A Rational Model of Innovation by Recombination

Bonan Zhao (bnz@princeton.edu)
Department of Computer Science
Princeton University

Natalia Vélez (nvelez@princeton.edu)
Department of Psychology
Princeton University

Thomas L. Griffiths (tomg@princeton.edu)
Department of Computer Science
Department of Psychology
Princeton University

Abstract

Human learning does not stop at solving a single problem. Instead, we seek new challenges, define new goals, and come up with new ideas. What drives people to disrupt the existing conceptual landscape and create new things? Here, we examine the decision to create new things under different levels of expected returns. We formalize innovation as stochastically recombining existing ideas, where successful and more complex combinations generate higher returns. This formalization allows us to cast innovation-seeking as a Markov decision process, and derive optimal policies under different settings. Data collected through an online behavioral experiment confirm our prediction that people should invest more time and effort in seeking innovations when they know the chances of success are high and the potential new ideas would be rewarding. However, people also deviate from being optimal, both innovating more and less than they should in different settings.

Keywords: discovery; innovation; Markov decision process; decision making; crafting game

Introduction

The ability to create new ideas, concepts, and technologies is a crowning achievement of human cognition. From making tools to developing theories about the world, we constantly enrich, expand, and even revolutionize our pool of available choices. Those changes in possibilities are fueled by our ability to create new things, i.e., innovation. Despite its importance, innovation poses a mysterious decision problem: given that the currently available choices are already carefully chosen by the rational agents and have good returns, and attempting new ones does not guarantee successes, how do rational agents know when they should innovate?

While there is a rich literature on the historical, philosophical, and empirical aspects of innovation (e.g. Basalla, 1988; Kuhn, 1970; Muthukrishna & Henrich, 2016; Youn, Strumsky, Bettencourt, & Lobo, 2015), rational analyses of when agents should consider creating new choices have been much rarer (Bramley, Zhao, Quillien, & Lucas, 2023). This is because rational models of cognition usually consider a given set of candidates (e.g., Callaway et al., 2022), or how to search in an open space of hypotheses (e.g., Piantadosi, Rule, & Tenenbaum, 2024). These assumptions about a preset space of candidate choices are useful for studying certain scientific questions, but nevertheless flatten the tension between the selection versus generation of those choices, which is a core process underlying innovation.

We propose a formalization of innovation in this paper, explicitly allowing the agent to create new choices by combining existing ones. We examine the rational solution to the problem of deciding when to innovate, based on considerations of the risk and reward of pursuing innovation in different settings. We show that with a finite number of opportunities to innovate, this problem corresponds to an optimal stopping problem (T. S. Ferguson, 2006). We then report a behavioral experiment where we manipulated the probability that new combinations will succeed, and the rate at which rewards increase for more complex ideas. Our results reveal that people are sensitive to the factors identified in our rational model, but also systematically under- and over-explore in different settings. We conclude with a discussion of how extensions of this formal framework could inform our understanding of how people making new choices, and potentially grow new knowledge, out of what has already been discovered.

Background

We are interested in when rational agents should innovate—i.e., creating new choices from recombining existing ones. This implicitly assumes that ideas are compositional, and evokes a decision-making problem that reflects the classic explore-exploit trade-off. We summarize both ideas in this section, and highlights how crafting games provide an ideal setup to study innovation.

Innovation as recombination of existing ideas

Our tools and ideas are deeply compositional (Stigler, 1955; Cornish, Dale, Kirby, & Christiansen, 2017). Introducing a steam engine to spinning mules led to a new generation of semi-automatic machines, and combining knowledge of neurons and logic gave birth to the first artificial neural networks (McCulloch & Pitts, 1943). Most research studying how new ideas spread operationalizes ideas as samples from a continuous distribution, and generating new ideas as drawing new samples (e.g. Mason, Jones, & Goldstone, 2008; Mesoudi, Chang, Murray, & Lu, 2015; Thompson & Griffiths, 2019). However, this representation cannot capture the compositional nature of innovation. In fact, analyses of patent application data suggest that new technologies usually come from combining existing technologies (Arts & Veugelers, 2015; J.-P. Ferguson & Carnabuci, 2017; Youn et al., 2015). Similarly, scientific breakthroughs are built on apply-

ing new methods to old questions (Kuhn, 1970). Recent work has explored settings in which ideas are represented as items and innovation is successful combination of existing items (Derex & Boyd, 2016; Brändle, Stocks, Tenenbaum, Gershman, & Schulz, 2023). This representation emphasizes the compositional and cumulative aspects of innovation, while being abstract enough to study the cognitive mechanisms and computational principles driving innovation.

Innovation and exploration

In a compositional and potentially open-ended space of ideas, rational agents are faced with the decision of when they should innovate, as opposed to just sticking with what they already have. On the face of it, this problem echoes the classic explore-exploit trade-off, usually studied using multi-armed bandit tasks (Cohen, McClure, & Yu, 2007; Sutton & Barto, 2018). In these tasks, we imagine a slot machine equipped with many arms and an agent who has to decide between exploiting arms with known rewards and exploring unknown arms, with the goal of maximizing total rewards collected from pulling the arms. Innovation in the compositional space of ideas, however, has a substantially different structure: discovering a new idea effectively increases the space of available choices, and investing in developing a particular idea could change the potential reward associated with this choice. That is, the expected return of selecting a choice depends on the future innovations it could bring, which in turn depends on the agent's decision of pursuing innovation for that line of development. In short, studying when agents should innovate requires a task that goes beyond multi-armed bandits (Brändle, Binz, & Schulz, 2021). Such a computational problem may also pose challenges to agents with limited computational resources (Anderson, 1990; Griffiths, Lieder, & Goodman, 2015), and it remains to be tested whether people can solve these problems rationally.

Crafting games

The compositional and open-ended views of ideas are nicely captured in crafting games, where people combine existing objects to make new objects. For instance, binding a sharp stone with a wooden handle may make a stone hatchet. Not all combinations work out, however, which reflects the risk and opportunity cost of pursuing innovation. Popular crafting games, such as Minecraft and Little Alchemy, have inspired research on autonomous exploration in people (Brändle et al., 2023) and artificial agents (G. Wang et al., 2023). Crafting games are also widely used for designing benchmarks for human-like generalization and reasoning (Hafner, 2022; J. X. Wang et al., 2021). Based on these crafting games, in the next section we formally define a discovery game as part of a framework for studying innovation. Like crafting games, the discovery game sits on an open-ended space of compositions and recombinations. Instead of dealing with concrete objects and providing knowledge bases of recipes as in many crafting games (e.g., G. Wang et al., 2023), in the discovery game an item represents an idea, innovations are both rewarding and risky, and the game tree may grow infinitely.

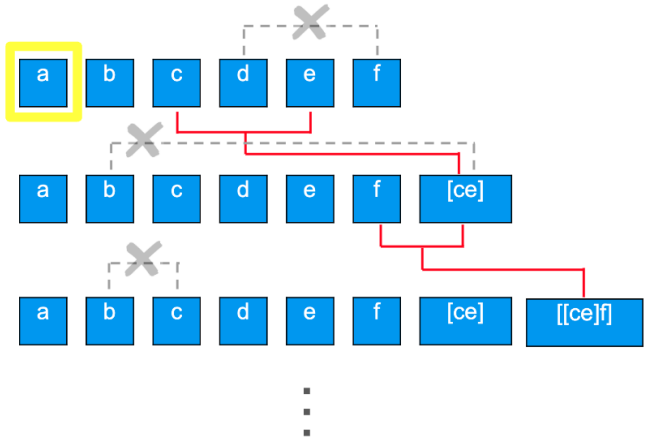

Figure 1: Visualization of the formal model. Blocks are game items (ideas). Each line is a round of the game. Yellow box marks cashing an existing item. The tree in red solid lines shows successful combinations (innovations), and gray dashed lines show failed combinations.

Formalizing innovation

We formalize ideas as items in a discovery game, and innovation as a successful recombination of existing ideas. In the discovery game, player can either benefit directly from an existing item (yellow box, Figure 1)—cashing in the item and collecting a reward—or try to combine existing items to produce a more rewarding one. A newly-discovered item will deliver greater reward, and the player can benefit from it by cashing in this item in a later step. Given the *chance* of discovery, and how much *increase* in returns a successful combination generates, we can derive rational solutions for when an agent should innovate.

Defining discovery games

A discovery game $\mathcal{G}$ is a tuple $\langle M, \mathcal{T}, A, R \rangle$, where M is a set of items, $\mathcal{T}$ the game tree recording successful combinations, A a set of actions, and R the reward function. Players may attempt to combine an item $m \in M$ with another item $n \in M$, denoted as $c(m,n)$. If $c(m,n) \in \mathcal{T}$, this is a successful combination and will produce a new item, say, $c(m,n) \Rightarrow o$. If $c(m,n) \notin \mathcal{T}$, the combination fails and no new item is discovered. There are many ways to parameterize the game, and we elaborate on some variations in the Discussion. To a first approximation, here we consider success rate p, defined as the probability of making a successful combination for a given item, i.e., $P(c(m,n) \in \mathcal{T}_m)$, where $\mathcal{T}_m$ is the subtree that only involves item m. We consider the situation where the success rate holds the same for all items. To capture the intuition that more complex ideas are more rewarding, we let the reward for an item to grow with the item's level. Items that cannot be produced by combining other items are base items, m^0, with base reward r. Climbing up the game tree increases levels: for

a combination $c(m^i, m^j) \Rightarrow m^k$, $k = \max(i, j) + 1$. The reward associated with item m^k is $w^k \cdot r$, where the reward increase rate $w > 1$.

Players can take two actions $A = \{use, combine\}$. Action *use* uses an item and receives the rewards associated with the item, $R(use(m^k)) = w^k r$, and action *combine* combines two items of choice. The immediate reward for taking this action is always zero, $R(combine(m,n)) = 0$. If the combination is successful, $c(m,n) \in \mathcal{T}$, $c(m,n) \Rightarrow o$, then later on the player may choose to benefit from this discovery by collecting rewards from this newly discovered item o, and $R(use(o)) = w \cdot \max(R(use(m)), R(use(n))) > \max(R(use(m)), R(use(n)))$.

Optimal policy

The above definitions form a transition matrix P: for a state $s = (k,t)$ where k records the highest level of currently available items, and step $t = 0, 1, \ldots$. With probability p, action *combine* discovers a new item and leads to state $s' = (k+1, t+1)$, and with probability $1-p$ action *combine* leads to state $(k, t+1)$. Action *use* always leads to state $(k, t+1)$. This forms a Markov decision process (MDP), and we can compute the optimal policy π^* following the optimal state-value function:

$$q^*(s,a) = R(s,a) + \gamma \sum_{s'} P(s,a,s') \max_a q^*(s',a). \qquad (1)$$

For a finite horizon of D steps in total, the optimal policy in this particular setting corresponds to an optimal stopping problem (T. S. Ferguson, 2006): one should keep attempting innovation (*combine*) until a switch point d, then focus on collecting the highest possible existing rewards (*use*). This is because it is always better to explore for x steps and then exploit than it is to alternate between exploring for x steps and spending the rest exploiting. The expected return for switching at step d is

$$\mathbb{E}_{\pi(d)} = (n-d) \left(\sum_{i=0}^{d} \binom{d}{i} (pw)^i (1-p)^{d-i} \right) r \qquad (2)$$

and the optimal switch point is $d^* = \arg\max_d \mathbb{E}_{\pi(d)}$. We can use the fact that it is always more rewarding to explore until the optimal switch point d^* to derive an analytical solution to Equation 2. Let $d' := D - d^* + 1$, interpreted as the number of steps left, and r_d for the most rewarding item m at step d^*:

$$\begin{aligned} (pwr_d + (1-p)r_d)(d'-1) &\geq d' r_d \\ pd'w - pw + d' - pd' - 1 + p &\geq d' \\ p(d'-1)(w-1) &\geq 1 \\ d' &\geq \frac{1}{p(w-1)} + 1. \end{aligned} \qquad (3)$$

Equation 3 implies that the optimal switch point depends on how may steps left, given the success rate p and the reward increase rate w. This future-looking aspect of the model follows from the exponential growth of the reward function, and may vary if the reward function is set in different ways.

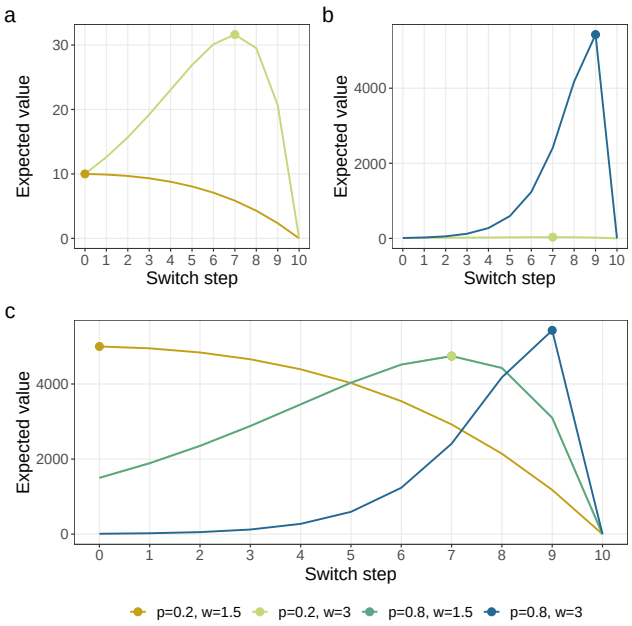

Figure 2: Expected returns for switching-once at different steps. Optimal switch point is marked in big dots. a. reward increase rate $w = \{1.5, 3\}$, success rate $p = 0.2$, and base reward 1. b. Success rate $p = \{0.2, 0.8\}$, reward increase rate $w = 3$, and base reward 1. c. $p = \{0.2, 0.8\} \times w = \{1.5, 3\}$ with scaled base rewards, see the Material section for explanation.

Here, a later switch point corresponds to more steps of attempting new combinations. Hence, our model predicts more exploration and greater rewards when the reward increase rate grows while holding success rate fixed (Figure 2a) or when success rate is increased while holding the reward increase rate fixed (Figure 2b). Below, we tested these predictions against human behavior in an online experiment.

Testing the model predictions

We implemented the discovery game presented above in an online experiment interface and tested the model predictions.

Methods

Participants 210 participants were recruited through Prolific Academic (97 females, $M_{\text{age}} = 38 \pm 12$). The sample size was determined by a power analysis aiming to to obtain .95 power to detect a medium effect size of .25 at the standard .05 alpha level. No participant was excluded from analysis. To ensure data quality, all participants had to complete two practice trials and pass a comprehension check before starting the main task. Participants were paid both for their time and a performance-based bonus. The task took 7 ± 2.5 minutes. The experiment was performed with approval by the Research Integrity & Assurance Committee of Princeton University (ref. IRB 10859). Preregistration for the experiment is available at https://osf.io/yph4f/. All participants gave informed consent before undertaking the experiment.

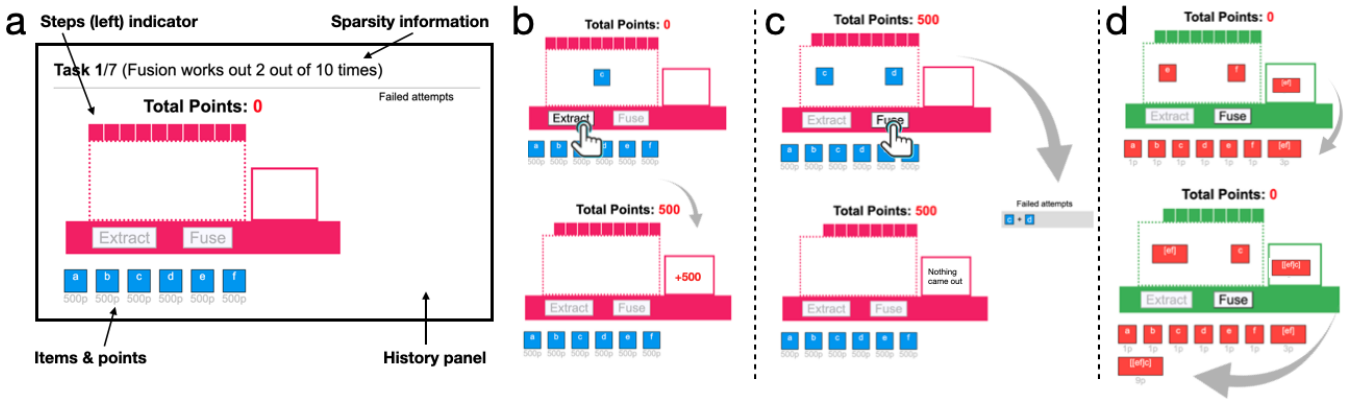

Figure 3: Experiment interface. a. Screenshot of a discovery game in the experiment. Arrows and bold text are for illustration only and not shown to participants. b. A usage of the "Extract" button, leading to gaining 500 points in this demo. Note that one step is consumed after clicking the button. c. A usage of the "Fuse" button that leads to nothing. d. Examples of successful fusion discoveries in another machine.

Materials, procedure & design Participants played the discovery game depicted in Figure 3a. In each round of the game, participants were shown a new machine with an "Extract" button and a "Fuse" button, six base items below the machine, and a counter indicating the number of actions left within the round, shown as a line of bars on the top rim of the machine. Clicking an item puts the item in the machine, and the machine can hold up to two items at a time. Participants were instructed that they could (1) collect points by putting a single item in the machine and clicking the "Extract" button (Figure 3b), or (2) make new items by putting two existing items in the machine and clicking the "Fuse" button (Figure 3c-d). We made it clear to participants that fusions succeed (i.e. lead to a new item) $x = 10 \times p$ out of 10 times, and that newly-discovered items are worth w times more points than the most rewarding item used to make them. Extracting points from an item or attempting a fusion each consume one available action; repeating a past unsuccessful fusion attempt does not consume an action. To make it easier for participants to keep track of actions they had already attempted, each item was labeled with the points that participants would gain from extraction, and a history of failed fusion attempts was displayed next to the machine. Participants' goal was to maximize the total number of points collected in each round. Bonus was calculated based on the total points collected.

Participants were randomly assigned to four between-subjects conditions that differed in the model-predicted optimal switch point. Within each of these conditions, we independently manipulated the success rate p to be high ($p = 0.8$) or low ($p = 0.2$), and the reward increase rate w to be high ($w = 3$) or low ($w = 1.5$). Together, these led to a 2×2 between-subject design: high-p-high-w (hh), high-p-low-w (hl), low-p-high-w (lh), and low-p-low-w (ll). To ensure that the total expected rewards are matched across conditions, the reward that participants receive for extracting base items was set to 1 point for the *ll* condition, 150 points for the *lh* and *hl* conditions, and 500 points for the *hh* condition (Figure 2c).

Participants completed seven independent rounds of the discovery game, each marked using different color-coded machines; p and w were held constant for all seven machines, but permissible combinations—i.e., $\mathcal{T}$—changed from round to round. After the task, participants completed a debriefing form where they provided demographic information, feedback, and self-reports of how they played the game.

Results

We analyzed the 7 rounds each participant played, totalling $210 \times 7 = 1470$ rounds. A button click in the experiment is a step in the model, extracting points corresponds to action *use*, and fusing corresponds to action *combine*. For the 10 actions in each round, the proportion of fusion attempts corresponds to how many steps of recombination the model predicts, as specified by the optimal switch point. All data and analysis are available at https://osf.io/8gwpv/.

Participants calibrated innovation-seeking to expected returns. As predicted by Equation 3, overall people seek more innovations when success rate p is higher and reward increase rate w is higher (Figure 4a). We ran a mixed-design ANOVA with success rate p and the reward increase rate w as primary factors and round as a repeated measure, and the results indicated a significant main effect of success rate ($F(1,206) = 46.954, p < 0.001, \eta^2 = 0.137$), and the reward increase rate ($F(1,206) = 10.905, p = 0.001, \eta^2 = 0.036$). We did not observe a significant effect of round ($F(5.2,1070.43) = 0.283, p = 0.928, \eta^2 = 0.000415$) nor an interaction between success rate and reward increase rate ($F(5.2,1070.43) = 2.049, p = 0.067, \eta^2 = 0.003$). As predicted, both factors independently encourage innovation seeking. Participants in higher p and higher w conditions also discovered more advanced items (higher levels) overall (Figure 4b). Both factors independently predict the highest item level that participants achieve in a round of game (success rate: $F(1,206) = 454.543, p < 0.001, \eta^2 = 0.583$; reward increase rate: $F(1,206) = 20.830, p < 0.001, \eta^2 = 0.060$).

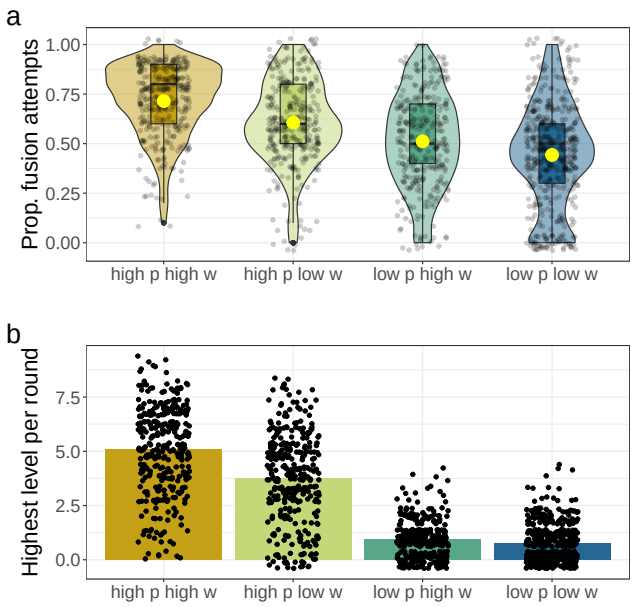

Figure 4: Behavioral experiment results. a. Proportion of fusion attempts per round in each condition. Yellow dots are group means. b. Highest item level discovered per round in each condition. Bars are group means.

Participants weighted p and w differently. Our model predicts that participants should attempt fusions at similar rates in the low-p-high-w and high-p-low-w conditions, because both have the same optimal switch point. However, participants attempted significantly more fusions in the high-p-low-w condition ($t(652.93) = -5.4, p < .001$, Cohen's $d = 0.44$). Figure 5a illustrates how often participants attempted to fuse over the course of a single round, and it shows that participants in the low-p-high-w condition consistently made fewer proportion of fusion attempts in each corresponding step of a round of the game. Together, these results suggest that though participants' fusion attempts are influenced by both success rate and reward increase rate, participants did not weigh these factors equally; lower success rates discouraged participants from innovating, even in the face of a high reward increase rate.

Most participants switch once from innovating to extracting. The model predicts that there is an optimal point when participants should switch from innovating and developing new items to extracting points from the items available. Overall, participants indeed started off by attempting to innovate, and then switched once from innovating to extracting points from the items available: 73.5% of rounds in the high-p-high-w adopted this "switch-once" strategy (chance level $10/2^{10} \approx 0.01$, $\chi^2(1, N = 336) = 17845, p < .001$, with simulated p-value based on 2000 replicates, same for below), along with 63.7% in high-p-low-w ($\chi^2(1, N = 336) = 13339, p < .001$), 64.7% in low-$p$-high-$w$ ($\chi^2(1, N = 329) = 13502, p < .001$),

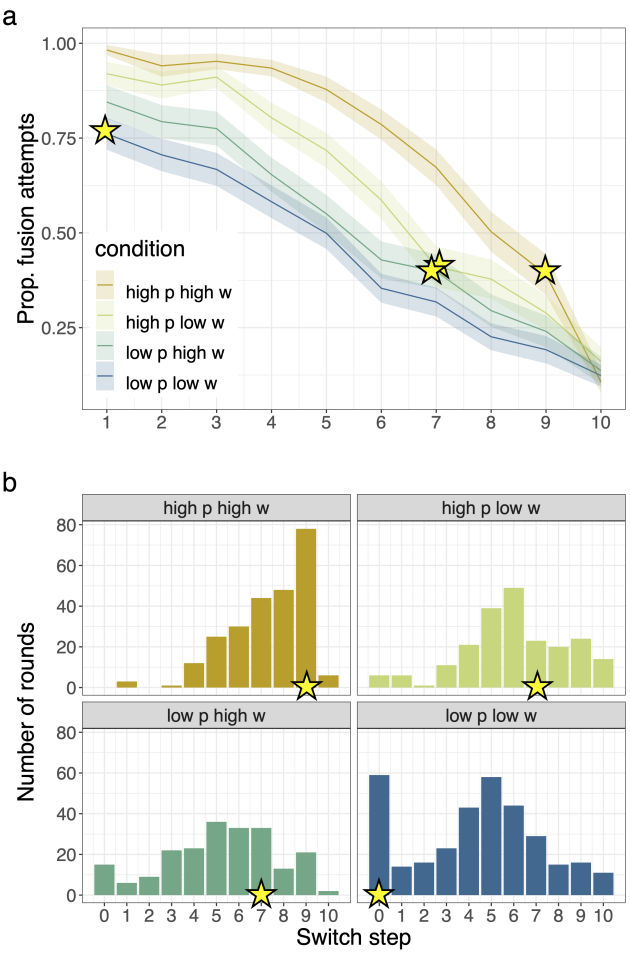

Figure 5: Comparing the timing of participants' switches to model-predicted optimal switching points (stars). a. Average frequency of fusion attempts in each step over rounds per condition. b. Histogram of switch points in each round that exhibits a switch-once strategy

and 69.9% in low-p-low-w ($\chi^2(1, N = 469) = 22513, p < .001$), all significantly above adopting a switch-once strategy by chance. The timing of this switch, however, did not always align with the model-predicted optimum. Figure 5b shows the distribution of the switch points in the rounds containing a single switch point. In the high-p-high-w condition, the most common switch point is the 9$^{\text{th}}$ step (32% of all rounds), which corresponds to the model-predicted optimum, and switch probabilities steadily increase from the 4$^{\text{th}}$ to the 9$^{\text{th}}$ steps. In the remaining conditions, however, participants adopted a mix of strategies. The most common switch point in the low-p-low-w condition is step zero (no fusion at all; 18% of rounds), which aligns with the model-predicted optimum; however, the timing of participants' switches was more variable overall, and there was a strong competing choice of switching at the 5$^{\text{th}}$ step (17.6%). In the other two conditions, the most commonly-selected switch point differ from the model-predicted optimum (step seven), and was distributed

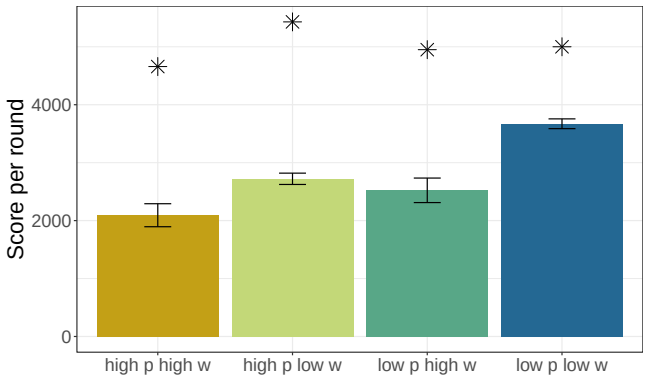

Figure 6: Average score per condition in the experiment. Asterisks mark the total expected points predicted by the theoretical optimal strategy.

rather evenly around the model-predicted optimum. The substantial over-exploration in the low-p-low-w condition and wide-spread under-exploration in the other three conditions are also reflected in the overall lower total scores than the theoretical optimal (Figure 6).

Discussion

Human innovation is fundamentally compositional and open-ended—each idea can give rise to new, increasingly complex ideas (Stanley, 2019) This creates a challenge for human decision making and planning: how much should we invest in attempting new innovations, versus capitalizing on ideas that are already available? Here, we formalized this decision in a discovery game inspired by crafting games, and examined rational solutions for when an agent should innovate. We tested model predictions in a simplified, yet open-ended crafting game. In line with our predictions, people make decisions about whether to attempt innovations by considering how likely it is that the attempt will work out, and how rewarding the discovery would be.

Although people considered both the success rate and reward increase rate in making innovation decisions, there are clear patterns of deviation from the optimal level of innovation predicted by the rational model. First, rather than considering both factors equally, participants seems to be more sensitive to success rate p. One possible explanation for this discrepancy is that participants may be risk averse (Arrow, 1965; Pratt, 1978), preferring to extract rewards from known options than to risk losing out on rewards by attempting innovations that are unlikely to work out. Interestingly, when both success rate p and reward increase rate w are low, the model predicts that any innovation attempt is sub-optimal, and yet in a substantial number of rounds we observed innovation attempts. This could be accounted for by the novelty bias (Gershman & Niv, 2015; Krebs, Schott, Schütze, & Düzel, 2009; Stojić, Schulz, Analytis, & Speekenbrink, 2020). However, instead of being driven by uncertainty, in this setup people actually have complete information, hence it remains to be seen whether such over-exploration is caused by mis-interpretation of the task, or some genuine bias about innovation seeking.

Moving forward, our simple parametrization of the task can be extended to capture other important features of real-world innovations. For example, not every new combination of ideas has the same probability of success, and real-world entrepreneurs and researchers often have to choose between developing small and incremental improvements that are likely to work and yield low rewards, or attempting big leaps that connect previously disparate ideas, carrying higher risk but also potentially bringing greater rewards (Fleming, 2001; J.-P. Ferguson & Carnabuci, 2017). We can study how people navigate this trade-off in a more controlled setting by introducing dependencies between the potential risk and returns, rather than varying the two independently and holding them fixed through the duration of the task. Future work may examine how people calibrate their innovation-seeking by estimating p from domain-specific knowledge, rather than being told p explicitly. Such extension will also allow modeling of "stepping stones"—items that are only meaningful as a precursor to some future discovery, rather than immediately useful for their own sake.

While the current work focuses on individual decisions about when to pursue innovations, individuals do not typically innovate alone. Our framework could be extended to capture how innovation-seeking decisions are made by teams or collections. In multiplayer games, teams may buffer themselves from the risk of pursuing innovations by dividing labor, while communication costs could impede the speed and quality of discovery (Almaatouq, Alsobay, Yin, & Watts, 2021; Ethiraj & Levinthal, 2004). Social network structures that specify how information flows have a big impact on how innovations spread (Mason et al., 2008; Derex & Boyd, 2016), and our framework provides a rich space for examining how individual cognitive mechanisms give rise to group-level dynamics (Muthukrishna & Henrich, 2016).

In short, this work is a first step toward answering how people renovate their current toolkit in a compositional space of options. By combining ideas from different literature in novel ways, we hope we have increased the rewards that might be derived from further studies of innovation.

Acknowledgments

This work was supported by a grant from the Templeton World Charity Foundation (TWCF 20648) to Griffiths and Vélez.

References

Almaatouq, A., Alsobay, M., Yin, M., & Watts, D. J. (2021). Task complexity moderates group synergy. *Proceedings of the National Academy of Sciences*, *118*(36), e2101062118.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Arrow, K. J. (1965). *Aspects of the theory of risk-bearing*. Yrjö Jahnssonin Säätiö, Helsinki.

Arts, S., & Veugelers, R. (2015). Technology familiarity, recombinant novelty, and breakthrough invention. *Industrial and Corporate Change*, *24*(6), 1215–1246.

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

Bramley, N. R., Zhao, B., Quillien, T., & Lucas, C. G. (2023). Local search and the evolution of world models. *Topics in Cognitive Science*.

Brändle, F., Binz, M., & Schulz, E. (2021). Exploration beyond bandits. In I. C. Dezza, E. Schulz, & C. M. Wu (Eds.), *The drive for knowledge: the science of human information seeking* (pp. 147–166). Cambridge University Press.

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.

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022). Rational use of cognitive resources in human planning. *Nature Human Behaviour*, *6*(8), 1112–1125.

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.

Cornish, H., Dale, R., Kirby, S., & Christiansen, M. H. (2017). Sequence memory constraints give rise to language-like structure through iterated learning. *PloS one*, *12*(1), e0168532.

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

Ethiraj, S. K., & Levinthal, D. (2004). Modularity and innovation in complex systems. *Management science*, *50*(2), 159–173.

Ferguson, J.-P., & Carnabuci, G. (2017). Risky recombinations: Institutional gatekeeping in the innovation process. *Organization Science*, *28*(1), 133–151.

Ferguson, T. S. (2006). *Optimal stopping and applications.* https://www.math.ucla.edu/~tom/Stopping/Contents.html.

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

Gershman, S. J., & Niv, Y. (2015). Novelty and inductive generalization in human reinforcement learning. *Topics in cognitive science*, *7*(3), 391–415.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Hafner, D. (2022). Benchmarking the spectrum of agent capabilities. In *Proceedings of the tenth international conference on learning representations.*

Krebs, R. M., Schott, B. H., Schütze, H., & Düzel, E. (2009). The novelty exploration bonus and its attentional modulation. *Neuropsychologia*, *47*(11), 2272–2281.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (Vol. 111). Chicago University of Chicago Press.

Mason, W. A., Jones, A., & Goldstone, R. L. (2008). Propagation of innovations in networked groups. *Journal of Experimental Psychology: General*, *137*(3), 422–433.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, *5*, 115–133.

Mesoudi, A., Chang, L., Murray, K., & Lu, H. J. (2015). Higher frequency of social learning in china than in the west shows cultural variation in the dynamics of cultural evolution. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1798), 20142209.

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

Piantadosi, S. T., Rule, J. S., & Tenenbaum, J. B. (2024). Learning as bayesian inference over programs. In T. L. Griffiths, N. Chater, & J. B. Tenenbaum (Eds.), *Bayesian models of cognition: Reverse-engineering the mind.* MIT Press.

Pratt, J. W. (1978). Risk aversion in the small and in the large. In *Uncertainty in economics* (pp. 59–79). Elsevier.

Stanley, K. O. (2019). Why open-endedness matters. *Artificial life*, *25*(3), 232–235.

Stigler, G. J. (1955). The nature and role of originality in scientific progress. *Economica*, *22*(88), 293–302.

Stojić, H., Schulz, E., Analytis, P. P., & Speekenbrink, M. (2020). It's new, but is it good? how generalization and uncertainty guide the exploration of novel options. *Journal of Experimental Psychology: General*, *149*(10), 1878—1907.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.

Thompson, B., & Griffiths, T. (2019). Inductive biases constrain cumulative cultural evolution. In *Proceedings of the 41st annual meeting of the cognitive science society* (pp. 1111–1117).

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

Wang, J. X., King, M., Porcel, N., Kurth-Nelson, Z., Zhu, T., Deck, C., ... others (2021). Alchemy: A benchmark and analysis toolkit for meta-reinforcement learning agents. *arXiv preprint arXiv:2102.02926*.

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.