Discovering Hidden Laws in Innovation by Recombination

Bonan Zhao (b.zhao@ed.ac.uk)
School of Informatics
University of Edinburgh

Elizabeth Mieczkowski
Department of Computer Science
Princeton University

Dilip Arumugam
Department of Computer Science
Princeton University

Natalia Vélez
Department of Psychology
Princeton University

Thomas L. Griffiths
Department of Computer Science
Department of Psychology
Princeton University

Abstract

Combining two things can create amazing new things—whether mixing water and flour or feeding large datasets into neural networks. Hypothesizing rules and theories for recombination, testing those hypotheses, and communicating our findings to each other are key cognitive mechanisms that allow us to navigate an open-ended world of possible combinations. However, in contrast to this open-ended and highly-complex search problem, cognition is constrained by its capacity. Using ideas from information theory, we hypothesize that the compressibility of recombination rules predicts how successfully people find and use these rules. In a combinatorial discovery game, we find that people indeed learn quicker and collect more points when the rules are more compressible. Interestingly, people use fewer words to communicate their findings when the rules are either too easy or too hard to compress, revealing an inverse-U shaped relationship between compressibility and communication effort.

Keywords: combination; discovery; innovation; rule learning; concept learning; active learning; exploration; communication channel; cultural transmission

Introduction

Innovation often involves recombination of existing things (Arthur, 2010; Basalla, 1988; Fleming, 2001; Youn et al., 2015). From tool-making to theory-building, new forms are forged from old components. Soaking cabbage with salty brine, for example, creates a new food (kimchi) that both introduces a novel flavor and can last much longer than fresh cabbage. However, not all combinations are born equal. While adding salty brine to chopped cabbage or tuna chunks brings desirable outcomes, preserving a slice of cake or a scoop of ice-cream in salty brine is unlikely to work out. Just as knowing the science of food preservation allows safer and more efficient practices, mastering the hidden rules behind successful combinations leads to more effective innovations (Arthur, 2010; Kuhn, 1997).

Picking out successful combinations from a space of virtually infinite possibilities presents a daunting challenge for individual, finite minds. Discovering hidden rules involves synthesizing generalizable hypotheses from limited observations (Goodman et al., 2008; Fränken et al., 2022; Zhao et al., 2022), and often requires collecting data and engaging in active learning (Bramley et al., 2017; Coenen et al., 2015; Gong et al., 2023). Moreover, balancing between exploiting known combinations and exploring new ones constitutes a classic exploration-exploitation trade-off (Cohen et al., 2007; Mehlhorn et al., 2015). Prior work suggests that cognitive constraints may bias individual learners to acquire simpler rules about which combinations are successful (Feldman, 2000; Goodman et al., 2008; Zhao, Lucas, & Bramley, 2024), and be subject to modifying guesses incrementally rather than adopting entirely new ones, even when better alternatives exist (Bramley et al., 2017; Fränken et al., 2022).

We develop a novel task to investigate how people actively explore, discover, and exploit innovation through recombination. In an interactive 2D world, participants can move freely, pick up, drop, and combine items, and harvest points by consuming items (Figure 1). This flexible setup creates a semi–open-ended discovery game that integrates a wide range of cognitive processes into a single coherent framework: from balancing exploration and exploitation to rule-based induction, and from reward maximization to optimal compression. We use this setting to examine how people find effective combinations. Drawing on classic findings in concept learning and information theory, we hypothesize that a set of hidden combinations that are easier to express in words are also easier to discover, and consequently enable more effective creation of new things, regardless of statistical proprieties like its size and sparsity. To foreshadow, while our key predictions are supported by empirical results, we also find an intriguing non-linear relationship between the compressibility of the hidden combinations and the lengths of participants' free-response texts describing their discoveries.

Background

Innovation by Recombination

Prior work in several fields—including cognitive science and organizational theory—has pointed to recombination as a key mechanism for the discovery of new ideas (Arthur, 2010; Basalla, 1988; Fleming, 2001; Lake et al., 2017; Youn et al., 2015). In recent studies of how people discover new combinations (Brändle et al., 2023; Vélez et al., 2024; Zhao, Vélez, & Griffiths, 2024), crafting games prove to be a powerful tool. In crafting games, players can collect resources and recombine them to grow an inventory of unique items. Commercially-available crafting games, such as Little Alchemy, Minecraft, and One Hour One Life, are increasingly used to understand how people explore in the absence of external rewards (Brändle et al., 2023), how to build agents

that can discover new technologies in vast physical environments (Wang et al., 2023), and how people collaborate to develop new technologies together (Vélez et al., 2024). Inspired by these works, here we present a "discovery game" that provides a customizable test environment for studying innovation by recombination (see also Zhao, Vélez, & Griffiths, 2024).

Discovery Games and Reinforcement Learning

Unlike running analyses on large-scale empirical game datasets (Brändle et al., 2023; Vélez et al., 2024), our discovery game allows the experimenter to design specific game rules, and study how people search for hidden laws with as little influence from domain specific knowledge as possible. The challenge in the discovery game is that labeled learning data is not always available, and players have to actively gather data, propose and update hypotheses (Coenen et al., 2015; Fränken et al., 2022; Gong et al., 2023), and meanwhile balance between exploring new combinations and exploiting existing ones to reap the rewards (Sutton & Barto, 1998). One algorithm that addresses this problem is Posterior Sampling for Reinforcement Learning (PSRL) (Strens, 2000; Osband et al., 2013). PSRL consumes as input a prior distribution over possible environments—e.g., a player's epistemic uncertainty (Der Kiureghian & Ditlevsen, 2009) of the hidden laws in the discovery game, and selects actions via Thompson Sampling (W. R. Thompson, 1933) in each episode. Theoretically, PSRL is known to achieve statistically-efficient reinforcement learning for a broad class of environments (Osband et al., 2013; Abbasi-Yadkori & Szepesvari, 2014; Agrawal & Jia, 2017).

Concept Learning and Compressibility

While PSRL offers a principled algorithmic model for how learning progresses in a discovery game, it assumes perfect information updating—an assumption that rarely holds in the real world. Given limited cognitive resources, people often need to compress raw observations into more succinct mental representations. This challenge to many aspects echoes the classic concept learning problem in cognitive psychology (Feldman, 2000; Goodman et al., 2008; Nosofsky et al., 1994; Shepard et al., 1961), where people learn to use a single concept to refer to many individual exemplars. One way to learn such concept-to-exemplar mappings is by describing what features of the exemplars suffice to define a concept. Feldman (2000) showed that the complexity of such descriptions is a reliable predictor of how hard it is for people to learn that concept (Shepard et al., 1961). In parallel, rate-distortion theory (RDT)—a subfield of information theory—offers a formal account of the trade-off between the capacity of a representation and its fidelity in compressing and recovering observations (Shannon, 1959; Berger, 1971). RDT has proven particularly useful for understanding concept formation (Imel & Zaslavsky, 2024; Zaslavsky et al., 2018) and capacity-limited learning (Arumugam et al., 2024; Prystawski et al., 2023). Recent work has also explored integrating RDT with PSRL (Prystawski et al., 2023), offering computational frameworks of how players transmit combination discoveries to future players through limited channels.

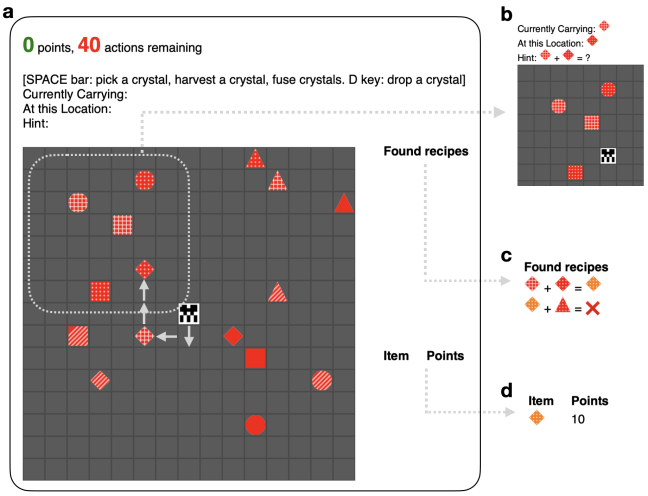

Figure 1: Demo interface of a discovery game. **a**. The GridWorld view with example actions (grey arrows). **b**. Information shown for the example actions. **c**. History of attempted recipes is updated automatically. **d**. Discovered items' points.

Theoretical Framework

In this section, we present the discovery game and our main theoretical proposal in detail.

Game Environment

The discovery game we focus upon here takes place in a 2D GridWorld (Figure 1a). Players can move around this GridWorld in four directions: `Up`, `Down`, `Left`, and `Right`. The environment is spawned with some game items to start. When a player moves on top of an item, they can `Pickup` the item, at which point it is added to their inventory. When holding an item in their inventory, the player may `Consume` the item and collect points from it, which also removes the item from their inventory. Alternatively, the player may `Drop` an item they are holding, which leaves the item in the player's inventory. If the player currently holds an item and moves on top of another item, they may try to `Combine` these two items. In this game, participants' move actions are controlled by the arrow keys, actions `Pickup`, `Consume` and `Combine` by the SPACE key, and action `Drop` by the D key.

A player's goal is to maximize the total points they collect in the game within a fixed number of T actions. `Consume` an item and `Combine` two items each count as one action. Moving around, picking up, or dropping items do not count towards T. The game is spawned with base-level (level $= 0$) items. When two items m and n are combined successfully, this creates a new item o that is one level higher than its parents, $\text{level}(o) = \max(\text{level}(m), \text{level}(n)) + 1$. The points of each item increases with its level. Combining items in itself does not earn any points. Once successfully combined, the

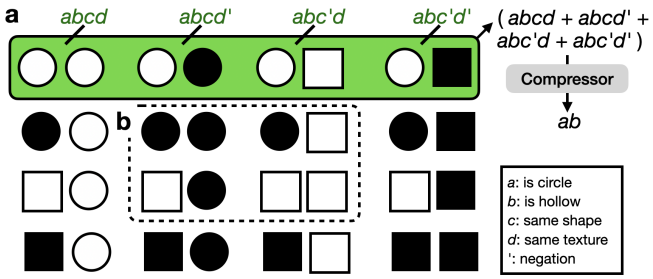

Figure 2: Illustration of the compression process. **a**. Compressing the example positive recipes in the shaded box. **b**. Difference in compressibility. The set in the shaded box can be compressed as *ab*, while the set in the dashed box cannot be further compressed.

used items are removed from the player's inventory, with the outcome item added to the inventory. The piece of information specifying what comes out of a combination is called a "recipe." We call the recipes where the outcome is a new item a "positive recipe," in contrast to recipes that result in a failed combination (i.e., nothing is produced). To maximize points, players have to discover and even predict recipes from attempting combinations.

The discovery game outlined above can be formalized as a finite-horizon, episodic Markov Decision Process (MDP) (Bellman, 1957; Puterman, 1994) $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \mu, T \rangle$, where $\mathcal{S}$ is the state space representing the player's current inventory of items, action space $\mathcal{A} = \{\texttt{Combine}, \texttt{Consume}\}$, reward function $\mathcal{R} : \mathcal{S} \times \mathcal{A} \to \mathbb{R}_{\geq 0}$, a deterministic transition function $\mathcal{P} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$, the initial state distribution $\mu \in \Delta(\mathcal{S})$, and T the horizon or maximum number of actions a player may take. As specified earlier, $\mathcal{R}(\texttt{Combine}(m,n)) = 0$ for any items m, n, and $\mathcal{R}(\texttt{Consume}(m)) = 10^{\text{Level}(m)}$. Although the transition function $\mathcal{P}$ is fully determined by some hidden laws that specify all the positive recipes, agents are not informed of these hidden laws at the start of the game. Hence, for any current inventory $s \in \mathcal{S}$, a player who tries to combine any two items $m, n \in s$ from will potentially obtain a new item o according to the collection of positive recipes $\mathcal{E}$:

$$\mathcal{P}(s, \texttt{Combine}(m,n)) = \begin{cases} (s \setminus \{m,n\}) \cup \{o\} & m,n \in s \text{ and } \langle m,n \rangle \in \mathcal{E} \\ s & \text{otherwise.} \end{cases} \tag{1}$$

Complexity of the Hidden Laws

A player attempting to learn all of the positive recipes contained in $\mathcal{E}$ faces a difficult exploration problem. As introduced earlier, PSRL may solve this challenge, but PSRL presumes perfect posterior belief updates given the entire history of interactions thus far, while people are likely too constrained by mental capacity to perform such exact updates. Inspired by Feldman (2000), we assume that players compress all the positive recipes they discovered so far into a mental message (Figure 2). Since each recipe involves two objects, we extend Feldman's original Boolean complexity measure to include a relational concept of "same." Specifically, we describe a recipe in two steps: (1) encode the feature values of the first object using a predefined dictionary (e.g., a for "is circle" and a' for "is not circle"), and (2) encode the feature values of the second object relative to the first (e.g., noting "same shape" if the two objects share the same shape). For each positive recipe $r \in \mathcal{E}$, we transcribe it into a description $l(r)$. We then form the description of the entire set $\mathcal{E}$, denoted E, by joining all individual descriptions $l(r)$ disjunctively. Let E represent the literal transcription of $\mathcal{E}$, defined as $E := \bigwedge_{r \in \mathcal{E}} l(r)$. Finally, we simplify E as much as possible to produce the maximally compressed message L^*. The compressibility of the set of positive recipes $\mathcal{E}$ is then

$$q(\mathcal{E}) = \text{len}(E) - \text{len}(L^*), \tag{2}$$

where the length function $\text{len}(\cdot)$ counts the number of symbols in a description. For $\mathcal{E}_1$ and $\mathcal{E}_2$ of the same size (i.e., containing the same number of positive recipes), the lengths of their literal transcriptions E_1 and E_2 are the same, but L_1^* and L_2^* may be of different lengths (see Figure 2b), leading to different compressibility, and therefore different conceptual difficulty.

In addition to this maximally compressed message L^*, players in theory can use many messages to represent their beliefs about the MDP—in this case, a message L about the set of positive recipes $\mathcal{E}$ suffices to construct the entire MDP because the states, actions, reward function, and horizon are all known. Following Prystawski et al. (2023), the expected loss or distortion incurred when a player encode their beliefs about MDP $\mathcal{M}$ into message L is

$$\mathbb{E}[d(\mathcal{M}, L)] = \mathbb{E}\left[V^*_{\mathcal{M}} - V^*_{\mathcal{M}_L}\right], \tag{3}$$

where $\mathcal{M}_L$ is the random variable representing a MDP consistent with message L. Overall, this expected distortion is the expected regret or performance shortfall between what a player could have achieved holding the full MDP $\mathcal{M}$ versus based on the limited information encoded in message L.

Recall that Equation 2 measures the compressibility of the true MDP. If we let $R \in \mathbb{R}_{\geq 0}$ be an individual player's rate limit or upper bound on how many bits of information they may use, then the minimum expected distortion an agent can achieve is subject to the rate-limit R, known as the distortion-rate function (Shannon, 1959):

$$\mathcal{D}(R) = \inf_{p(L|\mathcal{M}):\mathbb{I}(\mathcal{M};L) \leq R} \mathbb{E}[d(\mathcal{M}, L)], \tag{4}$$

where $\mathbb{I}(\mathcal{M}; L)$ is the mutual information quantifying how much information about the true MDP $\mathcal{M}$ are retained in the player's message L.

Here, greater compressibility $q(\mathcal{E})$ via shorter L^* implies that an agent is less limited in how much information about the true MDP $\mathcal{M}$ can be conveyed by its own message L.

This suggests that increased compressibility allows an agent to operate at a higher rate limit R. In contrast, a less compressible MDP, measured by $q(\mathcal{E})$, requires the player to optimize for making the best use of limited communication bandwidth subject to a lower rate limit R. Together, Equations 2 and 4 imply that the higher compressibility $q(\mathcal{E})$ is, the better performance (lower regret) we should expect. In terms of behavioral measures, we predict that the more compressible a hidden law is, the more total points people will collect. We additionally predict that people will use more words to describe less-compressible hidden laws, when communicating their findings to future players.

Testing Predictions

We test the above predictions in an online experiment, approved by the Research Integrity & Assurance committee at Princeton University (IRB ref. no. 11092). Pre-registration is available at `https://aspredicted.org/5znp-hgbp.pdf`.

Experiment

Participants 120 participants (54% female, $M_{\text{age}} = 35 \pm 11$) were recruited from Prolific Academic. Average task completion time was 16 minutes. Participants were paid \$1.50 for participation and a performance-based bonus, ranging from \$0.33 to \$1.00. All participants gave informed consent before participating. One participant was excluded from analysis because of a server connection error.

Materials Participants played a discovery game in a 20×20 GridWorld (Figure 1). Before beginning the game, they were told to harvest energy points from alien crystals in a faraway planet. These alien crystals varied along three dimensions—`shape` = { circle, square, diamond, triangle }, `texture` = { solid, dotted, lined, checkered }, and `color` = { red, orange, yellow, green, blue, purple }. The color dimension was reserved to indicate item points: let i be the index of a color in `color`, items of that color have points 10^i. For instance, each red item is worth 1 point, and each orange item is worth 10 points.

Each game is initialized with 16 red crystals that span all possible shape-texture combinations. We defined three hidden laws that determine which combinations lead to new, more rewarding items, summarized in Table 1. These hidden laws have a similar number of positive recipes, meaning that the success rates for these hidden laws are all similar to a stochastic learner. When a combination works out, the resulting crystal takes the same shape as the first crystal (held by the avatar), the same texture as the second crystal (on the grid where the avatar stands), and changes color to indicate its value. Each player's complete history—the combinations they had attempted and points collected from consuming items—were automatically updated and shown to the participants as part of the game's interface (Figure 1b-d).

Design and Procedure Participants were randomly assigned to one of the three between-subject conditions: high-, medium-, and low-compressibility (Table 1). After instructions, each participant played one game with $T = 40$ actions. Note that only consuming crystals and combining crystals count as actions. Picking up or dropping items, moving around, do not count. After the game phase, participants composed two messages to pass on what they learned in this game to a hypothetical future player. The first message is about how to better play the game, and the second about the rules or patterns they found. Participants were told they would be paid a bonus based on how well the next player who read their messages performs. The experiment then concluded with participants providing demographic information and feedback.

Compressibility	Description	# positive recipes
High	Same shape	48
Medium	Circle goes with square, triangle goes with diamond, textures must be different	48
Low	Circle goes with square, triangle goes with diamond, texture index of first crystal $\geq$ texture index of second crystal	46

Table 1: Hidden laws used in the experiment.

Results

Participants did better when the hidden laws are more compressible. As predicted, higher compressibility led to better performance. As illustrated in Figure 3a and Figure 3b, participants in the high-compressibility condition achieved more total points, and uncovered more advanced items, than participants in the medium-, and then the low-compressibility conditions. A one-way ANOVA revealed a significant effect of condition on log-scaled[1] total points ($F(2,116) = 13.97, p < 0.001$, $\eta^2 = 0.19, 95\%\text{CI: } [0.09, 1.00]$), as well as a significant effect of condition on the highest levels each participant managed to achieve ($F(2,116) = 16.55, p < 0.001, \eta^2 = 0.22, 95\%\text{CI: } [0.11, 1.00]$).

Breaking down by actions, participants in the high-compressibility condition constantly gathered more points throughout the task (Figure 3c), while this trend is slower for the medium- and low-compressibility conditions. A linear mixed-effects model with fixed effects for action indices, the conditions, their interactions, and random effects for individual participants revealed that while action index was obviously a significant factor ($t(4638) = 18.49, p < 0.001$), there were also significant interactions between action index and conditions (with low-compressibility as reference, $t(4638) = 16.95, p < 0.001$ for high-compressibility and $t(4638) = 9.93, p < 0.001$ for medium-compressibility),

[1] Since item points grow exponentially with levels, we transformed the total points to log scale for all the analyses throughout this paper.

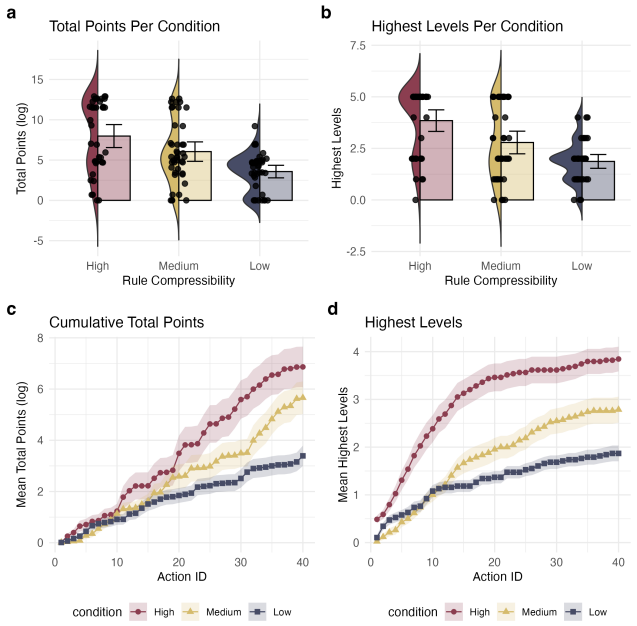

Figure 3: Participant performance. **a**. log-scaled total points collected per condition. **b**. highest levels created per condition. **c**. total points at each action step per condition; points are converted to log scales. **d**. highest levels created at each action step per condition.

indicating that the speed of point accumulation was significantly different among these conditions.

Similarly, for item levels, participants in the high-compressibility condition quickly unlocked more advanced items (Figure 3d), while those in the medium- and low-compressibility conditions struggled to uncover more advanced items. Figure 3d also indicates that the highest levels discovered in the high-compressibility condition plateaued towards the end of the experiment, and this is likely due to the hard limit on the highest levels participants could achieve at all (max level = 5). We ran a linear mixed-effects model with fixed effects for action indices, conditions and their interactions on the highest item levels at that point, with random effects for individual participants. This model suggested significant main effects of both action index ($t(4638) = 23.07, p < 0.001$) and the high-compressibility condition ($t(124) = 3.40, p = 0.0009$), indicating that both later actions and being in the high-compressibility condition are associated with a higher highest item level. We also found significant interaction effects between action index and the high-compressibility condition ($t(4638) = 15.73, p < 0.001$) as well as between action index and the medium-compressibility condition ($t(4638) = 14.89, p < 0.001$), suggesting that the effect of action index on highest item levels indeed differs depending on the condition.

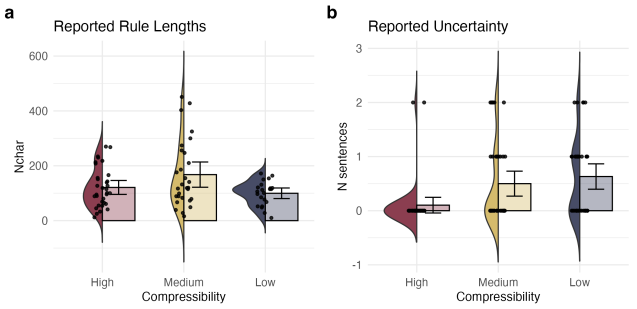

Figure 4: Self-report measurements. **a**. Total number of characters for sentences in the rules category, per condition. **b**. Number of Uncertainty sentences (labeled as NA by a customized GPT-4 prompt) per condition.

Inverse-U shape of communication effort and compressibility As these self-reports are noisy and unstructured, we pre-processed them with GPT-4. Using OpenAI API, we instructed GPT-4 to play the role of a helpful assistant, and sort each sentence in a participant's self-report into three categories: tips, rules, and NAs. Tips refer to descriptions of generic game rules (e.g., "more points are better"), and how to interact with the game, (e.g., "use D to drop items"). Rules include sentences talking about the possible hidden laws, (e.g., "same shapes combine"). Lastly, NAs are sentences expressing pure uncertainty or being lost (e.g., "i have no idea").

One of our key predictions is that as compressibility goes up, people would use more words to describe their findings about the hidden laws. In reality, we found an inverse-U shape (Figure 4a): for sentences classified as rules, participants in the medium-compressible condition used the most amount of words, whereas participants in both the high- and low-compressible conditions used fewer words. A one-way ANOVA revealed a significant effect of condition on the number of rule sentences, $F(2,116) = 3.17, p = 0.0457, \eta^2 = 0.05, 95\%\text{CI: } [0.00, 1.00]$, and a marginal effect of condition on the total number of characters of the rule sentences, $F(2,116) = 2.97, p = 0.05, \eta^2 = 0.05, 95\%\text{CI: } [0.00, 1.00]$. For the length measure (number of characters), a few outliers with excessively long sentences had an unbalanced impact on the group mean measures. We applied the Interquartile Range (IQR) method to exclude outliers ($n = 5$), and only considered those who did provide rule sentences. On this cleaned rule message dataset, a one-way ANOVA revealed a significant effect of condition on number of characters in rule sentences, $F(2,81) = 4.22, p = 0.01, \eta^2 = 0.09, 95\%\text{CI: } [0.01, 1.00]$. Although the effect size is modest, this effect is quite remarkable considering how noisy these self-reports are, especially the large variance of individual styles in self-reporting.

In line with objective performance, as compressibility went down, we saw more sentences about unknowns (Figure 4b). A one-way ANOVA revealed a significant effect of condition on Uncertain sentences (category NA), $F(2,116) =$

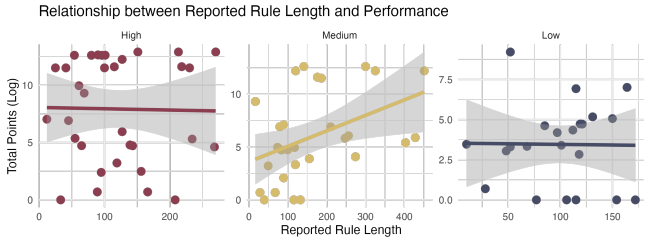

Figure 5: Relationship between reported rule lengths and total points collected (log-scaled) in each condition.

$6.98, p = 0.001, \eta^2 = 0.11, 95\%\text{CI: } [0.03, 1.00]$). A same but smaller effect holds for the total lengths of these uncertainty sentences, measured as the total number of characters, $F(2,116) = 5.99, p = 0.003, \eta^2 = 0.09, 95\%\text{CI: } [0.02, 1.00]$.

Communication effort interacts with performance and compressibility Having found an inverse-U relationship between compressibility and the lengths of participants' rule descriptions, we next examined whether participants who were more successful in the game wrote longer sentences to pass on their knowledge. In the high- and low-compressibility conditions, we did not observe strong correlation between performance and length of rule sentences composed (linear regression, high-compressibility condition: $R^2 = 0.0003, F(1,31) = 0.009, p = 0.9$, low-compressibility condition: $R^2 = 0.0002, F(1,20) = 0.003, p = 0.9$). However, in the medium-compressibility condition, rule sentence length was a positive predictor of performance (total points, log-scaled), $R^2 = 0.2, F(1,27) = 5.93, p = 0.02$. These results suggest that when the environment is too regular (high-compressibility) or too complex (low-compressibility), the relationship between performance (i.e., total points collected) and communicational effort (number of words used) was not necessarily correlated. However, there might exist some sweet spots, such as in the medium-compressibility condition, where people with better performance also put more effort into communicating and transmitting their discoveries.

Discussion

Humans have created countless new ideas by recombining old ones, from sprinkling salt on chocolate chip cookies to combining wings with bicycle parts to make the first airplane. Discovering the hidden laws common to successful combinations can be a powerful way to accelerate innovations. However, discovering those hidden laws presents a daunting challenge that taxes our individual capacities to learn generalizable concepts from experience and to communicate those concepts to others. Here, we found that participants who quickly identified the hidden laws in a discovery game were more efficient in discovering new, more powerful combinations of items, and collected more points in total. Our results also suggest that people rely on inductive biases when learning and communicating about the hidden laws of recombination. Specifically, in the experiment, we designed three hidden laws that had similar numbers of positive recipes, but differed in their compressibility. Overall, participants earned more points in environments governed by more compressible laws. Further, we found an inverse-U relationship between compressibility and communication effort when participants left messages about these hidden laws for future players. Participants used the most words to express the hidden laws in the medium-compressibility condition, and they used fewer words to express highly-compressible and uncompressible hidden laws.

Directly modeling people's behavior in the discovery game studied in this paper faces significant computational hurdles. A key facet of the discovery game is that labeled examples are not simply provided but rather must be discovered (Coenen et al., 2015; Bramley et al., 2017; Gong et al., 2023) through prudent exploration (Cohen et al., 2007; Mehlhorn et al., 2015). While CultuRL (Prystawski et al., 2023) offers one promising framework capturing this difficult exploration challenge subject to limited communication, even instantiating the base PSRL algorithm is computationally onerous for a discovery game due to the combinatorially-large state-action space. In future work, we aim to develop more efficient algorithms to tackle the inefficiencies of PSRL by leveraging known methods in inductive learning and approximate solutions.

As our results suggest, people drew on inductive biases to navigate this game. We operationalized these inductive biases with a compressibility measures (Feldman, 2000), which naturally connects with a Bayesian interpretation (Goodman et al., 2008). The sub-task of discovering the hidden law in our setup can be cast as a relational concept learning problem (Lake et al., 2015; Fränken et al., 2022; Zhao, Lucas, & Bramley, 2024), opening up the possibility of introducing abstract structures to the standard state space in MDPs (Li et al., 2006). Recent advances in Bayesian library learning further showcased how cognitively-contained learners may unlock increasingly complex concepts by boostrapping on past discoveries (Zhao, Lucas, & Bramley, 2024), providing a promising perspective to understand how knowledge of successful combinations may grow beyond capacity constraints.

Science is a process (Hull, 1990). The power of pushing observations through a communication channel is not just an efficient form of note-taking, but a way to enable the collective effort of searching for data and adapting theories. For example, physicists today do not have to rediscover Newtonian laws of physics, thanks to the fact that Newton had compressed and communicated his search results effectively. We are also keen to extend our work to examine how knowledge accumulates over multiple generation of learners, and how communication may scaffold, or shape, the search trajectories in communities (Derex & Boyd, 2016; Muthukrishna & Henrich, 2016; B. Thompson et al., 2022). By compressing the results of our experiment in this paper, we hope we may help others make their own discoveries.

Acknowledgments

This work was supported by a grant from the Templeton World Charity Foundation (TWCF 20648) to TG and NV, the National Defense Science and Engineering Graduate (NDSEG) Fellowship Program to EM, and ONR MURI N00014-24-1-2748 to TG and DA.

References

Abbasi-Yadkori, Y., & Szepesvari, C. (2014). Bayesian optimal control of smoothly parameterized systems: The lazy posterior sampling algorithm. *arXiv preprint arXiv:1406.3926*.

Agrawal, S., & Jia, R. (2017). Optimistic posterior sampling for reinforcement learning: Worst-case regret bounds. In *Advances in neural information processing systems* (pp. 1184–1194).

Arthur, W. B. (2010). *The nature of technology: What it is and how it evolves*. Penguin UK.

Arumugam, D., Ho, M. K., Goodman, N. D., & Van Roy, B. (2024). Bayesian reinforcement learning with limited cognitive load. *Open Mind*, *8*, 395–438.

Basalla, G. (1988). *The evolution of technology*. Cambridge University Press.

Bellman, R. (1957). A markovian decision process. *Journal of Mathematics and Mechanics*, 679–684.

Berger, T. (1971). *Rate distortion theory: A mathematical basis for data compression*. Prentice-Hall.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301.

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.

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

Cohen, J. D., McClure, S. M., & Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 933–942.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987.

Der Kiureghian, A., & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, *31*(2), 105–112.

Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*(6804), 630–633.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, *47*(1), 117–132.

Fränken, J.-P., Theodoropoulos, N. C., & Bramley, N. R. (2022). Algorithms of adaptation in inductive inference. *Cognitive Psychology*, *137*, 101506.

Gong, T., Gerstenberg, T., Mayrhofer, R., & Bramley, N. R. (2023). Active causal structure learning in continuous time. *Cognitive Psychology*, *140*, 101542.

Goodman, N. D., Tenenbaum, J. B., Feldman, J., & Griffiths, T. L. (2008). A rational analysis of rule-based concept learning. *Cognitive Science*, *32*(1), 108–154.

Hull, D. L. (1990). *Science as a process: An evolutionary account of the social and conceptual development of science*. University of Chicago Press.

Imel, N., & Zaslavsky, N. (2024). Optimal compression in human concept learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*, 1395–1401.

Kuhn, T. S. (1997). *The structure of scientific revolutions* (Vol. 962). University of Chicago Press.

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, e253.

Li, L., Walsh, T. J., & Littman, M. L. (2006). Towards a unified theory of state abstraction for MDPs. In *International Symposium on Artificial Intelligence and Mathematics* (Vol. 4, p. 5).

Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, K., Michael Dand Morgan, Braithwaite, V. A., Hausmann, D., Fiedler, K., & Gonzalez, C. (2015). Unpacking the exploration–exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, *2*(3), 191–215.

Muthukrishna, M., & Henrich, J. (2016). Innovation in the collective brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1690), 20150192.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*(1), 53–79.

Osband, I., Russo, D., & Van Roy, B. (2013). (More) efficient reinforcement learning via posterior sampling. *Advances in Neural Information Processing Systems*, *26*, 3003–3011.

Prystawski, B., Arumugam, D., & Goodman, N. (2023). Cultural reinforcement learning: a framework for modeling cumulative culture on a limited channel. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45).

Puterman, M. L. (1994). *Markov decision processes—discrete stochastic dynamic programming*. John Wiley & Sons.

Shannon, C. E. (1959). Coding theorems for a discrete source with a fidelity criterion. *Institute of Radio Engineers, International Convention Record*, *7*, 142–163.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1–42.

Strens, M. J. (2000). A bayesian framework for reinforcement learning. *Proceedings of the Seventeenth International Conference on Machine Learning*, 943–950.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning*. MIT Press.

Thompson, B., Van Opheusden, B., Sumers, T., & Griffiths, T. (2022). Complex cognitive algorithms preserved by selective social learning in experimental populations. *Science*, *376*(6588), 95–98.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, *25*(3/4), 285–294.

Vélez, N., Wu, C. M., Gershman, S. J., & Schulz, E. (2024). *The rise and fall of technological development in virtual communities.* OSF. https://doi. org/10.31234/osf. io/tz4dn.

Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., & Anandkumar, A. (2023). Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*.

Youn, H., Strumsky, D., Bettencourt, L. M., & Lobo, J. (2015). Invention as a combinatorial process: Evidence from us patents. *Journal of the Royal Society interface*, *12*(106), 20150272.

Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*, *115*(31), 7937–7942.

Zhao, B., Lucas, C. G., & Bramley, N. R. (2022). How do people generalize causal relations over objects? A non-parametric Bayesian account. *Computational Brain & Behavior*, *5*(1), 22–44.

Zhao, B., Lucas, C. G., & Bramley, N. R. (2024). A model of conceptual bootstrapping in human cognition. *Nature Human Behaviour*, *8*(1), 125–136.

Zhao, B., Vélez, N., & Griffiths, T. (2024). A rational model of innovation by recombination. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 46).