From Curiosity to Competence: How World Models Interact with the Dynamics of Exploration

Fryderyk Mantiuk[1,*]**(fryderyk.mantiuk@student.uni-tuebingen.de), Hanqi Zhou**[1,2,*]**(hanqi.zhou@uni-tuebingen.de) & Charley M. Wu**[1,2]

[1] Human and Machine Cognition Lab, University of Tübingen, Tübingen, Germany

[2] Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics

[*] shared authorship

Abstract

What drives an agent to explore the world while also maintaining control over the environment? From a child at play to scientists in the lab, intelligent agents must balance curiosity (the drive to seek knowledge) with competence (the drive to master and control the environment). Bridging cognitive theories of intrinsic motivation with reinforcement learning, we ask how evolving internal representations mediate the trade-off between curiosity (novelty or information gain) and competence (empowerment). We compare two model-based agents using handcrafted state abstractions (Tabular) or learning an internal world model (Dreamer). The Tabular agent shows curiosity and competence guide exploration in distinct patterns, while prioritizing both improves exploration. The Dreamer agent reveals a two-way interaction between exploration and representation learning, mirroring the developmental co-evolution of curiosity and competence. Our findings formalize adaptive exploration as a balance between pursuing the unknown and the controllable, offering insights for cognitive theories and efficient reinforcement learning.

Keywords: intrinsic motivation; exploration; representation learning; world models; reinforcement learning; curiosity

Introduction

Picture a child playing with two brand-new toys. One lights up predictably when a button is pressed, while the other flickers randomly, indifferent to the child's actions. Which toy captivates the child longer? The answer depends not just on novelty or control but on a deeper interplay between two fundamental drives: *curiosity*, which compels us to seek new knowledge, and *competence*, which drives us to leverage what we know to influence our environment (Meltzoff, Waismeyer, & Gopnik, 2012; Nussenbaum & Hartley, 2019).

This highlights a broader puzzle about human exploration: we are drawn to both the *unknown*, in our quest to reduce uncertainty, and the *predictable*, in our desire to exert influence over the world (Cogliati Dezza, Schulz, & Wu, 2022; Nelson, 2005; Ten, Oudeyer, Sakaki, & Murayama, 2024). Curiosity-related drives, such as *novelty* and *information gain*, push us to reduce uncertainty and build better mental models of the world (Gottlieb, Oudeyer, Lopes, & Baranes, 2013; Modirshanechi, Brea, & Gerstner, 2022; Ten et al., 2024). Competence-related drives, such as *empowerment* and skill learning, motivate us to predict and control outcomes (Aubret, Matignon, & Hassas, 2023; Gopnik, 2024; Klyubin, Polani, & Nehaniv, 2005; Poli, Serino, Mars, & Hunnius, 2020; Salge, Glackin, & Polani, 2014).

At first glance, these drives might seem sequential: first curiosity for learning, then competence to act effectively. But the reality is far more recursive. Consider a child learning to walk: only by developing competence in mobility can they access new environments to satisfy their curiosity. Conversely, the child's curiosity about more distant places may fuel their persistence in mastering locomotion. This bidirectional relationship raises a critical question: How do curiosity and competence co-evolve, as agents build and refine their understanding of the world?

Existing reinforcement learning (RL) agents struggle with this balance (Dulac-Arnold et al., 2021; Gruaz, Modirshanechi, & Brea, 2024; Ocana, Capobianco, & Nardi, 2023). Curiosity-driven agents, typically prioritizing novelty or information gain, often fall prey to the "noisy TV problem", getting distracted by random, uncontrollable stimuli that offer no meaningful opportunities for mastery (Burda, Edwards, Storkey, & Klimov, 2018; Pathak, Agrawal, Efros, & Darrell, 2017; Schmidhuber, 1991). Conversely, competence-focused agents, which maximize empowerment or control, often assume fixed world models, neglecting how exploration might reshape those models (Brändle, Stocks, Tenenbaum, Gershman, & Schulz, 2023; Du et al., 2023; Lidayan et al., 2025). Humans, in contrast, dynamically adjust their focus: a toddler confined to a crib may inspect every toy in meticulous detail, while the same child, once mobile, might prioritize breadth over depth, racing to explore new rooms. What enables this adaptive prioritization?

A key piece of the puzzle lies in the agent's "world model"—the internal representation of environmental dynamics that guides predictions and decisions (Ha & Schmidhuber, 2018; Hafner, Pasukonis, Ba, & Lillicrap, 2023). When uncertainty is high, curiosity dominates, driving exploration to refine the model. As confidence grows, competence takes precedence, leveraging the model to achieve goals. Yet this is not a one-way transition. Competence unlocks new frontiers for curiosity (e.g., learning to walk expands the horizon for exploration), while curiosity generates the knowledge needed for higher-order competence (e.g., understanding physics enables tool use). This creates a feedback loop: world models shape exploration strategies, which in turn reshape the models themselves.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

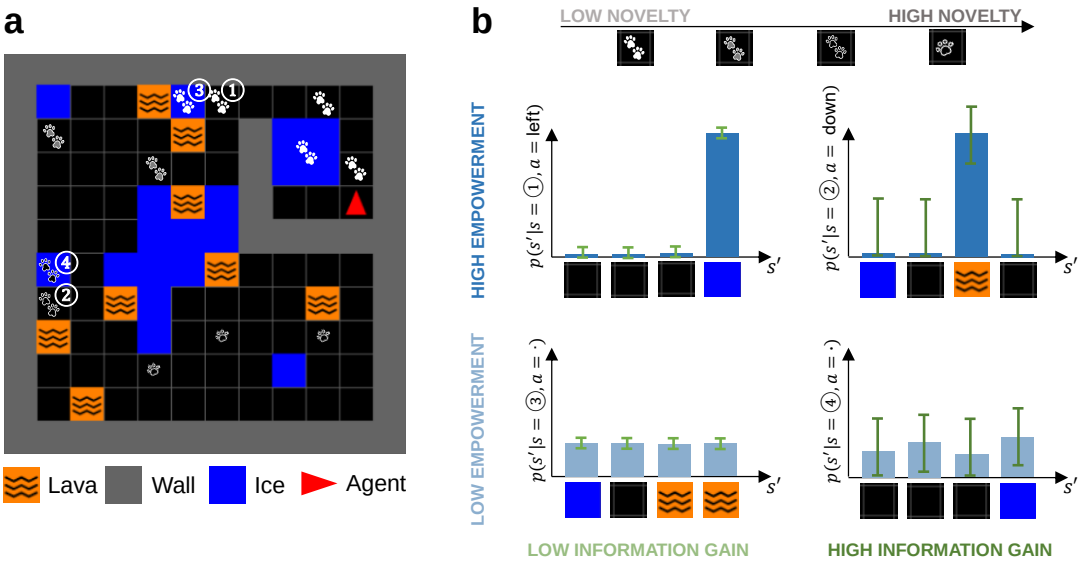

Figure 1: **Task overview**. **a**) The mixed-state playground has three different cells: lava, wall, and ice. The agent (red triangle) starts from the upper right corner and has to explore the environment while avoiding death (falling into lava). **b**) Novelty depends only on the visitation frequency for each state, regardless of the state types. Information gain is low for transitions which the agent has already learned to predict well (cells ① and ③), and high for transitions that the agent is less confident about (cells ② and ④). Empowerment is low in slippery ice cells where the agent cannot control where it will end up (cells ③ and ④), and high in states with many predictable action outcomes (cells ① and ②).

Goal and scope. In this work, we investigate how agents balance curiosity and competence during learning and exploration. We ask: under what conditions should exploration prioritize curiosity versus competence, and how do these priorities shift as world models evolve? By integrating insights from cognitive science and RL, we aim to shed light on the mechanisms that allow humans and artificial agents to navigate the fine line between learning and doing, uncertainty and mastery, and curiosity and competence.

Specifically, we compare three intrinsic motivations in sparse-reward environments: *novelty* (collecting diverse experiences), *information gain* (epistemic uncertainty reduction), and *empowerment* (perceived control over future states). We simulated and analyzed two different agents on grid worlds: (1) a **Tabular** Q-learning agent whose state representations are predefined by handcrafted features; and (2) a **Dreamer** world-model agent (Hafner et al., 2023) that learns a latent representation of the world during exploration.

With simulations in grid worlds, we analyze how each intrinsic motivation drives divergent exploration patterns, showing that they are neither functionally redundant nor universally optimal. Empirically, we show how environmental structure (stochastic dynamics and the presence of fatal danger) and the agent's evolving world model systematically modulate the efficacy of exploration strategies. Together, the findings clarify how safe and effective exploration could arise from balancing curiosity-driven information seeking with competence-driven control, each offering distinct advantages depending on the environmental context.

Methods

In this paper, we investigate the interactions between intrinsic motivations and model-based state representations using RL agents in a grid-based environment (MiniGrid; Chevalier-Boisvert et al., 2023). The following sections describe the key elements—the environment to explore, the agents that explore it, and intrinsic rewards that determine *how* the agents explore the environment.

Environment

Our environments are based on MiniGrid (Chevalier-Boisvert et al., 2023), a 2D grid-world framework which encourages exploration. To mimic real-world challenges, we first extend MiniGrid and design a heterogeneous playground (Fig. 1a) which has three state types: lava (imposing irreversible penalties, terminating the episode), ice (stochastic transitions), and barriers (constraining mobility; e.g., walls or moving balls). These states operationalize fundamental trade-offs between risk, uncertainty, and control that are also observed in naturalistic decision-making.

The agent starts in the upper right corner, requiring it to first exit through a bottleneck of walls and lava. If it succeeds, it could later find a larger area to explore in the bottom right, either via a shortcut through an icy and stochastic patch or by taking a longer but more reliable route.

Agent-environment interactions

Interactions between the agent and the environment are defined as a Partially Observable Markov Decision Process (POMDP) (Sutton, Barto, et al., 1998). At each time step,

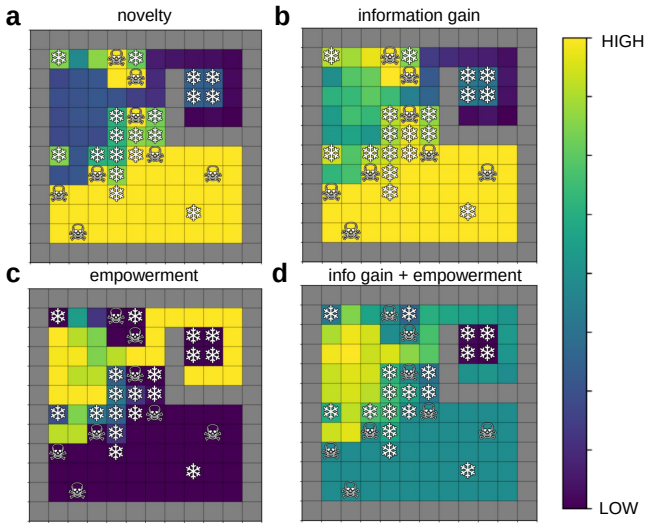

Figure 2: **Intrinsic motivation heatmaps**. A snapshot of each motivation metric in all tiles in Simulation 1 (skull represents fatal lava tiles, snowflake represents stochastic ice tiles). **a**) Novelty and **b**) information gain drive the agent to explore, but fail to recognize that the stochasticity of the slippery ice cannot be reduced through experience. **c**) Empowerment learns to avoid ice and lava, preferring the predictable dynamics of neutral states instead. **d**) Summing information gain and empowerment obtains the highest rewards in areas with controllable yet not fully explored dynamics, sensibly prioritizing reducible over irreducible uncertainty.

the agent receives an observation $o \in O$, which may include the agent's immediate surroundings, such as walls, objects, or other entities within its field of view.

Based on its internal representation z of the current observation o, the agent selects action $a \sim \pi(a \mid z)$ from a discrete action space given its policy π. The agent can face four directions and has three basic actions: move forward, or turn left or right. After executing an action, the agent transitions to a new state s' according to the environment's transition dynamics $p(s' \mid s,a)$. The agent may also receive an extrinsic reward r for reaching a goal state (see Simulation 2: generalization phase), although all agents initially need to rely on intrinsic rewards alone for the majority of the training, to build up their model of the environment.

Agents

A core challenge in studying exploration is that real-world agents must simultaneously learn *what parts of the world to focus on* (state representations) and *how the world works* (transition dynamics). Prior work often assumes that intrinsic motivations are computed on fixed and hand-crafted state representations (Du et al., 2023; Lidayan et al., 2025), or by learning these representations and policies (or value functions) separately (Burda et al., 2018; Ferrao & Cunha, 2025). However, this sidesteps a critical question: how do agents explore effectively when their understanding of "states" is formed by their experiences?

To evaluate how intrinsic motivations are impacted by changing internal representations, we conduct two progressive simulations: (1) a Tabular agent with predefined state representations to probe the interplay between different intrinsic motivations under fixed representations, and (2) a Dreamer agent that autonomously learns latent state representations from observations in the pixel space to further understand the role of these evolving representations.

Tabular agent. This first set of simulations provides a controlled baseline for isolating the dynamics of learning a transition model of the environment (i.e., model-based RL) and providing intuitions of the exploratory behavior under different motivations. Here, state representations are predefined, allowing the agent to focus on refining its transition model.

Our simple fixed representation reduces states to (position, orientation) tuples $z = (x, y, \theta) \in \mathbb{N}^3$ where x and y are the agent's grid coordinates and θ is the facing direction. On top of this, we can estimate a count-based transition model of the environment dynamics:

$$p(z' \mid z,a) = \frac{\alpha N(z,a,z') + 1}{\sum_{z' \in Z} N(z,a,z') + 1} \tag{1}$$

where $N(z,a,z')$ counts transitions from state z to z' under action a, scaled by an update factor $\alpha = 100|Z|$, which allows the agent to quickly update from the uninformative uniform prior towards the counts it actually experienced.

Following Gruaz et al. (2024), the agent learns Q-values with fast model-based updating through prioritized sweeping with the learned dynamics (Eq. 1):

$$Q(z,a) = \sum_{z' \in Z} p(z' \mid z,a) \left[r(z,a,z') + \gamma \max_{a'} Q(z',a') \right], \tag{2}$$

where the discount factor $\gamma = \sqrt[|Z|/2]{0.5}$ is defined by the number of states $|Z|$, and the immediate reward $r(z,a,z')$ is computed as the respective intrinsic objective of novelty, information gain or empowerment. To select actions, the agent uses a greedy policy over the Q-values, $p(a \mid z) = \arg\max_a Q(z,a)$.

Dreamer agent. To test how learned representations mediate exploration, we employ DreamerV3 (Hafner et al., 2023) as a world-model reinforcement learning agent that constructs discrete latent states from raw pixel observations without task-specific supervision. Specifically, the world model f_{wm} has two main components: an encoder $p_\phi(z' \mid z,a,o)$, which encodes raw observations o' into discrete latent states z, and a dynamics predictor $p_\phi(z' \mid z,a)$, which predicts z' from previous z and action a. These latent states distill sensory inputs into abstract variables (e.g., encoding "door states" as locked/unlocked), analogous to unsupervised category formation. For implementation and simulations, we largely follow the same architecture and hyperparameters as in Ferrao

and Cunha (2025), but use the partial pixel-based observation provided by MiniGrid, rather than the semantic observations which they used.

Intrinsic motivation

Intrinsic motivation drives agents to explore and learn, independent of extrinsic rewards. We focus on three key types: *novelty*, *information gain*, and *empowerment*, commonly used in both machine learning and psychology.

Novelty encourages exploration by prioritizing unfamiliar states (Fig. 2a; Berlyne, 1950; Dubey & Griffiths, 2020; Schwartenbeck, FitzGerald, Dolan, & Friston, 2013). We define novelty as the state surprise:

$$R_{\text{Novelty}}(z,a,z') = -\log \frac{N(z)}{\sum_{\tilde{z}\in\mathcal{Z}} N(\tilde{z})}, \tag{3}$$

where $N(z)$ is a count of how often the agent has seen the state representation z. This metric focuses on exploring less-visited states, but is agnostic about state-action dynamics (Fig. 1b). Since DreamerV3 has a discrete latent space, we can use the same count-based novelty for both Tabular agents and Dreamer agents.

Information gain drives agents to perform actions in states where they are uncertain about the outcome, aiming to improve their world model (Fig. 2b; Gottlieb et al., 2013; Nelson, 2005; Oudeyer & Kaplan, 2007; Still & Precup, 2012). This improvement is often quantified by how much the uncertainty in the world model is reduced after observing a transition. For the tabular world model, we compute the *predicted* information gain, following past work (Gruaz et al., 2024; Kolossa, Kopp, & Fingscheidt, 2015; Little & Sommer, 2013; Modirshanechi et al., 2022):

$$R_{\text{IG}_{\text{Tabular}}}(z,a,z') = \mathbb{E}_{\hat{z}'\sim p(Z'|z,a)}\left[\text{KL}\left(p(Z' \mid z,a,\hat{z}') \;||\; p(Z' \mid z,a)\right)\right] \tag{4}$$

The Dreamer agent requires a slightly different computation, since computing the expectation over all possible future states would require an intractable amount of forward passes through the dynamics predictor. Here, instead of computing the predicted information gain, we we can compute it *retrospectively* (after observing a transition), by using the KL divergence between posterior (latent state predicted by encoder that has access to the observation) and prior (latent state predicted by dynamics predictor before seeing the observation) as an information gain reward:

$$R_{\text{IG}_{\text{Dreamer}}}(z,a,z') = \text{KL}(p_\phi(z' \mid z,a,o') \;||\; p_\phi(z' \mid z,a)) \tag{5}$$

Empowerment reflects the degree of control an agent has over its environment (Fig. 2c; Du et al., 2023; Gopnik, 2024; Klyubin et al., 2005; Lidayan et al., 2025) by maximizing the mutual information between its actions and resulting states:

$$\mathfrak{E}(s) = \max_{\omega(a|s)} I\left(S';A \mid s\right). \tag{6}$$

Thus, empowerment encourages an agent to visit states where it has the greatest influence over future outcomes, where both a greater diversity of outcomes and less stochasticity in the mapping of actions to outcomes contribute to more empowerment (Fig. 1c). Empowerment thus prioritizes states where the agent has the highest agency, independent of uncertainty or visitation frequency.

Again, we compute this intrinsic motivation over the latent representations z rather than the observations o or true states s. In the Tabular agent, we compute 1-step empowerment using the Blahut-Arimoto algorithm, which iteratively optimizes the action distribution ω to maximize the mutual information between actions and resulting states (Dupuis, Yu, & Willems, 2004). The empowerment reward is then simply:

$$R_{\text{Emp}_{\text{Tabular}}}(z,a,z') = \mathfrak{E}(z) = \max_{\omega(a|z)} I\left(Z';A \mid z\right) \tag{7}$$

In the Dreamer agent, applying Blahut-Arimoto to find the optimal action distribution is intractable, so we assume a random policy $\omega = \mathcal{U}$ following Seitzer, Schölkopf, and Martius (2021) and compute the reward

$$R_{\text{Emp}_{\text{Dreamer}}}(z,a,z') = I(Z';A \mid z) = H(Z' \mid z) - H(Z' \mid A,z). \tag{8}$$

Combinations of empowerment with any of the two knowledge-seeking measures into a unified reward for the same agent have not been explored to the best of our knowledge. To evaluate whether competence-seeking in the form of empowerment could benefit from knowledge-seeking (and vice versa), we choose to combine information gain and empowerment in two simple ways, via a sum or a product. This combination is chosen for two reasons: Firstly, from a theoretical perspective, information gain aims to specifically improve the dynamics model—which is necessary to compute empowerment—while novelty only rewards seeing infrequently visited states. Secondly, from an empirical perspective, we observed that information gain resulted in a more thorough exploration than novelty in both Tabular and Dreamer experiments.

Results

Simulation 1: tabular agent

Each agent was evaluated over 10,000 steps and 5 random rounds using different intrinsic rewards: novelty (Eq. 3), information gain (Eq. 4 and Eq. 5), empowerment (Eq. 7 and Eq. 8), along with two combinations of information gain and empowerment (sum and product) to explore their synergy. This allows us to test whether agents can distinguish reducible (epistemic) uncertainty from irreducible (aleatoric) uncertainty in environmental dynamics, by prioritizing useful unknowns (controllable floor cells) over useless unpredictability (e.g., stochastic ice cells).

The results (Fig. 3a-c) show that the most thorough discovery is achieved by a pure information-gain-seeking agent

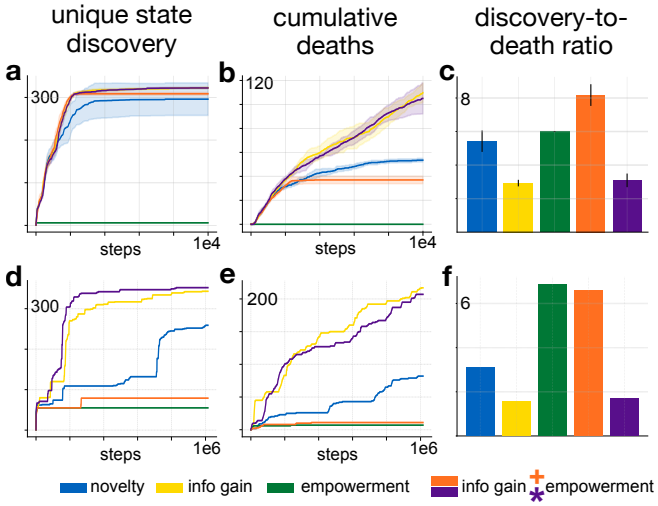

Figure 3: **Exploration patterns** of the Tabular (a-c) and Dreamer agents (d-f) on the mixed-state playground (Fig. 1). **a, d)** Total number of new states discovered over time. **b, e)** Total number of deaths over time. **c, f)** Ratio of discovered states to deaths after 1e4, and 1e6 steps respectively.

(Fig. 3a). However, this exploration comes at a cost (Fig. 3b): The agent often risks dying in its quest to perfectly predict the stochastic dynamics in ice cells, making it slip into lava. In contrast, empowerment fails to explore and stays in the starting state, avoiding death entirely.

Crucially, combining information gain and empowerment rewards via a naive sum results in a more balanced approach, leading to a higher discovery-to-death ratio (Fig. 3c): The jointly curiosity- and competence-seeking agent explores most of the environment while avoiding uncontrollability as soon as it recognizes it.

Simulation 2: dreamer agent

We now turn to simulation where the state representations are learned during exploration by a Dreamer world-model agent. Thus the intrinsic motivations are derived from dynamically evolving latent states, allowing for a bi-directional interaction. The Dreamer agent can test whether it bypasses perceptual distractions (e.g., a "noisy TV" generating irreducible randomness) to instead target controllable uncertainties, such as states where actions reliably alter transitions.

On the mixed-state playground. We begin by presenting results comparable to those of tabular agents, evaluating effective exploration (in terms of discovery and deaths) across different motivations in Figure 3d-f. Much like the Tabular agent, we can observe that an agent driven by empowerment alone exhibits poor exploration behavior (Fig. 3d). However, it is able to explore all the states in the corner room where it starts, in contrast to the Tabular agent. Yet, the Dreamer agent with empowerment is still unable to traverse the bottleneck and explore the rest of the environment, contrary to those driven by other motivations. The synergistic effects of

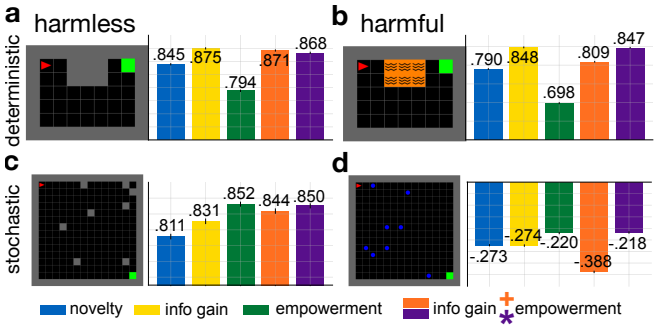

Figure 4: **Generalization performance** of the Dreamer agent on unary-state grid worlds. The unary-state environments are split into 2×2 categories, together with the agent's generalization performance under different intrinsic motivations.

the combinations of empowerment and information gain begin to mirror the Tabular setting, but are less pronounced.

On the unary-state grid. To better understand the learned world representation and its role in exploration behavior, we further isolate the impact of individual state types. We decompose the environment used in Simulation 1 into four variants with orthogonal features (Fig. 4): harmful vs. harmless and deterministic vs. stochastic. The first dimension modulates the safety of the environment, where harmful environments introduce states that result in reward penalties (lava and balls). The second dimension modulates the predictability of the environment: the stochastic environments introduce stochastic transitions in the placement of walls or path-blocking blue balls. This design allows us to test whether *novelty* distinguishes true semantic novelty from continually high combinatorial novelty from random configurations, whether *information gain* overfixates on inherently random (and unlearnable) transitions, and whether *empowerment* can maintain a locus of control when this requires more than simply standing still.

Our analysis has two phases. (1) Pretraining: The agent learns in four distinct environments using only intrinsic rewards. (2) Generalization: The agent is tested in novel environments (similar properties as pretraining but differing in layout or dynamics) with only extrinsic rewards.

Generalization performance is shown in Figure 4 (higher is better). Performance is highly context-specific: novelty and information gain excel in deterministic environments, whereas empowerment proves more effective in stochastic settings. However, hybrid approaches (combining information gain and empowerment through a sum or product) are generally equal to or better than each individual intrinsic motivation, suggesting their synergy may offer a promising compromise.

Exploration patterns show that *Novelty* had the lowest entropy of observations and actions during exploration (Fig. 5a), meaning it encountered the fewest new states and had the lowest diversity of actions. This is because the agent is satisfied

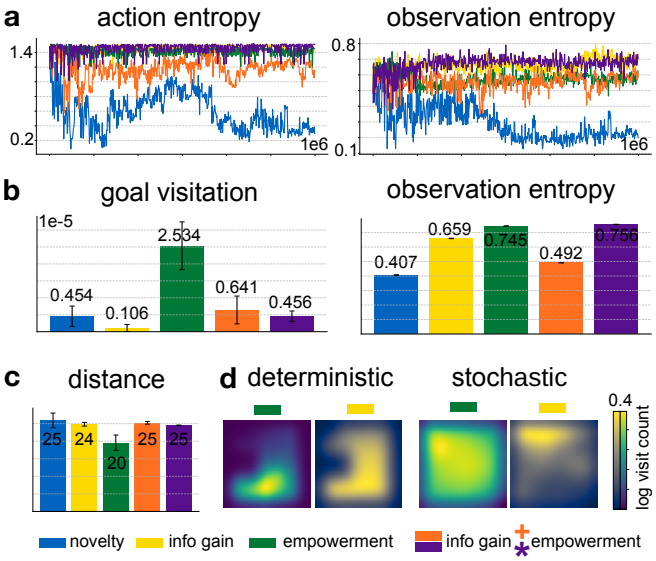

Figure 5: **Exploration patterns** of the Dreamer agent during pretraining. **a)** Observation and action entropy over time averaged over all environments. **b)** Extrinsic reward and observation entropy in the stochastic + harmless environment (averaged over time steps), which is the hardest for information gain. **c)** Distance between the agent and the lower border in deterministic environments. **d)** Heatmaps of state visitations for empowerment and information gain agents.

with a trivial form of novelty: the world model naturally develops changes in a few dimensions of its learned state representations every time, thus increasing the subjective sense of novelty without translating into concrete exploration.

Information gain improved exploration in deterministic contexts, but had trouble in stochastic ones (Fig. 4b). This aligns with our results in the tabular setting: contrary to what one might expect from theory, information gain often conflates (reducible) epistemic uncertainty with (irreducible) aleatoric uncertainty in practice. Focusing on stochastic + harmless environments (Fig. 5b), the information gain agent achieves minimal goal visits (during pretraining) despite having high observation entropy. This is due to the agent fixating on inherently unpredictable transitions (randomly moving walls) rather than on exploration.

Empowerment diverges fundamentally from novelty and information gain by prioritizing control over future states. In deterministic environments, this is maladaptive, with the agents getting stuck in a stable "comfort zone" (regions maximally distant from walls or lava or near the starting point). Quantitative analysis of wall-distance distributions further supports this, showing agents systematically position themselves to retain access to multiple future paths (Fig. 5c). Heatmaps of exploration patterns (Fig. 5d) also confirm this tendency, revealing a strong preference for obstacle-sparse regions in deterministic settings (i.e., bottom right). However, in stochastic environments, an emphasis on controllability becomes adaptive. Here, the empowerment agent abandons its comfort zone to instead roam continuously to maintain influence over outcomes. This contrasts starkly with information gain, which fixates on inherently irreducible aleatoric noise (e.g., tracking random object movements) and becomes trapped near the starting area (upper left corner).

Discussion

We investigated how intrinsic motivation mechanisms—specifically curiosity (novelty and information gain) and competence (empowerment)—guide exploration in model-based reinforcement learning (RL) agents. By comparing two architectures—a tabular agent with fixed state representations and a Dreamer agent with learned ones—we demonstrate that curiosity and competence play complementary roles in exploration and enhance generalization.

Our results show distinct trade-offs among intrinsic motivation strategies. *Novelty*-driven exploration could get stuck in local optima, going back and forth between a limited set of states. *Information gain* avoided this because it sought out reducible uncertainty which required exploring the entire environment, but it was slowed down in stochastic contexts by mistakenly fixating on irreducible uncertainty. *Empowerment* preferred deterministic dynamics—sometimes hindering exploration by staying in a controllable "comfort zone", but sometimes instead aiding exploration by actively avoiding harmful or risky stochasticity to preserve agency, thus avoiding the myopic distractions that limited curiosity-driven strategies. Hybrid strategies combining information gain and empowerment leverage this natural complementarity to achieve better exploration-safety balances in the tabular agent. This synergy was also present in the Dreamer agent with more robust generalization in novel environments, although the effects were less pronounced.

While our simplified grid-worlds enabled precise manipulation of state transitions, they exclude challenges posed by open-ended worlds. Additionally, although our computational agents were designed to isolate core principles of exploration, validating these mechanisms against human behavior (e.g., in video games like Crafter; Du et al., 2023; Lidayan et al., 2025) and over developmental timescales (Giron et al., 2023) remains essential to assess their psychological plausibility. Future work could extend our framework by (1) integrating adaptive motivation strategies that dynamically adapt the weighting of curiosity and competence based on environment type, (2) testing whether our observed empowerment-like "safety first" models of behavior could be used to describe human subjects under threat-of-shock paradigms, and (3) scaling these principles to real-world robotics tasks requiring robustness to environmental stochasticity.

In sum, we studied the co-evolution of world model learning and exploration strategies defined by different intrinsic motivation mechanisms. Our results revealed that although each intrinsic motivation had context-specific advantages, hybrid strategies yielded synergistic benefits, illustrating the complementarity of curiosity and competence.

Acknowledgments

We thank Alison Gopnik, Aly Lidayan, Eliza Kosoy, Alireza Modirshanechi, Georg Martius and Cansu Sancaktar for helpful discussions. The authors thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for supporting HZ. This research is supported as part of the LEAD Graduate School & Research Network, which is funded by the Ministry of Science, Research and the Arts of the state of Baden-Württemberg within the framework of the sustainability funding for the projects of the Excellence Initiative II. This work is supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy–EXC2064/1–390727645, and funded by the DFG under Germany's Excellence Strategy – EXC 2117 – 422037984.

References

Aubret, A., Matignon, L., & Hassas, S. (2023). An information-theoretic perspective on intrinsic motivation in reinforcement learning: A survey. *Entropy*, *25*(2), 327.

Berlyne, D. E. (1950). Novelty and curiosity as determinants of exploratory behaviour. *British journal of psychology*, *41*(1), 68.

Brändle, F., Stocks, L. J., Tenenbaum, J. B., Gershman, S. J., & Schulz, E. (2023). Empowerment contributes to exploration behaviour in a creative video game. *Nature Human Behaviour*, *7*(9), 1481–1489.

Burda, Y., Edwards, H., Storkey, A., & Klimov, O. (2018). Exploration by random network distillation. *arXiv preprint arXiv:1810.12894*.

Chevalier-Boisvert, M., Dai, B., Towers, M., de Lazcano, R., Willems, L., Lahlou, S., ... Terry, J. (2023). Minigrid & miniworld: Modular & customizable reinforcement learning environments for goal-oriented tasks. *CoRR*, *abs/2306.13831*.

Cogliati Dezza, I., Schulz, E., & Wu, C. M. (Eds.). (2022). *The drive for knowledge: The science of human information-seeking*. Cambridge University Press. doi: 10.1017/9781009026949

Du, Y., Kosoy, E., Dayan, A., Rufova, M., Abbeel, P., & Gopnik, A. (2023). What can ai learn from human exploration? intrinsically-motivated humans and agents in open-world exploration. In *Neurips 2023 workshop: Information-theoretic principles in cognitive systems.*

Dubey, R., & Griffiths, T. L. (2020). Reconciling novelty and complexity through a rational analysis of curiosity. *Psychological Review*, *127*(3), 455.

Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., & Hester, T. (2021). Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Machine Learning*, *110*(9), 2419–2468.

Dupuis, F., Yu, W., & Willems, F. M. (2004). Blahut-arimoto algorithms for computing channel capacity and rate-distortion with side information. In *International symposium oninformation theory, 2004. isit 2004. proceedings.* (p. 179).

Ferrao, J. L., & Cunha, R. F. (2025). World model agents with change-based intrinsic motivation. In *Northern lights deep learning conference* (pp. 66–74).

Giron, A. P., Ciranka, S., Schulz, E., van den Bos, W., Ruggeri, A., Meder, B., & Wu, C. M. (2023). Developmental changes in exploration resemble stochastic optimization. *Nature Human Behaviour*. doi: 10.1038/s41562-023-01662-1

Gopnik, A. (2024). Empowerment as causal learning, causal learning as empowerment: A bridge between bayesian causal hypothesis testing and reinforcement learning.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, *17*(11), 585–593.

Gruaz, L., Modirshanechi, A., & Brea, J. (2024). Merits of curiosity: a simulation study. *PsyArXiv*.

Ha, D., & Schmidhuber, J. (2018). World models. *arXiv preprint arXiv:1803.10122*.

Hafner, D., Pasukonis, J., Ba, J., & Lillicrap, T. (2023). Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*.

Klyubin, A. S., Polani, D., & Nehaniv, C. L. (2005). Empowerment: A universal agent-centric measure of control. In *2005 ieee congress on evolutionary computation* (Vol. 1, pp. 128–135).

Kolossa, A., Kopp, B., & Fingscheidt, T. (2015). A computational analysis of the neural bases of bayesian inference. *Neuroimage*, *106*, 222–237.

Lidayan, A., Du, Y., Kosoy, E., Rufova, M., Abbeel, P., & Gopnik, A. (2025). Intrinsically-motivated humans and agents in open-world exploration. *arXiv preprint arXiv:2503.23631*.

Little, D. Y., & Sommer, F. T. (2013). Learning and exploration in action-perception loops. *Frontiers in neural circuits*, *7*, 37.

Meltzoff, A. N., Waismeyer, A., & Gopnik, A. (2012). Learning about causes from people: observational causal learning in 24-month-old infants. *Developmental psychology*, *48*(5), 1215.

Modirshanechi, A., Brea, J., & Gerstner, W. (2022). A taxonomy of surprise definitions. *Journal of Mathematical Psychology*, *110*, 102712.

Nelson, J. D. (2005). Finding useful questions: on bayesian diagnosticity, probability, impact, and information gain. *Psychological review*, *112*(4), 979.

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental cognitive neuroscience*, *40*, 100733.

Ocana, J. M. C., Capobianco, R., & Nardi, D. (2023). An overview of environmental features that impact deep reinforcement learning in sparse-reward domains. *Journal of Artificial Intelligence Research*, *76*, 1181–1218.

Oudeyer, P.-Y., & Kaplan, F. (2007). What is intrinsic motivation? a typology of computational approaches. *Frontiers in neurorobotics*, *1*, 108.

Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning* (pp. 2778–2787).

Poli, F., Serino, G., Mars, R., & Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science advances*, *6*(39), eabb5053.

Salge, C., Glackin, C., & Polani, D. (2014). Empowerment–an introduction. *Guided Self-Organization: Inception*, 67–114.

Schmidhuber, J. (1991). Curious model-building control systems. In *Proc. international joint conference on neural networks* (pp. 1458–1463).

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology*, *4*, 710.

Seitzer, M., Schölkopf, B., & Martius, G. (2021). Causal influence detection for improving efficiency in reinforcement learning. In A. Beygelzimer, Y. Dauphin, P. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems.* Retrieved from https://openreview.net/forum?id=DXJl9826dm

Still, S., & Precup, D. (2012). An information-theoretic approach to curiosity-driven reinforcement learning. *Theory in Biosciences*, *131*, 139–148.

Sutton, R. S., Barto, A. G., et al. (1998). *Reinforcement learning: An introduction* (Vol. 1) (No. 1). MIT press Cambridge.

Ten, A., Oudeyer, P.-Y., Sakaki, M., & Murayama, K. (2024). The curious u: Integrating theories linking knowledge and information-seeking behavior. *PsyArXiv*. Retrieved from osf.io/preprints/psyarxiv/s8mkj doi: 10.31234/osf.io/s8mkj