Title
Slow But Motivated: Social Learning in Open-ended Intellectual Innovation

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Gong, Siyi
Jiang, Kaiwen
Kreiss, Elisa
et al.

Publication Date
2026

Slow But Motivated: Social Learning in Open-ended Intellectual Innovation

Siyi Gong (siyi.gong@ucla.edu)[1], Kaiwen Jiang[2], Elisa Kreiss (ekreiss@ucla.edu)[1*] &
Tao Gao (taogao@ucla.edu)[1,3*]
[1]Department of Communication, University of California, Los Angeles
[2]Department of Statistics and Probability, Michigan State University
[3]Department of Statistics & Data Science, University of California, Los Angeles
[*]Co-senior authors

Abstract

Social learning is known for preserving existing cultural designs, but is it also important for generating novel, effective solutions? Building on theoretical and empirical work, we identify intellectual innovation as a knowledge-based form of innovation that has been largely overlooked and examine how social learning shapes it. We introduce a riddle-crafting paradigm combined with a closed diffusion-chain design, enabling open-ended creation, multidimensional measurement, and quantifiable performance. In Experiment 1, participants in social and asocial chains composed and refined riddles distinguishing poisonous from edible mushrooms; in Experiment 2, independent participants attempted to solve them. Although the two modes showed comparable objective solvability, social riddles were more efficient and diverse in form and content. Social participants also showed greater motivation to improve, with hints of incremental gains that may emerge more fully over extended generations. Therefore, socially transmitted intellectual innovations become better tailored for communication, open to mutation, and poised for cumulative refinement.

Keywords: intellectual innovation; social learning; cultural evolution; collective problem-solving

Introduction

The cumulative advancement of human culture is suggested to be powered by two complementary "engines": imitation and innovation (Legare & Nielsen, 2015). While high-fidelity imitation stabilizes and preserves useful modifications (Tennie et al., 2009; Tomasello et al., 1993), innovation generates new variants of cultural traits that allow individuals and groups to better adapt to changing environments (Laland, 1992). As such, innovation is defined by the dual properties of *novelty* and *improved performance*: it is a new or modified solution that more effectively addresses a new or existing problem in a novel manner (Carr et al., 2016; Fogarty et al., 2015; Legare & Nielsen, 2015; Mesoudi & Thornton, 2018).

While innovation is often framed as an asocial activity distinct from social learning (e.g., Kendal et al., 2005), other accounts emphasize that innovation frequently relies on socially scaffolded processes (Carr et al., 2016; Mesoudi & Thornton, 2018). This raises a central question: What role does social learning play in generating novel and functionally effective designs? The cognitive niche account proposes that humans' unique cognitive capacities—especially technical reasoning—are the primary engines of improvement (Osiurak, 2014; Osiurak & Badets, 2016, 2017; Osiurak & Reynaud, 2020; Pinker, 2010). In contrast, the cultural niche account argues that social learning expands access to diverse ideas and solutions, enabling individuals to stumble upon unexpected improvements, recombine existing elements, and make incremental refinements (Boyd et al., 2011; Henrich, 2016; Muthukrishna & Henrich, 2016).

To rigorously test these competing perspectives, empirical studies typically use micro-society paradigms (Baum et al., 2004; Jacobs & Campbell, 1961), such as the closed diffusion-chain method (Bartlett, 1995[1932]) or closed-group designs, which simulate cultural "generations" within a controlled laboratory setting. These designs compare innovations with objective performance metrics (e.g., tool manufacture) produced by social diffusion chains against a carefully matched asocial baseline. Social learning shows an advantage when transmission across individuals yields better outcomes than repeated efforts by the same individual.

Yet, evidence from these studies remains mixed and complex: Some show that social learning facilitates innovation in adults (Derex & Boyd, 2015; Derex et al., 2013, 2015; Mesoudi, 2008; Mesoudi & O'Brien, 2008; Wasielewski, 2014) and children (McGuigan et al., 2017). Others indicate that this advantage is not universal and depends on the specific task (Lucas et al., 2020; Zwirner & Thornton, 2015), form of learning (Derex et al., 2013), the problem space (Derex et al., 2015; Yahosseini & Moussaïd, 2020), and individuals' skill levels (Yahosseini & Moussaïd, 2020).

We identify two key limitations in the existing literature. First, empirical research has focused predominantly on innovations that directly modify the physical world, such as tool manufacture. Yet innovation extends well beyond instrumental skills. An important and largely overlooked form is *intellectual innovation*—innovation rooted in ideas, concepts, strategies, and knowledge. Such innovations, which include those in social conventions, strategic reasoning, and information acquisition, offer essential advantages for human adaptation to complex ecological and social environments. For example, the *buffalo jump* technique, developed by many Native American Plains groups, involved coordinating hunters to drive bison off cliffs. This cooperative strategy dramatically increased hunting efficiency without requiring new physical tools. Second, many tasks are closed-space and unidimensional, prescribing a fixed set of permissible solutions—for example, constraining artifact construction to specific ingredient combinations (Derex & Boyd, 2015). Real-world innovation, by contrast, arises in open-ended design spaces where

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

possibilities are not exhaustively defined in advance.

In this paper, we investigate the role of social learning in intellectual innovation by introducing a riddle-crafting paradigm. As a form of folklore, riddles pair an enigma with a hidden solution. As such, crafting clever clues to conceal the referent requires substantial cognitive effort, making riddle crafting a strong testbed for intellectual innovation. Moreover, the arbitrariness and productivity of natural language create a minimally constrained design space in which genuinely novel ideas can emerge. This paradigm enables us to capture the flexibility, diversity, and generativity of human innovation beyond a single performance metric. Crucially, riddles also provide a clear ground truth—solvability—which offers an objective anchor for evaluating performance and an efficient way to calibrate task difficulty.

The study implemented two transmission modes within a closed diffusion-chain structure to examine the effect of social learning on this task. The asocial mode had the same individual compose and refine a riddle across three generations, while the social mode had a riddle created by one participant and successively refined by new participants. In Experiment 1, we examine the innovation process and outcome to identify the effects of social learning on riddles' efficiency, novelty, and subjective solvability. In Experiment 2, we evaluate the riddles' objective solvability to investigate how accurate participants' self-evaluations align with objective performance.

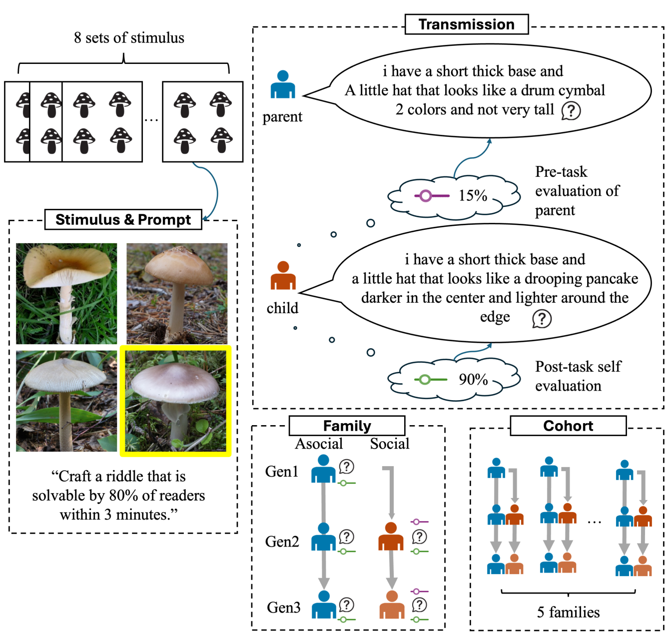

Figure 1: Illustration of the structure of Experiment 1.

Experiment 1

Experiment 1 investigates how intellectual innovation emerges and develops across generations in riddle crafting for the two transmission modes. The experiment was pre-registered on OSF and received approval from the University of California, Los Angeles Office of the Human Research Protection Program. All methods were performed in accordance with the relevant guidelines and regulations. The general structure of the experiment is shown in Fig.1.

Method

Participants Participants were recruited online via Prolific, limited to users located in the US, who have English as their first language, and within the age range of 18-50. Data from 120 participants ($N_{asocial}$=40, N_{social}=80) were collected after applying exclusion criteria (see pre-registration). The experiment lasted approximately 10 minutes for the asocial transmission mode and 8 minutes for social, with a compensation rate of approximately $15 per hour.

Materials A total of 8 stimuli sets were created. Each set consisted of four mushroom images, each depicting a distinct fungal species, three edible and one poisonous. All mushroom pictures were collected from publicly available sources. All images were cropped to a square format and approximately 129 px in width. The images within each set were arranged in a 2×2 grid and had high visual similarity, with their order randomized for each participant. The poisonous mushroom image was highlighted with a 6-pt yellow border.

Design This experiment contained two transmission modes (asocial and social) and three generations of transmissions. Each condition was defined by a specific combination of mode and generation. Stimulus and transmission mode were fully manipulated between participants. In the asocial mode, the same participant incrementally improved their own riddle, whereas in the social condition, a new participant took over improving a riddle provided by a different participant.

Participants in the asocial mode each completed all three generations for a single stimulus. Stimulus sets were counterbalanced such that each set was presented to five asocial participants. Asocial Generation 1 riddles directly seeded Social Generation 2 to ensure an equivalent baseline. Since riddles were first conceived in Generation 1, we consider this generation neutral with respect to transmission mode.

Therefore, there are five unique conditions, each producing 40 riddles (8 stimulus sets × 5 participants each), thus yielding 200 unique riddles in total (120 asocial, 80 social). For organizational clarity, we use the following terms:

Cohort: *all 25 riddles associated with the same stimulus, pooled across transmission modes.*

Family: *all five riddles ultimately descended from the same Generation 1 seed under a given stimulus, consisting of one neutral, two asocial, and two social riddles.*

Parent riddle: *the previous riddle that directly seeds the current riddle within a family.*

Procedure Before beginning the task, participants were familiarized with three well-known folk riddles and their solutions. In Generation 1, participants were given stimuli and instructions to compose a riddle that could be solved by just

80% of readers within 3 minutes. They had 7 minutes to compose the riddle. After completing the composition task, they were asked how they would rate the solvability of their own riddle by answering (*perceived solvability*): "According to your estimation, what percentage of people will be able to solve it within 3 minutes?" Ratings were provided on a 0–100 slider with a step size of 1, initialized at 0.

Asocial mode: Proceeding to Generation 2, the participants from Generation 1 were asked to revise their riddle as they see fit to better align with the task instructions within 5 minutes. After the refinement, they were again asked to rate the self-perceived solvability of the refined riddle. Generation 3 followed the same procedure of Generation 2.

Social mode: The social mode followed a similar pipeline, but with each generation completed by a different participant. When introducing the task, they were informed that another prolific user had already crafted a riddle that they would receive. They were then asked to rate the solvability of that *parent riddle*, alongside the corresponding stimulus set. During the rating, they had the option to reveal the poisonous mushroom to inform their judgment. After the refinement task, participants again evaluated their own riddles' solvability.

Measures We looked at three theoretical constructs of riddles to assess the innovation process and outcomes.

Efficiency: A riddle's efficiency is estimated based on (1) its length in terms of words and characters, (2) the time (in minutes) it took to craft it, and (3) the density of information, computed as the Shannon entropy (Shannon, 1948) of each riddle divided by its token count.

Novelty: The novelty of each riddle was quantified on three levels of scope. *Local drift* captured how much a riddle differed from its immediate parent, and lexical and semantic differences were computed using Jaccard distance (Jaccard, 1901) and cosine distance between sentence-transformer embeddings (`all-MiniLM-L6-v2`; Devlin et al., 2019; Wang et al., 2020). Higher values indicate greater parent–child divergence (i.e., more changes made).

Second, *family-level diffusion* measures how similar each riddle was to all other riddles within its lineage family. Lexical similarity was quantified using Self-BLEU (Zhu et al., 2018), and semantic similarity was computed using the cosine similarity of their sentence embeddings. Higher values indicate greater similarity to the family (i.e., less novelty).

Third, *population dispersion* assessed how far each riddle deviated from its cohort. For each stimulus, we constructed its lexical centroid using TF–IDF vectors (Leskovec et al., 2020) and its semantic centroid using sentence embeddings. We then computed the cosine distance between each riddle and its corresponding centroid. Higher values reflect greater dispersion (i.e., more diversity).

Subjective solving performance: All riddles received a self-perceived solvability rating from the crafters themselves. To assess how closely each riddle was perceived to meet the desired solvability, we calculated the *subjective deviation*–its absolute deviated proportion from the target ($abs(percentage - 80)/100$). Lower values indicate riddles whose perceived solvability was closer to the desired level.

In addition, we measured how much crafters believed their riddles improved. *Subjective improvement* was defined as the change in absolute deviation from the 80% solvability constraint between a crafter's rating of their refined riddle and their rating of the parent riddle ($abs(parent - 80) - abs(child - 80)$). Larger values indicate a stronger belief that the refined riddle approached the desired solvability.

Results

Each measure was analyzed using a mixed-effects model (Baayen, 2008; Baayen et al., 2008) implemented in R (R Core Team, 2023) via the lme4 package (Bates et al., 2015). Models were estimated using restricted maximum likelihood (REML). All models included fixed effects of transmission mode (asocial, social), generation (Generation 2, Generation 3), and their interaction.[1] Random effects were specified for crafters, families, and stimulus sets. Post-hoc pairwise comparisons for main effects and simple effects were conducted using estimated marginal means (emmeans; Lenth, 2025). All model results are displayed in Table 1.

Efficiency We first examined whether riddles differed in efficiency—measured by length, information density, and crafting duration—to understand how participants approached the task and whether the two transmission modes produced systematically different forms of riddle design.

Across generations, asocial riddles were generally less efficient than social riddles: they were reliably longer (via word and character counts), contained less information (via information density), and took less time to compose (via craft duration). Importantly, these differences emerged clearly by Generation 3, whereas Generation 2 showed no reliable divergence between the modes. Looking within each mode, asocial riddles grew longer and less information-dense across generations, suggesting an accumulation of edits without substantial compression. In contrast, social riddles maintained relatively stable length and density, despite being repeatedly revised by different individuals.

These results highlight distinct editing trajectories across transmission modes. Asocial participants tended to elaborate on their own riddles, gradually expanding them in ways that made refinement easier but added little new information. In contrast, social participants invested more time reworking the riddle, producing versions that were more concise while retaining information. Together, these patterns suggest that the two modes fostered different styles of intellectual innovation, each with its own characteristic trade-offs.

Novelty Having established that the two transmission modes differ in their editing characteristics, we next examined

[1] Generation 1 data were excluded because they served as the neutral baseline and would have led to an imbalance in observations. Generation thus was treated as a categorical factor with two levels.

	Contrast: Social - Asocial			Contrast: Gen 3 - Gen 2		
	Overall	Gen 2	Gen 3	Overall	Asocial	Social
Efficiency Measures						
Word count (↓)	−8.43**	−5.60	−11.2***	−0.425	2.40*	−3.25
Character count (↓)	−38.6**	−26.2	−51.1***	−1.39	11.1*	−13.8
Craft duration (–)	0.672**	0.400	0.944**	−0.0544	−0.327	0.218
Information density (↑)	0.0297**	0.0206	0.0387**	−0.00008	−0.00912**	0.00895
Novelty Measures						
Jaccard distance from parent (↑)	0.334***	0.215***	0.452***	−0.0426	−0.161***	0.076
Semantic distance from parent (↑)	0.271***	0.183***	0.360***	0.00553	−0.083*	0.094
Self-BLEU from family (↓)	−40.8***	−32.2***	−49.5***	−16.4***	−7.79***	−25.10***
Semantic similarity from family (↓)	−0.179***	−0.120**	−0.238***	−0.0876***	−0.0285*	−0.1466***
TF-IDF distance from cohort (↑)	0.102***	0.0781***	0.1254***	0.0425***	0.0188*	0.0662***
Semantic distance from cohort (↑)	0.0448*	0.0493**	0.0402	0.0138	0.0184	0.00927
Subjective solving Performance						
Subjective deviation (↓)	−0.0123	−0.0600*	0.0355	0.0205	−0.0272	0.0683*
Subjective improvement (↑)	15.4***	19.2***	11.6**	−0.0125	3.80	−3.83
Objective solving performance						
Solving duration (–)	−0.0743**	−0.0587	−0.0898**	0.00901	0.0246	−0.00655
Objective deviation (↓)	−0.0238	0.0100	−0.0575	−0.0288	0.0050	−0.0625
Objective improvement (↑)	2.88	−1.00	6.75	−0.125	−4.00	3.75
Exact proportion (↑)	0.396	−0.184	0.977	0.185	−0.395	0.765

Table 1: Mixed-effects model results. Each row presents the emmeans output for one dependent variable. The leftmost column lists the dependent variables; arrows indicate a desirable direction of change, and dashes indicate cases where no clear directional expectation applies. The middle three columns report contrasts between social and asocial modes; the right three columns report contrasts between Generations 3 and 2. Within each contrast, the "Overall" subcolumn shows the main effect, and the remaining subcolumns show the corresponding simple effects. Table entries are model coefficients, with significance indicated by stars (* $p < .05$; ** $p < .01$; *** $p < .001$).

whether these editing differences led to different types of mutations (i.e., changes in narrative). To do so, we examined the riddles' local drift, family-level diffusion, and population dispersion.

Overall, riddles from both modes generally became more diverse from their parent (via Jaccard and semantic distances from parent), family (via Self-BLEU and Semantic similarity from family), and cohort (via TF-IDF and Semantic distances from cohort). Moreover, we found that asocial riddles became more similar to their parents over generations (see Fig.2A), suggesting progressively narrower edits and reduced exploration. However, social riddles consistently exhibited greater novelty and diversity than asocial riddles (also see Fig.2A). Across both lexical and semantic measures, social riddles tended to diverge more from their parent, family, and cohort. These differences were already visible in Generation 2 and became more pronounced in Generation 3.

Together, these findings suggest that participants in the asocial mode made smaller, incremental edits, producing a more monotonic and conservative evolution of their riddles over time. On the contrary, participants in the social mode made broader, more global transformations to the riddle they inherited. This contrast highlights how different transmission modes foster distinct innovation dynamics: social chains encourage restructuring and diversification, whereas asocial chains tend toward refinement and stabilization.

Subjective solving performance Lastly, we aimed to understand whether differential editing patterns and mutation dynamics relate to differences in how participants perceived the solvability of their own riddles and their parent riddles.

As shown in Fig.2B, overall, participants generally believed that their riddles were situated around the 80% solvability mark, with more riddles rated on the easier side (above 80%) than on the difficult side (below 80%). Results on subjective deviation show that, at Generation 2, asocial riddles were rated as closer to the desired solvability than social riddles, though this difference disappeared by Generation 3. Moreover, across generations, subjective ratings of social riddles gradually diverged from the target solvability from Generation 2 to 3, whereas those of asocial riddles remained close.

More interesting patterns emerge when we examine the subjective perceived *improvement* in solvability (see Fig.2C). We found a stark difference between modes: social participants believed their riddles improved substantially relative to their parents, whereas asocial participants reported a comparable amount of improvement and decline from their previous composition. These patterns were robust across generations.

From these findings, we see that although social riddles were not any closer to the target solvability than asocial riddles, they showed substantially larger self-perceived improvement. This pattern may arise because social participants appeared to believe that their parent riddles performed poorly, leading them to feel that they had "corrected" the trajectory. In other words, they may not have viewed their own refinements as objectively strong, but they consistently perceived them as markedly better than the riddles they inherited.

Taken together, these findings suggest a strong motivational contrast between modes. Participants in the social mode appeared motivated to start fresh, made substantial edits, and were optimistic about their changes. In contrast, participants working on their own riddles were more reluctant to make substantial edits and did not consistently believe their changes were helpful.

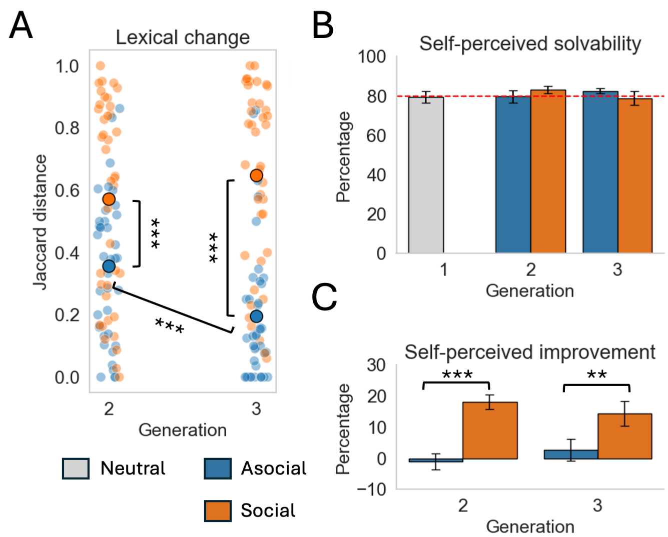

Figure 2: Results from Experiment 1. Transmission modes are consistently coded by colors: grey for neutral Generation 1, blue for the asocial mode, and orange for the social mode. All error bars in black indicate the respective standard error. Significance is indicated by stars (* $p < .05$; ** $p < .01$; *** $p < .001$). (A) Jaccard distance from parent riddles by mode. Mean values for each mode and generation are shown as bordered points on top. (B) Self-perceived solvability by mode, with the red dashed line marking the target solvability. (C) Self-perceived improvement by mode.

Experiment 2

Observing the quality differences between riddles generated from the two transmission modes, we next examined whether riddles from Experiment 1 also differed in their *actual* (i.e., objective) solvability. To address this question, we conducted a riddle-solving experiment with a separate participant pool using the riddles generated from Experiment 1. The experiment was pre-registered and received approval from the University of California, Los Angeles Office of the Human Research Protection Program, following relevant guidelines and regulations.

Method

Participants Additional participants were recruited from Prolific using the same eligibility criteria as in Experiment 1. Data from 250 participants were used for analysis after excluding those who failed attention checks. The experiment lasted approximately 7 minutes, with a compensation rate of approximately $25 per hour.

Design Each of the 200 riddles produced in Experiment 1 was solved by 10 independent solvers. The 5 conditions and 8 stimulus sets in Experiment 1 were counterbalanced across every group of 25 participants using a Latin square design. As a result, each participant solved 8 riddles, each from a unique stimulus.

Procedure Each participant completed 8 consecutive solving trials. In each trial, they viewed a riddle alongside its stimulus set and had 3 minutes to identify the poisonous mushroom by clicking on a mushroom among four, and to confirm their choice by clicking a confirmation button. A trial was coded as solved if they selected the right mushroom, or failed if they chose the wrong one or made no selection within the time limit. Participants' accuracy scores were not given as feedback and did not influence their payment.

Measures The solvability level of each riddle is measured by (1) its average solving duration (in minutes) across five independent solvers and (2) the proportion of successful solves: (a) the riddle's absolute deviation from the target solvability and (b) the *exact proportion*—whether it reached the exact 80% solvability. These success measures capture both continuous and discrete aspects of accuracy. Additionally, we compared the solvability between each riddle and its parent.

Results

All continuous outcomes were analyzed using mixed-effects models following the same specifications as in Experiment 1. Exact proportion was analyzed using a binomial Generalized Linear Model. All model results are displayed in Table 1.

Objective solvability First, riddles in both transmission modes showed a similarly wide range of solvability outcomes, with no reliable statistical differences in absolute deviation, objective improvement, or exact proportion across generations. Nevertheless, the descriptive patterns suggest that social riddles continued to drift toward the target solvability (see Fig.3A) with an increased number of exact 80% matches compared to riddles from the asocial mode (see Fig.3B). These observations call for further investigations.

In line with this qualitative trend, by Generation 3, social riddles took significantly less time to solve than asocial riddles. Correlations with efficiency measures further showed that, across both modes, riddles that were solved more quickly typically had shorter length ($r = 0.623$, $p < 0.001$) and higher information density ($r = -0.588$, $p < 0.001$). Together, these patterns indicate that despite undergoing more substantial revisions, social riddles remained concise yet informative, enabling solvers to reach decisions more efficiently without compromising performance. While we observed an initial trend toward enhancement, further investigation is warranted.

Subjective vs. objective solvability With objective solving performance in hand, we next examined how accurately crafters judged the solvability of their own and parent riddles. Overall, self-perceived solvability did not reliably predict objective solvability ($r = 0.111$, $p = 0.118$), whereas other-perceived solvability did ($r = 0.369$, $p < 0.001$). This suggests that participants were generally poor judges of their *own* riddles but more accurate when evaluating someone else's—a benefit of a "fresh judgment."

Additionally, although self-perceived solvability did not align with exact objective solvability, self-perceived *improvement* did correlate with objective *improvement* (overall: $r = 0.222$, $p = 0.005$), an effect driven primarily by the asocial mode (see Fig. 3C; asocial: $r = 0.240$, $p = 0.032$; social: $r = 0.203$, $p = 0.071$). These findings suggest that asocial crafters appeared to correctly recognize that they made minimal changes and showed some sensitivity to whether those changes were helpful, whereas social crafters tended to believe they improved the parent riddles more than they actually did.

Discussion

This study took an initial step into understanding intellectual innovation by examining its dynamics with social learning. We employed a novel riddle-crafting paradigm combined with a closed diffusion-chain design, implementing a social condition alongside an asocial baseline across three generations of riddle composition and refinement. Participants' innovation processes and outcomes were evaluated in terms of efficiency, novelty, and solving performance.

Overall, the social and asocial chains resulted in different patterns of innovation outcomes. Compared to asocial riddles, social riddles were more communicatively efficient: they were shorter, denser, took longer to craft, yet were easier for solvers to make a decision. They were also consistently more diverse and innovative across local drift, family-level diffusion, and population dispersion. Importantly, these advantages emerged without compromising performance: participants in both modes believed they performed well and achieved comparable levels of objective solvability.

These findings suggest that social transmission may support a distinct form of intellectual innovation—one that adapts more readily to communicative pressures, remains open to mutation, and preserves accumulated progress. A speculative theme that emerges from our results is that social learners are more willing to make changes: they preserve what works, but they also inject new mutations whenever they see room for improvement. Subsequent learners can then correct or

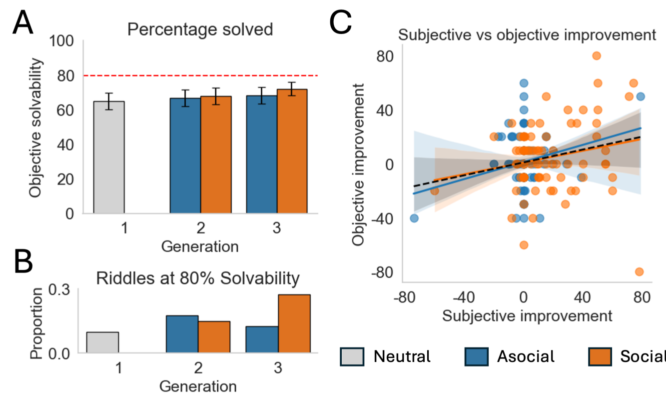

Figure 3: Results from Experiment 2 combined with Experiment 1. Transmission modes are consistently coded by colors: grey for the neutral Generation 1, blue for the asocial mode, and orange for the social mode. (A) Objective solvability by mode, with the red dashed line marking the target solvability. Error bars in black indicate the standard error. (B) Proportion of riddles that reached exactly 80% solve by mode. (C) Correlations between self-predicted and objective improvement by mode. Linear regressions (solid and dashed lines) with 95% confidence intervals (shaded) are shown for asocial (blue), social (orange), and combined data (black).

refine these changes. If this pattern holds, such willingness to alter and refine designs across individuals may diversify the cultural repertoire and gradually ratchet innovations forward, enriching the collective knowledge base over time.

Moreover, the two transmission modes exhibited distinct innovation trajectories. When innovating alone, individuals tended to become anchored to their previous designs, quickly settling into a local optimum and believing that no further improvement was possible or needed. In contrast, social transmission provided a fresh and moderately accurate perspective on prior outcomes, continuously motivating participants to revise and enhance their designs, resulting in more efficient and diverse innovations. Although objective improvements have so far been slow and variable, the shift from motivation to measurable gains is likely incremental and requires patience. Importantly, social participants showed a sustained impetus to improve and had not yet reached a plateau—suggesting that, given more generations, this momentum may eventually yield qualitative advances. This possibility motivates future work extending social chains across additional generations to determine when, and whether, such improvements emerge.

Some nuances in our findings also highlight the limits of self-evaluation and the illusion of self-improvement: participants in both modes never had access to external judgments, which likely contributed to the slow rate of improvement. Future work should examine whether introducing real-time external feedback can accelerate progress, and whether such feedback provides comparable or differential benefits across the two transmission modes.

Acknowledgments

Authors' contributions Conceptualization: S.G., T.G., E.K.; Methodology: S.G., T.G., E.K., K.J.; Formal Analysis: S.G., K.J.; Data Collection: S.G.; Writing: S.G., T.G., E.K., K.J.; Editing: S.G., T.G., E.K., K.J..

Declaration of AI use Portions of the writing were refined with the assistance of AI-based editing tools (ChatGPT) to improve clarity. AI tools were also used for minor coding support. All scientific ideas, analyses, and interpretations originate from the authors.

References

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using r*. Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items [Special Issue: Emerging Data Analysis]. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bartlett, F. C. (1995[1932]). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baum, W. M., Richerson, P. J., Efferson, C. M., & Paciotti, B. M. (2004). Cultural evolution in laboratory microsocieties including traditions of rule giving and rule following. *Evolution and Human Behavior*, *25*(5), 305–326.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918–10925.

Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What is innovation, how does it develop, and who does it? *Child Development*, *87*(5), 1505–1519.

Derex, M., & Boyd, R. (2015). The foundations of the human cultural niche. *Nature Communications*, *6*(1), 8398.

Derex, M., Feron, R., Godelle, B., & Raymond, M. (2015). Social learning and the replication process: An experimental investigation. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1808), 20150719.

Derex, M., Godelle, B., & Raymond, M. (2013). Social learners require process information to outperform individual learners. *Evolution*, *67*(3), 688–697.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, volume 1 (long and short papers)*, 4171–4186.

Fogarty, L., Creanza, N., & Feldman, M. W. (2015). Cultural evolutionary perspectives on creativity and human innovation. *Trends in ecology & evolution*, *30*(12), 736–754.

Henrich, J. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, *37*, 547–579.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *The Journal of Abnormal and Social Psychology*, *62*(3), 649.

Kendal, R., Coe, R., & Laland, K. N. (2005). Age differences in neophilia, exploration, and innovation in family groups of callitrichid monkeys. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *66*(2), 167–188.

Laland, K. N. (1992). A theoretical investigation of the role of social transmission in evolution. *Ethology and Sociobiology*, *13*(2), 87–113.

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences*, *19*(11), 688–699.

Lenth, R. V. (2025). *Emmeans: Estimated marginal means, aka least-squares means* [R package version 1.11.2-00002]. https://rvlenth.github.io/emmeans/

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). *Mining of massive data sets*. Cambridge University Press.

Lucas, A. J., Kings, M., Whittle, D., Davey, E., Happé, F., Caldwell, C. A., & Thornton, A. (2020). The value of teaching increases with tool complexity in cumulative cultural evolution. *Proceedings of the Royal Society B*, *287*(1939), 20201885.

McGuigan, N., Burdett, E., Burgess, V., Dean, L., Lucas, A., Vale, G., & Whiten, A. (2017). Innovation and social transmission in experimental micro-societies: Exploring the scope of cumulative culture in young children. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1735), 20160425.

Mesoudi, A. (2008). An experimental simulation of the "copy-successful-individuals" cultural learning strategy: Adaptive landscapes, producer–scrounger dynamics, and informational access costs. *Evolution and Human Behavior*, *29*(5), 350–363.

Mesoudi, A., & O'Brien, M. J. (2008). The cultural transmission of great basin projectile-point technology i: An experimental simulation. *American Antiquity*, *73*(1), 3–28.

Mesoudi, A., & Thornton, A. (2018). What is cumulative cultural evolution? *Proceedings of the Royal Society B*, *285*(1880), 20180712.

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

Osiurak, F. (2014). What neuropsychology tells us about human tool use? the four constraints theory (4ct): Mechanics, space, time, and effort. *Neuropsychology Review*, *24*, 88–115.

Osiurak, F., & Badets, A. (2016). Tool use and affordance: Manipulation-based versus reasoning-based approaches. *Psychological Review*, *123*(5), 534.

Osiurak, F., & Badets, A. (2017). Use of tools and misuse of embodied cognition: Reply to buxbaum (2017). *Psychological Review*, *124*(3), 361–368.

Osiurak, F., & Reynaud, E. (2020). The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*, *43*, e156.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(supplement_2), 8993–8999.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1528), 2405–2415.

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, *16*(3), 495–511.

Wang, W., Wei, F., Dong, L., Bao, H., Yang, N., & Zhou, M. (2020). Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems*, *33*, 5776–5788.

Wasielewski, H. (2014). Imitation is necessary for cumulative cultural evolution in an unfamiliar, opaque task. *Human Nature*, *25*(1), 161–179.

Yahosseini, K. S., & Moussaïd, M. (2020). Comparing groups of independent solvers and transmission chains as methods for collective problem-solving. *Scientific Reports*, *10*(1), 3060.

Zhu, Y., Lu, S., Zheng, L., Guo, J., Zhang, W., Wang, J., & Yu, Y. (2018). Texygen: A benchmarking platform for text generation models. *The 41st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1097–1100.

Zwirner, E., & Thornton, A. (2015). Cognitive requirements of cumulative culture: Teaching is useful but not essential. *Scientific Reports*, *5*(1), 16781.