this section we describe the practical instantiations of the two components of our curious neural agent: a *disentangled world model* which fits the forward dynamics and a *progress-driven controller* which acts to maximize γ-Progress reward.

Disentangled World Model. We assume that the agent has access to an oracle encoder $e : \mathcal{O} \rightarrow \mathcal{X}$ that maps an image observation $\mathbf{o}_t \in \mathcal{O}$ to a disentangled object-oriented feature vector $\mathbf{x}_t = (\mathbf{x}_t^{ext}, \mathbf{x}_t^{aux}, \mathbf{x}_t^{ego})$ where $\mathbf{x}_t^{ext} = (\tilde{\mathbf{c}}_t, \mathbf{m}_t) = (\tilde{\mathbf{c}}_{t,1}, \ldots \tilde{\mathbf{c}}_{t,n_{ext}}, \mathbf{m}_{t,1}, \ldots, \mathbf{m}_{t,n_{ext}})$ contains information about the external agents; namely the observability masks $\mathbf{m}_{t,i}$ ($\mathbf{m}_{t,i} = 1$ if external agent i is in curious agent's view at time t, else $\mathbf{m}_{t,i} = 0$) and masked position coordinates $\tilde{\mathbf{c}}_{t,i} = \mathbf{c}_{t,i}$ if $\mathbf{m}_{t,i} = 1$ and else $\tilde{\mathbf{c}}_{t,i} = \hat{\mathbf{c}}_{t,i}$. Here, $\mathbf{c}_{t,i}$ is the true global coordinate of external agent i and $\hat{\mathbf{c}}_{t,i}$ is the model's predicted coordinate of external agent i where $i = 1, \ldots, n_{ext}$. Note that the partial observability of the environment is preserved under the oracle encoder since it provides coordinates only for external agents in view. $\mathbf{x}_t^{aux}$ contains coordinates of auxiliary objects, and $\mathbf{x}_t^{ego}$ contains the ego-centric orientation of the curious agent.

Our disentangled world model ω_θ is an ensemble of component networks $\{\omega_{\theta^k}\}_{k=1}^{N_{cc}}$ where each ω_{θ^k} independently predicts the forward dynamics for a subset $I_k \subseteq \{1, ..., \dim(\mathbf{x}^{ext})\}$ of the input dimensions of $\mathbf{x}^{ext}$ corresponding to a minimal interdependent group in the world. For example, $\mathbf{x}_{t:t+\tau, I_k}^{ext}$ may correspond to the masked coordinates and observability masks of the chaser and runner external agents for times $t, t+1, ..., t+\tau$. We assume $\{I_k\}_{k=1}^{n_{cc}}$ is given as prior knowledge but future work may integrate disentanglement learning into our pipeline. A component network ω_{θ^k} takes as input

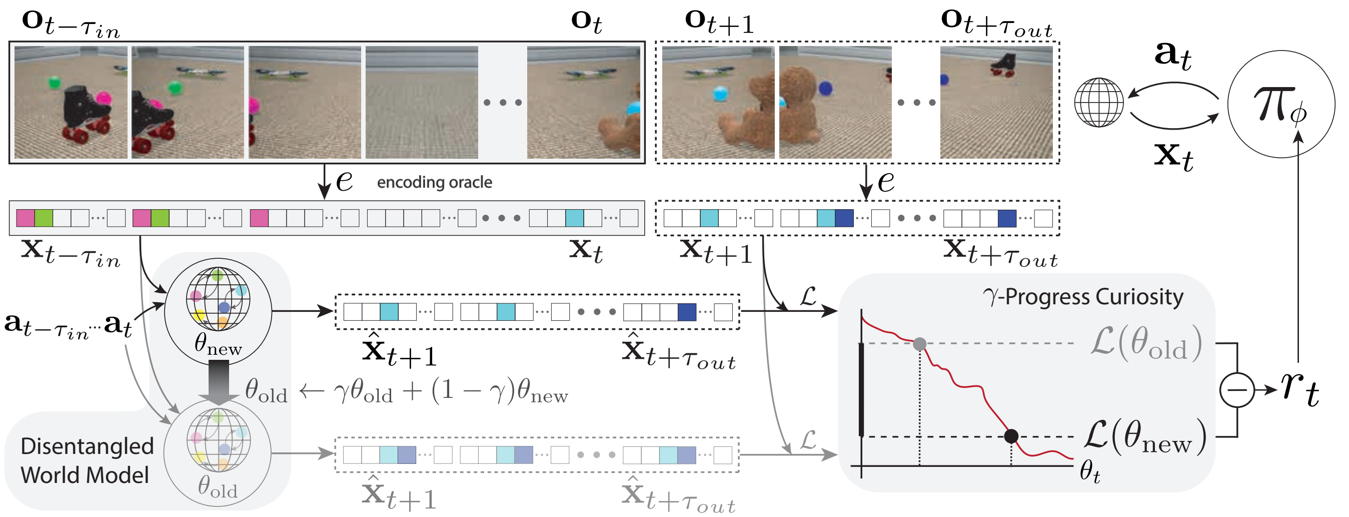

$\mathbf{o}_{t-\tau_{in}}$
$\mathbf{o}_t$
$\mathbf{o}_{t+1}$
$\mathbf{o}_{t+\tau_{out}}$
$\mathbf{a}_t$
$\mathbf{x}_t$
π_ϕ
e
encoding oracle
e
$\mathbf{x}_{t-\tau_{in}}$
$\mathbf{x}_t$
$\mathbf{x}_{t+1}$
$\mathbf{x}_{t+\tau_{out}}$
$\mathbf{a}_{t-\tau_{in}} \cdots \mathbf{a}_t$
θ_{new}
$\hat{\mathbf{x}}_{t+1}$
$\hat{\mathbf{x}}_{t+\tau_{out}}$
$\mathcal{L}$
γ-Progress Curiosity
$\mathcal{L}(\theta_{\text{old}})$
$\mathcal{L}(\theta_{\text{new}})$
θ_t
r_t
$\theta_{\text{old}} \leftarrow \gamma\theta_{\text{old}} + (1-\gamma)\theta_{\text{new}}$
Disentangled World Model
θ_{old}
$\hat{\mathbf{x}}_{t+1}$
$\hat{\mathbf{x}}_{t+\tau_{out}}$
$\mathcal{L}$

Figure 2: **Curious Neural Agent Architecture**. The curious neural agent consists of a *disentangled world model* and a *progress-driven controller*. The disentangled world model contains independent component networks that each learn the dynamics of one external agent behavior. The curious agent's observations $\mathbf{o}_t$ are passed through an encoding oracle e that returns an object-oriented representation $\mathbf{x}_t$ containing the positions of external agents that are in view, auxiliary object positions, and the curious agent's orientation. Both the new (solid) and old (faded) models take as input $\mathbf{x}_{t-\tau_{in}:t}$, route appropriate behavior-wise inputs to each component network, and jointly predict $\hat{\mathbf{x}}_{t:t+\tau_{out}}$. The old model weights, θ_{old}, are slowly updated to the new model weights θ_{new}. The controller, π_ϕ, is optimized to maximize γ-Progress reward: the difference $\mathcal{L}(\theta_{old}) - \mathcal{L}(\theta_{new})$.

$(\mathbf{x}^{ext}_{t-\tau_{in}:t,I_k}, \mathbf{x}^{aux}_{t-\tau_{in}:t}, \mathbf{x}^{ego}_{t-\tau_{in}:t}, \mathbf{a}_{t-\tau_{in}:t+\tau_{out}})$, where $\mathbf{a}$ denotes the curious agent's actions, and outputs $\hat{\mathbf{x}}^{ext}_{t:t+\tau_{out},I_k}$. The outputs of the component network are concatenated to get the final output $\hat{\mathbf{x}}^{ext}_{t:t+\tau_{out}} = (\hat{\mathbf{c}}_{t:t+\tau_{out}}, \hat{\mathbf{m}}_{t:t+\tau_{out}})$.

The world model loss is:

$$\mathcal{L}(\theta, \mathbf{x}_{t-\tau_{in}:t+\tau_{out}}, \mathbf{a}_{t-\tau_{in}:t+\tau_{out}}) = \sum_{t'=t}^{t+\tau_{out}} \sum_{i=1}^{N_{ext}} \mathbf{m}_{t',i} \cdot \|\hat{\mathbf{c}}_{t',i} - \tilde{\mathbf{c}}_{t',i}\|_2 + \mathcal{L}_{ce}(\hat{\mathbf{m}}_{t',i}, \mathbf{m}_{t',i}) \quad (1)$$

where $\mathcal{L}_{ce}$ is cross-entropy loss. Each component network ω_{θ^k} is a two-layer Long Short-Term Memory (LSTM) followed by a two-layer Multi Layer Perceptron (MLP), with number of hidden units adapted to the number of external agents modeled.

Progress-driven Controller. We propose **γ-Progress**, a scalable progress-based curiosity signal which approximates learning progress by the difference in the losses of an old model and a new model. The old model weights, θ_{old}, lag behind those of the new model, θ_{new}, with a simple update rule: $\theta_{old} \leftarrow \gamma\theta_{old} + (1-\gamma)\theta_{new}$, where γ is scalar mixing constant.

The curiosity reward is:

$$R(\mathbf{x}_t) = \mathcal{L}(\theta_{new}, \mathbf{x}_{t-\tau_{in}-\tau_{out}:t}, \mathbf{a}_{t-\tau_{in}-\tau_{out}:t}) - \mathcal{L}(\theta_{old}, \mathbf{x}_{t-\tau_{in}-\tau_{out}:t}, \mathbf{a}_{t-\tau_{in}-\tau_{out}:t}) \quad (2)$$

Our controller π_ϕ follows an ε-*greedy* sampling scheme with respect to a Q-function Q_ϕ trained with the curiosity reward in Eq. 2. Q_ϕ is parametrized by a two-layer MLP with 512 hidden units that takes as input $\mathbf{x}_{t-2:t}$ and outputs estimated state-action values for all 9 possible actions. Q_ϕ is updated with the DQN (Mnih et al., 2013) learning algorithm.

Results

In this section, we evaluate how the world model performance depends on specific choices in the curiosity signal and the disentangled architecture. We also analyze attentional behavior underlying performance increases.

Performance

We evaluate our method's ability to learn interactively on two metrics: *end performance* (inverse of final world model loss) and *sample complexity* (rate of reduction in loss with respect to the number of environment interactions). The environment is instantiated with four external agent types: static, periodic, noise, and animate. We compare AWML performance for the following methods:

γ-Progress (Ours) is our proposed variant of progress curiosity, with θ_{old}, a geometric mixture of all past models as in Eq.2.

δ-Progress (Achiam & Sastry, 2017; Graves, Bellemare, Menick, Munos, & Kavukcuoglu, 2017) is another variant of progress curiosity which uses one past model δ steps behind the current model to compute progress. It requires careful tuning of the δ parameter and is intractable in practice as memory usage grows $O(\delta)$. We use $\delta = 1$, found to be best performing by a hyperparameter search across practical values of δ.

RND (Burda et al., 2018) is a novelty-based method that trains a predictor network to match the outputs of a random state encoder. States for which the predictor fails to match the random encoder are deemed "novel", and receive high reward.

Disagreement (Pathak et al., 2019) assumes that future world model loss reduction is proportional to the prediction variance of an ensemble of N world models. We use $N = 3$, as we found $N > 3$ impractical due to memory constraints.

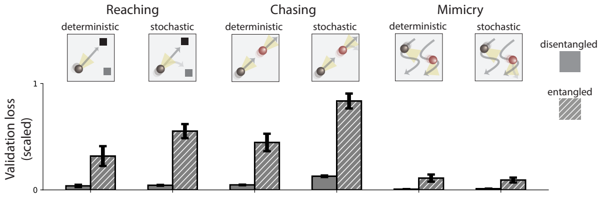


Figure 3: **Asymptotic Model Performance**. Final performance of the disentangled model and entangled ablation.

Adversarial (Pathak et al., 2017; Stadie, Levine, & Abbeel, 2015) uses current world model loss as the reward and is therefore susceptible to the white noise problem. We use the ℓ_2 prediction loss of the model.

Random chooses actions uniformly at random among the 9 possible rotations.

Figure 4 shows end performance (first row) and sample complexity (second row). γ-Progress has (1) higher end performance on all baselines and tasks, and (2) lower sample complexity than Disagreement, Adversarial, and Random baselines on all behaviors, and RND and δ-Progress on all but one behavior, tying on stochastic chasing. Upon visual inspection of model predictions, we found prediction quality correlates with performance. See visualizations at https://bit.ly/2uf7lEY.

World model architecture ablation

To evaluate the importance of disentanglement in world model architecture, independently of controller choice, we produce datasets for offline training for each task (excluding peekaboo, since the behavior is dependent on the observer's choices, no policy-independent offline training dataset can be constructed). We then train the world model to convergence. We compare the loss of our disentangled world model to an *entangled* LSTM architecture that instead takes as input and predicts the coordinate and observability information of all external agents together. As seen in Figure 3, the disentangled architecture significantly outperforms the entangled ablation.

Behavior analysis

Because animate attention is an important component of social behavior in the real world, we sought to determine to what extent the artificial system also exhibited the behavior. As seen in Figure 5, the γ-Progress agents spend substantially more time attending to animate behaviors than do alternative policies. This increased animate-inanimate attention differential often corresponds to a characteristic attentional "bump" that occurs early as the γ-Progress curious agent focuses on animate external behaviors quickly before eventually "losing interest" as prediction accuracy is achieved.

Baselines display two distinct modes in failing to exhibit animate attention. The first is *attentional indifference*, in which the curious agent finds no particular external behavior interesting. The second is *white noise fixation*, where the observer is captivated by the noise external agents. δ-Progress, a direct information gain measure, had no white noise failure but frequently led to attentional indifference as the new and old world models, separated by a fixed time difference, were often too similar to generate a useful curiosity signal. Non-progress-based curiosity signals exhibited both kinds of failure modes but were more dominated by white noise. RND suffers from white noise due to the fact that our noise behaviors have the most diffuse visited state distribution. We also observe that for noise behaviors, a world model ensemble does not collectively converge to a single mean prediction, and as a result Disagreement finds the behavior highly interesting. Finally, Adversarial fails since noise behaviors yield the highest prediction errors. Overall, emergence of animate attention is highly correlated with prediction performance, suggesting that γ-Progress succeeds because its ability to flexibly estimate information gain allows it to focus on more informative interactions.

Discussion and future directions

In this work, we address the problem of how to design an agent that can efficiently learn to make effective world models of

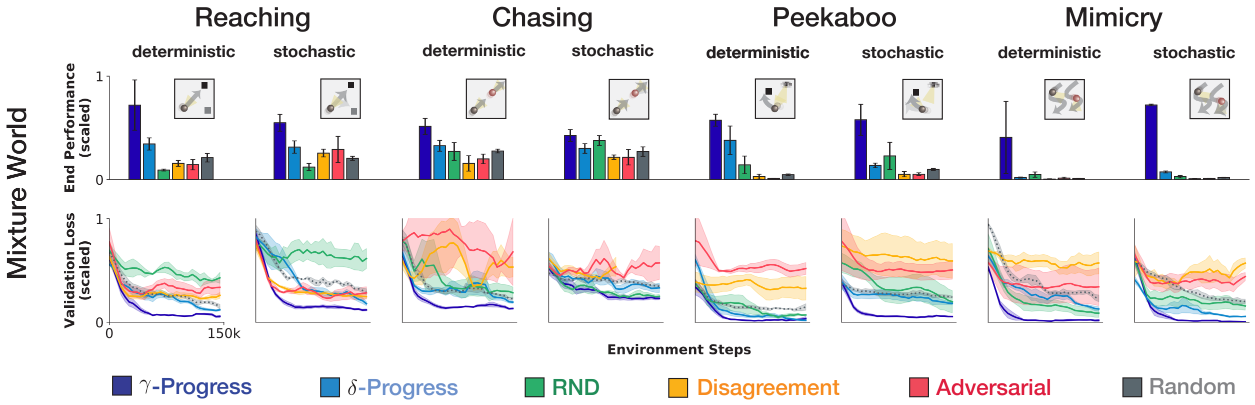
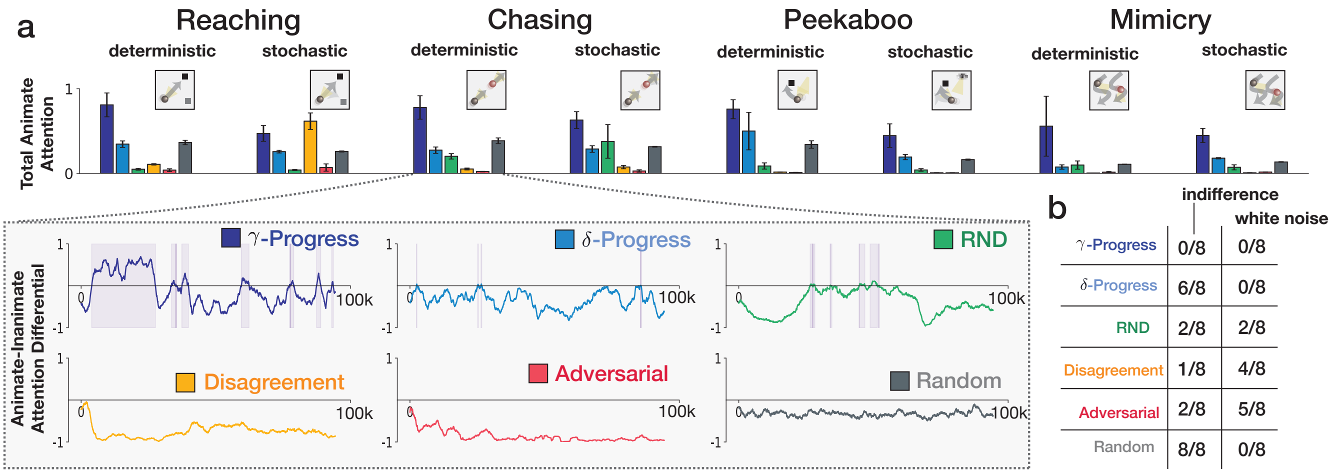


Figure 4: **Learning Efficiency**. Results shown for 8 experiments with the animate external agent varied according to the column labels. End performance (top row) is the average of the last 5 validation losses. Sample complexity plots (bottom row) show validation losses every 5000 environment steps. Error bars/regions are computed over 5 seeds. γ-Progress achieves lower sample complexity than all baselines on 7/8 behaviors while tying with RND and δ-Progress on the stochastic chasing. Notably, γ-Progress also outperforms baselines in end performance. See prediction visualizations at https://bit.ly/2uf7lEY



	indifference	white noise
γ-Progress	0/8	0/8
δ-Progress	6/8	0/8
RND	2/8	2/8
Disagreement	1/8	4/8
Adversarial	2/8	5/8
Random	8/8	0/8

Figure 5: **Attention Patterns**. a) The bar plot shows the total animate attention, which is the ratio between the number of time steps an animate external agent was visible and the number of time steps a noise external agent was visible. The zoom-in box plots show the differences between mean attention to the animate external agents and the mean of attention to the other agents in a 500 step window, with periods of animate preference highlighted in purple. Results are averaged across 5 runs. γ-Progress displays strong animate attention while baselines are either indifferent, e.g δ-Progress, or fixating on white noise, e.g Adversarial. b) Fraction of indifference and white noise failures, out of eight tasks.

"typical social scenes" — environments that are rich with other entities, including both inanimate objects of static and dynamic varieties, and animate agents. We show that an architecture with a disentangled world model and a controller based on γ-Progress curiosity is one possible solution.

Success of our curious neural agent is driven in large part by the emergence of animate attention. A standard hypothesis is that the brain has a built-in animacy-detection module (Leslie, 1994). Our modeling results suggest that animate attention may instead arise from a more general curiosity-driven learning process.

Intuitively, the disentangled architecture performs better because it ignores spurious correlations between causally-unrelated events in the agent's data stream. Formalizing this intuition and explaining why this is particularly salient in our current environment, in contrast to some other situations (Locatello et al., 2018), is an important future direction. Interestingly, the disentangled architecture shares a key feature with Theory of Mind, which involves the ability to predict the behaviors of other agents as a function of inferred mental states, such as beliefs, desires, and goals (Astington et al., 1990; Premack & Woodruff, 1978; Wellman, 1992). A core, though often unstated, assumption behind Theory of Mind is the agent-centric allocation of computational resources. Our disentangled model builds this in as a key feature, suggesting that at least one possible function of Theory of Mind may be to enable statistical disentangling. This certainly requires considerable follow-up work to substantiate.

Human behavior To quantify how the emergent behavior matches that of humans, we have run a pilot human subject experiment (Figure 6a) in which we conveyed static, periodic, animate, and noise stimuli to twelve human participants via spherical robots moving along a mat, while measuring patterns of attention via a mobile eye tracker. We find average fixation proportions favoring the animate stimuli. Furthermore, a γ-Progress network tasked to predict the virtual robot trajectories produces a similar aggregate attentional fixation pattern. In follow-up work, we aim to make a finer model comparison to the behavior of humans shown a diverse array of stimuli.

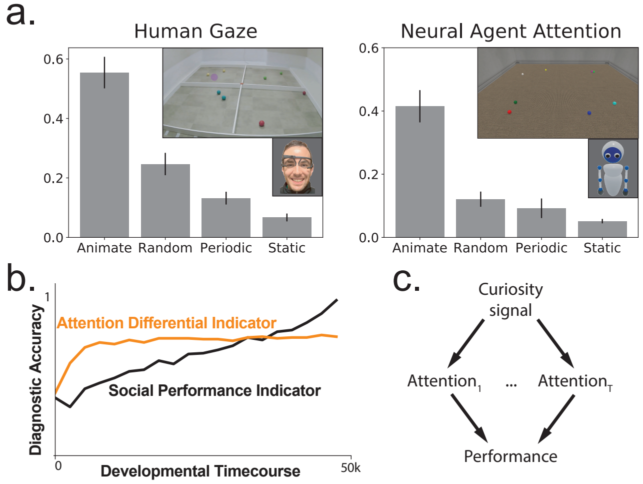


Figure 6: **Modeling human behavior**. (a) Human adults wear a mobile eye tracker while watching displays consisting of four sets of spherical robots travelling along a mat. Human and model fixation proportions are similar. (b) Accuracy of early indicators of final performance, as a function of time, and (c) factor analysis hypothesis: curiosity signal determines attention, which determines final performance.

Early indicator analysis Eventually, we would like to use curiosity-driven learning as a model for intrinsic motivation in early childhood. In this interpretation, the attention timecourse is a readily observable behavioral metric, and performance represents a more difficult-to-obtain measure of social acuity. Variation in curiosity signal would, in this account, be a latent correlate of developmental variability. For example, Autism Spectrum Disorder is characterized by both difference in low-level facial attention (Jones & Klin, 2013; Constantino et al., 2017) and high-level social acuity (Hus & Lord, 2014).

Motivated by this example, we sought to determine whether the easily-measurable low-level attention could be used as an early indicator of high-level social prediction performance. To perform an early indicator analysis, we thus train two models: (1) $\text{PERF}_{\leq T}$, which takes performance before time T as input, and (2) $\text{ATT}_{\leq T}$, which takes attention before time T as input. As seen in Figure 6b, $\text{ATT}_{\leq T}$ is an effective predictor of late social performance, and in fact, throughout most of the timecourse, a more accurate indicator than direct measurement of early-stage model performance itself. The overall situation is conveyed by the factor diagram in Figure 6c. Translating this modeling result into a real-world experimental population could lead to substantial improvements in affordable, early-deployable diagnostics of developmental variability.

References

Achiam, J., & Sastry, S. (2017). Surprise-based intrinsic motivation for deep reinforcement learning. *arXiv preprint arXiv:1703.01732*.

Astington, J. W., Harris, P. L., & Olson, D. R. (1990). *Developing theories of mind.* CUP Archive.

Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., ... Pascanu, R. (2018, June). Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261*.

Burda, Y., Edwards, H., Storkey, A., & Klimov, O. (2018). *Exploration by random network distillation.*

Constantino, J. N., Kennon-McGill, S., Weichselbaum, C., Marrus, N., Haider, A., Glowinski, A. L., ... Jones, W. (2017, Jul). Infant viewing of social scenes is under genetic control and is atypical in autism. *Nature*, *547*(7663), 340–344.

Foster, M., et al. (1973). Visual attention to non-contingent and contingent stimuli in early infancy.

Frankenhuis, W. E., House, B., Barrett, H. C., & Johnson, S. P. (2013). Infants' perception of chasing. *Cognition*, *126*(2), 224–233.

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

Graves, A., Bellemare, M. G., Menick, J., Munos, R., & Kavukcuoglu, K. (2017). Automated curriculum learning for neural networks. In *Proceedings of the 34th international conference on machine learning-volume 70* (pp. 1311–1320).

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

Hus, V., & Lord, C. (2014). The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *Journal of autism and developmental disorders*, *44*(8), 1996–2012.

Johnson, S. C. (2003). Detecting agents. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 549–559.

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399.

Leslie, A. M. (1994). Tomm, toby, and agency: Core architecture and domain specificity. *Mapping the mind: Domain specificity in cognition and culture*, *29*.

Locatello, F., Bauer, S., Lucic, M., Rätsch, G., Gelly, S., Schölkopf, B., & Bachem, O. (2018). Challenging common assumptions in the unsupervised learning of disentangled representations. *arXiv preprint arXiv:1811.12359*.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), 255–260.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. *arXiv preprint arXiv:1705.05363*.

Pathak, D., Gandhi, D., & Gupta, A. (2019). Self-supervised exploration via disagreement. *arXiv:1906.04161*.

Piaget, J. (1952). The origins of intelligence in children. , *8*.

Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*(1), 1–16.

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

Schmidhuber, J. (2010, Sept). Formal theory of creativity, fun, and intrinsic motivation (1990 – 2010). *IEEE Transactions on Autonomous Mental Development*, *2*(3), 230-247.

Smith, K., Mei, L., Yao, S., Wu, J., Spelke, E., Tenenbaum, J., & Ullman, T. (2019). Modeling expectation violation in intuitive physics with coarse probabilistic object representations. In *Advances in neural information processing systems* (pp. 8983–8993).

Stadie, B., Levine, S., & Abbeel, P. (2015). Incentivizing exploration in reinforcement learning with deep predictive models. *arXiv preprint arXiv:1507.00814*.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94.

Wellman, H. M. (1992). *The child's theory of mind.* The MIT Press.