Understand Intrinsic Motivation in Causal Learning Through Intrinsically Motivated Exploration

Fei Dai (f_d@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94704 USA

Alison Gopnik (gopnik@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA 94704 USA

Abstract

Intrinsic motivation plays a crucial role in shaping exploration and learning, yet its specific contributions to causal discovery remain underexplored. This study examines the impact of three intrinsic motivation metrics—entropy, information gain, and empowerment—on causal learning outcomes. Across two experiments, participants engaged in interactive tasks requiring them to infer causal structures through exploration. Results indicate that information gain and empowerment significantly predict learning success, whereas broad, undirected exploration (entropy) does not. These findings suggest that learners optimize causal discovery by prioritizing actions that maximize information and control, rather than engaging in indiscriminate exploration. Our study offers insights into how strategic exploration facilitates causal reasoning and how these principles can be applied to machine learning.

Keywords: Causal learning; intrinsic motivation; exploration

Introduction

Causal learning is a fundamental cognitive process that allows individuals to infer relationships between events, predict future outcomes, and exert control over their environment. Research in cognitive science has extensively examined causal reasoning through structured paradigms, such as Bayesian inference and causal graphical models, which describe how learners infer causal structures from patterns of data (Pearl, 2000; Gopnik et al., 2004; Spirtes, Glymour, & Scheines, 2000). While these frameworks offer powerful formal accounts of causal learning, they often assume that learners are provided with well-defined problems and optimal data. However, in naturalistic environments, causal discovery is not a passive process but an active, self-driven endeavor. Learners must decide what to explore, which interventions to perform, and how to allocate their cognitive resources efficiently. The mechanisms underlying these exploratory choices remain an open question in the current literature.

Emerging research suggests that intrinsic motivation—internal drivers such as curiosity, the need for control (empowerment), and the desire to reduce uncertainty (information gain)—plays a crucial role in guiding causal exploration (Schulz, 2012; Oudeyer, Kaplan, & Hafner, 2007; Brändle, Binz, & Schulz, 2021). Theoretically, intrinsic motivation serves as a computationally efficient heuristic for exploration: rather than exhaustively testing all possible causal relationships, learners prioritize actions that maximize their ability to generate meaningful interventions and acquire new information. Developmental studies show that even young children spontaneously experiment with objects in ways that resemble optimal experimental design, suggesting that intrinsic motivation supports the efficient extraction of causal knowledge (Gopnik & Wellman, 2012; Gopnik, 2024). Similarly, computational models in reinforcement learning demonstrate that agents equipped with intrinsic motivation mechanisms outperform traditional reward-driven agents in sparse or ambiguous environments, reinforcing the idea that internally guided exploration is critical for learning (Houthooft et al., 2016; Schmidhuber, 1991). Despite these theoretical insights, empirical investigations linking intrinsic motivation directly to causal learning outcomes remain limited, particularly in ecologically valid settings where learners must balance exploration with goal-directed behavior.

Although prior research has explored the impact of active interventions on causal inference (Sobel & Kushnir, 2006; Coenen, Rehder, & Gureckis, 2015), it has largely overlooked the specific mechanisms that drive learners to engage in exploration in the first place. Studies on intrinsic motivation have examined curiosity-driven behavior in open-ended tasks (Sun, Gomez, & Schmidhuber, 2011) but have not explicitly connected these behaviors to the quality of causal learning. Additionally, while reinforcement learning frameworks provide strong theoretical justifications for intrinsic motivation, most implementations focus on artificial agents rather than human learners (Bellemare et al., 2016; Burda, Edwards, Storkey, & Klimov, 2018). This leaves open several important questions: How do different forms of intrinsic motivation interact to shape exploration? Which intrinsic rewards best predict learning success? And how can these insights be applied to both human cognition and artificial intelligence systems?

The present study addresses these gaps by systematically quantifying the relationship between intrinsic motivation and causal learning outcomes. We propose a theoretical framework in which three key intrinsic motivation metrics—entropy, information gain, and empowerment—serve as predictors of exploratory behavior and learning success. Entropy captures the breadth of exploration, reflecting whether learners sample the environment evenly or focus on a narrower subset of possibilities. Information gain quantifies how much learners refine their causal beliefs through exploration, aligning with Bayesian models of active learn-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society* ©2025 the author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

ing (Lindley, 1956; Gottlieb, Oudeyer, Lopes, & Baranes, 2013). Empowerment, a concept rooted in control theory and reinforcement learning, represents the extent to which learners can influence future states, which has been hypothesized to guide exploration in both humans and artificial agents (Klyubin, Polani, & Nehaniv, 2005). By integrating these measures, our framework provides a more nuanced understanding of how intrinsic motivation shapes exploratory decisions in causal learning contexts.

Method

Information Theoretic Metrics

To quantify participants' exploratory behavior and its impact on causal learning, we employ three intrinsic motivation metrics: entropy, information gain, and empowerment. Each metric captures a distinct aspect of exploration.

Entropy quantifies the diversity of participants' exploration by measuring how evenly they distribute their interactions across possible states. A higher entropy value indicates broader, more uniform exploration, whereas a lower entropy value suggests a more focused, selective approach (Matusch, Ba, & Hafner, 2020; Schmidhuber, 1991; Oudeyer et al., 2007; Bellemare et al., 2016; Burda et al., 2018). In the current experiment, entropy was computed based on the participant's experience during exploration, providing insight into whether they engaged in systematic hypothesis testing or simply sampled broadly without a structured strategyg.

$$\text{Entropy} = \sum_{s} -p(s)\log(p(s)) \tag{1}$$

Empowerment captures the extent to which participants can influence their environment through their actions, making it a measure of perceived control. This metric is rooted in control theory and reinforcement learning, where agents benefit from maximizing their ability to predictably affect future states. Empowerment is defined as the mutual information between actions (A) and the resulting future states (S):

$$\begin{aligned}\text{Empowerment} &= \max_{p(a)} I(S;A) \\ &= \max_{p(a)} \sum_{s,a} p(s|a)p(a)\log\frac{p(s|a)}{\sum_a p(s|a)p(a)}\end{aligned} \tag{2}$$

In this equation, S denotes future states, A represents actions, and $p(s|a)$ is the transition probability from action a to state s. Maximizing empowerment encourages agents to seek states where their actions have the most predictable and significant impact, thereby enhancing their influence over the environment (Klyubin et al., 2005).

The calculation of expected information gain differs between Experiment 1 and Experiment 2. This distinction reflects the different cognitive processes involved in open-ended exploration (Experiment 1) versus structured hypothesis testing (Experiment 2).

In Experiment 1, information gain is computed using a novelty-based approach, because participants begin with no defined set of candidate rules, we approximate expected IG by rewarding novel transitions. This formulation, adapted from Du et al. (2023), aligns with reinforcement learning models that reward exploration of novel state-action pairs as a mechanism for efficient learning, which is calculated as:

$$\text{Information Gain} = \sum_{(s,a)} \frac{\log(1+N_{(s,a)})}{N_{(s,a)}} \tag{3}$$

where N(s,a) represents the number of times state-action pair (s,a) has been visited. This approach assigns higher information gain to actions that introduce new transitions, assuming that each unique state transition provides valuable learning while repeated visits yield diminishing returns (Sun et al., 2011; Friston et al., 2017). By prioritizing novelty, this measure captures the role of curiosity-driven exploration in discovering causal relationships.

In Experiment 2, information gain follows a Bayesian framework, which focuses on uncertainty reduction rather than novelty. Learners face a known hypothesis space (feature-relevance combinations), we compute classical expected entropy reduction (Bramley, Lagnado, & Speekenbrink, 2015; Coenen et al., 2015). Information gain is defined as the expected reduction in entropy over a participant's belief about the causal structure:

$$\text{Information Gain}(a,o) = H(G) - H(G|a,o) \tag{4}$$

where H(G) represents the prior uncertainty over possible causal structures, and H(G | a, o) is the updated uncertainty after taking action a and observing outcome o . This approach aligns with Bayesian models of active learning (Lindley, 1956) and suggests that optimal interventions are those that maximize informativeness, helping learners refine their causal models more efficiently (Tong & Koller, 2001).

Experiment 1

Participants The first experiment recruited 40 adult participants (56.1% female, ages 18-50, M=34.39 $\pm$ 8.96) through the online platform Prolific. Participants were fluent in English and reported no prior familiarity with the study's tasks or materials. Each participant received \$3.5 for their time, with the opportunity to earn a performance-based bonus of up to \$1, incentivizing engagement with the causal prediction task.

Materials Participants engaged in an interactive exploratio inspired by Brändle, Stocks, Tenenbaum, Gershman, and Schulz (2023), designed to encourage open-ended discovery. The task began with an inventory of four distinct elements, each defined by color (red/green/blue/yellow), shape (flower/crystal/mushroom/leaf), and quantity (1–5). On each trial, participants selected two elements and combined them in a virtual pot. If the hidden rule ("the larger-quantity element inherits the smaller's color") was satisfied, a new element reflecting that color appeared and was automatically added to the inventory; if not, no new element appeared, but the failed combination was logged. The original pair remained available for further tests.

Procedure The exploration phase lasted up to ten minutes, during which participants could freely perform combinations. They were instructed to discover the underlying rule by trial and error. Crucially, participants could end the session early by clicking a "Done" button; the trial terminated when the 10-minute limit was reached or the participant chose to stop. After the exploration session, participants completed a causal-prediction test to assess learning: they answered twelve multiple-choice questions about the outcomes of novel element combinations (elements not seen together before). Accuracy on this test (not the exploration phase itself) served as the primary measure of causal learning. Participants received a fixed payment plus a bonus based on test performance to motivate engagement.

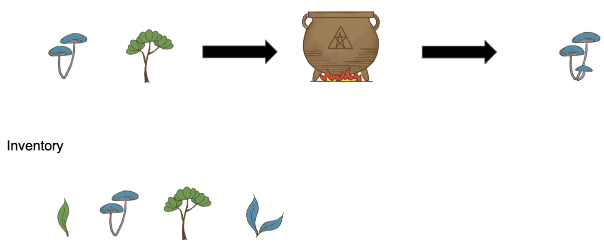


Figure 1: The interactive exploration task: Participants can select two elements from the inventory, and place them into the magic pot, and the outcome of either creating an element or not will be shown.

Results

All information theoretic metrics were standardized to z-scores within each experiment (mean = 0, SD = 1). We first report trial-level choice models, then accuracy models, for each experiment. Variance-inflation factors (VIFs) for all main-effects predictors were below 1.2, and pairwise correlations among information theoretic metrics ranged from –0.28 to 0.42, indicating no problematic collinearity.

We compared four nested mixed-effects logistic models predicting whether a participant chose action A on each free-exploration trial (N=329,639 trials, 41 subjects): M_0 (intercept-only), M_1 (+ main effects ΔEIG, ΔEmpowerment, ΔEntropy), M_2 (+ all two-way interactions), and M_3 (+ three-way interaction). Model comparison via likelihood-ratio tests indicated that M_1 improved over M_0 (ΔAIC = –529, $\chi^2(3)$=534.9, $p<0.001$), and M_2 further improved over M_1 (ΔAIC = –38, $\chi^2(3)$=44.3, $p<0.001$), whereas adding the three-way term (M_3) did not yield a significant improvement ($\chi^2(1)$=0.49, p=0.48). Thus, M_2 was selected as the final choice model.

Table 1 reports the fixed-effect estimates for M_2. Notably, empowerment exhibited the largest effect (ΔEmpowerment: estimate = 0.402, OR = 1.49, $p<2\times10^{-16}$), followed by information gain (ΔEIG: estimate = 0.113, OR = 1.12, $p<0.001$). Entropy also contributed positively (ΔEntropy: estimate = 0.098, OR = 1.10, $p<1.3\times10^{-7}$). The significant interaction between ΔEIG and ΔEmpowerment (estimate = 0.147, OR = 1.16, $p<3.7\times10^{-11}$) indicates that tests combining high expected information gain and empowerment were chosen disproportionately.

We then predicted each of the 12 post-exploration causal-prediction questions (total trials = 41 × 12 = 492) from subjects' mean information theoretic metrics via mixed-effects logistic regression. Comparing A_0 (intercept-only) to A_1 (main effects ΔEIG, ΔEmpowerment, ΔEntropy), we found a marginal omnibus improvement (ΔAIC = –0.65, $\chi^2(3)$=6.66, p=0.084), with empowerment emerging as the only significant predictor. Adding pairwise interactions (A_2) did not improve fit ($\chi^2(4)$=3.89, p=.42), so A_1 was retained.

Table 2 shows that only empowerment significantly predicted accuracy (ΔEmpowerment: estimate = 0.347, OR = 1.41, p=.012), whereas information gain and entropy had no reliable effects (both $p>.67$). In other words, although all three metrics guided exploration, only empowerment translated into improved causal-prediction performance.

Discussion

In Experiment 1, free exploration in the alchemy environment was guided by all three intrinsic-motivation metrics—but only one of these translated into improved causal inference. At the choice stage, participants systematically favored test actions with higher expected information gain (IG) and empowerment, and even showed a modest preference for high-entropy options (Table 1). The significant IG×Empowerment interaction (OR = 1.16) further indicates that learners prioritized actions that were both maximally informative and offered a strong sense of control over outcomes.

However, when those same participants later made twelve causal-prediction judgments, only their empowerment-driven exploration correlated with accuracy (OR = 1.41, Table 2). Information gain and entropy, despite having steered sampling choices, did not predict who would answer causal questions correctly. This dissociation suggests that while entropy and IG capture broad curiosity and hypothesis-testing drives, respectively, empowerment may uniquely foster deeper learning by orienting learners toward interventions that yield clear, discriminative evidence.

Experiment 2

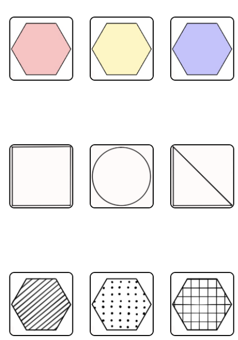
Participants 40 participants (57.5% are female, ages 21-40, M=30.8±5.14) were recruited via the online platform Prolific. All participants were adults fluent in English, with no prior exposure to the experimental materials. Participants were compensated \$7 for their participation, with the option of earning up to \$1.5 bonus based on performance in the task.

Parameter	Estimate	SE	z	p	OR
(Intercept)	–4.8374	0.1136	–42.60	$< 2 \times 10^{-16}$	0.008
ΔEIG	0.1128	0.0301	3.74	0.00018	1.12
ΔEmpowerment	0.4018	0.0219	18.34	$< 2 \times 10^{-16}$	1.49
ΔEntropy	0.0979	0.0185	5.29	1.3×10^{-7}	1.10
ΔEIG×ΔEmpowerment	0.1472	0.0223	6.62	3.7×10^{-11}	1.16
ΔEIG×ΔEntropy	0.0266	0.0354	0.75	0.45	1.03
ΔEmpowerment×ΔEntropy	–0.0270	0.0203	–1.33	0.18	0.97

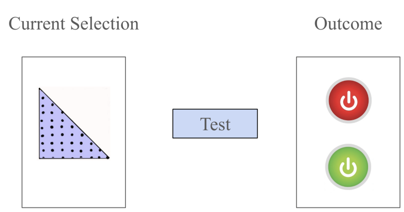
Table 1: Experiment 1 Choice Model (M_2): Mixed-Effects Logistic Regression

Parameter	Estimate	SE	z	p	OR
(Intercept)	1.1065	0.1266	8.74	$< 2 \times 10^{-16}$	3.02
ΔEIG	–0.0594	0.1410	–0.42	0.673	0.94
ΔEmpowerment	0.3472	0.1386	2.51	0.0122	1.41
ΔEntropy	–0.0109	0.1245	–0.09	0.930	0.99

Table 2: Experiment 1 Accuracy Model (A_1): Mixed-Effects Logistic Regression



(a) Selecting features



(b) Testing and outcome displayed

Figure 2: Experiment 2: (a) Selecting features from each dimension. (b) Detector activation outcome.

Materials Participants engaged in a task adapted from Song, Baah, Cai, and Niv (2022), designed to examine exploration and causal learning in a probabilistic context. The stimuli for Experiment 2 were three-dimensional objects that participants configured by selecting features across three dimensions, such as color, shape, and texture. Each object was tested on a "blicket detector," a device that probabilistically determined whether the object was a "blicket" based on its features.

The experiment consisted of 18 games per participant, with each game containing 30 trials. The games were evenly distributed across three levels of complexity, where either one (1D), two (2D), or three (3D) dimensions were relevant in determining blicket activation. Each game was further categorized into one of two conditions: 1) Informed Condition (9 games) – Participants were explicitly told how many dimensions were relevant to blicket activation (1D, 2D, or 3D), but not which specific features were responsible. 2) Uninformed Condition (9 games) – Participants were not informed about the number of relevant dimensions. Each condition contained three games per complexity level, ensuring a balanced design. The blicket detector provided probabilistic feedback based on the presence of rewarding features. The probability of activation depended on the number of correct features in the constructed object: the activation probability is 80% for selecting all dimensions correctly, and 20% for selecting at least on incorrect dimensions.

Each trial consisted of: 1) Feature Selection (5s): Participants configured an object by selecting one feature per dimension or leaving some dimensions unselected (in which case the computer randomly assigned features). 2) Object Presentation (0.5s): The constructed object was displayed on the screen. 3) Outcome Feedback (1.5s): The blicket detector provided probabilistic feedback (activation or no activation), determined by the selected features. Participants implicitly tested hypotheses during the task by observing activation feedback and adjusting their configurations accordingly. This iterative process enabled participants to refine their understanding of which features or dimensions defined blickets. The task setup encouraged trial-and-error exploration, promoting a naturalistic learning experience. To evaluate exploration quality, three metrics were calculated: entropy, information gain, and empowerment.

Procedure Participants began each trial by constructing an object, selecting one feature from each dimension (e.g., a blue, triangular, smooth object). To ensure participants understood the task, they completed a brief training session at the start, which familiarized them with the interface and the probabilistic nature of the detector. The instructions emphasized that their goal was to discover the features defining blickets rather than maximizing a score. They then completed a training phase on the probabilistic "blicket" task interface and then passed a comprehension quiz: True/False and multiple-choice questions about example objects' activation probabilities under sample rules (e.g. "If red is the only blicket-defining feature, what is the chance that a red dotted circle activates the detector?"). Participants who failed more than 3 attempts were not permitted to continue. The activation feedback followed probabilistic rules, with certain feature combinations more likely to trigger the detector. Participants were instructed that the detector's response was not deterministic and were encouraged to explore different combinations to deduce the underlying rules (Figure 3). At the end of each game, participants were asked to explicitly identify which features or dimensions they believed defined blickets.

Results

All information theoretic metrics were standardized to z-scores within each experiment (mean = 0, SD = 1). We first report trial-level choice models, then accuracy models, for each experiment. Variance-inflation factors (VIFs) for all main-effects predictors were below 1.2, and pairwise correlations among information theoretic metrics ranged from –0.02 to 0.12, indicating no problematic collinearity.

In the blicket-detector task, each of 583,200 choice trials was analyzed with mixed-effects logistic models M_0–M_3 as above, now also permitting a negative effect of entropy. Model comparison showed M_1 over M_0 (ΔAIC = –5,155, $\chi^2(3)$=5,161, p<0.001), M_2 over M_1 (ΔAIC = –98, $\chi^2(3)$=103.9, p<0.001), and M_3 over M_2 (ΔAIC = –24, $\chi^2(1)$=26.6, p<0.001). We therefore selected the full three-way model M_3.

As Table 3 indicates, information gain (ΔEIG: estimate = 0.306, OR = 1.36, $p<2\times10^{-16}$) and empowerment (ΔEmpowerment: estimate = 0.445, OR = 1.56, $p<2\times10^{-16}$) both strongly increased choice odds, while entropy now produced avoidance (ΔEntropy: estimate = –0.135, OR = 0.87, $p<2\times10^{-16}$). Both two-way interactions ΔEIG$\times\Delta$Empowerment (OR = 1.13, $p<2\times10^{-16}$) and ΔEmpowerment$\times\Delta$Entropy (OR = 1.11, $p<1.9\times10^{-9}$) were significant, and the three-way interaction also contributed (OR = 1.07, $p<2.6\times10^{-8}$), revealing a more complex interplay of drives under goal-directed testing.

We reshaped each subject's binary judgments of individual features (correct/incorrect; N=2,160 feature-trials, 40 subjects) and fit nested GLMMs with predictors ΔEIG , ΔEmpowerment , ΔEntropy , training condition, and number of relevant dimensions. Comparing A_0 (intercept-only) to A_1 (+ main effects and covariates) yielded a substantial improvement (ΔAIC = –92, $\chi^2(5)$=102.2, p<0.001), and A_2 (+ two-way interactions) further improved fit (ΔAIC = –10, $\chi^2(3)$=16.0, p=0.001), so A_2 was selected.

Table 4 shows that higher expected information gain (ΔEIG : estimate = 0.507, OR = 1.66, $p<24\times10^{-16}$) and empowerment (ΔEmpowerment: estimate = 0.740, OR = 2.10, $p<5.2\times10^{-6}$) both raised the odds of correct feature identification, with a significant ΔEIG$\times\Delta$Empowermentz interaction (OR = 1.20, p=.001). Entropy remained non-predictive (p=.52), while training condition had a marginal effect (OR = 1.19, p=.059) and task complexity was not significant (p=.188).

Discussion

In Experiment 2's structured blicket-detector task, the profile of intrinsic drives shifted in meaningful ways. As Table 3 shows, information gain (OR = 1.36) and empowerment (OR = 1.56) remained strong predictors of trial-level test selection, but participants now avoided high-entropy actions (OR = 0.87), reflecting a move away from pure novelty toward targeted hypothesis testing. Moreover, both two-way (IG×Empowerment, Empowerment×Entropy) and three-way interactions were significant, revealing a nuanced interplay: learners sought tests that jointly balanced informativeness, control, and appropriate uncertainty.

When we examined feature-level causal judgments, both IG and empowerment again emerged as key drivers of accuracy (OR = 1.66 and 2.10, respectively; Table 4), with a robust IG×Empowerment synergy (OR = 1.20). Entropy remained non-predictive, and training condition had only a marginal effect. Thus, under explicit causal-learning demands, participants appear to reweight their intrinsic drives—down-regulating entropy in favor of strategic information gain and empowerment—while still relying on these same metrics to achieve correct inferences.

General Discussion

The present study investigated the role of intrinsic motivation in causal learning, specifically examining the effects of entropy, information gain, and empowerment on exploratory behaviors and learning outcomes. Consistent with prior research (Schulz, 2012; Oudeyer et al., 2007), our results suggest that intrinsic motivation serves as a key driver of strategic exploration. Participants who implicitly maximized expected information gain and empowerment demonstrated higher accuracy, indicating that information-seeking behavior supports learning causal relationships, while entropy, which captures broad exploration, does not show a strong association with causal inference accuracy. Importantly, although our participants had an explicit causal goal, they did not receive extrinsic rewards for correct actions. Instead, their exploration seems to have been guided by internal estimates of each test's utility—that is, by the expected learning value of each action. In other words, learners appeared to rely on their own predictions of information gain rather than external feedback.

Parameter	Estimate	SE	z	p	OR
(Intercept)	–3.3261	0.0071	–466.9	$< 2 \times 10^{-16}$	0.036
ΔEIG	0.3063	0.0060	50.8	$< 2 \times 10^{-16}$	1.36
ΔEmpowerment	0.4450	0.0136	32.7	$< 2 \times 10^{-16}$	1.56
ΔEntropy	–0.1353	0.0070	–19.2	$< 2 \times 10^{-16}$	0.87
ΔEIG×ΔEmpowerment	0.1244	0.0084	14.7	$< 2 \times 10^{-16}$	1.13
ΔEIG×ΔEntropy	–0.0009	0.0064	–0.14	0.89	1.00
ΔEmpowerment×ΔEntropy	0.1003	0.0167	6.00	1.9×10^{-9}	1.11
ΔEIG×ΔEmpowerment×ΔEntropy	0.0660	0.0118	5.57	2.6×10^{-8}	1.07

Table 3: Experiment 2 Choice Model (M_3): Mixed-Effects Logistic Regression

Parameter	Estimate	SE	z	p	OR
(Intercept)	–0.5979	0.1571	–3.805	0.00014	0.55
ΔEIG	0.5071	0.0572	8.862	$< 2 \times 10^{-16}$	1.66
ΔEmpowerment	0.7403	0.1625	4.557	5.2×10^{-6}	2.10
ΔEntropy	–0.0332	0.0514	–0.646	0.518	0.97
informed	0.1756	0.0930	1.887	0.059	1.19
numRelevantDimensions	0.0762	0.0578	1.318	0.188	1.08
ΔEIG×ΔEmpowerment	0.1834	0.0564	3.251	0.0011	1.20
ΔEIG×ΔEntropy	0.0235	0.0468	0.502	0.615	1.02
ΔEmpowerment×ΔEntropy	0.0053	0.0280	0.188	0.851	1.01

Table 4: Experiment 2 Feature-Level Accuracy Model (A_2): Mixed-Effects Logistic Regression

These findings align with active-learning theories, which posit that learners prioritize interventions yielding maximal expected information gain (Lindley, 1956; Schulz, 2012; Oudeyer et al., 2007), and with control-theoretic accounts that highlight empowerment as an intrinsic reward for actions that expand an agent's influence over future states (Klyubin et al., 2005; Houthooft et al., 2016; Schmidhuber, 1991). In our tasks, participants appeared to select tests not for novelty alone, but for how much they expected each test to reduce uncertainty or afford clear, discriminative feedback. This combination of information-seeking and control-oriented drives supports more effective causal discovery than undirected exploration, mirroring results from developmental work on children's active interventions (Kidd & Hayden, 2015; Gottlieb et al., 2013) and from computational models of intrinsic motivation in artificial agents (Bellemare et al., 2016; Burda et al., 2018).

Importantly, although entropy often benefits exploration in reinforcement-learning benchmarks (Bellemare et al., 2016; Burda et al., 2018), our human participants showed that quality of exploration matters more than sheer quantity. High-entropy strategies corresponded to sampling many different options evenly, but without strategic guidance this did not translate into better causal inferences. Instead, actions that maximized expected IG or empowerment—targeted hypothesis tests and choices that isolated specific causes—proved most beneficial.

Several limitations warrant future work. First, our tasks provided an explicit external goal ("discover the rule") and performance bonuses, which may have introduced extrinsic incentives that partially overlapped with intrinsic drives. Free-play paradigms, with minimal instruction and no performance payoffs, would help disentangle pure curiosity from goal-directed exploration (Pelz & Kidd, 2020). Second, our three metrics—entropy, IG, and empowerment—capture key facets of exploration but do not encompass other motivational factors such as surprise, metacognitive confidence, or social curiosity. Integrating additional measures (e.g., prediction-error gaps, eye-tracking, verbal protocols) and employing computational belief-tracking could reveal how multiple drives interact moment-by-moment (Gureckis & Markant, 2012; Dubey, Griffiths, & Lombrozo, 2019).

Looking ahead, a unified model that dynamically weights entropy, IG, and empowerment according to context could better predict human exploration and inform the design of intelligent tutoring systems or autonomous agents. By grounding exploration in both information-theoretic and control-oriented principles, we can advance our understanding of the complex, multi-faceted motivations that underlie effective causal learning.

Acknowledgments

This work was funded by the DOD ONR MURI Co-PI Self-Learning Perception through Real World Interaction funding and Templeton World Charity Foundation (TWCF-2021-20639). The authors are grateful to the members of the Cog-

nitive1945Development and Learning Lab at UC Berkeley.

References

Bellemare, M., Srinivasan, S., Ostrovski, G., Schaul, T., Saxton, D., & Munos, R. (2016). Unifying count-based exploration and intrinsic motivation. *Advances in neural information processing systems*, *29*.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 708.

Brändle, F., Binz, M., & Schulz, E. (2021). Exploration beyond bandits. *The drive for knowledge: The science of human information seeking*, 147–168.

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*.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive psychology*, *79*, 102–133.

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., & Lombrozo, T. (2019). If it's important, then i am curious: A value intervention to induce curiosity. In *Cogsci* (pp. 282–288).

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017). Active inference, curiosity and insight. *Neural computation*, *29*(10), 2633–2683.

Gopnik, A. (2024). Empowerment gain as causal learning, causal learning as empowerment gain: A bridge between bayesian causal hypothesis testing and reinforcement learning. presented at the 29th meeting of the philosophy of science association, new orleans, dec 2024.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological review*, *111*(1), 3.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: causal models, bayesian learning mechanisms, and the theory theory. *Psychological bulletin*, *138*(6), 1085.

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.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Houthooft, R., Chen, X., Duan, Y., Schulman, J., De Turck, F., & Abbeel, P. (2016). Vime: Variational information maximizing exploration. *Advances in neural information processing systems*, *29*.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*(3), 449–460.

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).

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*(4), 986–1005.

Matusch, B., Ba, J., & Hafner, D. (2020). Evaluating agents without rewards. *arXiv preprint arXiv:2012.11538*.

Oudeyer, P.-Y., Kaplan, F., & Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, *11*(2), 265–286.

Pearl, J. (2000). Causal inference without counterfactuals: Comment. *Journal of the American Statistical Association*, *95*(450), 428–431.

Pelz, M., & Kidd, C. (2020). The elaboration of exploratory play. *Philosophical transactions of the royal society B*, *375*(1803), 20190503.

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

Schulz, L. (2012). The origins of inquiry: Inductive inference and exploration in early childhood. *Trends in cognitive sciences*, *16*(7), 382–389.

Sobel, D. M., & Kushnir, T. (2006). The importance of decision making in causal learning from interventions. *Memory & Cognition*, *34*, 411–419.

Song, M., Baah, P. A., Cai, M. B., & Niv, Y. (2022). Humans combine value learning and hypothesis testing strategically in multi-dimensional probabilistic reward learning. *PLoS computational biology*, *18*(11), e1010699.

Spirtes, P., Glymour, C. N., & Scheines, R. (2000). *Causation, prediction, and search.* MIT press.

Sun, Y., Gomez, F., & Schmidhuber, J. (2011). Planning to be surprised: Optimal bayesian exploration in dynamic environments. In *Artificial general intelligence: 4th international conference, agi 2011, mountain view, ca, usa, august 3-6, 2011. proceedings 4* (pp. 41–51).

Tong, S., & Koller, D. (2001). Active learning for structure in bayesian networks. In *International joint conference on artificial intelligence* (Vol. 17, pp. 863–869).