Main Manuscript for

The roles of cultural transmission and causal reasoning in the cultural evolution of technology

Ali Seyhun Saral[1]§; Paula Ibáñez de Aldecoa[1]; Maxime Derex[1,2]§*
[1] Institute for Advanced Study in Toulouse, Toulouse, France
[2] Toulouse School of Economics, CNRS, France.
§ These authors contributed equally
***Email:** maxime.derex@iast.fr

Abstract

Humans are uniquely capable of producing highly efficient tools, but the extent to which this capacity depends on individual reasoning abilities remains unclear. In particular, the respective roles of causal and technical reasoning versus cultural transmission in driving technological improvement are the subject of long-standing debate. We address this question by directly manipulating causal and technical reasoning in transmission chains, where participants sequentially inherit and refine prior solutions. Across two transmission chain experiments, participants (n = 900) completed tasks under three conditions: (1) technical reasoning, involving physical tasks and intuitive physics; (2) causal reasoning, where similar causal structures could be exploited without physical context; and (3) pure cultural transmission, in which causal structures were removed. We found that cumulative improvement occurred across generations even in the absence of causal structure, demonstrating that cultural transmission alone can drive technological improvement. While causal reasoning accelerated early improvement by helping participants focus on promising regions of the design space, its impact diminished over time. Notably, technical reasoning offered no added benefit over causal reasoning. These results highlight the beneficial yet dispensable role of causal reasoning in the improvement of culturally evolving technology, and challenge the view that improvements guided by causal models depend on domain-specific rather than domain-general reasoning abilities.

Significance Statement

Human adaptation to diverse habitats has been enabled by complex technologies, raising key questions about the role of individual reasoning abilities in their development. This study introduces a novel methodology for rigorously evaluating how causal and technical reasoning, compared to cultural transmission, contribute to technological improvement. Across two experiments, we show that cultural transmission alone can lead to improvement, even in the absence of causal reasoning. Causal reasoning accelerates early improvement, although its influence diminishes over time. Technical reasoning (i.e., reasoning grounded in intuitive physics) offers no advantage compared to causal reasoning. By disentangling the roles of reasoning and cultural transmission in technological evolution, these findings offer new insight into the cognitive and cultural processes that have shaped the evolution of human technology.

Main Text

Introduction

Causal reasoning is often seen as a defining feature of human cognition [1-3]. Humans effortlessly think in terms of causal relationships and can easily identify "difference-making" variables, which are variables that influence other variables [3]. Identifying those variables allows humans to design interventions and adjust them to achieve desired outcomes. For example, when improving a basic hunting tool like a spear, one may identify the sharpness of the point as a key "difference-making" variable and could decide to make it sharper to increase the spear's perforating capability.

The ability to exploit "difference-making" variables is sometimes presented as the key factor underlying humans' capacity to design efficient tools [4-6]. For instance, causal reasoning has been argued to support tool invention and iterative improvement through feedback during an individual's lifetime [4, 6]. Related accounts have emphasized technical reasoning, defined as a specific form of causal reasoning directed at understanding the physical world, as central to technological improvement [5].

Yet, the effectiveness of causal reasoning has been challenged in the context of complex, high-dimensional technologies commonly found in human populations [7-9]. In such systems, even small changes to one feature can produce unforeseen consequences in others. For example, adjusting a bowstring to increase arrow speed might inadvertently increase vibration, resulting in a worse overall outcome. This interdependence between features creates what some researchers refer to as the "complexity barrier" to artifact improvement [10] and often demands computational resources beyond human cognitive capacity [7, 11].

These limitations suggest that alternative or complementary mechanisms may be required for the emergence of highly optimized tools. One proposed mechanism is payoff-biased social learning, in which individuals adopt solutions that have previously led to successful outcomes in others [12-14]. When such a selective process operates across generations, it may yield highly efficient solutions even when individuals lack an understanding of why those solutions work [7, 15] see also [16]. In support of this, experimental work has shown that the gradual improvement of a technology across generations is not necessarily accompanied by increased understanding [17]. Similarly, interviews with Hadza bowyers suggest that their understanding of how relevant variables affect bow efficiency is limited [18].

Nonetheless, neither line of evidence fully dismisses the role of causal reasoning in technological improvement [19]. For instance, participants in experimental studies tend to direct their exploration toward specific regions of the parameter space, suggesting that causal reasoning may help guide exploration toward promising regions [17, 19]. Moreover, other experiments found that improvements can coincide with increased understanding [20].

Critically, because existing studies have focused on whether improvement across generations correlates with increased understanding [17, 20], they have not allowed direct assessment of the actual contribution of causal reasoning. As a result, a major gap remains in understanding the role of causal reasoning in technological evolution, with significant implications for inferring the cognitive and cultural processes that have shaped the evolution of human technology.

Here, we present a novel methodology that allowed us to experimentally manipulate participants' ability to rely on either causal or technical reasoning while solving a task within transmission chains. This approach enabled us to test whether cultural transmission on its own can lead to improvement across generations and to assess the extent to which both forms of reasoning further enhance this process.

Our methodology builds on a physical task used in prior research [17] and uses the same four adjustable parameters across three experimental conditions (Figure 1) , varying only the information available to participants. These conditions were implemented in transmission chains where participants sequentially inherit and refine prior solutions [21, 22].

In the Technical Reasoning condition, the input variables represented the four weights of the wheel task from [17]. Participants adjusted weights along the radial spokes of a digital wheel to select a configuration, which determined two physical properties: the wheel's moment of inertia and the position of its centre of mass. These properties determined the causal structure of the task, which participants could learn and exploit by observing covariation between inputs (i.e., weight positions) and output (i.e., their payoff, as determined by the time it takes the wheel to travel down the track). In this treatment, participants could observe the wheel's motion, potentially allowing them to draw on intuitive physics to refine their causal model of how weight positions govern its mechanical behaviour (Figure 1.a).

In the Causal Reasoning condition, participants were informed that the task involved adjusting four sliders, with no physical context provided, rendering the task fully abstract (Figure 1.b). Participants could not observe the wheel's motion and therefore could not draw on intuitive physics to refine their causal model of the task. However, the task retained the same underlying payoff structure as the wheel task, enabling them to exploit identical covariation patterns between input variables (i.e., slider positions) and output variables (i.e., payoffs).

Finally, in the Pure Cultural Transmission condition, the task was fully abstract and lacked any underlying causal structure. In this treatment, slider configurations were assigned fixed payoffs randomly drawn from the original payoff distribution, breaking the mapping between inputs and outputs and preventing participants from exploiting causal variables (Figure 1.c). As a result, any improvement observed across generations could only arise from the combined effects of random exploration and cultural transmission.

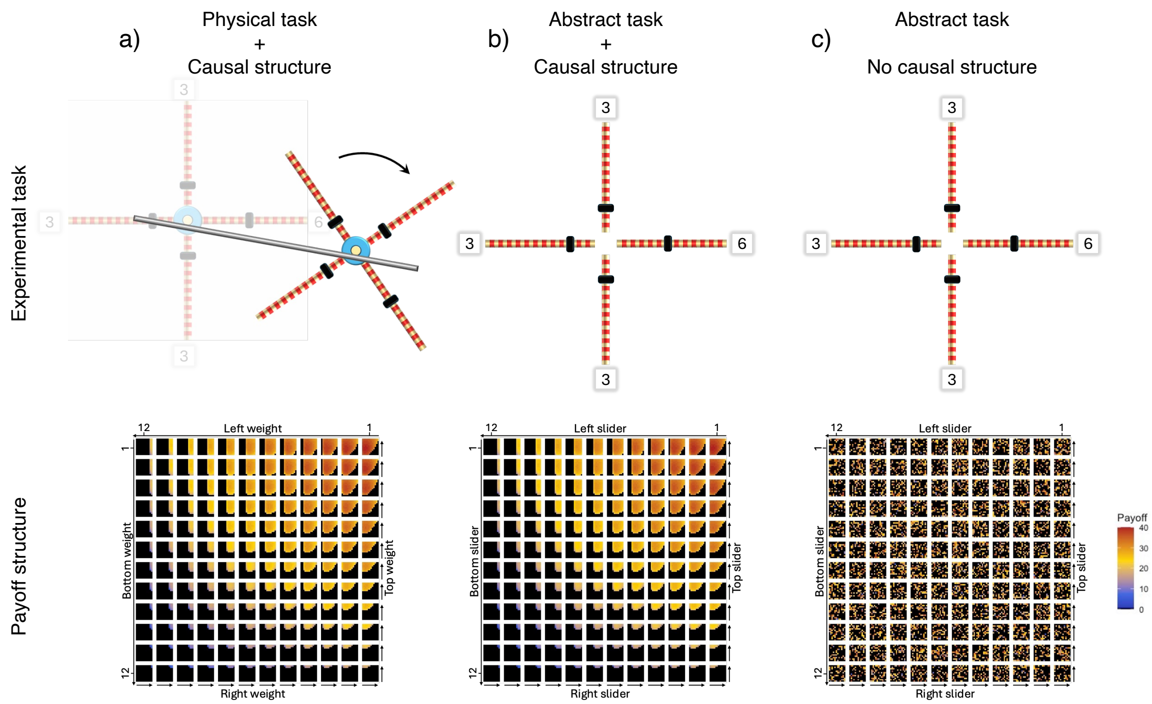

Figure 1: *Experimental conditions. a) Technical reasoning: Participants are informed that the task involves a wheel rolling down a track. After selecting the positions of the four weights, participants observe the wheel's motion on an inclined track and then receive a payoff, determined by the time it takes the wheel to travel down the track. The task features a stable mapping between weight positions and payoff, mediated by physical properties (moment of inertia and centre of mass), allowing participants to rely on intuitive physics. b) Causal reasoning: Participants are informed that the task involves adjusting four sliders, with no physical context provided. After selecting the positions of the four sliders, participants receive a payoff. Although participants cannot rely on intuitive physics, the task retains the same underlying payoff structure as the wheel task, enabling them to exploit covariation patterns between slider positions and payoffs. c) Pure Cultural Transmission: Participants are informed that the task involves adjusting four sliders. After selecting the positions of the four sliders, participants receive a payoff. Here, the causal structure is removed: slider configurations are assigned fixed payoffs randomly drawn from the original payoff distribution, breaking the mapping between inputs and outputs and preventing participants from exploiting causal variables. In the bottom plots, payoff is shown as a function of the positions of the top, right, bottom, and left weights/sliders (with 1 corresponding to the innermost position). The plot is organized as a grid of squares, where each square represents a unique combination of left and bottom slider values: the left slider varies across columns of the outer grid, and the bottom slider varies across its rows. Variation in the right and top sliders is shown within each square: the right slider varies across columns, and the top slider across rows. Black cells indicate configurations with a payoff of zero, corresponding to wheel configurations that fail to descend the rails upon release.*

The experiment was pre-registered and structured as follows: All participants (n = 450) were randomly assigned to one of three experimental conditions and placed into transmission chains consisting of five individuals (30 chains per condition). Each participant completed five trials with the goal of maximizing their own cumulative payoff and the payoff of the next participant in the chain. Before each trial, all participants, except those in the first generation, could consult the two most recent configurations produced within their chain along with their associated payoffs.

In line with the view that cultural transmission can drive technological improvement without causal reasoning, we predicted that cumulative improvement would occur across generations in all conditions, including the Pure Cultural Transmission condition. However, we expected the pace of improvement to be faster in the Causal Reasoning condition. We did not make specific predictions regarding the relative pace of improvement in the Technical and Causal Reasoning conditions.

Results

The average payoff increased across generations in all conditions (Pure Cultural Transmission: Generation 95% CI [0.60, 1.98], mean = 1.29; Causal Reasoning: Generation 95% CI [0.85, 2.36], mean = 1.61; Technical Reasoning: Generation 95% CI [0.50, 1.95], mean = 1.23). Among participants who completed an abstract version of the task, performance was higher when the causal structure was present (Causal Structure 95% CI [6.93, 14.87], mean = 10.95; Causal Structure x Generation 95% CI [−0.69, 1.41], mean = 0.37). Performance did not differ between participants presented with the physical version of the task and those presented with the abstract version when the causal structure was present (Physical Task 95% CI [-1.73, 3.85], mean = 1.05; Physical Task x Generation 95% CI [-1.19, 0.44], mean = -0.39).

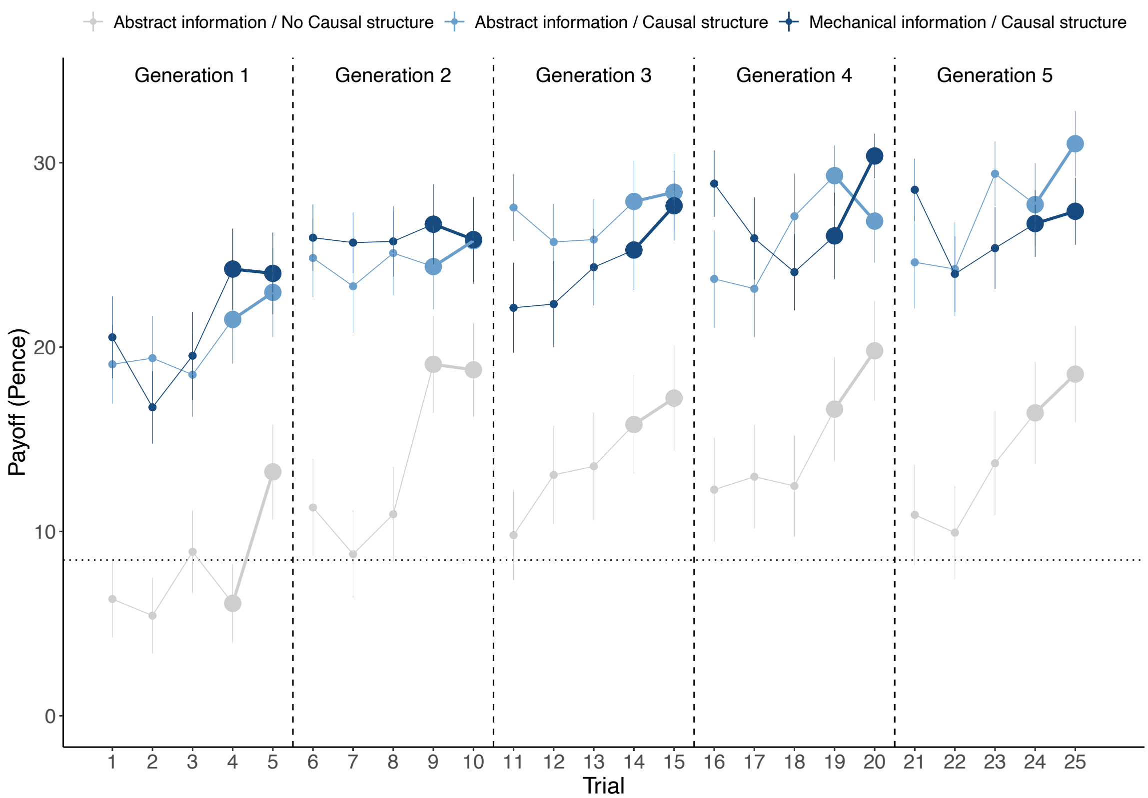

Figure 2. *In Experiment 1, payoff increased across generations in all conditions. Vertical dashed lines separate successive generations, with large dots representing culturally transmitted trials. The horizontal dotted line marks the mean of the payoff distribution. Error bars represent the standard error of the mean.*

Although these results align with our pre-registered prediction that participants in the Causal Reasoning condition would outperform those in the Pure Cultural Transmission condition, the observed statistical differences do not fully match our expectations. Specifically, we expected the Causal Structure × Generation interaction to yield a reliably positive effect, rather than a main effect of Causal Structure alone. This prediction was based on the assumption that participants facing any version of the abstract task would initially be naive about what might constitute a better-than-random configuration. However, our results show that first-generation participants in both the Technical and Causal Reasoning conditions produce configurations with payoffs more than twice the average of the distribution (Figure 2).

Exploratory analyses revealed that this difference stems from first-generation participants' tendency to produce balanced configurations (i.e., configurations with weights/sliders positioned at equal distances from the axis) during their early trials across all conditions. Balanced configurations are associated with above-average payoffs in both the Technical and Causal Reasoning conditions because balanced configurations always roll down the inclined rails, guaranteeing a non-zero payoff with a probability of 1 and an average payoff of 22.6. In contrast, unbalanced configurations do not consistently roll down, yielding a non-zero payoff with a lower probability of 0.32 and a lower average payoff of 8.44. This contrasts with the Pure Cultural Transmission condition, where balanced and unbalanced configurations yield similar average payoffs, as all configurations are randomly assigned a payoff drawn from the wheel's payoff distribution.

This leaves open the possibility that our results underestimate the role of technical reasoning. For instance, participants in the Technical Reasoning condition may have been more likely to produce balanced configurations due to useful priors grounded in intuitive physics, while those in the Causal Reasoning condition may have benefited from an irrelevant aesthetic or cognitive preference for balanced patterns. Additionally, participants in the Causal Reasoning condition may have outperformed those in the Pure Cultural Transmission condition due to an artifactual head start, rather than an effective exploitation of the task's underlying causal structure.

To address these possibilities, we conducted Experiment 2, which used a modified version of the task calibrated so that balanced and unbalanced configurations had similar average payoffs, with the greatest potential gain coming from producing unbalanced configurations (i.e., configurations with weights/sliders positioned at unequal distances from the axis). To achieve this, we recruited 450 additional participants and provided them with a wheel featuring unequal weights (Supplementary Figure 1) and an inverted goal: to produce wheels that moved as slowly as possible while still descending the rails. Figure 3 illustrates how these changes effectively shifted the most rewarding solutions to another area of the parameter space while maintaining a causal structure comparable to that of Experiment 1.

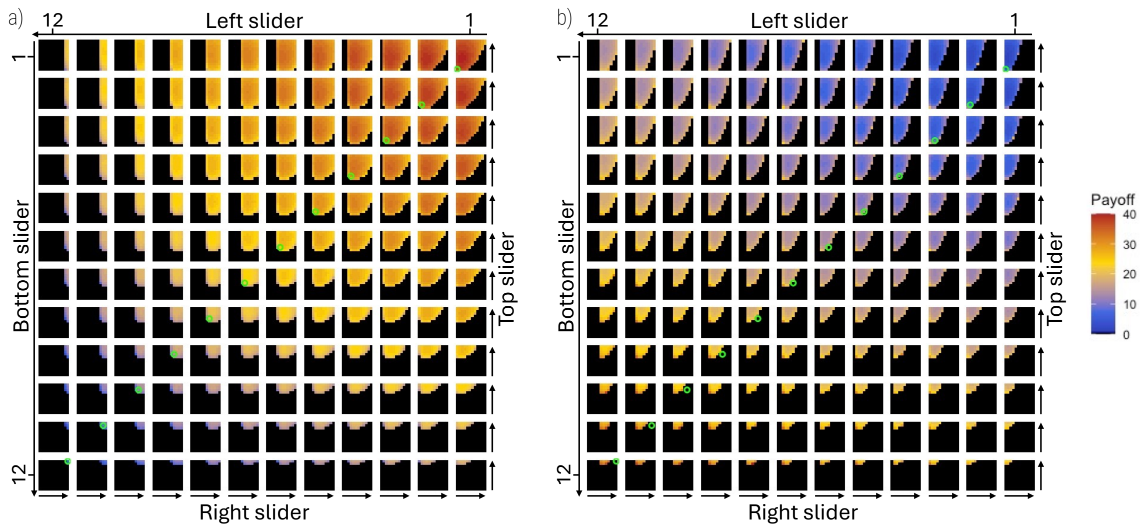

Figure 3: *Comparison of the payoff structure in Experiments 1 and 2. a) Experiment 1: The mapping between weight positions and payoff is mediated by the physical properties of a wheel featuring weights of equal mass. Since the goal is to minimize the time it takes the wheel to complete the track, the task rewards configurations with a low moment of inertia (i.e., weights positioned closer to the axis) and a centre of mass located in the upper right side of the wheel (toward the slope). Because the weights are equal, all balanced configurations (i.e., configurations with weights/sliders positioned at equal distances from the axis) roll down the rails and yield a positive payoff (lime green circles), ranging from low to high. The best configuration in this experimental treatment was: Top = 5; Right = 3; Bottom = 1; Left = 1. b) Experiment 2: The wheel features weights of unequal mass, with the right-side weight being heavier. Since the goal is to maximise the time it takes the wheel to complete the track, the task rewards configurations with a high moment of inertia (i.e., weights positioned far from the axis) and a centre of mass located on the left side of the wheel (away from the slope). Because the weights are unequal, balanced configurations do not consistently roll down the rails and may yield either zero payoff or a low positive payoff (lime green circles). The best-performing configuration in this treatment was: Top = 11; Right = 6; Bottom = 12; Left = 8. Note how in Experiment 2 the most rewarding solutions (orange/red cells) tend to cluster in riskier regions of the parameter space, where non-viable solutions (black cells) are more common.*

Results from Experiment 2 indicate that the implemented changes successfully eliminated the early above-average payoffs observed in the Causal Reasoning condition of Experiment 1. Crucially, the qualitative pattern of results remained consistent across experiments, with average payoff increasing across generations in all conditions (Pure Cultural Transmission: Generation 95% CI [0.51, 1.43], mean = 0.97; Causal Reasoning: Generation 95% CI [0.42, 1.63], mean = 1.01; Technical Reasoning: Generation 95% CI [0.13, 1.39], mean = 0.74). Among participants who faced an abstract version of the task, performance was higher when the causal structure was present rather than absent (Causal Structure 95% CI [1.83, 5.95], mean = 3.86; Causal Structure x Generation 95% CI [−0.57, 0.66], mean = 0.05). Performance did not differ between participants presented with the physical version of the task and those presented with the abstract version when the causal structure was present (Physical Task 95% CI [-1.79, 2.60], mean = 0.43; Physical Task x Generation 95% CI [-0.90, 0.37], mean = -0.26).

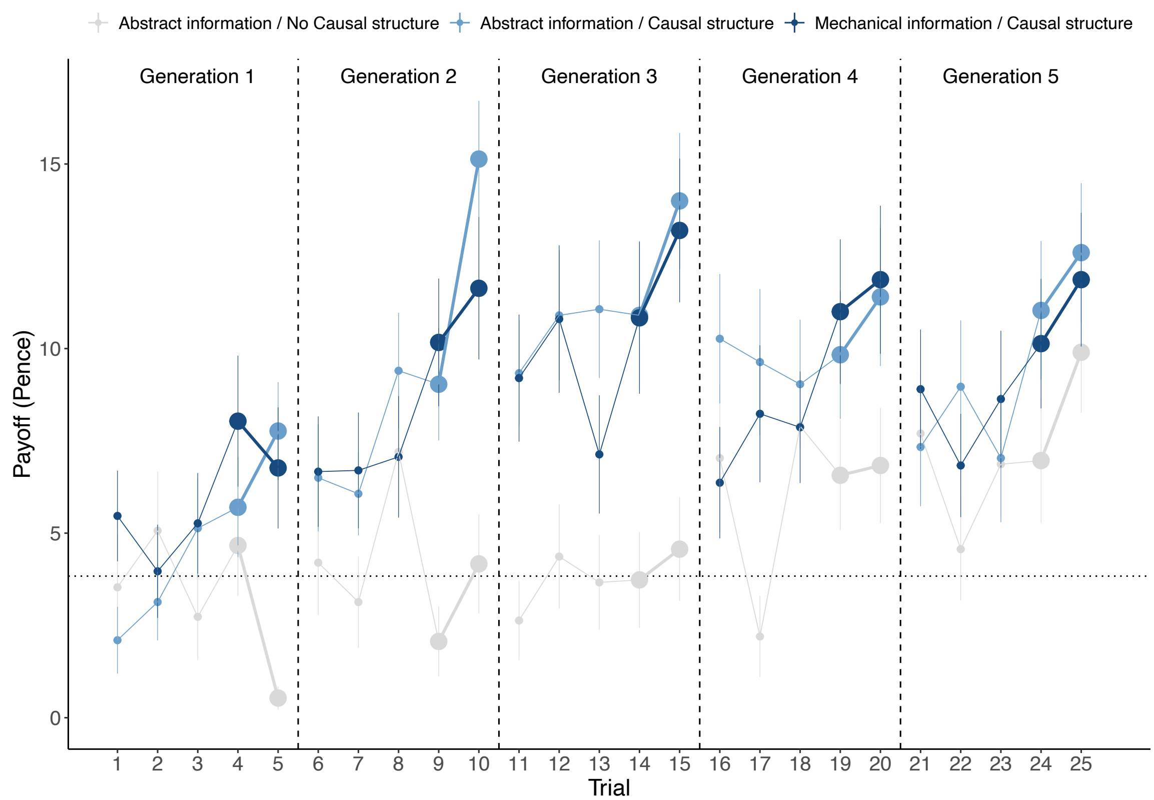

Figure 4. *In Experiment 2, payoff increased across generations in all conditions. Vertical dashed lines indicate successive generations, with large dots representing culturally transmitted trials. The horizontal dotted line marks the mean of the payoff distribution. Error bars represent the standard error of the mean.*

As shown in Figure 4, the higher performance of participants who faced a version of the task with an underlying causal structure appears to be driven primarily by the first three generations of learners, after which performance begins to plateau. Exploratory analyses indicate that this effect stems from a feature of the payoff structure of Experiment 2 where moving toward more rewarding solutions in the Technical and Causal Reasoning conditions brings learners into riskier regions of the parameter space where non-viable solutions are more common (Figure 3.b). This contrasts with the payoff structure of Experiment 1, where moving toward more rewarding solutions brings learners into safer regions of the solution space (Figure 3.a). This means that in the two conditions where the causal structure was present (i.e., the Technical and Causal Reasoning conditions), cumulative improvement progressively became more challenging in Experiment 2 compared to Experiment 1 (Supplementary Figures 2 and 3).

While patterns of cumulative improvement did not differ between the Technical and Causal Reasoning conditions in either experiment (Figures 2 and 4), our results indicate that exposure to the physical version of the task influenced how naïve learners explored the parameter space. In Experiment 1, first-generation participants frequently produced balanced configurations across all conditions (Figure 5.a). However, while balanced configurations dominated in both the Causal Reasoning and Pure Cultural Transmission conditions, participants in the Technical Reasoning condition generated more unbalanced configurations with their centroid (i.e., their centre of mass) in the top-right quadrant. A similar pattern emerges in Experiment 2: balanced configurations were again the most frequent in both abstract conditions, whereas participants in the Technical Reasoning condition tended to produce unbalanced configurations with their centre of mass in the top-left quadrant (Figure 5.b). These findings indicate that awareness of the task's mechanical nature influenced participants, biasing initial exploration toward more accelerating wheels in Experiment 1 and more decelerating wheels in Experiment 2.

Two findings help explain why the biased exploration observed in the Technical Reasoning condition did not result in faster cumulative improvement compared to the Causal Reasoning condition. First, contrary to our predictions, facing the physical task did not reduce the extent of exploration compared to the abstract task. In fact, exploration was broader in the Technical Reasoning condition in both experiments, even when considering physical properties directly relevant to the wheel's dynamics, such as the centre of mass and inertia (Supplementary Figures 4 and 5). This indicates that while exploration in the Technical Reasoning condition is influenced by physical cues, participants still vary their configurations substantially, reflecting the vagueness and partial nature of their underlying causal model. Second, our results show that participants do not need physical cues to effectively leverage the task's causal structure. Even when presented with the task in its abstract form, participants were able to converge on promising regions of the parameter space: unbalanced configurations with their centroid in the top-right quadrant in Experiment 1 and in the top-left quadrant in Experiment 2 (Figure 5c-d).

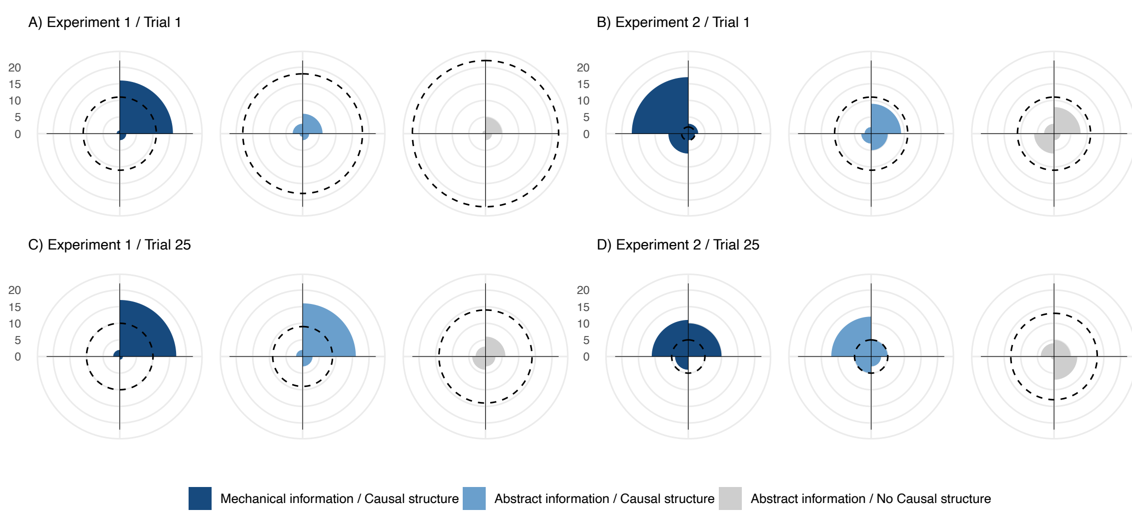

Figure 5. *Initial versus final exploration patterns across conditions in Experiments 1 and 2. Radial plots show the number of configurations whose centroid falls within each quadrant of the wheel (upper-right, bottom-right, bottom-left, and upper-left). The dotted circle indicates the number of configurations with weights/sliders positioned at equal distances from the axis (i.e. balanced configurations). Top row: Panels A and B depict the first trial of first-generation participants (i.e., Trial 1) in Experiments 1 and 2, respectively. Participants in the Technical Reasoning condition appropriately biased their initial exploration toward unbalanced configurations, favoring accelerating wheels in Experiment 1 (with a centre of mass in the upper right quadrant due to their*

top and right weights being positioned farther from the axis than their bottom and left weights) and decelerating wheels in Experiment 2 (with a centre of mass in the upper left quadrant due to their top and left weights being positioned farther from the axis than their bottom and right weights). In contrast, balanced configurations dominated the early exploration in both the Causal Reasoning and Pure Cultural Transmission conditions. Bottom row: Panels C and D depict the last trial of fifth-generation participants (i.e., Trial 25) in Experiments 1 and 2, respectively. By the end of the experiment, participants in the Causal Reasoning condition had converged toward exploration patterns similar to those observed in the Technical Reasoning condition.

Discussion

Our study set out to examine the role of reasoning abilities in technological improvement. Unlike previous experiments that tested whether performance gains across generations were correlated with increased understanding [17, 20], we introduced a novel methodology that allowed us to experimentally manipulate participants' ability to rely on either causal or technical reasoning while solving a task, enabling a rigorous evaluation of their contribution to technological improvement.

In our baseline, Pure Cultural Transmission condition, the causal structure was absent, preventing participants from correctly predicting which configurations may yield above-average performance. As a result, between-generation improvement could only occur through the retention of modifications that happened to outperform inherited solutions by chance. Despite this, improvement across generations was observed in both experiments, confirming that causal reasoning is not necessary for cultural improvement to occur [7, 17, 18].

This Pure Cultural Transmission condition provides a baseline against which to compare the pace of cultural improvement when participants can rely on causal reasoning to exploit patterns of covariation between input and output variables. Our results reveal that the presence of a causal structure enables participants to effectively detect performance gradients and preferentially explore promising regions of the parameter space. This is reflected in end-of-experiment exploration patterns, which were biased toward upper-right-offset configurations in Experiment 1 and upper-left-offset configurations in Experiment 2. These results demonstrate that causal reasoning can promote the emergence of beneficial modifications, thereby accelerating the pace of cultural evolution [5, 23].

Yet, our results reveal that the benefits of causal reasoning vary significantly depending on the circumstances. This is particularly evident in Experiment 2, where the presence of a causal structure is highly beneficial during the early stages of cultural evolution but less so in later stages. This contrasts with Experiment 1, where the benefits of a causal structure are maintained during the later stages of the experiment. This difference stems from the distinct payoff structures used in Experiments 1 and 2. In both experiments, causal reasoning helped participants bias their exploration toward regions of the parameter space where the highest-rewarding solutions lie. In Experiment 1, these regions also happen to be safe, with non-viable configurations being uncommon (Figure 3.a). In contrast, in Experiment 2, an exploration bias toward high-reward areas inadvertently led participants into riskier regions of the parameter space, where non-viable configurations are more frequent (Figure 3.b). This difference in payoff structure between Experiments 1 and 2 reveals the limitations of causal reasoning in driving cultural improvement. When learners are initially uncertain about what constitutes a good solution, even partial causal models can be helpful because they allow individuals to exploit the most salient performance gradients and accelerate the discovery of beneficial modifications. However, as solutions improve and the room for further improvements narrows, these partial models become less useful.

These findings align with the view that individuals encode causal patterns in a highly sparse manner [24], such that their causal models lack the detailed mechanistic structure needed to support fine-grained predictions about the effects of specific changes [25]. Taken together, our results help reconcile opposing views about the role of causal reasoning in cultural evolution. On one hand, they demonstrate the contribution of causal reasoning to technological improvement; on the other, they underscore its limitations [17, 26]. Most notably, our experiments reveal how causal reasoning becomes relatively ineffective as learners approach the 'barrier' to artifact improvement [10] (see also [27]). Once this barrier is reached, attempts to improve would often lead to worse outcomes, as shown by the early performance of late-generation participants, who frequently produced solutions less efficient than those they had inherited (Figure 4, see also [17]). This means that further improvement would largely depend on cultural evolutionary processes in which chance modifications are selectively retained [15], a process whose effectiveness depends heavily on factors such as group size and connectedness [15, 28, 29]. This helps explain why between-generation improvement was sometimes modest or absent in our experiments, as typical one-person-per-generation transmission chains provide limited opportunities for cultural evolutionary processes to operate. In our setup, participants inherited only two configurations from a single cultural demonstrator, and the continuation of the cumulative process depended entirely on one individual per generation. In more realistic conditions, multiple learners help buffer against the risk of cultural loss, and the presence of more individuals increases the likelihood that beneficial modifications will arise [13, 15, 30-32].

A striking finding of our study is the lack of benefits from being aware of the task's mechanical nature and observing the wheel’s motion compared to the abstract condition. This indicates that the causal models formed in the Technical Reasoning condition were not more fine-grained and remained similarly limited in their ability to predict the payoff of untested configurations. The fact that participants did not require physical cues to leverage the task’s causal structure raises the question of whether causal models about physical phenomena depend on a content-specific form of reasoning or instead rely on more domain-general reasoning abilities [33, 34].

Yet, it is noteworthy that awareness of the task's mechanical nature influenced exploration among naïve participants, even though they had not previously interacted with our experimental apparatus. Participants biased their initial exploration in meaningful ways, likely drawing on prior experience with analogous systems and transferring relevant, albeit imprecise, causal models to a novel task. This highlights the potential role of analogical reasoning in cultural evolution [35]. The benefits of this form of reasoning may have been limited in our setup because the task involved optimizing an existing solution (i.e., improving toward a fixed optimum). While this is a critical component of cumulative cultural evolution, it may not fully capture the challenges associated with open-ended cumulative cultural evolution, through which increasingly complex and ever more efficient solutions can emerge [36-40]. For instance, open-ended evolutionary dynamics may depend on specific types of modifications not captured by our experimental task, such as cultural exaptation (in which a solution originally developed for one purpose is repurposed in a new context [8, 37]) and combinatorial innovation (where elements from different domains are combined to generate novel and more effective solutions [27, 41, 42]). Future work should extend our methodology to tasks that not only involve improving existing technologies but also creating new ones, an activity that may demand reasoning abilities beyond those required for marginal improvement alone [36, 38].

Materials and Methods

Ethics Statement. All methods were approved by the Toulouse School of Economics Review Board for Ethical Standards in Research (11-23-16). All participants provided informed consent before taking part in the experiment.

Participants. Two experiments were conducted with a total of 900 participants (450 women), all recruited via the online platform Prolific. The subjects ranged in age from 18–51 years (mean = 28.7 years; s.d. = 4.8 years).

Experimental conditions. All participants were randomly assigned to one of three experimental conditions. In all conditions, participants completed an optimization task involving four weights/sliders that could be positioned at one of 12 discrete locations, creating a parameter space of 20,736 unique configurations. Each participant completed five trials with the goal of maximizing their cumulative payoff and received feedback after each trial. In the Technical Reasoning condition, participants completed a computerized version of the wheel task used in Derex et al. (2021), which accurately simulated the physics of the original task (Supplementary Figures 6 and 7). In each trial, participants set the position of four weights, which determined the wheel's motion upon release (see 'Task versions' below for details about the physics of the wheel). After validating their configuration, they observed the wheel's motion and received feedback in the form of a payoff once the wheel reached the end of the rails (or once the simulation stopped for wheels that failed to descend). In the Causal Reasoning and Pure Cultural Transmission conditions, participants set the position of four sliders arranged like the spokes of the wheel (Figure 1). After validating their configuration, they received feedback in the form of a payoff. In the Causal Reasoning condition, the configuration-payoff mapping was identical to that in the Technical Reasoning condition, meaning that the same configuration produced the same payoff in both conditions. In the Pure Cultural Transmission condition, the configuration-payoff mapping was randomized, so that each configuration was mapped to a payoff outcome from the Technical Reasoning condition. These mappings were randomized between chains, but remained consistent within a given chain. Screenshots showing what participants saw in each condition are publicly available at https://www.github.com/aseyq/causal-reasoning.

Experimental design. In each condition, participants were placed into one of 30 independent transmission chains, each consisting of five individuals. No statistical methods were used to pre-determine sample sizes; however, the number of independent chains in our study was more than twice that reported in previous publications [17, 20]. Each participant completed five trials and could, before each trial, consult the two most recent configurations produced within their chain along with their associated payoffs. . After three trials, participants were reminded that their last two configurations would be transmitted to the next participant in the chain. Participants earned between 0 and 0.4 pounds per trial, depending on the performance of their configuration. All participants, except those in the first generation, received social information in the form of the last two configurations and associated payoffs of the preceding participant in their chain.

Task versions. In Experiment 1, the wheel's weights were of equal visual dimensions, indicating equal mass, as in the original physical task [17]. The goal for participants in the Technical Reasoning condition was to minimize the time it took for the wheel to descend a simulated 0.92 m inclined track. In Experiment 2, the wheel featured weights of unequal visual dimensions, indicating unequal mass, with the right-side weight having a simulated mass twice as heavy as the other three weights (Supplementary Figure 1). The goal was inverted, requiring participants to maximize the descent time over a 0.33 m inclined track. In both task versions, the dynamics of the wheel depended on two variables: its moment of inertia and the position of its centre of mass. The wheel's moment of inertia is determined by how mass is distributed relative to its axis of rotation. Wheels with a smaller moment of inertia (i.e., weights positioned closer to the axis) require less torque to increase angular momentum and spin faster. Additionally, the position of the centre of mass affects the wheel's initial acceleration. When the centre of mass is located in the wheel's upper-right quadrant (assuming the wheel descends from left to right), more potential energy is converted into angular kinetic energy, resulting in greater increases in angular momentum. In Experiment 1, wheels featured equal weights, meaning that unbalanced configurations had their centre of mass positioned away from the axis of rotation. In Experiment 2, wheels had unequal weights, so that balanced configurations systematically had their centre of mass positioned away from the axis of rotation. Configurations that resulted in a non-rolling wheel (due to the position of its centre of mass) received a performance score of zero in both task versions. Across treatments and conditions, the maximum payoff per trial was 0.4 pounds. Sliders in the abstract experimental conditions had the same visual dimensions as those in the Technical Reasoning condition in both task versions.

Procedure. After indicating their interest in the study and providing initial consent on Prolific, participants connected to the experimental interface. They were then presented with the study's consent form and could proceed with the experiment if they approved. Participants were randomly assigned to one experimental condition and one sex-segregated transmission chain. On-screen instructions followed. In the abstract conditions, participants were informed that they were tasked with "optimizing a configuration by setting the position of four movable sliders." In the Technical Reasoning condition, they were told to "optimize a wheel traveling down an inclined rail by setting the position of four movable weights." Regardless of the condition, participants were informed that there was a relationship between the configuration and the reward they would earn and that repeating the same configuration would result in the same payoff. Participants were also informed that they were part of a chain and that the task was collective. They were told they would complete five trials and that their last two configurations would be transmitted to the next participant in the chain, and were informed of their position in the transmission chain. Participants were further notified that their final earnings would depend on both their own cumulative payoff and the payoff of the next participant in the chain. They were not informed of the payoff associated with the best possible configuration. Participants had to answer two comprehension questions and could not proceed if they provided incorrect answers twice.

Analyses. We used Bayesian multilevel models fitted in R [43] using the *rethinking* package [44]. Results are reported as posterior means with 95% credible intervals. Codes used to analyse the data are publicly available at https://www.github.com/aseyq/causal-reasoning. Full model outputs are provided in the Supplementary Information section. No data points were excluded from the analyses.

Payoff across generations. To assess whether payoff improved across generations within a given experimental condition, we fitted a linear model with 'Payoff' as the outcome variable, 'Generation' as the predictor variable, and 'Participant Identity' and 'Chain Identity' as random effects. To determine whether payoff was higher or improved more rapidly across generations in one condition compared to another, we fitted a linear model with 'Payoff' as the outcome variable, 'Generation,' 'Condition,' and the interaction term 'Generation: Condition' as predictor variables, with 'Participant Identity' and 'Chain Identity' as random effects. In both types of analyses, the dataset included all participants' five trials and all five generations.

Exploration. To assess whether facing the physical task reduced exploration compared to the abstract task, we calculated four-dimensional Euclidean distances between each configuration produced by first-generation participants and the average configuration of that condition. We then tested whether these distances differed between the Technical and Causal Reasoning conditions by fitting a linear model with ‘Distance’ as the outcome variable, ‘Condition’ as the predictor variable, and ‘Participant Identity’ as a random effect. Similar models were run to assess exploration in terms of the position of the centroid of the configurations (i.e., the center of mass of the wheel in Experiment 1) and their compactness (i.e., the inertia of the wheel in Experiment 1). For the former analyses, we calculated two-dimensional Euclidean distances based on the centroid coordinates, determined by subtracting the position of the left slider/weight from that of the right slider/weight and the position of the bottom slider/weight from that of the top slider/weight. For analyses of compactness, we calculated one-dimensional Euclidean distances based on compactness level, determined by summing the positions of the four sliders/weights. In these analyses, only data from first-generation participants were included to ensure that exploration patterns reflected participants’ intuitive search strategies rather than being influenced by information inherited from previous participants. In Figure 5, configurations were considered balanced when the coordinates of the centroid (x, y) both equaled zero. The centroid was classified as falling within the upper-right quadrant when $x \geq 0$ and $y \geq 0$, the bottom-right quadrant when $x \geq 0$ and $y < 0$, the bottom-left quadrant when $x \leq 0$ and $y < 0$, and the upper-left quadrant when $x \leq 0$ and $y \geq 0$.

Deviation from the pre-registered analyses. Compared to the pre-registered analyses, the models investigating ‘Payoff’ across generations do not include ‘Trial’ or the interaction term ‘Trial:Condition’ as predictor variables. Under the original model specification, the model had difficulty allocating variance between the ‘Trial’ and ‘Generation’ predictors, leading to misleading parameter estimates for both main effects and interaction terms. To address this issue and align the analyses more closely with our primary focus on ‘Generation’, we removed the ‘Trial’ variable from all models.

Pre-registration. Both experiments were pre-registered on AsPredicted prior to data collection. Experiment 1 was pre-registered under ID #169017 (Technical and Causal Reasoning conditions) and ID #173,042 (Pure Cultural Transmission condition), and Experiment 2 under ID #212,783. Each pre-registration specified the experimental design, hypotheses, sample size, exclusion criteria, and planned analyses.

Data and Code Availability. The data and scripts for the analyses are publicly available at: https://www.github.com/aseyq/causal-reasoning.

Acknowledgments. We thank Antoine Jacquet for providing the equations describing the wheel’s motion. We are also grateful to Jean-François Bonnefon, Rob Boyd, Léo Fitouchi, and Ben Pitt for their helpful discussions and comments on earlier drafts of this paper. Finally, we thank the editor and two anonymous reviewers for their detailed comments, which greatly improved the manuscript.

M.D. acknowledges Institute for Advanced Study in Toulouse (IAST) funding from the French National Research Agency (ANR) under grant no. ANR−17-EURE−0010 (Investissements d'Avenir programme) and ANR funding through the ANR−21-CE28−0019−01 project.

References

1. Wolpert, L. and C. Pasternak, *Causal belief makes us human.* What makes us human, 2007: p. 164-181.
2. Penn, D.C., K.J. Holyoak, and D.J. Povinelli, *Darwin's mistake: Explaining the discontinuity between human and nonhuman minds.* Behavioral and Brain Sciences, 2008. **31**(2): p. 109-130.
3. Goddu, M.K. and A. Gopnik, *The development of human causal learning and reasoning.* Nature Reviews Psychology, 2024. **3**(5): p. 319-339.
4. Wolpert, L., *Causal belief and the origins of technology.* Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 2003. **361**(1809): p. 1709-1719.
5. Osiurak, F. and E. Reynaud, *The elephant in the room: What matters cognitively in cumulative technological culture.* Behavioral and Brain Sciences, 2020. **43**: p. e156.
6. Pinker, S., *The cognitive niche: Coevolution of intelligence, sociality, and language.* Proceedings of the National Academy of Sciences, 2010. **107**: p. 8993-8999.
7. Boyd, R., P.J. Richerson, and J. Henrich, *The cultural niche: Why social learning is essential for human adaptation.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**: p. 10918-10925.
8. Boyd, R., P.J. Richerson, and J. Henrich, *The cultural evolution of technology: facts and theories.* Cultural evolution: society, technology, language, and religion, 2013: p. 119-142.
9. Henrich, J., *The secret of our success: how culture is driving human evolution, domesticating our species, and making us smarter*. 2015: Princeton University Press.
10. Lehmann, L. and J.Y. Wakano, *The handaxe and the microscope: individual and social learning in a multidimensional model of adaptation.* Evolution and Human Behavior, 2013. **34**(2): p. 109-117.
11. Wilson, R.A. and F. Keil, *The Shadows and Shallows of Explanation.* Minds and Machines, 1998. **8**(1): p. 137-159.
12. Kendal, R.L., et al., *Social Learning Strategies: Bridge-Building between Fields.* Trends in Cognitive Sciences, 2018. **22**(7): p. 651-665.
13. Thompson, B., et al., *Complex cognitive algorithms preserved by selective social learning in experimental populations.* Science, 2022. **376**(6588): p. 95-98.
14. Miu, E., et al., *Innovation and cumulative culture through tweaks and leaps in online programming contests.* Nature Communications, 2018. **9**(1): p. 2321.
15. Henrich, J., *Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses - The Tasmanian case.* American Antiquity, 2004. **69**(2): p. 197-214.

16. Nia, H.T., et al., *The evolution of air resonance power efficiency in the violin and its ancestors.* Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2015. **471**(2175): p. 20140905.
17. Derex, M., et al., *Causal understanding is not necessary for the improvement of culturally evolving technology.* Nature Human Behaviour, 2019. **3**(5): p. 446-452.
18. Harris, J.A., R. Boyd, and B.M. Wood, *The role of causal knowledge in the evolution of traditional technology.* Current Biology, 2021.
19. Kendal, R.L., *Explaining human technology.* Nature Human Behaviour, 2019. **3**(5): p. 422-423.
20. Osiurak, F., et al., *Technical reasoning is important for cumulative technological culture.* Nature Human Behaviour, 2021.
21. Caldwell, C.A. and A.E. Millen, *Studying cumulative cultural evolution in the laboratory.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2008. **363**(1509): p. 3529-3539.
22. Mesoudi, A. and A. Whiten, *The multiple roles of cultural transmission experiments in understanding human cultural evolution.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2008. **363**(1509): p. 3489-3501.
23. Derex, M. and R. Boyd, *The foundations of the human cultural niche.* Nat Commun, 2015. **6**: p. 8398.
24. Rozenblit, L. and F. Keil, *The misunderstood limits of folk science: an illusion of explanatory depth.* Cognitive science, 2002. **26**(5): p. 521-562.
25. Osiurak, F., N. Claidière, and G. Federico, *Bringing cumulative technological culture beyond copying versus reasoning.* Trends in Cognitive Sciences, 2022.
26. Derex, M., et al., *Social learning preserves both useful and useless theories by canalizing learners' exploration.* Proceedings of the Royal Society B: Biological Sciences, 2025. **292**(2039): p. 20242499.
27. Miu, E. and T.J.H. Morgan, *Cultural adaptation is maximised when intelligent individuals rarely think for themselves.* Evolutionary Human Sciences, 2020. **2**: p. e43.
28. Derex, M. and A. Mesoudi, *Cumulative Cultural Evolution within Evolving Population Structures.* Trends in Cognitive Sciences, 2020. **24**(8): p. 654-667.
29. Creanza, N., O. Kolodny, and M.W. Feldman, *Greater than the sum of its parts? Modelling population contact and interaction of cultural repertoires.* Journal of The Royal Society Interface, 2017. **14**(130).
30. Derex, M., et al., *Experimental evidence for the influence of group size on cultural complexity.* Nature, 2013. **503**: p. 389-391.
31. Muthukrishna, M., et al., *Sociality influences cultural complexity.* Proceedings of the Royal Society B: Biological Sciences, 2014. **281**(1774): p. 20132511.
32. Kempe, M. and A. Mesoudi, *An experimental demonstration of the effect of group size on cultural accumulation.* Evolution and Human Behavior, 2014. **35**(4): p. 285-290.
33. Heyes, C., *The cognitive reality of causal understanding.* Trends in Cognitive Sciences, 2023. **27**(4): p. 337-338.
34. Cantlon, J.F. and S.T. Piantadosi, *Uniquely human intelligence arose from expanded information capacity.* Nature Reviews Psychology, 2024.

35. Brand, C.O., A. Mesoudi, and P.E. Smaldino, *Analogy as a Catalyst for Cumulative Cultural Evolution.* Trends in Cognitive Sciences, 2021. **25**(6): p. 450-461.
36. Derex, M., *Human cumulative culture and the exploitation of natural phenomena.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2021. **377**: p. 20200311.
37. Winters, J., *Escaping optimization traps: the role of cultural adaptation and cultural exaptation in facilitating open-ended cumulative dynamics.* Palgrave Communications, 2019. **5**(1): p. 149.
38. Morgan, T.J.H. and M.W. Feldman, *Human culture is uniquely open-ended rather than uniquely cumulative.* Nature Human Behaviour, 2024.
39. Tennie, C., L.M. Hopper, and C.P. Van Schaik, *On the Origin of Cumulative Culture: Consideration of the Role of Copying in Culture-Dependent Traits and a Reappraisal of the Zone of Latent Solutions Hypothesis*, in *Chimpanzees in Context*. 2020, University of Chicago Press.
40. Mesoudi, A. and A. Thornton, *What is cumulative cultural evolution?* Proceedings of the Royal Society B: Biological Sciences, 2018. **285**(1880): p. 20180712.
41. Lewis, H.M. and K.N. Laland, *Transmission fidelity is the key to the build-up of cumulative culture.* Philosophical Transactions of the Royal Society B-Biological Sciences, 2012. **367**(1599): p. 2171-2180.
42. Charbonneau, M., *Modularity and Recombination in Technological Evolution.* Philosophy & Technology, 2016. **29**(4): p. 373-392.
43. Team, R.D.C., *R: A Language and Environment for Statistical Computing*. 2011, R Foundation for Statistical Computing: Vienna, Austria.
44. McElreath, R., *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2016: CRC Press.