



The evolution of human cognition

Citation

Park, Peter S. 2023. The evolution of human cognition. Doctoral dissertation, Harvard University Graduate School of Arts and Sciences.

Permanent link

<https://nrs.harvard.edu/URN-3:HUL.INSTREPOS:37375827>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The Harvard Kenneth C. Griffin
Graduate School of Arts and Sciences



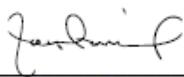
DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Department of Mathematics
have examined a dissertation entitled


The Evolution of Human Cognition

presented by **Peter S. Park**

candidate for the degree of Doctor of Philosophy and hereby
certify that it is worthy of acceptance.

Signature 


Typed name: Professor Joseph Henrich

Signature 

Typed name: Professor Martin Nowak

Signature 

Typed name: Professor Christian Hilbe (Max Planck Institute)

Signature 

Typed name: Professor Feng Fu (Dartmouth)

Date: May 10, 2023

The evolution of human cognition

a dissertation presented

by

Peter S. Park

to

The Department of Mathematics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Mathematics

Harvard University

Cambridge, Massachusetts

May 2023

©2023 – Peter S. Park
all rights reserved.

The evolution of human cognition

Abstract

The evolution of human cognition presents a number of empirical puzzles. Why did humans evolve cognitive biases, behaviors that cause systematic errors in judgement? Why are humans the only animal species to cooperate in very large groups of non-kin? And what accounts for the consistent emergence of egalitarian social norms in evolutionarily relevant foragers and likely in most ancestral humans? In this dissertation, I present three evolutionary models of human cognition that set out to explain these and other puzzles from plausible first principles. An overarching theme of the three studies is that integrating the social component of ancestral human decision-making into traditional evolutionary models, which primarily focus on individual decision-making, may be essential for successfully explaining the various evolutionary puzzles posed by human cognition.

Contents

Title page	i
Copyright	ii
Abstract	iii
Contents	iv
Dedication	vi
Acknowledgements	vii
0 Introduction	1
0.1 First evolutionary puzzle of human behavior: Cognitive biases	2
0.2 Second evolutionary puzzle: Cooperation in large groups of non-kin	5
0.3 Third evolutionary puzzle: Egalitarianism	10
1 The evolution of cognitive biases in human learning	14
1.1 Introduction	15
1.2 The model	22
1.3 Results	41
1.4 Discussion	73
1.5 Appendix	97
2 Cooperation in alternating interactions with memory constraints	138
2.1 Introduction	139
2.2 Results	144
2.3 Discussion	160
2.4 Methods	163
2.5 Supplementary Note 1: Baseline model	171
2.6 Supplementary Note 2: Equilibrium analysis for alternating games	176
2.7 Supplementary Note 3: Extensions of the baseline model	203
2.8 Supplementary Note 4: Proofs of the analytical results	232
3 A theory of specialization, exchange, and innovation in human groups	243
3.1 Introduction	244
3.2 The model	247

3.3	Results	253
3.4	Discussion	278
3.5	Supplementary Information: The model	286
3.6	Supplementary Information: Empirical results	316
	References	356

Dedicated to Alice.

Acknowledgments

I would like to express my heartfelt gratitude to several people who have been instrumental to my scientific growth in the course of my Ph.D. research program.

First and foremost, I would like to thank my advisor, Joe Henrich, for his continual support and guidance. Your to-the-point and interdisciplinary approach to research has been an immense inspiration, and I am extremely grateful for the opportunity to learn from you: both as a Ph.D. student and a collaborator.

I would also like to express my appreciation to my mentors—Feng Fu, Christian Hilbe, Eric Maskin, and Martin Nowak—for their valuable input and constructive feedback on my research work. Your expertise in the field of game theory and evolutionary dynamics—and for Christian and Martin, your invaluable contributions to our research collaborations—have been tremendously helpful for refining my research.

To fellow members of the Henrich lab—Mohammad Atari, Rahul Bhui, Damian Blasi, Matt Cashman, Cammie Curtin, Helen Davis, Jennifer Devereaux, Joseph Dexter, Tommy Flint, Patricia Greenfield, Kevin Hong, Daniel Kelly, Ivan Kroupin, Graham Noblit, Rachna Reddy, Amar Sarkar, Slava Savitskiy, Manvir Singh, Max Winkler, Mona Xue, Tian Chen Zeng, and others—I am grateful for the camaraderie and support throughout my graduate school journey. Your insights, discussions, and feedback have been invaluable in shaping my ideas and improving my work.

I would also like to thank my Princeton friends residing across the country—Sam Cabot, Sunny Feng, Siddhartha Jayanti, Kevin Griffin, and Leo Tinone—for providing a much-needed respite from the demands of research via board games.

To my Harvard friends spanning the math department and the Effective Altruism student group—Chris Bakerlee, Holly Elmore, Elliot Glazer, Trevor Levin, Andrew Liu, Tina Torkaman, and Zechen Zhang—thank you for providing a supportive community and a sounding board for my ideas (on both research and issues of society).

To my family—Mom, Dad, and John—thank you for your unwavering love and support throughout my graduate school journey. Your encouragement and belief in me have been a constant source of motivation.

Finally, I would like to express my deepest gratitude to my wife, Alice, for all her love, support, and understanding throughout my Ph.D. journey. Your presence in my life has been the anchor that kept me grounded and motivated to pursue my dreams.

0

Introduction

Human cognition poses a number of evolutionary puzzles:

1. Why did humans evolve cognitive biases, behaviors that cause systematic errors in judgement?²³⁴
2. Why are humans the only animal species to cooperate in very large groups of non-kin?²⁹

3. What accounts for the consistent emergence of egalitarian social norms in evolutionarily relevant foragers, and likely, in most ancestral humans?^{19,63,127}

In this dissertation, I present mathematical evolutionary models—constructed by myself and my collaborators—that set out to explain these behavioral patterns from plausible first principles. I argue that incorporating the social element of ancestral humans’ decision-making—into traditional evolutionary models that are more centered on individual decision-making—is crucial for causally explaining the aforementioned empirical puzzles. This dissertation thereby represents an effort to understand the evolutionary roots of human cognition, cooperation, and social norms in an interconnected way.

0.1 First evolutionary puzzle of human behavior: Cognitive biases

The Cultural Brain Hypothesis posits that adaptive, socially exchanged, and inter-generationally accumulated knowledge comprised the primary selection pressure on ancestral human cognition^{10,90,103,126,160,163,193,223,237,244}. In other words, the ancestral humans who tended to survive were not those who were better at learning from their individual observations; they were those who were better at learning knowledge from their fellow group members. This was true because learning knowledge from group members entailed learning from their ancestors’ knowledge: from the group’s inter-generationally accumulated culture. Over time, this culture tends to become more adaptive (at least in the local environment) because the best knowledge is more likely to be obtained, retained, and passed on to the next generation. This overall stochastic process is called cumulative cultural evolution.

At the moment, arguably the most influential paradigm for the evolution of human

cognition is not the Cultural Brain Hypothesis, but a competing theory: the Machiavellian Intelligence Hypothesis³⁵. This hypothesis posits that human cognition’s unprecedented characteristics—its unusual complexity, adaptability, and propensity to cooperate—evolved primarily due to selection pressures favoring better strategic thinking, rather than those favoring the accumulation of adaptive knowledge.

However, the Machiavellian Intelligence Hypothesis cannot explain why humans have evolved to make cognitively biased decisions even in fitness-relevant domains. When learning from high-variance payoff data, unbiased Bayesian updating is evolutionarily optimal¹⁵¹. In line with this, many non-human animal species’ foraging and reproductive decisions are consistent with the predictions of Bayesian updating models²³⁶. But human learning systematically deviates from Bayesian updating in various ways. These deviations are called cognitive biases²³⁴, and include:

1. persistent underinference from observations¹⁶ (e.g., overconfident beliefs that do not meaningfully learn from past experience: as shown, for example, by non-learning leaders who repeatedly underprepare for catastrophes like pandemics),
2. the hard-easy effect¹⁵⁷ (overconfidence on difficult tasks, underconfidence on easy tasks), and
3. non-monotonic confidence²⁰³ (task confidence is not monotonically increasing in the experience level).

A rational strategy of individual learning would base its estimate of future task-payoffs on an unbiased, Bayesian aggregation of the past payoff data: say, starting from an initial prior ω that has not ruled out the true state $\varphi \in \Theta$. Here, Θ denotes the state space. With prior probability one, the learner’s estimate of their expected

payoff-acquisition ability

$$\int_{\hat{\theta} \in \Theta} \mathbb{E}[\hat{\theta}] d\omega_{s_1, \dots, s_n}(\hat{\theta}), \quad (1)$$

would converge to the true expected payoff $\mathbb{E}[\varphi]$ as the number of observations n goes to infinity. (Here, ω_{s_1, \dots, s_n} denotes the Bayesian-updated prior after observing the sequence of payoff data s_1, \dots, s_n .) While the true expected payoff $\mathbb{E}[\varphi]$ is unobservable, it will with probability one coincide with the mean of the past payoff data

$$\frac{s_1 + \dots + s_n}{n} \quad (2)$$

as $n \rightarrow \infty$, due to the Law of Large Numbers. Thus, we should be skeptical of a learner's purported Bayesianness if their estimate of their expected payoff-acquisition ability does not appear to converge to the mean of the past payoff data (2).

The fact that human learning is cognitively biased (rather than Bayesian-rational) is evolutionarily puzzling when assuming the Machiavellian Intelligence Hypothesis, since flawed learning from individual observations constitutes a strategic weakness.

On the other hand, the three aforementioned cognitive biases can be explained in the framework of the Cultural Brain Hypothesis. In Chapter 1, I present the decision-theoretic model of my Journal of Theoretical Biology paper, which represents how ancestral humans might have learned various foraging tasks over time and over repeated attempts¹⁷⁸. In this model, an ancestral human learner (the student) attempts to learn task-specific knowledge from a role model (the teacher). However, the student does not meaningfully retain their past payoff data, because doing so would result in onerous fitness costs from overcommitting attention (e.g., risks of ambushes and accidental injuries due to a lack of situational awareness). Instead, the student's evolutionarily optimal estimate of their expected marginal payoff—their confidence—

is a function of setting-specific sources of information: specific to the setting of social learning, not individual learning. The evolutionary mismatch between (a) this estimate of confidence that is specialized to ancestral social learning and (b) contemporary settings of individual learning can manifest as all three of the aforementioned cognitive biases.

Our evolutionary model of human cognitive biases has the following practical implication. Cost-benefit analyses in public policy often assume that the highest price a person is willing to pay for an item—their willingness to pay—reveals an unbiased aggregate of their private payoff observations regarding it⁸⁰. However, this may not be true when these observations have high variance. In this case, when a person is buying an item to use for a certain task, their willingness to pay for the item may be a non-monotonic function of the true value acquirable from the item at the person’s level of task experience. It follows that the person’s willingness to pay for an item may not even be a sufficient statistic of the item’s true value to them; in general, the latter cannot be predicted from the former. Cost-benefit analyses may thus be improved by more directly estimating an item’s true value to each person (e.g., by estimating their mean past payoff from the item), rather than by using the person’s willingness to pay as a proxy. This implication casts doubt on several central dictums of economics, such as the dictum that economic surplus (defined in terms of willingness to pay) is the measure of success that public policy should pursue.

0.2 Second evolutionary puzzle: Cooperation in large groups of non-kin

Humans’ extensive cooperation with non-kin is unique among social animal species²⁹. In many realistic settings, this level of cooperation requires altruism: actions that

benefit another individual at a cost to oneself. Trivers²³³ proposed direct reciprocity as an evolutionary explanation of such altruism. In this framing, human cooperation results from an evolutionarily stable strategy of bestowing altruism conditional on the other player also bestowing altruism, which results in an equilibrium of a self-sustaining quid pro quo. Such a strategy would need to be evolutionarily stable, because otherwise, a strategically rational player would eventually find another strategy that can invade the self-cooperating strategy. It has been hypothesized that unprecedented intelligence is what allows humans to consistently evolve the desirable strategy of reciprocal altruism, more consistently than other social animal species. This notion is laid out in Hypothesis 2.1 of the study of Proto et al.¹⁹¹:

Higher-Intelligence subjects (i) find a better strategy—that is, with higher payoff—and conceive a larger set of strategies in a given environment and (ii) are more consistent in their implementation...[H]igher-Intelligence subjects will achieve, in general, higher rates of cooperation.

Probing the veracity of the hypothesis of Proto et al. may be especially important, given that increasingly intelligent AI models may become increasingly embedded in society—consider, for example, OpenAI’s mission of creating “highly autonomous systems that outperform humans at most economically valuable work”¹⁷⁷—and that researchers have non-empirically posited in line with the hypothesis of Proto et al. that these AI models will convergently evolve human-level cooperative tendencies²⁵².

Suppose that humans’ unusually extensive cooperation with non-kin is indeed best understood as a largely independent combination of dyadic reciprocal relationships: of favor-sharing and goodwill between various pairs of people²¹⁸. At first glance, this favor-sharing does seem achievable by a stable cooperative strategy in a set-

ting of direct reciprocity. This is because in the most commonly used model of direct reciprocity—the repeated prisoner’s dilemma—there exist evolutionarily stable strategies that are self-cooperative⁶⁵, as long as we make the realistic assumption that actions in the game are played erroneously with a positive probability ε .

However, several empirical findings cast doubt on this proposed explanation. First, the simplest evolutionary stable strategy that can sustain cooperation, Win-Stay-Lose-Shift, is rarely observed in human-subject experiments of the repeated prisoner’s dilemma^{45,66}. Second, non-cooperation (Always Defect) is also an evolutionarily stable strategy of the repeated prisoner’s dilemma¹³⁷. Third, humans perform strategically worse than chimpanzees in some simple games¹⁴⁴, which casts doubt on the hypothesis that increased strategic ability was the mechanism by which unprecedented human intelligence led to unprecedented cooperation. And finally, data from social animal species generally do not identify frequent reciprocity opportunities to be a strong predictor of cooperative behavior^{78,145,243}.

Motivated by the inability of the traditional repeated prisoner’s dilemma to explain the aforementioned findings, my coauthors Martin Nowak, Christian Hilbe, and I analyzed a more biologically realistic variant of the repeated prisoner’s dilemma. The traditionally considered variant of the repeated prisoner’s dilemma assumes that the players’ interactions are simultaneous, an often non-realistic assumption. For example, consider someone who incurs a cost to help another individual, in the hopes of receiving a similar favor in the future. This more realistic setting of cooperation is asynchronous rather than simultaneous.

In Chapter 2, I present my Nature Communications paper (coauthored with Martin Nowak and Christian Hilbe) analyzing a more realistic variant of the repeated prisoner’s dilemma, in which interactions are alternating rather than simultaneous¹⁸¹.

The key lemma for this paper was inspired by William Press and Freeman Dyson’s pioneering result for the traditionally studied simultaneous game¹⁹⁰. Their result was that a player can generally afford to have the same memory size as their opponent’s strategy, without any loss to their or their co-player’s payoff. Press and Dyson’s result is tight for the simultaneous game, but our key lemma improves on this result for the alternating version of the game.

Lemma 0.1. Consider an alternating prisoner’s dilemma between two players with strategies \mathbf{p} and \mathbf{q} , with memory sizes P and Q respectively. Without loss of generality, suppose that $Q \geq P$. Then, the player using strategy \mathbf{q} can switch to a strategy of memory size $< P$ without changing either player’s payoff.

In particular, whenever a strategy has at least as high a memory size as the resident strategy, there exists a neutral mutation to the former which possesses a strictly lower memory size than the latter.

Using this key lemma, Nowak, Hilbe, and I found that the assumption that interactions are alternating makes cooperation more volatile and more difficult to evolve (as long as we again assume a positive probability ε of implementation error). Our result is that among the memory-1 strategies—strategies whose conditional probability of cooperating solely depends on each player’s most recent action—there are no evolutionarily stable strategies that can sustain cooperation. In fact, the only evolutionarily stable memory-1 strategy is Always Defect, the strategy that never cooperates. In other words, sustaining cooperation via direct reciprocity may be more volatile and more difficult to evolve than was traditionally thought.

Moreover, Hilbe and I are in the process of generalizing this to a similar result for all finite-memory strategies, rather than just memory-1 strategies¹⁷⁹. Again using the

key lemma, we proved the preliminary result that if higher-memory strategies require a higher fitness cost than lower-memory ones, then Always Defect is the unique Nash-equilibrium strategy. This result is true for agents of arbitrarily high memory sizes, and also implies that Always Defect is the only evolutionarily stable strategy they can use (since every evolutionarily stable strategy is also a Nash-equilibrium strategy).

The proof is as follows. Consider a population with an arbitrary resident strategy. Our key lemma implies that unless the resident strategy has memory size 0 (i.e., action is always given by the same probability distribution), it can be strictly invaded by a mutant with a strictly lower memory size. The reason that this mutant strategy has a strictly higher payoff than the resident strategy is due to our assumption that all else equal, higher-memory strategies require a higher fitness cost than lower-memory ones. Moreover, among strategies with memory size 0—the only strategies that are not vulnerable mutants of the aforementioned type—the only Nash-equilibrium strategy is Always Defect. Indeed, a strategy that places any positive probability mass on the action Cooperate will be strictly invaded by an Always Defect mutant. Thus, Always Defect is the only resident strategy that can emerge in the long run. Any resident population will continue to be invaded by lower-memory mutants until it becomes memory-0, whereupon it will be invaded by increasingly non-cooperative strategies, the overall culmination of which is necessarily Always Defect.

The above argument works for a wide class of asynchronous interaction structures, but a crucial assumption is that every probabilistic strategy of finite memory size—and in particular, the lower-memory mutant strategy that can invade the resident population—is achievable as a phenotype. This constitutes a formal demonstration of Joe Henrich’s conjecture regarding the evolution of altruistic cooperation⁸⁷. Specifically, Henrich conjectured that “all genetic evolutionary explanations to the altruism

dilemma are successful to the degree that they allow natural selection to operate on statistically reliable patterns or regularities in the environment,” and that cooperation is difficult to evolve in the absence of “constraints that maintain this regularity.” Our more realistic asynchronous variant of the repeated prisoner’s dilemma casts doubt on the hypothesis of Proto et al.¹⁹¹ in favor of Henrich’s conjecture. Higher-intelligence subjects’ ability to access a larger set of strategies will—all else equal—select against cooperative tendencies, rather than select for them. This is because an unconstrained strategy space is precisely what allows for Always Defect to robustly dominate: for cooperation to robustly be selected against.

It follows that humans’ unusually cooperative tendencies may not be best understood as a largely independent combination of dyadic reciprocal relationships that are enabled by humans’ purportedly unconstrained strategic ability. More sociality-based mechanisms of cooperation, such as indirect reciprocity^{173,216} and group selection^{25,87,197,230,247}, may play a pivotal causal role that direct reciprocity by itself may not be able to play.

0.3 Third evolutionary puzzle: Egalitarianism

Despite its pivotal importance, cumulative cultural evolution is challenging to study. This is because it involves so many different mechanisms, all of which interact together in complex and mysterious ways. How does each individual choose which group member they will learn knowledge from? What exact knowledge will they learn? How and when does new knowledge get discovered?

We would like to know the answers to these questions, as well as how they interact to produce cumulative cultural evolution. A reasonable (albeit challenging) plan

would be to construct a parsimonious model of the system that is based on correct first principles, incorporates all of the relevant mechanisms, and is still analytically tractable enough to make robust predictions.

In Chapter 3, I present my joint work with Slava Savitskiy and Joe Henrich¹⁸², a model of cumulative cultural evolution that takes a first stab at incorporating all of the relevant mechanisms (though it almost certainly misses some important ones). Our model is inspired by Paul Krugman’s Nobel-prize-winning model of specialization, cooperative production, and exchange¹²⁵. We added to Krugman’s model even more mechanisms of cumulative cultural learning posited by the Cultural Brain Hypothesis, such as knowledge discovery, intergenerational knowledge accumulation, and social norms.

The resulting trade model yields ten scientific predictions about human societies’ specialization and trade, some of which are counterintuitive. Yet, we find preliminary empirical corroboration for nine of the ten predictions, and we propose future empirical tests for the remaining prediction.

To illustrate, our model predicts that egalitarian societies robustly innovate the optimal number of specializations in the long run, a prediction that finds preliminary empirical support in the Ethnographic Atlas^{12,22,76,118,121,158} and the Western North American Indians^{112–114,118} ethnographic datasets. This helps explain the consistent emergence of egalitarian social norms in contemporary evolutionarily relevant forager societies, and likely in most ancestral human societies^{19,63,127}. We propose that egalitarianism was evolutionarily adaptive in spite of its free-rider problem²⁰, because among ancestral humans whose primary selection pressure was the optimal social accumulation of specialized knowledge, egalitarianism robustly helped facilitate this.

Our model also predicts that the penalty to long-run innovation imposed by non-

egalitarianism can be offset by an intermediate degree of overconfidence. This prediction finds preliminary corroboration in the study of Cieřlik et al.⁴⁰ of how countries' GDP varies with their average level of entrepreneurial overconfidence. Our model thus adds to the case that cognitive biases like overconfidence—which are evolutionarily puzzling when viewed as systematic errors in how people learn from their individual observations—may actually be behavioral byproducts of a selection pressure towards the social learning of specialized knowledge¹⁷⁸. Even if overconfidence is selected against at the level of individual selection, it can be selected for at the level of group selection due to its optimizing effects on human societies' collective brains.

Our model also helps resolve the evolutionary puzzle of why humans cooperate in large groups of non-kin. While direct reciprocity has traditionally been proposed as an evolutionary explanation of altruism²³³, recall that reciprocal cooperation is much more difficult to sustain when the repeated prisoner's dilemma is modified to be more biologically realistic¹⁸¹. Also, among social animals, a large amount of reciprocity opportunities is not as strong of a predictor of altruism as proposed by the theory of direct reciprocity^{78,145,243}. This overall suggests that direct reciprocity, by itself, may be insufficient to evolutionarily explain human ultrasocial cooperation.

Cultural group selection is a more viable evolutionary explanation^{25,87,197}. This differs from genetic group selection, which is likely inconsistent with the negligibility of genetic differences between competing human groups³⁹. Indeed, culturally transmissible phenotypes are characterized by far higher variation between groups than genetically transmissible phenotypes¹⁴.

Our model helps explain why among groups, cultural variation is larger than genetic variation. It does so by identifying a source of cultural variation other than random mutations. Our model hypothesizes that when different groups begin to trade,

their members' pursuit of their own individual fitness result in the groups specializing away from each other. Intentional specialization at the level of individual decisions is a more viable cause of large phenotypic variation than random changes like genetic mutations, especially when the gene flow between groups is substantial. Cultural group selection can act on this large phenotypic variation in a way where group-cooperative phenotypes are selected, despite their sizeable negative effect on individual fitness in excess of what can be sufficiently explained by direct reciprocity.

Overall, preliminary empirical evidence suggests that complex specialization and trade is not a modern outlier of Pleistocene human societies. It is an adaptation that emerged in the likely egalitarian foraging societies of the Holocene which comprised most of human evolutionary history^{30,32,133}. This adaptation encompasses various selection pressures towards the optimal social accumulation of specialized knowledge, in ways that are crucial to resolving multiple puzzles in the evolution of human cognition. Such puzzles include the robustness of egalitarian sharing norms in evolutionary relevant societies, the evolution of overconfidence, and altruistic cooperation in large groups of non-kin.

What would I eliminate if I had a magic wand? Overconfidence.²¹⁴

Daniel Kahneman

1

The evolution of cognitive biases in human learning

Abstract: Cognitive biases like underinference, the hard-easy effect, and recurrently non-monotonic confidence are evolutionarily puzzling when viewed as persistent flaws in how people learn from environmental feedback. To explain these empirically robust cognitive biases from an evolutionary perspective, we propose a model of ancestral human learning

based on the cultural-evolutionary-theoretic hypothesis that the primary selection pressure acting on ancestral human cognition pertained not to learning individually from environmental feedback, but to socially learning task-specific knowledge. In our model—which is inspired by classical Bayesian models—an ancestral human learner (the student) attempts to learn task-specific knowledge from a role model, with the option of switching between different tasks and role models. Suppose that the student’s method of learning from their role model is a priori uncertain—in that it can either be successful imitation learning or de facto innovation learning—and the ecological fitness costs of meaningfully retaining environmental feedback are high. Then, the student’s fitness-maximizing strategy does not retain their environmental feedback and—depending on the choice of model parameters—can be characterized by all of the aforementioned cognitive biases. Specifically, in order for the evolutionarily optimal estimate of confidence in this learning environment to be recurrently non-monotonic, it is necessary (as long as the environment’s marginal payoff function satisfies a plausible quantitative condition) that a positive proportion of ancestral humans’ attempted imitation learning was unknowingly implemented as de facto innovation learning. Moreover, an ecologically rational strategy of selective social learning can plausibly cause the evolutionarily optimal estimate of confidence to be recurrently non-monotonic in the empirically documented way: general increase with an intermediate period of decrease.

1.1 Introduction

Humans have evolved to meaningfully incorporate into their beliefs the low-variance, essentially deterministic environmental feedback they observe—the domain of causal inference—so as to improve future decisions¹⁸⁶. For example, people often learn to pay credit card bills¹ and return rented videos⁸¹ on time after first paying late fees.

However, the same cannot be said when the variance is high. In the domain of high-variance environmental feedback, unbiased Bayesian updating should in theory be normatively rational⁴³ and even evolutionarily optimal¹⁵¹ in many settings. In line with this, a review of 11 empirical studies of animal foraging and reproductive decisions—spanning eight species of birds, three of non-human mammals, one of fish, and one of insects—found the behavior of all but one of the species to be consistent with the predictions of Bayesian updating models²³⁶. For humans, however, learning in settings of high-variance environmental feedback deviates from Bayesian updating in various ways²³⁴. These deviations, referred to in the literature as cognitive biases, result from evolved tendencies by which humans systematically fail to learn meaningfully from high-variance environmental feedback.

A myriad of cognitive biases are apparent from the insightful experiments of Sanchez and Dunning^{203,204} on human learning. In each variant of their experiment, subjects learned a new task possessing a payoff structure with fixed uncertainty: classifying profiles with lists of properties (for example, symptoms) into categories (for example, made-up diseases). The subjects attempted this task 60 times while simultaneously reporting their confidence: their self-estimate of the probability that their answer is correct. After each of their 60 answers, they received immediate feedback. Despite this, the subjects did not learn from their environmental feedback in a Bayesian-rational manner, as one can see from the following patterns in the data (see Figures 1–4 of Sanchez and Dunning, 2018²⁰³ and Figures 1–3 of Sanchez and Dunning, 2020²⁰⁴).

1. The subjects' confidence graph—that of their average self-estimate as a function of trial number—was non-monotonic. Specifically, the confidence graph was comprised of three phases: a beginning phase of increase, an intermedi-

ate phase of decrease, and a final phase that returned to increase. This pattern agrees with the finding of the well-known experiment of Kruger and Dunning¹²⁴ on confidence as a function of true ability—as well as its replications—that the former variable can be a non-monotonic function of the latter (see Figures 4–6 of Burson et al., 2006³⁴; Figures 5–7 of Haun, 2000⁸⁵; and Figures 2–3 of Kruger and Dunning, 1999¹²⁴). This also agrees with the work of Hoffman and Burks¹⁰⁰ investigating truckers’ self-estimates of the number of miles driven each week, which found their average to be non-monotonic with respect to the level of experience and the average of the true value, monotonically increasing in the level of experience (see Figure 1 of Hoffman and Burks, 2020¹⁰⁰).

2. The average difference between confidence and the environmental feedback eventually became positive—signifying overconfidence—and proceeded to increase instead of decaying to zero. This pattern is consistent with the extensive evidence on overconfidence in the cognitive bias literature: for example, as a cause of wars^{51,109}, stock market bubbles^{2,206}, and underpreparation for catastrophes^{139,208}. Consistently becoming overconfident compared to the environmental feedback, by itself, likely suffices to contradict Bayesian rationality⁷.
3. The confidence graphs from all variants of the experiment were essentially indistinguishable from each other, even though the subjects of each experimental variant on average performed differently and thus received different environmental feedback. The confidence graph in essence only depended on the number of past observations, the level of experience. This pattern is consistent with two well-documented cognitive biases: underinference¹⁶, the tendency to insufficiently update one’s belief in the direction of new evidence compared to

Bayesian inference; and the hard-easy effect^{131,157}, the tendency to be overconfident on difficult tasks and underconfident on easy tasks. Indeed, a pre-determined confidence function—one that depends not on past environmental feedback, but only on other types of information like one’s level of experience—would generically differ from the Bayesian aggregate of the past environmental feedback. The difference between the two would generically persist, manifesting as both underinference and—depending on the hard-easy effect—either persistent overconfidence or underconfidence.

These three non-Bayesian patterns robustly replicated in all six variants of the Sanchez–Dunning experiment (2018, 2020), including the variant that used the incentive-compatible Becker–DeGroot–Marschak method¹³ to monetarily incentivize accurate answers. The non-Bayesian inaccuracy of subjects’ learning¹⁰⁸ and the persistence of this inaccuracy in the face of monetary incentivization⁵⁶ have also been documented in replications of the Kruger–Dunning experiment; these phenomena have been found in the aforementioned work of Hoffman and Burks¹⁰⁰ on truckers’ self-estimates of productivity, as well. Note that the Kruger–Dunning experiment is similar in objective and design to the Sanchez–Dunning experiment. A crucial difference, however, is that accurate environmental feedback is immediately provided by the experimenter in the latter, but not in the former. The Sanchez–Dunning experiment thus compellingly raises the question of why humans have evolved to underinfer from freely available environmental feedback, even when meaningfully learning from it is made easy and monetarily advantageous.

How did our evolutionary past select for cognitive biases, traits that systematically cause errors in judgement? To solve this puzzle, we appeal to cultural evolutionary theory’s extensive body of evidence that humans primarily rely on learning

from their fellow group members, rather than from the environmental feedback itself^{26–28,38,129}. This evidence informs and is informed by a central hypothesis of cultural evolutionary theory: that adaptive, socially exchanged, and intergenerationally accumulated knowledge—relevant to fitness-relevant tasks like foraging, reproduction, and warfare—comprised the primary selection pressure acting on ancestral human cognition^{10,90,103,126,160,163,193,223,237,244}.

In this paper, we construct an evolutionary model of human learning based on this cultural-evolutionary-theoretic hypothesis: one in which an ancestral human learns primarily via knowledge learned from group members, rather than via environmental feedback. The model is constructed by modifying a classical Bayesian model of repeated task-learning to veridically represent the hypothesized setting of social, knowledge-based task-learning. Another key modification we add is our assumption that the cognitively constrained agent of our model—representing an ancestral human learner—faces selection pressures against meaningful retention of high-variance environmental feedback, due to onerous ecological fitness costs of overcommitting attention (e.g., increased risks from ambushes and accidental injury caused by a lack of situational awareness). It follows from this assumption that the confidence function comprising the agent’s fitness-maximizing strategy is characterized by discrete confidence levels and systematic deviations from classical Bayesian inference (i.e., from unbiased incorporation of environmental feedback), consistent with the empirical finding of Lisi et al.¹³⁴. Specifically, this confidence function is characterized by various cognitive biases like underinference, the hard-easy effect, and—depending on the parameters of our model—recurrent non-monotonicity.

We begin by describing in Subsection 1.2.1 a finite-outcome-space version of the classical Bayesian decision-theoretic model. This general model serves both as an in-

spiration for our evolutionary model and as a *reductio ad absurdum* argument that humans may not learn from high-variance environmental feedback via classical Bayesian inference. The contradiction is as follows. Classical Bayesian inference is effective because a Bayesian-updating prior (that has not a priori ruled out any possibility) is almost surely guaranteed to eventually converge to the truth: the property of consistency. However, this property is in contradiction with the aforementioned findings from the cognitive biases literature: first, that a human learner’s prior (such as that of their ability) can persistently deviate from their past observations; and second, that it can be recurrently non-monotonic with respect to the number of observations, regardless of the actual observations themselves.

We then resolve these empirical contradictions by presenting in Subsection 1.2.2 our evolutionary model: a modification of the classical Bayesian model, adapted to represent the knowledge-based learning environment of ancestral humans in the context of high-variance payoff observations. In our modified Bayesian model, the agent learns a task over repeated attempts, each of which generates a payoff. When the expected cost of retaining high-variance payoff observations—due to onerous ecological fitness costs from overcommitting attention—is sufficiently high, the agent’s optimal learning strategy does not update their prior of their payoff-acquisition ability in the given task (confidence) with respect to the payoff observations. Instead, the agent updates their confidence as a function of information in the complement of payoff observations: in our model, knowledge and the speed of learning. The consequent unavailability of payoff data—the key departure from classical Bayesian decision theory—generates our first desired conclusion: that evolved confidence generically deviates from the past payoff observations in a recurrent manner. This conclusion is a special case of a more general phenomenon: a given learning strategy’s systematic departure

from classical Bayesian updating when the ancestral learning environment for which it is ecologically rational differs from the contemporary learning environment in which it actually operates^{71,72,150}. Persistent underinference and the hard-easy effect follow from the recurrent nature of this evolutionary optimal confidence function.

The second desired conclusion—that this recurrent, evolutionarily optimal confidence function can be non-monotonic—follows from incorporating the cultural-evolutionary-theoretic hypothesis that the agent’s learning occurs via attempted imitation of a role model. This non-monotonicity can occur due to a dichotomy between successful imitation learning and de facto innovation learning: two learning methods whose classification is a priori uncertain to the agent.

The details of this dichotomy and of other aspects of our model are presented in Section 1.2. The predictions of this model are then made mathematical precise in the theorem statements presented in Section 1.3. The proofs of the theorems can be found in the Appendix.

We thus find that several classes of cognitive biases can be parsimoniously explained as evolutionary byproducts of the idiosyncratically knowledge-based and social nature of ancestral humans’ hypothesized learning environment. Often thought of as structural flaws in humans’ individual learning, cognitive biases may instead be evolutionarily rooted in two hypothesized characteristics of our ancestral environment: first, the primarily knowledge-based and social—not individual—nature of human learning in natural settings, as theorized by cultural evolutionary theory; and second, ecological fitness costs of meaningfully retaining environmental feedback—due to cognitive constraints—and the consequent pressure to rely instead on setting-specific sources of information, as theorized by the ecological rationality hypothesis^{71,72}.

1.2 The model

1.2.1 Classical Bayesian model

Suppose that an agent repeatedly attempts a task. Each yields a random payoff that is contained in a finite set of values $S \subset \mathbb{R}$. The finiteness of S constitutes the realistic assumption that the agent, due to cognitive constraints, categorizes observations into finitely many bins. The payoff from each task attempt is drawn i.i.d. from a fixed probability distribution $\varphi \in \Phi \subseteq \mathcal{P}(S)$, which would depend on the agent’s ability to acquire payoffs, the abundance of the environment, and various other factors. Here, $\mathcal{P}(S)$ denotes the set (which can be thought of as a state space) of all probability distributions on S , and $\Phi \subseteq \mathcal{P}(S)$ denotes the subset of probability distributions that may feasibly occur in a given setting.

For the purpose of maximizing payoff, the agent is incentivized to accurately predict the expected value of the future task attempt’s payoff. This was likely the case for ancestral human foragers, who by default engaged repeatedly in a highly specialized foraging role¹⁰², but also faced incentives to be opportunistic: to accurately appraise—and based on the result of said appraisal, possibly procure—additional foraging opportunities as they arise¹⁸. We model this dichotomy as follows. We assume that before each task attempt, the agent has the choice of forgoing a fraction r of the time spent on it (corresponding to the same fraction of the task attempt’s entire payoff) for a payoff whose value is observed beforehand. The opportunity-cost payoff is κ , where κ drawn from a fixed distribution $\psi \in \mathcal{P}(S)$ whose support is all of S . It follows that the agent maximizes the immediate payoff by taking the payoff from the task attempt if its mean $r\mathbb{E}[\varphi]$ is greater than κ , take the opportunity cost if $r\mathbb{E}[\varphi]$ is

less than rc , and take either option when $r\mathbb{E}[\varphi]$ is equal to rc .

The agent thus benefits from accurately estimating the task attempt’s expected payoff $\mathbb{E}[\varphi]$. This can likely be achieved by a small number of observations—even just one—when φ has low variance. Under our assumption that payoffs are observationally categorized by the agent into finitely many bins, assuming further that the payoffs have low variance amounts to the condition that nearly all payoffs (i.e., close to probability one) fall in a single bin $s \in S$. Consequently, the agent can productively use causal inference, in the sense that assuming every future task attempt will yield the previously observed payoff of s will nearly always be correct. The payoff-maximizing strategy is to choose the higher value between the task attempt’s expected payoff $r\mathbb{E}[\varphi] \approx rs$; and the observed opportunity cost rc .

The discernment of the payoff distribution φ —and more specifically, its expected value $\mathbb{E}[\varphi]$ —is more difficult when φ has high variance. In this domain, more than one bin in S occurs with significant probability. Consequently, the agent will in general need to learn from a large sample size of payoffs in order to asymptotically determine the true state φ from the set of a priori possible states Φ .

Suppose that the true state φ is initially drawn from a probability distribution $\xi \in \mathcal{P}(\Phi)$. Then, Bayes’ theorem states that the probability distribution of φ conditional on the previous payoff observations being s_1, s_2, \dots, s_n is given by

$$\xi_{s_1, \dots, s_n} = \mathcal{B}_{s_n} \circ \dots \circ \mathcal{B}_{s_2} \circ \mathcal{B}_{s_1}(\xi), \tag{1.1}$$

where $\mathcal{B}_x : \mathcal{P}(\Phi) \rightarrow \mathcal{P}(\Phi)$ is the Bayes'-rule map

$$\mathcal{B}_x(\omega)(\theta) = \frac{\theta(x)\omega(\theta)}{\int_{\hat{\theta} \in \Phi} \hat{\theta}(x)\omega(\hat{\theta})d\hat{\theta}}. \quad (1.2)$$

Consequently, the payoff-maximizing choice of whether to forgo part of the task-attempt payoff is to compare its expected value

$$r \int_{\varphi \in \Phi} \mathbb{E}[\varphi] d\xi_{s_1, \dots, s_n}(\varphi) \quad (1.3)$$

with the observed opportunity cost rc . In summary, the agent's evolutionarily optimal strategy overall is to begin with the prior ξ , update it via the Bayes' rule map \mathcal{B} , in terms of each task attempt's observed payoff s , and decide whether to forgo part of the n th task attempt for an observed opportunity cost by using the prior $\xi_{s_1, \dots, s_{n-1}}$ at that point in time.

Bayesian inference can be effective even without explicit knowledge of the true distribution ξ from which the state φ is drawn. An obvious obstruction to this effectiveness is Cromwell's rule: if a state is not contained in the support of the prior ω , then this will persist in ω_{s_1, \dots, s_n} for any sequence of observations s_1, \dots, s_n . It turns out that Cromwell's rule is the only such obstruction when the outcome space S is finite. Specifically, suppose that the true state φ is contained in the support of the prior ω . Then, as $n \rightarrow \infty$, the n th Bayesian update of ω

$$\omega_n = \omega_{s_1, \dots, s_n} \quad (1.4)$$

will converge to the one-point distribution

$$\chi_\varphi(\theta) = \begin{cases} 1 & \text{if } \theta = \varphi \\ 0 & \text{otherwise} \end{cases} \quad (1.5)$$

with prior probability one: the property of consistency^{54,62}. In other words, even an agent with a misspecified initial prior—for example, one that evolved in a past environment with a different distribution of φ —will in all likelihood eventually converge to the true state φ , as long as the initial prior is not too restrictive.

The property of consistency yields a practical test to reject the null hypothesis that a given learner is Bayesian in the classical sense. We can do so if the learner’s prior does not converge to the (one-point distribution on the) true state as the number of observations goes to infinity. A special case of this test is provided by checking whether a learner’s estimate of their expected payoff-acquisition ability converges to the true expected payoff. Indeed, suppose that the learner’s prior were updated via classical Bayesian inference while starting from an initial prior ω that has not ruled out the true state φ . Then, with prior probability one, the learner’s estimate of their expected payoff-acquisition ability

$$\int_{\hat{\theta} \in \Theta} \mathbb{E}[\hat{\theta}] d\omega_n(\hat{\theta}), \quad (1.6)$$

would converge to the true expected payoff

$$\mathbb{E}[\varphi] \quad (1.7)$$

as the number of observations n goes to infinity. While the true expected payoff (1.7)

is unobservable, it will with probability one coincide with the mean of the past payoff data

$$\frac{s_1 + \cdots + s_n}{n} \tag{1.8}$$

as $n \rightarrow \infty$, due to the law of large numbers. We should thus be skeptical of a learner's Bayesianness if their estimate (1.6) of their expected payoff-acquisition ability does not appear to converge to the mean of the past payoff data (1.8). Note that this practical test for falsifying a learner's Bayesianness is not new; it is essentially a corollary of standard Bayesian statistics.

To illustrate, consider a gambler who, over repeated attempts, continues to be mistaken about the expected value of a fixed probabilistic lottery. They may persistently believe that the expected payoff from betting their money on a negative-expected-value lottery is positive, even after gambling on it a large number of times while observing the resulting payoff data. Then, we can be reasonably certain that the gambler is not, in the classical sense, Bayesian-updating with respect to their payoff data. We hypothesize that the persistent deviation of the gambler's prior from the true state is caused by the high variance of the payoff data. Other learners who may fail our test for classical Bayesianness include professionals whose priors of their performance persistently deviate from the true value^{100,183}, traders and managers who persistently overestimate future returns on their investments^{11,140}, and gymgoers who repeatedly overpay on membership fees based on persistently overoptimistic priors of their attendance rate⁴⁸. Such field evidence against the hypothesis that human learning from high-variance payoff data is classically Bayesian corroborates the extensive lab evidence of the relevant cognitive biases.

1.2.2 Evolutionary model of ancestral human learning

To resolve the predictive inadequacies of the classical Bayesian paradigm, we modify it in the following way. We assume that the agent estimates their payoff-acquisition ability as a function of task-specific knowledge, and not necessarily of the previously observed payoff data. Our evolutionary model incorporates two veridical sources of uncertainty which are sufficient to generate recurrent non-monotonicity. First, tasks vary in difficulty, a value that represents the total amount of knowledge required to completely learn the task. The agent’s marginal payoff is a bivariate function of the difficulty value and their current level of knowledge: the subset of the total knowledge they have learned so far.

Second, tasks vary in the method used to learn the relevant knowledge: imitation and innovation. We incorporate into our model the cultural-evolutionary-theoretic finding that the primary source of an ancestral human’s task-specific knowledge was learning from role models who were ostensibly proficient in the task—imitation—rather than learning individually from environmental feedback—innovation^{26,38}. The superior efficiency of imitation learning, especially in the context of intergenerational knowledge accumulation, is hypothesized to have enabled humans’ unprecedented evolutionary success.

The dichotomy between imitation learning and innovation learning is confusing at first glance, given that in our model, the student always attempts to imitate a role model. This dichotomy occurs because the helpfulness of role models in providing a genuinely new path forward via imitation learning is not guaranteed. A student may successfully learn via imitation of their role model, as planned. It is also possible that the role model’s ostensible proficiency in the task does not translate to productive im-

imitation learning, in which case the student learns by de facto innovation. Specifically, the role model may not actually be providing a new learning path that the student would not have accessed if they were to instead learn by innovation. In the context of direct teaching, for instance, this may be due either to the method of teaching (a teacher may use an open-ended or ambiguous teaching method, such as the Socratic method, without actually guiding students to think in a new way) or to the teacher’s own limitations (which may not be discernible to students when their environmental feedback has high variance). It would be difficult for the student to deduce from high-variance environmental feedback whether their role model is meaningfully providing them with a new learning path to imitate.

Throughout this paper, the term “task” will denote a student’s package comprised of a repeated knowledge-intensive task that produces fitness-aiding payoffs (i.e., foraging for food), their choice of role model for it, and the learning method by which the student obtains the relevant knowledge: classified into imitation learning and innovation learning. The student’s task package can be thought of as a pair (j, a) for the type of learning method $j \in \{im, in\}$ with which the student learns the task from the teacher (where $j = im$ denotes imitation and $j = in$ denotes innovation) and the difficulty value $a \in (0, \infty) \cup \{\infty\}$ of the task.

The difficulty value $a \in (0, \infty) \cup \{\infty\}$ of a task denotes the amount a of knowledge the student needs to completely learn it, given the specifics of the task package (the teacher, the learning method, and the task itself). A task with the difficulty value $a = \infty$ represents an impossible one, in that the specifics of the task prevent the student from learning it to completion. Suppose the student currently knows $b \leq a$ of the total amount of knowledge required to completely learn the task. The values of b and a determine the marginal payoff $f(a, b)$, which we assume is strictly increasing

in b , strictly decreasing in a , and continuously differentiable. By scaling the marginal payoff values to have minimum 0 and maximum 1, we can suppose that the function $f(a, \cdot)$ maps the domain $[0, a]$ to the range $[0, 1]$. We assume that completely learning a task guarantees the maximum marginal payoff: $f(a, a) = 1$ for every a . Moreover, we assume that impossible tasks—unable to be meaningfully learned—always yield the minimum marginal payoff: $f(\infty, b) = 0$ for all b .

One example of a marginal payoff function

$$f: \{(a, b) \in ((0, \infty) \cup \{\infty\}) \times [0, \infty) : b \leq a\} \rightarrow [0, 1] \quad (1.9)$$

satisfying these conditions is

$$f(a, b) = \left(\frac{b}{a}\right)^\lambda \quad (1.10)$$

for $\lambda > 0$, which is extended to the point at infinity $a = \infty$ as

$$f(\infty, b) = \lim_{a \rightarrow \infty} \left(\frac{b}{a}\right)^\lambda = 0. \quad (1.11)$$

This family of functions is characterized by polynomial growth in b . Another example of such a marginal payoff function is

$$f(a, b) = \zeta^{a-b} \quad (1.12)$$

for $\zeta \in (0, 1)$, which is also extended to the point at infinity $a = \infty$ as

$$f(\infty, b) = \lim_{a \rightarrow \infty} \zeta^{a-b} = 0. \quad (1.13)$$

This family of functions is characterized by exponential growth in b .

We assume that the risk of an infinitely difficult task $a = \infty$ only exists when $j = in$. In the other case of $j = im$, the learnability of the given task is guaranteed by the teacher already having learned it completely. However, when $j = in$, the teacher may not have actually learned the task completely despite serving as the student's role model. The lack of guarantee of the given task's learnability leads to a nontrivial probability of an unfortunate setting: one in which the student squanders time on attempting to learn an impossible task from a teacher, one or both of whom have not yet realized the said impossibility. The exclusivity of unlearnability to innovation learning can be seen by the comparison between solving an exam problem and solving a research problem. The former—imitation learning—is guaranteed to complete in finite time, because the teacher has solved the problem before assigning it as an exam question. However, the latter—innovation learning—is not guaranteed to complete in finite time. Indeed, a research problem, by definition, is one that has not yet been solved by anyone, so it may a priori be impossible to solve. Overall, we assume that the difficulty values of tasks with learning method $j = im$ are distributed as a regular exponential distribution (i.e., with p.d.f. $\mu_{im}(a) = \eta^a \log \frac{1}{\eta}$ for finite a and $\mu_{im}(\infty) = 0$, where $0 < \eta < 1$), whereas the distribution of difficulty values of tasks with learning method $j = in$ is assumed instead to have a positive probability p on $a = \infty$ (i.e., with p.d.f. $\mu_{in}(a) = (1 - p)\eta^a \log \frac{1}{\eta}$ for finite a and $\mu_{in}(\infty) = p$). The overall distribution of tasks (j, a) on

$$\mathcal{U} = \{im, in\} \times ((0, \infty) \cup \{\infty\}), \quad (1.14)$$

defined by the p.d.f.

$$\mu(j, a) = \begin{cases} q\mu_{im}(a), & \text{if } j = im, \\ (1 - q)\mu_{in}(a) & \text{if } j = in; \end{cases} \quad (1.15)$$

places probability q on the task's learning type being imitation and $1 - q$ on that being innovation.

Other than the risk of unlearnability, the second way in which tasks of learning method $j = im$ differ from those of learning method $j = in$ is in the speed of learning. Regardless of the learning method, the student learns knowledge in discrete jumps, each following a task attempt. Let $B(t)$ denote the knowledge level after the t th task attempt, where $B(0) = 0$, meaning that the initially naive student has knowledge $b = B(0) = 0$ of the task when starting out. The discrete knowledge levels $0 = B(0) < B(1) < \dots$ are assumed to satisfy $\lim_{t \rightarrow \infty} B(t) = \infty$. The amount of time the t th task attempt takes for the student is assumed to differ between the two learning types. Let $\Delta_{im}(t)$ (respectively, $\Delta_{in}(t)$) denote the amount of time the t th task attempt takes when engaged in imitation learning (respectively, innovation learning); we require for both $j \in \{im, in\}$ that $\lim_{k \rightarrow \infty} \sum_{t=1}^k \Delta_j(t) = \infty$. Then, we assume that imitation is (weakly) faster than innovation: that $\Delta_{im}(t) \leq \Delta_{in}(t)$ for all $t \in \mathbb{N} \setminus \{0\}$. Moreover, we denote by

$$T_j(i) = \sum_{n=1}^i \Delta_j(n) \tag{1.16}$$

the total amount of time that a task of learning type j occupies until the end of the i th attempt.

With sufficient time in a fixed environment, natural selection is likely to maximize the objective function (fitness) within the space of feasible policies (fitness landscape). A policy is defined by a function $\pi : \mathcal{H} \rightarrow \mathcal{A}$, where \mathcal{A} denotes the space of feasible actions;

$$\mathcal{H} = \{(O_1, A_1, \dots, O_{T-1}, A_{T-1}, O_T) : O_i \in \mathcal{O}, A_i \in \mathcal{A}, \text{ and the history is feasible}\}, \tag{1.17}$$

the space of feasible histories; \mathcal{O} , the space of feasible observations; and a history

$$b = (O_1, A_1, \dots, O_{T-1}, A_{T-1}, O_T) \quad (1.18)$$

is called feasible if its sequence of observations and actions can occur in the model. It remains to specify the student's action space \mathcal{A} , observation space \mathcal{O} , and the objective function $V(\pi)$ on the space of policies π .

The student's objective function $V(\pi)$ is the expectation of the total payoff. Most of it comes from the payoffs yielded by the student's task attempts. Suppose that the student finishes a task attempt of time length Δ while at level of knowledge b for a task of difficulty value a . At time T that ends a learning period, the student obtains an expected payoff proportional to $f(a, b)$, scaling with the length Δ of the learning period, and simultaneously accounting for exponential time-discounting. The marginal payoff is obtained as a high-variance probabilistic lottery $\phi(a, b) \in \mathcal{P}(S)$ with expected value $\mathbb{E}[\phi(a, b)] = f(a, b)$. Specifically, a payoff value \bar{s} is drawn independently from $\phi(a, b)$ to determine the payoff of the task attempt

$$v(a, b, \Delta, T) = \bar{s} \int_{T-\Delta}^T \delta^t dt, \quad (1.19)$$

where $\delta \in (0, 1)$ denotes the factor of exponential time-discounting. We see that the expected payoff yielded by the task attempt is

$$\mathbb{E}[v(a, b, \Delta, T)] = f(a, b) \int_{T-\Delta}^T \delta^t dt = \begin{cases} f(a, B(i)) \int_{T-\Delta}^T \delta^t dt & \text{if } b = B(i) < a, \\ \int_{T-\Delta}^T \delta^t dt & \text{if } b = a, \end{cases} \quad (1.20)$$

Instantaneously after the acquisition of this payoff at time T , the student's level of

knowledge jumps to the next discrete level of knowledge $B(\cdot)$ or to the maximum level of knowledge a for the task, whichever is smaller. The expected sum of the student's task-attempt payoffs over all time $T \in [0, \infty)$ is the main component of the student's objective function $V(\pi)$.

There are three auxiliary components of the student's objective function $V(\pi)$. The first such component is as follows. After obtaining the payoff of expected value $v(a, b, \Delta, T)$, the student has the option of committing the observed payoff value to memory. Doing so requires the student to pay an expected cost of $-C_{\text{retain}}$, which represents various ecological fitness risks that result from overcommitting attention to the retention of high-variance payoff data. Due to the exponential time-discounting, the true value of the expected cost as applied to the student's objective function $V(\pi)$ is

$$-\delta^T C_{\text{retain}}, \tag{1.21}$$

where T denotes the ending time of the task attempt that has yielded the given payoff.

The second auxiliary component of the student's objective function $V(\pi)$ relates to a choice (described in Subsection 1.2.1) that the student makes before every task attempt: whether to allocate a fraction r of the task attempt's time—and the corresponding fraction of its payoff—to an alternative foraging opportunity unrelated to the task. Like in the classical Bayesian model of Subsection 1.2.1, the marginal payoff $s \in S$ of the alternative foraging opportunity is drawn i.i.d. from a distribution $\psi \in \mathcal{P}(S)$ and known to the student prior to their decision. If the student chooses to forgo a fraction of the task attempt's time for this alternative foraging opportunity,

their payoff is changed from (1.19) to

$$rs \int_{T-\Delta}^T \delta^t dt + (1-r)v(a, b, \Delta, T) = (rs + (1-r)\bar{s}) \int_{T-\Delta}^T \delta^t dt. \quad (1.22)$$

These unrelated foraging opportunities allow the student to increase their expected payoff $V(\pi)$ strictly above the baseline level provided by the sum of the task-attempt payoffs $v(a, b, \Delta, T)$. Consequently, the student is incentivized to accurately estimate each task-attempt's payoff—as best as allowed by their informational constraints—prior to deciding whether to exploit an unrelated foraging opportunity instead.

The third auxiliary component of the student's objective function $V(\pi)$ relates to the student's other choice of action. In between task attempts, the student not only chooses whether to exploit an unrelated foraging opportunity before each task attempt, but also chooses whether to quit on their current task package for an alternative one. If the student chooses to cut their losses on a given foraging task and/or their role model for it, they can choose a new task package (j, a) . All of the student's task packages (j, a) , including the initial one and any intermediate ones assigned after quitting, are drawn i.i.d. from the probability distribution μ defined in (1.15).

In addition to the option of quitting the current task, the student is also assumed to situationally possess the option of paying a fitness cost to ascertain their current task package's learning method $j \in \{im, in\}$, on which they can base their specific decision. We propose that humans carry out this ascertainment via a mental experiment to measure the length of time $\Delta_j(t)$, which may be sufficient to distinguish the speeds of the two learning methods. Specifically, our assumption that $\Delta_{im}(t) \leq \Delta_{in}(t)$ can be divided into two possibilities: $\Delta_{im}(t) < \Delta_{in}(t)$ and $\Delta_{im}(t) = \Delta_{in}(t)$. In the case of the former, a time-measurement experiment can identify the learning type j . In the case

of the latter, however, it cannot. Each mental time-measurement experiment requires the student to pay an expected cost $-C_{\text{identify}}$, again due to various ecological fitness costs that can result from overloading a cognitively constrained forager’s decision-making. Due to the exponential time-discounting, the true value of the expected cost as applied to the student’s objective function $V(\pi)$ is

$$-\delta^T C_{\text{identify}}, \tag{1.23}$$

where T denotes the ending time of the task attempt during which the time-measurement experiment was performed.

We have introduced all components of the student’s objective function $V(\pi)$, as well as all components of the student’s action space \mathcal{A} . Unlike the classical Bayesian model of Subsection 1.2.1, our model is characterized by a potential tradeoff between earlier and later payoffs. In the classical Bayesian model, each of the agent’s actions was only relevant to maximizing the payoff of the corresponding task attempt, not to any future ones. Thus, the relative weights of each task attempt’s payoff do not affect the agent’s decision problem. In contrast, in our model, the student has two actions—quitting the current task and identifying the learning type via a time-measurement experiment—that reduces payoffs in the short-term for a potential gain in long-term payoffs. Thus, specifying the relative weights of each task attempt’s payoff is essential for the prescription of the optimal policy π . As is standard, we have set these relative weights to be exponentially decaying in time, which aids model tractability and captures the evolutionary fact that earlier payoffs are likelier to be relevant to fitness than later payoffs.

Formally, the student's actions are of the form

$$A_t = (x_{\text{forgo}}(t), x_{\text{identify}}(t), x_{\text{retain}}(t), x_{\text{quit}}(t)), \quad (1.24)$$

where

$$x_{\text{forgo}}(t) : S \rightarrow \{\text{true}, \text{false}\} \quad (1.25)$$

denotes the choice of whether to forgo a fraction of the t th task attempt's time to exploiting an alternative foraging opportunity of a known marginal payoff $s \in S$;

$$x_{\text{identify}}(t) : S \rightarrow \{\text{true}, \text{false}\} \quad (1.26)$$

denotes the choice of whether to pay an expected cost of $-C_{\text{identify}}$ to identify the learning type $j \in \{\text{im}, \text{in}\}$ during the t th task attempt via a time-measurement experiment, given the alternating foraging opportunity's previously drawn marginal payoff s ;

$$x_{\text{retain}}(t) : S \times S \rightarrow \{\text{true}, \text{false}\} \quad (1.27)$$

denotes the choice of whether to retain the observation of the t th task attempt's payoff given $(s, \bar{s}) \in S \times S$, where s is given as above and \bar{s} denotes the task-specific marginal payoff; and

$$x_{\text{quit}}(t) \in \{\mathcal{K}(s, \bar{s}, j, c) : S \times (S \cup \{\text{null}\}) \times \{\text{im}, \text{in}, \text{null}\} \times \{\text{true}, \text{false}\} \rightarrow \{\text{true}, \text{false}\}\}$$

denotes the choice of whether to quit the current task after the t th task attempt.

When the student has not performed the identification of the learning type j during the current task attempt, $x_{\text{identify}}(t) = \text{false}$, then the value $x_{\text{quit}}(t)$ takes the form

of a boolean-valued function $\mathcal{K}(s, \bar{s}, \text{null}, c)$: a function of the alternative foraging opportunity's marginal payoff s ; of the task's yielded marginal payoff \bar{s} (which may be unretained and thus given by $\bar{s} = \text{null}$); and whether or not the level of knowledge has caught up to the task difficulty a , denoted by

$$c \in \{\text{true}, \text{false}\}. \quad (1.28)$$

If $c = \text{true}$, then we say that learning has completed during this task attempt. In the opposite case of $x_{\text{identify}}(t) = \text{true}$, $x_{\text{quit}}(t)$ takes the form of a boolean-valued function $\mathcal{K}(s, \bar{s}, j, c)$ for $j \in \{\text{im}, \text{in}\}$, representing the decision whether to quit conditional on the identified learning type being imitation or innovation, on the payoff observation, and on whether learning has completed during this task attempt. We also note the feasibility constraint that the value $x_{\text{identify}}(t)$ is required to satisfy the feasibility constraint that $x_{\text{identify}}(t) = \text{true}$ is only possible if $\Delta_{\text{im}}(t) < \Delta_{\text{in}}(t)$ rather than $\Delta_{\text{im}}(t) = \Delta_{\text{in}}(t)$.

The student's observations are of the form

$$O_t = (b(t), x_{\text{type}}(t), x_{\text{payoff}}(t)), \quad (1.29)$$

where

$$b(t) \in [0, \infty) \quad (1.30)$$

denotes the level of knowledge after the t th task attempt;

$$x_{\text{payoff}}(t) \in S \cup \{\text{null}\} \quad (1.31)$$

denotes the student’s observed payoff value (if the payoff observation was not retained, then we use the denotation “null”); and

$$x_{\text{type}}(t) \in \{\text{null}, \text{im}, \text{in}\} \quad (1.32)$$

denotes whether the student has carried out a mental identification of the learning type during the t th task attempt (if this is false, then we use the denotation “null”), and if so, whether the result was imitation (“im”) or innovation (“in”).

In summary, Table 1.1 provides the list of parameters comprising our learning model, and Table 1.2 presents a step-by-step algorithm for the model. The expected payoff of the policy π (correcting for time-discounting) during the time remaining after a history b is given by

$$V_b(\pi) = \mathbb{E} \left[\sum_{k=0}^{\infty} \left(\left(r x_{\text{forgo}}(k) s(k) + (1 - r x_{\text{forgo}}(k)) \bar{s}(k) \right) \int_{T(k)}^{T(k+1)} \delta^t dt - \delta^{T(k+1)} (C_{\text{retain}} x_{\text{retain}}(k) + C_{\text{identify}} x_{\text{identify}}(k)) \right) \right], \quad (1.33)$$

where we have abused notation by having $\bar{s}(k)$, $s(k)$, and the choices $x_{\square}(k)$ denote the values of \bar{s} , s , and the choices x_{\square} during the k th learning period from the present, letting $T(k)$ denote the ending time of the k th learning period from the present, and setting the boolean values of the choices $x_{\square}(k)$ to be 0 when false and 1 when true.

Given a choice of parameters, the corresponding model parametrization \mathbf{M} can be solved numerically with dynamic-programming-type methods. However, we instead pursue an analytic study to demonstrate desired facts about the model that hold more generally, regardless of the specific choice of parameters. The results of this

Model parameters of the modified Bayesian model

1. the marginal payoff distribution $\phi(a, b) \in \mathcal{P}(S)$ and its expected value $f(a, b)$, for every $a \in (0, \infty) \cup \{\infty\}$ and finite $0 \leq b \leq a$,
 2. the discrete knowledge jumps $\{B(i) : i \in \mathbb{N}, i > 0\}$,
 3. the learning period lengths $\{\Delta_j(i) : i \in \mathbb{N}, i > 0\}$ for $j \in \{im, in\}$,
 4. the exponential discount factor δ of time,
 5. the proportion p of infinite-difficulty tasks among all innovation-learning tasks,
 6. the proportion q of imitation-learning tasks among all tasks,
 7. the exponential discount factor η of the distribution of task difficulty values,
 8. the fraction of time r of task attempts that can be devoted to alternative foraging opportunities,
 9. the distribution $\psi \in \mathcal{P}(S)$ of the marginal payoffs of alternative foraging opportunities,
 10. the expected cost $-C_{\text{retain}}$ of retaining a payoff observation, and
 11. the expected cost $-C_{\text{identify}}$ of a mental time-measurement experiment to identify the learning type j .
-

Table 1.1. List of model parameters of the Bayesian model modified to represent ancestral human learning, presented in Subsection 1.2.2. These model parameters are required to satisfy the conditions discussed earlier in this subsection.

Algorithmic description of the modified Bayesian model

1. The student draws from the distribution μ the task (j, a) , the value of which is unknown to them. The attempt number specific to the task, i , is set to zero, and their level of knowledge b is set to zero. The time value T is set to zero.
2. The student carries out the i th attempt of the current task, which constitutes the following.
 - First, the student draws from the distribution ψ a random marginal payoff $s \in S$ whose value is known to them, and decide whether to forgo a fraction r of the task attempt for this alternative marginal payoff.
 - Second, the student decide whether to pay an expected cost $-C_{\text{identify}}$ for a time-measurement experiment to identify $\Delta_j(i)$, which is only possible if $\Delta_{im}(i) < \Delta_{in}(i)$ rather than $\Delta_{im}(i) = \Delta_{in}(i)$.
 - Third, they spend the time $\Delta_j(i)$ on the task attempt (T is incremented by this amount), at the end of which they receive a payoff of

$$\begin{cases} \delta^T (rs + (1-r)\bar{s}) \int_0^{\Delta_j(i)} \delta^t dt & \text{if the student had decided to forgo,} \\ \delta^{T+\bar{s}} \int_0^{\Delta_j(i)} \delta^t dt & \text{otherwise,} \end{cases} \quad (1.34)$$

where $\bar{s} \in S$ is drawn from the distribution $\phi(a, b)$. The student chooses whether to retain the observation \bar{s} of the payoff value.

- Fourth, if the student had performed a time-measurement experiment during this learning attempt, then they learn the value $\Delta_j(i)$ and thereby, the learning type j .
 - Fifth, b discretely jumps to the next level— $B(i + 1)$ or a , whichever is smaller—and the index i is incremented by one.
 - Finally, the student chooses whether to quit the current task. If so, they draw a new task (j, a) from μ (independently with respect to the previously drawn tasks), b is set to zero, and i is set to zero. Otherwise, they continue to learn the task attempt at the new level of experience $i + 1$.
3. Step 2 is infinitely repeated.
-

Table 1.2. An algorithmic description of the Bayesian model modified to represent ancestral human learning, presented in Subsection 1.2.2.

investigation are documented in Section 1.3.

1.3 Results

We denote the space of feasible policies of the model described in Subsection 1.2.2 by Π . A policy π is called optimal if it maximizes the expected payoff in the remaining time at any feasible history b :

$$\pi \in \arg \max_{\pi \in \Pi} V_b(\pi). \quad (1.35)$$

In the following, we obtain results on properties necessarily possessed by any optimal policy π , which can help simultaneously explain the various empirically documented deviations of human confidence from a classically Bayesian estimate of past payoff data.

First, if the magnitude C_{retain} of the expected cost of retaining payoff observations is sufficiently large, then no optimal policy π ever retains payoff observations. This can be seen, for example, by taking

$$C_{\text{retain}} > \int_0^\infty \delta^t \max(S) dt, \quad (1.36)$$

an upper bound—for any time T at which a task attempt ends—to the payoff (accounting for time-discounting) that can be obtained during the remaining time. The upper bound (1.36) is obtained when the student receives the maximal marginal payoff $\max(S)$ for every task, and does not pay any cost to retaining payoff observations or identifying the task’s learning type. If C_{retain} were larger than this maximum possible expected payoff in the remaining time, then the information yielded by paying a

cost of that magnitude would clearly never be worth it.

Throughout this paper, we assume that the magnitude C_{retain} of the expected cost of “observing” (in the ecological setting, retaining in memory) payoff data is great enough that the student does not ever do so: so that the optimal choice $x_{\text{retain}}(t)$ is always given by

$$x_{\text{retain}}(t) = \text{false}. \quad (1.37)$$

This is functionally equivalent to assuming that the payoff data is unavailable to the student.

The second characteristic that an optimal policy π must possess is the following. Every action $\pi(b)$ of an optimal policy in response to a history b might as well solely depend on the information of b relevant to the current task (j, a) , and not on the other information (relevant to the previous tasks); this follows from the assumption that the student’s tasks are statistically independent. Specifically, the choices of $x_{\text{forgo}}(t)$, $x_{\text{identify}}(t)$, and $x_{\text{quit}}(t)$ should only depend on the conditional distribution $\mu_{\text{cond}}(b)$ of the current task’s value (j, a) , conditional on the information contained in the past history b . This information, which allows the student to rule out (via Bayes’ formula of conditional probability) certain task values (j, a) from the initial conditional distribution of μ , includes two components. For one thing, if there has been a time-measurement experiment on the current task, say with result $j \in \{im, in\}$, then the student can rule out all task values (j', a) with $j' \neq j$.

For another, the student’s past sequence of knowledge levels on the task, $b(0), b(1), \dots, b(i-1)$, allows the student to rule out task values. If the sequence ends in one or more instances of $b = a \notin \{B(i) : i \in \mathbb{N}\}$, then the student knows that their level of knowledge b has caught up to the maximum value a . In other words, all task values (j', a') with

$a' \neq a$ can be ruled out. However, if the sequence has been completely consistent with the discrete knowledge values $\{B(i) : i \in \mathbb{N}\}$ of the model, then the only task values (j', a') that can be ruled out are those with $a' \leq b = B(i)$. (Without loss of generality, we assume that the probability-zero event that the task difficulty a drawn from μ precisely equals one of the model's discrete knowledge levels $B(i)$, rather than falling between them, does not occur.)

Third, in an optimal policy π , every decision $x_{\text{forgo}}(t)$ whether to forgo a fraction of a task attempt's time for a known marginal payoff of $s \in S$ must be of the form described in Subsection 1.2.1: forgo if the task attempt's expected marginal payoff

$$\mathbb{E}_{(j,a) \rightsquigarrow \mu_{\text{cond}}(b)} [f(a, b)] \tag{1.38}$$

is greater than the alternative marginal payoff s , and do not forgo if the latter is greater than the former (when they are equal, both choices are optimal). In other words, the student should choose the payoff that is greater in expectation. We call the quantity (1.38) the expected marginal payoff function or the confidence function. We propose that the evolutionary pressure to optimally exploit alternative foraging opportunities shaped ancestral humans' task-specific notion of confidence to track the task's expected marginal payoff (1.38), conditional on both the information known so far and the parameters of the ancestral environment.

The task's expected marginal payoff (1.38) is a function of the student's two relevant pieces of information: their level of knowledge b and their information set on the learning type j (whether they have ruled out the event $\{j = im\}$, the event $\{j = in\}$, or neither). Specifically, the confidence function can be written as $\hat{g}(E_b, E_j)$ mapping the

domain

$$\begin{aligned} & (\{b = B(i) : i \in \mathbb{N}\} \cup \{b = a \neq B(i)\}) \\ & \times \{\{j = im \text{ ruled out}\}, \{j = in \text{ ruled out}\}, \{\text{neither } j \text{ ruled out}\}\} \ni (E_b, E_j) \end{aligned}$$

to the range of marginal payoffs $[0, 1]$, where E_b denotes the information set regarding the student's information on b and E_j , the information set regarding the student's information on j . We compute that the confidence function (1.38) is generally given by

$$\hat{g}(E_b, E_j) = \begin{cases} 1 & \text{if } E_b = \{b = a \neq B(i)\}, \\ g_{im}(B(i)) & \text{if } E_b = \{b = B(i) < a\} \text{ and } E_j = \{j = in \text{ ruled out}\}, \\ g_{in}(B(i)) & \text{if } E_b = \{b = B(i) < a\} \text{ and } E_j = \{j = im \text{ ruled out}\}, \\ g_u(B(i)) & \text{if } E_b = \{b = B(i) < a\} \text{ and } E_j = \{\text{neither } j \text{ ruled out}\}, \end{cases} \quad (1.39)$$

for

$$g_{im}(b) = \frac{\int_{a>b} f(a, b) d\mu_{im}(a)}{\int_{a>b} d\mu_{im}(a)}, \quad (1.40)$$

$$g_{in}(b) = \frac{\int_{a>b} f(a, b) d\mu_{in}(a)}{\int_{a>b} d\mu_{in}(a)}, \quad (1.41)$$

and

$$g_u(b) = \frac{\int_{a>b} f(a, b) d\bar{\mu}(a)}{\int_{a>b} d\bar{\mu}(a)}, \quad (1.42)$$

where $\bar{\mu}$ denotes the probability distribution $P \circ \mu : [0, \infty) \rightarrow [0, 1]$ for the projection map $P(j, a) = a$. We call g_{im} , g_{in} , and g_u the imitation-learning confidence function, the innovation-learning confidence function, and the unconditional confidence function,

respectively.

Let ρ_y be a distribution of the form $\rho_y(a) = (1 - y)\eta^a \log \frac{1}{y}$ for finite a and $\rho_y(\infty) = y$, where $y \in [0, 1)$. Define the generalized confidence function $g_{\rho_y} : [0, \infty) \rightarrow [0, 1]$ by

$$g_{\rho_y}(b) = \frac{\int_{a>b} f(a, b) d\rho_y(a)}{\int_{a>b} d\rho_y(a)}. \quad (1.43)$$

Then, we see that

$$\mu_{im} = \rho_0, \mu_{in} = \rho_p, \text{ and } \bar{\mu} = \rho_{(1-q)p}, \quad (1.44)$$

and therefore,

$$g_{im}(b) = g_{\rho_0}(b), g_{in}(b) = g_{\rho_p}(b), \text{ and } g_u(b) = g_{\rho_{(1-q)p}}(b). \quad (1.45)$$

One can then verify the following fact.

Proposition 1.1. For any $b > 0$, the value of the generalized confidence function, $g_{\rho_y}(b)$, is strictly monotonically decreasing in y . In particular, the innovation-learning confidence function $g_{in}(b)$ is at most the unconditional confidence function $g_u(b)$, which is at most the imitation-learning confidence function $g_{im}(b)$. Specifically, we have

$$g_{in}(b) \leq g_u(b) \leq g_{im}(b), \quad (1.46)$$

where the first inequality occurs with equality if and only if $q = 0$ (or $p = 0$, if this is allowed); and the second inequality, if and only if $q = 1$ (or $p = 0$, if this is allowed).

In other words, the evolutionarily optimal estimate of confidence at a level of knowledge b (conditional on learning not yet having completed) is decreasing in the proportion y of unlearnable tasks. This is due to the fact that the risk of unlearnability, of

the task difficulty $a = \infty$, has a reduction effect on the expected marginal payoff. This risk occurs with the highest probability within the distribution of task difficulties $a > b$ conditional on $j = in$, occurs with zero probability within the distribution conditional on $j = im$, and occurs with an in-between probability value within the distribution that is unconditional of the learning type j . Thus, the reduction effect on the confidence function also falls in this order. This phenomenon is illustrated in the plots of the three confidence functions for several model parametrizations in Figure 1.1.

Another consequence of the risk of unlearnability is non-monotonicity. Specifically, we will show that $g_{im}(b)$ is monotonically increasing in b under a non-restrictive assumption on the marginal payoff function $f(a, b)$. Note that if all tasks were learned by imitation rather than innovation ($q = 1$), then the confidence function (1.38) is of the form

$$\hat{g}(E_b, E_j) = \begin{cases} 1 & \text{if } E_b = \{b = a \neq B(i)\}, \\ g_{im}(B(i)) & \text{if } E_b = \{b = B(i)\}. \end{cases} \quad (1.47)$$

and consequently, monotonically increasing in the level of experience i . In other words, if human confidence evolved in an environment where all tasks were learned by imitation, then we should expect it to be monotonically increasing in the level of knowledge: and thereby, the level of experience. The empirically documented confidence is non-monotonic in the level of experience, and thus unlikely to have evolved in such an environment.

On the other hand, we will show that due to the nontrivial risk of unlearnability, the confidence functions $g_{in}(b)$ and $g_u(b)$ each decay to zero as $b \rightarrow \infty$. This opens up the possibility for the confidence function (1.38) to be non-monotonic in the empirically documented way: general increase with an intermediate period of decrease

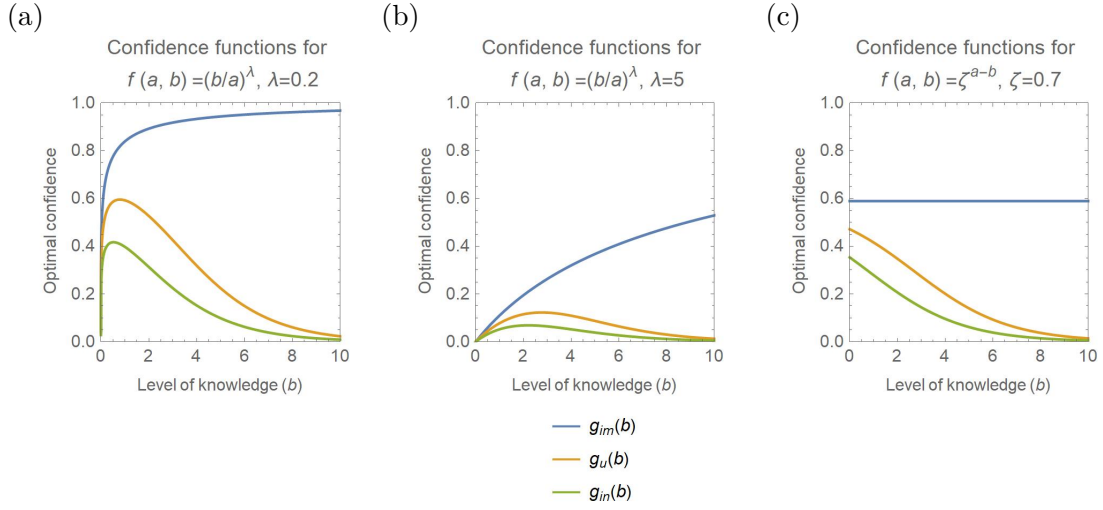


Figure 1.1. The imitation-learning confidence function $g_{im}(b)$, the innovation-learning confidence function $g_{in}(b)$, and the unconditional confidence function $g_u(b)$ for model parameter choices $p = 0.4$, $q = 0.5$, $\eta = 0.6$, and varying payoff function $f(a, b)$; note that the other model parameters do not affect these confidence functions. Consistent with Proposition 1.1, we have the inequalities $g_{in}(b) < g_u(b) < g_{im}(b)$. Also, consistent with Proposition 1.2(a), when the payoff function $f(a, b)$ satisfies Assumption 1—panels (a) and (b)—the imitation-learning confidence function $g_{im}(b)$ is strictly increasing. The payoff function of panel (c) does not satisfy Assumption 1. As a result, the corresponding imitation-learning confidence function $g_{im}(b)$ is not necessarily strictly increasing (in fact, it is constant). Finally, consistent with Proposition 1.2(b), the confidence functions $g_{in}(b)$ and $g_u(b)$ are eventually decaying to zero.

with respect to the level of experience. Whether this non-monotonicity evolves depends on the two remaining actions prescribed by the student’s optimal policy π : identifying the learning type, $x_{\text{identify}}(t)$; and quitting, $x_{\text{quit}}(t)$.

1.3.1 Imitation learning alone cannot explain non-monotonic confidence

Under reasonable assumptions on the model parameters, whether each of the confidence functions $g_{im}(b)$, $g_{in}(b)$, and $g_u(b)$ is monotonic is determined by the presence of the risk of unlearnable tasks. Since the distribution μ_{im} has zero probability on the event $\{a = \infty\}$, its associated confidence function $g_{im}(b)$ is monotonically increasing in b , as long as the payoff function $f(a, b)$ satisfies the following condition:

Assumption 1. For all $m > 0$ and $a \geq m$, the payoff function $f(a, b)$ satisfies

$$\frac{\partial}{\partial a} f(a, a - m) > 0. \tag{1.48}$$

We argue that Assumption 1 is plausible because a fixed amount m of knowledge constitutes a larger fraction of the total knowledge of an easy task than a difficult task; consequently, the argument goes, not knowing it should cause a harsher penalty in the former case. However, whether this claim generally holds is a question that should be studied empirically. Note that the assumption is satisfied by the example family of payoff functions (1.10), but not by the example family of payoff functions (1.12). Our aforementioned argument would then suggest that the former family (polynomial growth) is plausible as the marginal payoff function of ancestral learning environments, but not the latter family (exponential growth).

On the other hand, since the distributions μ_{in} and $\bar{\mu}$ have a positive probability on

the event $\{a = \infty\}$, their associated confidence functions $g_{in}(b)$ and $g_u(b)$ are non-monotonic. Specifically, both $g_{in}(b)$ and $g_u(b)$ decay to zero for all sufficiently large b . In fact, the functions are strictly decreasing to zero for all sufficiently large b , as long as the following condition holds.

Assumption 2. As $b \rightarrow \infty$, the payoff function $f(a, b)$ satisfies

$$\int_{a>b} \frac{\partial}{\partial b} f(a, b) \eta^a da \ll \eta^b. \quad (1.49)$$

Here, the notation $F(b) \ll G(b)$ denotes the asymptotic condition that $F(b)/G(b) \rightarrow 0$ as the input variable $b \rightarrow \infty$. Note that Assumption 2 is satisfied by the family of payoff functions (1.10) for any parameter $\eta \in (0, 1)$.

We summarize the above discussion in the following theorem statement.

Proposition 1.2. The generalized confidence function g_{ρ_y} satisfies the following:

- a) If $y = 0$, then we have $\frac{d}{db} g_{\rho_y}(b) > 0$ for all $b \geq 0$, as long as Assumption 1 holds.
- b) If $0 < y < 1$, then we unconditionally have $g_{\rho_y}(b) \rightarrow 0$ as $b \rightarrow \infty$.
- c) If $0 < y < 1$, then we have $\frac{d}{db} g_{\rho_y}(b) < 0$ for all sufficiently large b , as long as Assumption 2 holds.

The expected marginal payoff of a task is monotonically increasing when there is no risk that the task is unlearnable, as is the case when it is learned by innovation. In other words, since $\mu_{in} = \rho_0$, the function g_{in} should be monotonically increasing. However, there is a nontrivial probability y of unlearnability when the learning type of the task is either uncertain or fully determined as innovation: $\mu_{in} = \rho_p$ and $\bar{\mu} = \rho_{(1-q)p}$. In this case, the corresponding expected marginal payoffs (g_{in} and g_u , respectively) both eventually monotonically decrease to zero.

We have plotted in Figure 1.1 confidence functions of an example model parametrization with varying marginal payoff function $f(a, b)$, which illustrate the conclusions of Proposition 1.2. We note in particular that functions of the form $f(a, b) = (b/a)^\lambda$ —detailed in (1.10)—satisfy Assumption 1. Thus, by Proposition 1.2(a), any model parametrization with this choice of marginal payoff function will have a strictly increasing imitation-learning confidence function $g_{im}(b)$. However, functions of the form $f(a, b) = \zeta^{a-b}$ —detailed in (1.12)—do not satisfy Assumption 1, which opens up the possibility that $g_{im}(b)$ will not be strictly increasing. In fact, we then can apply the change of variables $\bar{a} = a - b$ to see that the imitation-learning confidence function

$$\begin{aligned} g_{im}(b) &= \frac{\left(\log \frac{1}{\eta}\right) \int_{a>b} \zeta^{a-b} \eta^a da}{\left(\log \frac{1}{\eta}\right) \int_{a>b} \eta^a da} = \left(\log \frac{1}{\eta}\right) \int_{a>b} \zeta^{a-b} \eta^{a-b} da \\ &= \left(\log \frac{1}{\eta}\right) \int_0^\infty \zeta^{\bar{a}} \eta^{\bar{a}} d\bar{a} \end{aligned} \tag{1.50}$$

is constant with respect to b . Thus, we see that Assumption 1 constitutes a nontrivial necessary condition for $g_{im}(b)$ to be strictly increasing.

1.3.2 Analyzing a subfamily of model parametrizations via approximation

We have solved for the optimal choice of $x_{\text{forgo}}(t)$, the decision of when to forgo a proportion of the task payoff for an alternative foraging opportunity. Assuming the policy π always uses this optimal choice, the only other components of π that can vary are $x_{\text{identify}}(t)$, the decision whether to perform a time-measurement experiment to identify the learning type j ; and $x_{\text{quit}}(t)$, the decision whether to quit. Recall that the only information that is relevant for the optimal choice of these components is the pair of information sets E_b and E_j regarding the student's current task. We abuse no-

tation by letting

$$\pi(E_b, E_j) = (x_{\text{identify}}, x_{\text{quit}}) \quad (1.51)$$

denote the action of the optimal policy π (omitting the components x_{retain} and x_{forgo} , which have already been solved previously) at the pair of information sets (E_b, E_j) .

We proceed to define a tractable subfamily of parametrizations of our model for which the optimal estimate of confidence, as a function of the level of experience i , displays the empirically documented non-monotonicity: general increase with an intermediate period of decrease. Whether this non-monotonicity occurs would depend, in general, on the action components $x_{\text{retain}}(k)$ and $x_{\text{forgo}}(k)$ of the optimal policy π . Our subfamily of model parametrizations will be constructed—via approximation—to have the appropriate optimal action components x_{retain} and x_{forgo} that guarantee the desired non-monotonicity.

Let us fix all choices of model parameters with the exception of the discrete knowledge jumps $\{B(i, n) : i \in \mathbb{N}, i > 0\}$, the learning period lengths $\{\Delta_j(i, n) : i \in \mathbb{N}, i > 0\}$ —and the corresponding cumulative learning period lengths $\{T_j(i, n) : i \in \mathbb{N}, i > 0\}$ —for $j \in \{im, in\}$, the fraction of time $r(n)$ of task attempts that can be devoted to alternative foraging opportunities, and the expected cost $C_{\text{identify}}(n)$ of a time-measurement experiment to identify the learning type j . This gives a sequence of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ varying with n . We will construct $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ so that as $n \rightarrow \infty$, the imitation-learning knowledge function and the innovation-learning knowledge function, defined respectively by

$$L_{im,n}(t) = B(\max\{i : T_{im}(i, n) \leq t\}) \quad (1.52)$$

and

$$L_{in,n}(t) = B(\max\{i : T_{in}(i, n) \leq t\}), \quad (1.53)$$

can be well-approximated by the continuous imitation-learning knowledge function

$$L_{im,\infty}(t) : [0, \infty) \rightarrow [0, \infty), \quad (1.54)$$

and the continuous innovation-learning knowledge function

$$L_{in,\infty}(t) : [0, \infty) \rightarrow [0, \infty), \quad (1.55)$$

respectively. The knowledge functions $L_{im,\infty}(t)$ and $L_{in,\infty}(t)$ are required to be bijective, continuous, and piecewise continuously differentiable such that their respective derivatives $\frac{d}{dt}L_{im,\infty}(t)$ and $\frac{d}{dt}L_{in,\infty}(t)$ are positive whenever they are well-defined. We will describe the context of this continuous approximation in Subsection 1.3.4.

We now formally define the continuous learning model, a continuous approximation of our discrete learning model defined in Subsection 1.2.2. Suppose that instead of obtaining discrete payoffs at the end of discrete task attempts, the student obtains a flow payoff

$$\delta f(a(t), b(t))dt, \quad (1.56)$$

based on the task difficulty a and the student's level of knowledge b . The term $a(t)$ denotes the difficulty level of the task that is being learned at time t , and thus has zero derivative everywhere except for the discrete set of points of time at which tasks are quit. When a task is quit at time t , and at the starting time $t = 0$, a task is drawn i.i.d. from the distribution μ as in the model of Subsection 1.2.2; and if $t > 0$, the

term $a(t)$ is updated to the newly drawn task difficulty.

The term $b(t)$ denotes the student's level of knowledge, and in the continuous learning model, updates continuously in the amount of time t . Specifically, we have

$$b(t) = \begin{cases} L_{im,\infty}(\bar{t}) & \text{if } j = im \text{ and } L_{im,\infty}(\bar{t}) < a, \\ a & \text{if } j = im \text{ and } L_{im,\infty}(\bar{t}) \geq a, \\ L_{in,\infty}(\bar{t}) & \text{if } j = in \text{ and } L_{in,\infty}(\bar{t}) < a, \\ a & \text{if } j = in \text{ and } L_{in,\infty}(\bar{t}) \geq a; \end{cases} \quad (1.57)$$

where

$$\bar{t} = t - T_{\text{start}}(t) \quad (1.58)$$

denotes the length of the time period $[T_{\text{start}}(t), t]$ spent learning the current task (at time t) and

$$T_{\text{start}}(t) \quad (1.59)$$

denotes the time at which the current task has been drawn.

We further suppose that in the continuous learning model, there is no option to exploit alternative foraging opportunities. Similarly, we suppose that the learning type of a task is not information that can be learned by paying a cost. The justification for these assumptions is that these quantities—the payoff difference due to alternative foraging opportunities and the costs of identifying the learning type—become negligible as $n \rightarrow \infty$ in the continuous approximation.

Finally, we suppose that the option to quit for an opportunity-cost task satisfies the following. For a positive constant β , the student can—when learning has not yet completed—either quit all tasks (both $j = im$ and $j = in$) at any level of experience $b \in$

$(0, \infty)$ without identifying the task type, or quit $j = im$ tasks at a level of experience $b_{im} \in [\beta, \infty) \cup \{\infty\}$ and $j = in$ tasks at a level of experience $b_{in} \geq [\beta, \infty) \cup \{\infty\}$.

The student's strategy space in the continuous learning model pertains entirely to quitting, and is given by

$$\mathcal{A}_\infty = \mathcal{Q}^\infty = \left(((0, \infty) \cup \infty) \cup ([\beta, \infty) \cup \{\infty\})^2 \right)^\infty \quad (1.60)$$

for

$$\mathcal{Q} = ((0, \infty) \cup \infty) \cup ([\beta, \infty) \cup \{\infty\})^2. \quad (1.61)$$

Here, the first subset $(0, \infty)$ denotes the set of quitting strategies b that quit all tasks at any level of experience $b > 0$ without identifying the learning type, and the second subset $([\beta, \infty) \cup \{\infty\})^2$ denotes the set of quitting strategies (b_{im}, b_{in}) that quit $j = im$ tasks at a level of experience $b_{im} \in [\beta, \infty) \cup \{\infty\}$ and $j = in$ tasks at a level of experience $b_{in} \geq [\beta, \infty) \cup \{\infty\}$. The action

$$(\mathbf{b}_1, \mathbf{b}_2, \dots) \in \mathcal{A}_\infty \quad (1.62)$$

denotes the overall strategy that quits the i th task using the strategy action \mathbf{b}_i for $i \in \mathbb{N}$. The total payoff in the continuous learning model is given by

$$\int_0^\infty \delta^t f(a(t), b(t)) dt, \quad (1.63)$$

where $a(t)$ is the difficulty value of the task being learned at time t (which discretely changes whenever a new task is drawn), and $b(t)$ is the student's level of knowledge of this task.

In summary, Table 1.3 provides the list of parameters comprising our continuous learning model, and Table 1.4 provides a step-by-step algorithm for the model. The student's objective is to maximize the expected payoff, the expected value of (1.63):

$$V_{\infty}((\mathbf{b}_1, \mathbf{b}_2, \dots)) = \mathbb{E} \left[\int_0^{\infty} \delta^t f(a(t), b(t)) dt \right]. \quad (1.64)$$

Decision theory yields that the maximal expected payoff $V_{\infty}((\mathbf{b}_1, \mathbf{b}_2, \dots))$ is obtained by a strategy that acts in the same way for every history sharing the same information set. In particular, the maximal payoff is obtained by a strategy that uses the same quitting strategy $\mathbf{b} \in \mathcal{Q}$ for every drawn task, corresponding to the strategy

$$(\mathbf{b}, \mathbf{b}, \dots) \in \mathcal{A}_{\infty}. \quad (1.65)$$

The expected total payoff of such a quitting strategy \mathbf{b} is given by the function

$$V_{\infty}(\mathbf{b}) = \begin{cases} V_{\infty,u}(b) & \text{if } \mathbf{b} = b, \\ V_{\infty,c}(b_{im}, b_{in}) & \text{if } \mathbf{b} = (b_{im}, b_{in}). \end{cases} \quad (1.66)$$

Here, the value function $V_{\infty,u}(b)$ is defined by

$$V_{\infty,u}(b) = qV_{im,\infty,u}(b) + (1-q)V_{in,\infty,u}(b), \quad (1.67)$$

where $(V_{im,\infty,u}(b), V_{in,\infty,u}(b))$ is the solution to the system of equations

Model parameters of the continuous learning model

1. the marginal payoff function $f(a, b)$,
 2. the imitation-learning knowledge function $L_{im,\infty}(t)$,
 3. the innovation-learning knowledge function $L_{in,\infty}(t)$,
 4. the exponential discount factor δ of time,
 5. the proportion p of infinite-difficulty tasks among all innovation-learning tasks,
 6. the proportion q of imitation-learning tasks among all tasks,
 7. the exponential discount factor η of the distribution of task difficulty values,
 8. the constant β constraining the student's quitting.
-

Table 1.3. List of model parameters of the continuous learning model, which approximates our modified Bayesian model of ancestral human learning. The continuous learning model is presented in Subsection 1.3.2.

Algorithmic description of the continuous learning model

1. Time is set to $T = 0$.
2. The student draws from the distribution μ the task (j, a) , the value of which is unknown to them.
3. If the student's quitting strategy is $\mathbf{b} = b$, then they receive a payoff of

$$\int_T^{T+L_{j,\infty}^{-1}(b)} \delta f(a, L_{j,\infty}(t)) dt, \quad (1.68)$$

and T is incremented by $L_{j,\infty}^{-1}(b)$. If the student's quitting strategy is $\mathbf{b} = (b_{im}, b_{in})$, then they receive a payoff of

$$\int_T^{T+L_{j,\infty}^{-1}(b_j)} \delta f(a, L_{j,\infty}(t)) dt. \quad (1.69)$$

and time is incremented by $L_{j,\infty}^{-1}(b_j)$.

4. If $T = \infty$, the algorithm is complete. If T is finite, return to Step 2 and repeat it along with the following steps.
-

Table 1.4. An algorithmic description of the continuous learning model, which approximates our modified Bayesian model of ancestral human learning. The continuous learning model is presented in Subsection 1.3.2.

$$\begin{aligned}
V_{im} &= \int_0^b \left(\int_0^{L_{im,\infty}^{-1}(a)} \delta^t f(a, L_{im,\infty}(t)) dt + \int_{L_{im,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{im}(a) \\
&\quad + \int_{a>b} \left(\int_0^{L_{im,\infty}^{-1}(b)} \delta^t f(a, L_{im,\infty}(t)) dt + \delta^{L_{im,\infty}^{-1}(b)} (qV_{im} + (1-q)V_{in}) \right) d\mu_{im}(a),
\end{aligned} \tag{1.70}$$

and

$$\begin{aligned}
V_{in} &= \int_0^b \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{in}(a) \\
&\quad + \int_{a>b} \left(\int_0^{L_{in,\infty}^{-1}(b)} \delta^t f(a, L_{in,\infty}(t)) dt + \delta^{L_{in,\infty}^{-1}(b)} (qV_{im} + (1-q)V_{in}) \right) d\mu_{in}(a);
\end{aligned} \tag{1.71}$$

while the value function $V_{\infty,c}(b_{im}, b_{in})$ is defined by

$$V_{\infty,c}(b_{im}, b_{in}) = qV_{im,\infty,c}(b_{im}, b_{in}) + (1-q)V_{in,\infty,c}(b_{im}, b_{in}), \tag{1.72}$$

where $(V_{im,\infty,c}(b_{im}, b_{in}), V_{in,\infty,c}(b_{im}, b_{in}))$ is the solution to the system of equations

$$\begin{aligned}
V_{im} &= \int_0^{b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(a)} \delta^t f(a, L_{im,\infty}(t)) dt + \int_{L_{im,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{im}(a) \\
&\quad + \int_{a>b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(b_{im})} \delta^t f(a, L_{im,\infty}(t)) dt + \delta^{L_{im,\infty}^{-1}(b_{im})} (qV_{im} + (1-q)V_{in}) \right) d\mu_{im}(a),
\end{aligned} \tag{1.73}$$

and

$$\begin{aligned}
V_{in} = & \int_0^{b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_{in}(a) \\
& + \int_{a>b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(b_{in})} \delta^t f(a, L_{in,\infty}(t)) dt + \delta^{L_{in,\infty}^{-1}(b_{in})} (qV_{im} + (1-q)V_{in}) \right) d\mu_{in}(a).
\end{aligned} \tag{1.74}$$

In fact, we can explicitly solve for these value functions.

Lemma 1.3. The value functions $V_{im,\infty,c}$, $V_{in,\infty,c}$, $V_{im,\infty,u}$, and $V_{in,\infty,u}$ are given by

$$((V_{im,\infty,c}(b_{im}, b_{in}), V_{in,\infty,c}(b_{im}, b_{in})) = (\hat{V}_{im}(b_{im}, b_{in}), \hat{V}_{in}(b_{im}, b_{in})) \tag{1.75}$$

and

$$((V_{im,\infty,u}(b), V_{in,\infty,u}(b)) = (\hat{V}_{im}(b, b), \hat{V}_{in}(b, b)). \tag{1.76}$$

Here, the functions $\hat{V}_{im}, \hat{V}_{in} : ((0, \infty) \cup \{\infty\})^2 \rightarrow [0, \infty)$ are defined by

$$\hat{V}_{im}(b_{im}, b_{in}) = \frac{\mathfrak{d}\mathfrak{e} - \mathfrak{b}\mathfrak{f}}{\mathfrak{g}} \tag{1.77}$$

and

$$\hat{V}_{in}(b_{im}, b_{in}) = \frac{\mathfrak{a}\mathfrak{f} - \mathfrak{c}\mathfrak{e}}{\mathfrak{g}} \tag{1.78}$$

for

$$\mathfrak{a} = 1 - q\delta^{L_{im,\infty}^{-1}(b_{im})}\eta^{b_{im}} \tag{1.79}$$

$$\mathfrak{b} = -(1-q)\delta^{L_{im,\infty}^{-1}(b_{im})}\eta^{b_{im}}, \tag{1.80}$$

$$\mathbf{c} = -q\delta^{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right), \quad (1.81)$$

$$\mathbf{d} = 1 - (1-q)\delta^{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right), \quad (1.82)$$

$$\begin{aligned} \mathbf{e} = & \int_0^{b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{in}(a) \\ & + \int_{a>b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(b_{in})} \delta^t f(a, L_{in,\infty}(t)) dt \right) d\mu_{in}(a), \end{aligned} \quad (1.83)$$

$$\begin{aligned} \mathbf{f} = & \int_0^{b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{in}(a) \\ & + \int_{a>b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(b_{in})} \delta^t f(a, L_{in,\infty}(t)) dt \right) d\mu_{in}(a), \end{aligned} \quad (1.84)$$

and

$$\mathbf{g} = 1 - \delta^{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right) + q \left(\delta^{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right) - \delta^{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} \right). \quad (1.85)$$

In particular, we have $V_{\infty}(b_{im}, b_{in}) = \hat{V}_{\infty}(b_{im}, b_{in})$ and $V_{\infty}(b) = \hat{V}_{\infty}(b)$ for

$$\hat{V}_{\infty} = q\hat{V}_{im} + (1-q)\hat{V}_{in}. \quad (1.86)$$

Note that it makes sense to view the space of quitting strategies \mathcal{Q} as the domain

$$\bar{\mathcal{Q}} = \{(b, b) : b \in (0, \infty) \cup \{\infty\}\} \cup \{(b_{im}, b_{in}) : b_{im}, b_{in} \geq \beta\} \subset ([0, \infty) \cup \infty)^2, \quad (1.87)$$

representing the space of strategies that use the same quitting strategy for every task. Note that the two subsets above nontrivially intersect. This has the meaning that the strategy $\mathbf{b} = b \geq \beta$ that quits without identifying the learning type obtains the same payoff as the strategy $\mathbf{b} = (b, b)$ that identifies the learning type before quitting, due to our assumption that the cost of identifying the learning type limits to zero in the continuous approximation.

We formalize the aforementioned assumptions regarding the approximation of the discrete learning models $\mathbf{M}(n)$ by the continuous learning model $\mathbf{M}(\infty)$. A sequence of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ is said to converge to the continuous model parametrization $\mathbf{M}(\infty)$ if:

1. The sequence of functions $\{L_{j,n}\}_{n>0}$ monotonically converges (increasing with respect to n) to $L_{j,\infty}$ in a way such that $L_{j,\infty}(T(i, n)) = B(i, n)$ for all n and i .
2. The parameters $\delta, f(a, b), p, q,$ and η are shared by all $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ and $\mathbf{M}(\infty)$.
3. We have $\Delta_{im}(i, n) = \Delta_{in}(i, n)$ for all i such that $B(i, n) < \beta$, and $\Delta_{im}(i, n) < \Delta_{in}(i, n)$ for all i such that $B(i, n) \geq \beta$.
4. The parameters $r(n)$ and $C_{\text{identify}}(n)$ are monotonically decreasing to zero such that

$$r(n) \ll C_{\text{identify}}(n) \ll 1. \tag{1.88}$$

The first condition constitutes the assumption that the student's accumulation of knowledge is sufficiently fine, and thus can be approximated by a continuous knowledge function. The second condition specifies the shared parameters between the approximated model parametrizations and the approximating continuous learning model. The third condition constitutes the assumption that the speeds of imitation

and innovation are too similar to distinguish in the early stages of learning ($b < \beta$), but branch off so that they become distinguishable in the later stages ($b \geq \beta$). This branch-off can occur, for example, if the respective speeds of learning increase over time—as they did in the experimental variant of Sanchez and Dunning²⁰⁴ that measured learning speeds—such that the rate of increase is faster for imitation than it is for innovation. And finally, the fourth condition represents the assumption that the additional payoffs from alternative foraging opportunities are negligible compared to the ecological fitness cost of identifying a given task’s learning type, which is negligible compared to task payoffs.

This notion of convergence is key to our approach of continuous approximation. Recall that the optimal payoff of our original discrete learning model is achieved by a policy π whose choice of action $\pi(b)$ is the same for all histories of the same pair of information sets (E_b, E_j) . For such a policy π , define

$$i_{\text{identify}} = \min\{i : \pi(\{b = B(i)\}, \{\text{neither } j \text{ ruled out yet}\}) = (\text{true}, x_{\text{quit}})\}, \quad (1.89)$$

the level of experience at which the learning type j is identified. If the policy π (conditional on learning not having completed) quits earlier than i_{identify} , say at level of experience

$$i_{\text{quit,u}} = \min\{i < i_{\text{identify}} : \pi(\{b = B(i)\}, \{\text{neither } j \text{ ruled out yet}\}) = (\text{false}, \text{true})\}, \quad (1.90)$$

then we say that the quitting strategy of π is representable by $\mathbf{b} = B(i_{\text{quit,u}})$. If the policy π (conditional on learning not having completed) quits at or later than i_{identify} ,

then we define

$$i_{\text{quit,im}} = \min\{i \geq i_{\text{identify}} : \pi(\{b = B(i)\}, \{j = in \text{ ruled out}\}) = (\text{false}, \text{true})\}, \quad (1.91)$$

and

$$i_{\text{quit,in}} = \min\{i \geq i_{\text{identify}} : \pi(\{b = B(i)\}, \{j = im \text{ ruled out}\}) = (\text{false}, \text{true})\}, \quad (1.92)$$

which denote the earliest levels of experience $i \geq i_{\text{identify}}$ at which tasks of learning type j are quit (conditional on learning not having completed). Then, we say that the quitting strategy of π is representable by $\mathbf{b} = (B(i_{\text{quit,im}}), B(i_{\text{quit,in}}))$.

Assuming these conditions hold, we have the following approximation result:

Proposition 1.4. Suppose we have a sequence of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ that converges to the continuous learning model $\mathbf{M}(\infty)$. Let V_n denote the payoff function corresponding to $\mathbf{M}(n)$. For every $\varepsilon > 0$, there exists N sufficiently large that for all $n \geq N$, we have

$$|V_n(\pi) - V_\infty(\mathbf{b}(\pi))| < \varepsilon \quad (1.93)$$

whenever π is representable as $\mathbf{b}(\pi)$.

The intuition is that since the magnitude of the cost of identifying the learning type C_{identify} is negligible compare to the main term, and the additional payoffs from alternative foraging opportunities are even more negligible, the main term of the payoff $V_n(\pi)$ —comprised of payoffs obtained from the task—will asymptotically dominate. In the proof of Proposition 1.4, we will construct a function $\hat{V}_n(b_{im}, b_{in})$ that can represent this main term. A key step in the proof that the inequality (1.93) holds will be that the constructed function with b placed in both inputs, $\hat{V}_n(b, b)$, is continuous

at $b = 0$, and that the same holds for $\hat{V}_\infty(b, b)$. This allows us to apply Dini’s theorem that for a sequence of continuous functions on a compact space that monotonically converges to another continuous function on the compact space, the convergence is uniform. Dini’s theorem, a tool we will use several times in this paper, is the reason we have defined the notion of convergence of model parametrizations $\mathbf{M}(n)$ in terms of monotonic convergence of the knowledge functions $L_{j,n}$.

Through Proposition 1.4, we have essentially reduced the problem of studying the action components x_{identify} and x_{quit} in sufficiently fine model parametrizations $\mathbf{M}(n)$ to looking at the analogous problem in the continuous approximation $\mathbf{M}(\infty)$. We proceed to analyze the latter in the following subsections to gain an insight on the optimal choice of whether to quit the status-quo task in a sufficiently fine model parametrization $\mathbf{M}(n)$, i.e., with n sufficiently large. The advantage of studying the continuous learning model $\mathbf{M}(\infty)$ is that it is significantly more tractable. For it, we can obtain quite general results about the optimal quitting strategy \mathbf{b} , which can manifest in the evolutionarily optimal estimate of confidence in the approximated model parametrizations $\mathbf{M}(n)$.

1.3.3 Dichotomy of quitting strategies based on the learning type

We begin by proving that tasks that are known to be learned by imitation are never optimally quit in the continuous learning model, as long as Assumption 1 holds and the knowledge function $L_{im,\infty}(t)$ is convex. The intuition is the following. First, the optimal expected marginal payoff is increasing in the level of knowledge when the task is known to be learned by imitation, due to Assumption 1. Second, tasks learned by imitation are learned at least as fast at higher levels of knowledge, by the assumption

of the convexity of $L_{im,\infty}(t)$. Finally, tasks learned by innovation in expectation yield less payoff than tasks learned by imitation. Thus, quitting at any level of knowledge $b > 0$ has three negative effects on expected payoff—reducing the expected marginal payoff, slowing down learning, and replacing the current imitation-learning task with an on-expectation inferior innovation-learning task—and is thus suboptimal.

Proposition 1.5. In a continuous learning model, every $\mathbf{b} = (b_{im}, b_{in})$ that maximizes the payoff $V_\infty(\mathbf{b})$ must have $b_{im} = \infty$, as long as Assumption 1 holds and the imitation-learning knowledge function $L_{im,\infty}(t)$ is convex.

As a result, the problem of finding the quitting strategy $\mathbf{b} = (b_{im}, b_{in})$ that maximizes the value function $V_{\infty,c}(b_{im}, b_{in})$ becomes a one-dimensional maximization problem

$$\max_{b_{in} \in [\beta, \infty) \cup \{\infty\}} V_{\infty,c}(\infty, b_{in}). \quad (1.94)$$

Note that the convexity of a knowledge function $L_{j,\infty}(t)$ constitutes the assumption that knowledge-learning is (weakly) faster in its later stages. If true, this may reflect a dynamic where potential advances in task-specific knowledge are limited by the amount of previously held knowledge, so that such advances are more likely to arise from the substantial knowledge base in the late stages of learning than from the lacking knowledge base in the early stages of learning. However, the opposite assumption of a concave knowledge function $L_{j,\infty}(t)$, the assumption that knowledge-learning is (weakly) faster in its earlier stages, is also plausible. If true, this may reflect a dynamic where there are more “low-hanging fruits” in the early stages of learning than in the late stages. Empirical studies can help quantitatively investigate aspects of knowledge accumulation as a function of time: in particular, which of the two aforementioned dynamics dominates at any given stage of learning.

Next, we prove an unconditional result: that tasks known to be either learned by innovation or of ambiguous learning type are always optimally quit at an intermediate level of knowledge.

Proposition 1.6. In a continuous learning model, every $\mathbf{b} = (b_{im}, b_{in})$ that maximizes the value function $V_\infty(\mathbf{b})$ satisfies $b_{in} < \infty$. Also, every $\mathbf{b} = b \in (0, \infty) \cup \{\infty\}$ that maximizes the value of the function $V_\infty(\mathbf{b})$ satisfies $b < \infty$.

The intuition is that these tasks, in contrast to tasks known to be learned by imitation, come with a risk of unlearnability that asymptotically dominates as the level of knowledge becomes sufficiently high. As a result, conditional on learning not yet having completed, the expected payoff from staying the course asymptotically decays to the point of being overtaken by that yielded by switching to an opportunity-cost task.

1.3.4 Implications for the evolutionarily optimal estimate of confidence

Consider the evolutionarily optimal estimate of confidence $\hat{g}(E_b, E_j)$, defined in (1.39), for a model parametrization $\mathbf{M}(n)$ for a sufficiently large n . Unlike in the continuous limit $\mathbf{M}(\infty)$, the model parametrization $\mathbf{M}(n)$ is characterized by alternative foraging opportunities, whose exploitation factors into the payoff function $V_n(\pi)$. Thus, the student in the model $\mathbf{M}(n)$ is predicted to evolve the optimal estimate of confidence $\hat{g}(E_b, E_j)$.

The possible values of confidence as a function of the level of knowledge b (conditional on learning not having completed yet, $b < a$) are $g_{im}(b)$, $g_{in}(b)$, and $g_u(b)$. Under Assumption 1 and the assumption that the imitation-learning knowledge function $L_{im,\infty}(t)$ is convex, tasks learned by imitation are never quit. Consequently, there are

two possibilities for how a payoff-maximizing strategy \mathbf{b} in the approximating continuous learning model $\mathcal{M}(\infty)$ will learn tasks.

The first possibility, corresponding to the case that $\mathbf{b} = b'$, is that tasks are learned until a level of knowledge b' and quit if learning has not completed by then. In this case, the optimal estimate of confidence $\hat{g}(E_b, E_j)$, conditional on $b < a$, is given by $g_u(b)$ for $b < b'$, and tasks are never learned to a higher level of knowledge than b' . This conclusion seems empirically untenable for two reasons. First, there are many instances of human learning of tasks that continues on to high levels of experience and knowledge without quitting. Second, the function $g_u(b)$ has been shown in Proposition 1.2 to eventually decay to zero for b sufficiently high, which contradicts the empirical pattern that confidence is generally increasing in the level of experience (albeit with an intermediate period of decrease).

The second possibility, corresponding to the case that $\mathbf{b} = (\infty, b_{im})$ for $b_{im} \in [\beta, \infty)$, is that tasks are learned until a level of knowledge b_{im} , at which point tasks of innovation-learning type are quit if learning has not completed by then and tasks of imitation-learning type are learned to completion. Recall that we have assumed that the additional payoff obtainable from alternative foraging opportunities, which scale with r , is negligible compared to the cost of identifying the learning type $-C_{\text{identify}}$. A consequence of this assumption is that in the limit $n \rightarrow \infty$, the only possible upside of identifying the learning type is to enable differentiated choices pertaining to quitting that differ between the two learning types. Moreover, the negligibility of the cost $-C_{\text{identify}}$ in comparison to payoffs from the task necessitate that this cost is paid at the latest possible time which allows for the optimal such differentiated quitting strategy to be played: specifically, during the task attempt i_{identify} for which $b_{im} = B(i_{\text{identify}}, n)$ is payoff-maximizing among the possible quitting points

$\{B(i, n)\}_{n \in \mathbb{N}, n > 0}$.

Because of this, when the strategy of the form $\mathbf{b} = (\infty, b_{im})$ is used, the optimal estimate of confidence $\hat{g}(E_b, E_j)$, conditional on $b < a$, is given by $g_u(b)$ for $b < b_{im}$ and by $g_{im}(b)$ for $b \geq b_{im}$. Since $g_u(b)$ eventually decays to zero and $g_{im}(b)$ is monotonically increasing, their piecewise combination (conditional on learning not yet having completed),

$$g(b) = \begin{cases} g_u(b) & \text{if } b \leq b_{im}, \\ g_{im}(b) & \text{if } b > b_{im}, \end{cases} \quad (1.95)$$

can be non-monotonic in the empirically observed way: generally increasing with an intermediate period of decrease.

In order for the evolutionarily optimal estimate of confidence $\hat{g}(E_b, E_j)$ to be empirically tenable, the payoff-maximizing strategy seems to need to be of the form $\mathbf{b} = (\infty, b_{im})$, and not $\mathbf{b} = b$. To show the plausibility of the former possibility, we construct model parameters p (the proportion of unlearnable tasks among all tasks learned by innovation) and q (the proportion of tasks learned by imitation among all tasks) for which this is true. We do this by showing that both p and q can be taken sufficiently small in our continuous learning model $\mathcal{M}(\infty)$ so that any strategy maximizing $V_\infty(\mathbf{b})$ among the subset of strategies of the form $\mathbf{b} = b$ quits at an arbitrarily late level of knowledge b . In particular, this can be done so that b is at least β , at which point we can appeal to Proposition 1.5 to see that the best strategy of the form $\mathbf{b} = b$ is suboptimal in the overall set of strategies $\bar{\mathcal{Q}}$. Depending on the choice of model parameters (e.g., see Figure 1.2), the decreasing behavior at the tail end of the component function $g_u(b)$ can be captured in the piecewise function $g(b)$, where it is followed by the monotonic increase of the component function $g_{im}(b)$. Thus, it is

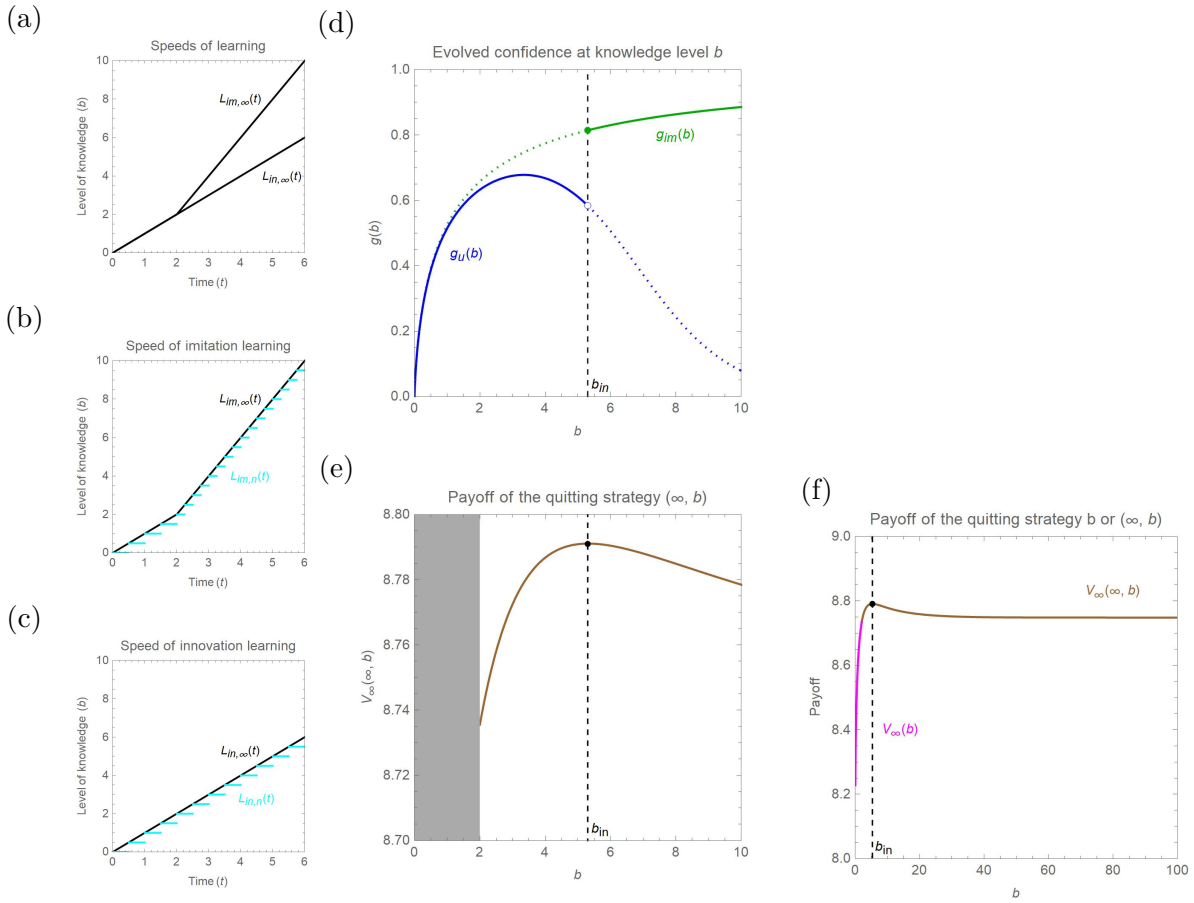


Figure 1.2. Plots of quantities relevant to the family of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ and the approximating continuous learning model $\mathbf{M}(\infty)$, presented in Table 1.5. Panel (a) plots the knowledge functions $L_{im,\infty}(t)$ and $L_{in,\infty}(t)$ of $\mathbf{M}(\infty)$, panel (b) shows how $L_{im,\infty}(t)$ approximates the imitation-learning knowledge functions $L_{im,n}(t)$ of $\mathbf{M}(n)$ ($n = 3$ is pictured), panel (c) shows how $L_{in,\infty}(t)$ approximates the imitation-learning knowledge functions $L_{in,n}(t)$ of $\mathbf{M}(n)$ ($n = 3$ is pictured), panel (d) plots the evolutionarily optimal estimate of confidence $g(b)$ (conditional on learning not yet having completed) in $\mathbf{M}(\infty)$, panel (e) plots the payoff $V_\infty(b)$ of the quitting strategy $\mathbf{b} = (\infty, b)$ for $b \geq \beta$, and panel (f) increases the domain and additionally plots the payoff $V_\infty(b)$ of the quitting strategy $\mathbf{b} = b$ for $b < \beta$. The value of the local-maximizing (in fact, ostensibly global-maximizing) value $b_{in} \approx 5.32$ is such that the confidence function $g(b)$ when using the quitting strategy $\mathbf{b} = (\infty, b_{in})$ is non-monotonic in the desired way: general increase with an intermediate period of decrease.

theoretically plausible that the evolutionarily optimal estimate of confidence conditional on learning not yet having completed, $g(b)$, is generally increasing with an intermediate period of decrease.

Corollary 1.7. Suppose Assumption 1 holds and the imitation-learning knowledge function $L_{im,\infty}(t)$ is convex. In the continuous learning model, fix all parameter choices except those of p and q . For every $\gamma \geq 0$, there exist choice of p and q such that the following simultaneously hold.

- a) Any quitting strategy $\mathbf{b} = (\infty, b_{in})$ maximizing V_∞ must satisfy

$$b_{in} > \gamma. \tag{1.96}$$

- b) Any quitting strategy $\mathbf{b} = b$ maximizing $V_\infty(b)$ (where we include the limiting strategy $\mathbf{b} = b \rightarrow 0$ in the domain) must satisfy

$$b > \gamma. \tag{1.97}$$

To prove this, we will use the following lemma, a comparative-statics result which is also of independent interest. It is comprised of two intuitive facts. First, the payoff value is decreasing in the proportion p of unlearnable tasks among those learned by innovation, which makes sense because unlearnable tasks yield the minimum possible payoff. Second, the payoff value is increasing in the proportion q of tasks learned by imitation, which makes sense because these tasks on expectation yield higher payoffs than those learned by innovation.

Lemma 1.8. For any fixed $(b_{im}, b_{in}) \in \bar{Q} \cup \{(0, 0)\}$, the following are true.

a) We have

$$\frac{\partial}{\partial p} \hat{V}_\infty(b_{im}, b_{in}) \leq 0, \quad (1.98)$$

with equality if and only if $q = 1$.

b) If Assumption 1 holds and the imitation-learning knowledge function $L_{im,\infty}(t)$ is convex, then we have

$$\frac{\partial}{\partial q} \hat{V}_\infty(\infty, b_{in}) > 0. \quad (1.99)$$

1.3.5 An example showing the plausibility of non-monotonic confidence

We conclude by constructing a family of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ whose approximating continuous learning model $\mathbf{M}(\infty)$ can be used to show that the confidence function $g(b)$ that evolves in a sufficiently fine model parametrization $\mathbf{M}(n)$ can plausibly be non-monotonic in the desired way: general increase with an intermediate period of decrease. The choice of parameters for $\mathbf{M}(n)$ is presented in Table 1.5. Then, the family of model parametrizations $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ is approximable by the continuous learning model $\mathbf{M}(\infty)$, which has knowledge functions $L_{im,n}(t)$ and $L_{in,n}(t)$ that are determined—by the values $\Delta_j(i, n)$ and $B(i, n)$ —to be

$$L_{im,\infty}(t) = \begin{cases} t & \text{if } t < 2, \\ 2(t-1) & \text{if } t \geq 2, \end{cases} \quad (1.100)$$

which is convex; and

$$L_{in,\infty}(t) = t. \quad (1.101)$$

Example family of model parametrizations $\mathbf{M}(n)$ of the modified Bayesian model

1. The time-discount factor is $\delta = 0.9$.
2. The marginal payoff function is $f(a, b) = b/a$.
3. The proportion of unlearnable tasks among those learned by innovation is $p = 0.01$.
4. The proportion of tasks that are learned by imitation is $q = 0.01$.
5. The decay factor of task difficulty values is $\eta = 0.5$.
6. The learning period lengths are given by

$$\Delta_{im}(i, n) = \begin{cases} \frac{2}{n+1} & \text{if } i < n + 1, \\ \frac{1}{n+1} & \text{if } i \geq n + 1, \end{cases} \quad (1.102)$$

and

$$\Delta_{in}(i, n) = \frac{2}{n + 1}. \quad (1.103)$$

7. The knowledge jump values are given by $B(i, n) = \frac{2i}{n+1}$.
 8. The expected cost of a time-measurement experiment to identify the learning type is $-C_{\text{identify}}$ for $C_{\text{identify}} = \frac{1}{n+1}$.
 9. The fraction of time of task attempts that can be devoted to alternative foraging opportunities is given by $r = e^{-(n+1)}$.
 10. The distribution ψ of the marginal payoffs of alternative foraging opportunities is arbitrary.
 11. The distributions $\phi(a, b)$ can be arbitrarily chosen, as long as we have $\mathbb{E}[\phi(a, b)] = f(a, b)$.
 12. As we have assumed throughout the paper, the expected cost of retaining a payoff observation, $-C_{\text{retain}}$, has sufficiently high magnitude C_{retain} so that payoff data are never retained: e.g., large enough so that the inequality (1.36) holds.
-

Table 1.5. An example family of parametrizations $\mathbf{M}(n)$ of our modified Bayesian model of ancestral human learning. The continuous learning model approximating this family, $\mathbf{M}(\infty)$, is characterized by a non-monotonic confidence function (see Figure 2). It follows that for sufficiently large n , the evolutionarily optimal confidence function of the model parametrization $\mathbf{M}(n)$ is also non-monotonic.

Also, the threshold for learning-type identification is determined—by the values $\Delta_j(i, n)$ —to be $\beta = 2$. Moreover, all other parameters are shared with the model parametrizations $\mathbf{M}(n)$. Plots relevant to the family $\{\mathbf{M}(n)\}_{n \in \mathbb{N}}$ and its approximating continuous learning model $\mathbf{M}(\infty)$ are shown in Figure 1.2.

We use Mathematica 12.2’s NMaximize function to find a local-maximizing, potentially global-maximizing quitting strategy $\mathbf{b} = (\infty, b_{im}) \in \bar{Q}$ for $b_{im} \approx 5.32$. That the quitting strategy $\mathbf{b} = (\infty, b_{im})$ is local-maximizing and ostensibly global-maximizing is illustrated in Figure 1.2(f)’s plot of the global-maximum candidates $V_\infty(b)$ for $b < 2$ and $V_\infty(\infty, b)$ for $b \geq 2$, within the domain $0 \leq b \leq 100$. Thus, it is plausible that the quitting strategy $\mathbf{b} = (\infty, b_{im})$ evolves, and consequently, that $b = b_{im}$ is the cutoff point for the (limiting) piecewise-defined confidence function $g(b)$ that is optimal when using the quitting strategy $\mathbf{b} = (\infty, b_{im})$. As shown in Figure 1.2(d), this cutoff point makes the confidence function $g(b)$ is non-monotonic in the desired way: general increase with an intermediate period of decrease. By Proposition 1.4, this type of non-monotonic pattern will manifest in the corresponding confidence functions $g(b)$ of the model parametrizations $\mathbf{M}(n)$ for sufficiently large n , thereby illustrating via example the theoretical plausibility of this pattern’s evolution.

1.4 Discussion

Classical Bayesian models are often used to represent task-learning over repeated attempts, each of which yields an observable payoff²⁰⁵. In this paper, we have described a practical test for rejecting the null hypothesis that a learner is meaningfully learning from their environmental feedback in the sense of classical Bayesian updating.

The test—essentially a corollary of standard Bayesian statistics—is to check whether

the learner’s estimate of their expected payoff-acquisition ability is converging to the mean of the past payoff data.

However, there is extensive empirical evidence of people’s persistent failures to meaningfully learn from high-variance environmental feedback. This manifests in cognitive biases like underinference, the hard-easy effect, and recurrently non-monotonic confidence. Our test thus suggests that we should consider rejecting the null hypothesis that humans by default meaningfully learn (in the sense of classical Bayesian updating) from high-variance payoff data. Indeed, the version of the classical Bayesian model we have presented in Subsection 1.2.1 is specialized to repeated task-learning and incorporates the realistic assumption that a cognitive biological agent bins observations into finitely many bins. Under this assumption, tasks that yield low-variance payoff data are easily learned via deterministic causal inference, because it is likely that nearly all payoff data will fall in a single observational bin. However, learning tasks that yield high-variance payoff data requires a large number of observations for classical Bayesian inference to reliably learn the true state. Overcommitting attention to meaningfully retain a large number of high-variance observations could result in onerous ecological fitness costs, which we hypothesize is the causal mechanism behind the proposed non-selection of classically Bayesian learning strategies in settings of high-variance payoff data.

Next, we have modified the classical Bayesian model to represent ancestral humans’ learning environment in a way that can evolutionary explain the puzzling predictive inadequacies of classical Bayesian updating models (when applied to humans). When the ecological fitness cost of retaining payoff data is high, the optimal strategy does not retain them, in contrast to the Bayesian principle that free information should always be taken. The optimal strategy then instead relies on setting-specific sources of

information, as theorized by the ecological rationality hypothesis. The informational setting of ancestral human learning is hypothesized by cultural evolutionary theory to be one where social learning of task-specific knowledge is paramount.

Our modified Bayesian model seeks to represent this hypothesized learning environment. In it, a student attempts to learn a fitness-relevant task via attempted imitation of a role model, with the option of switching between tasks and role models (between task packages). The main term of the student's payoff function is comprised of payoffs yielded by task attempts, which are obtained in the form of high-variance probabilistic lotteries and thus unfeasible to meaningfully retain. However, the payoff function also has a secondary term comprised of the ecological fitness cost of identifying the learning type (we hypothesize that this is accomplished via a mental time-measurement experiment to distinguish learning speeds), as well as a tertiary term comprised of additional payoffs obtained by devoting a fraction of a task attempt's time to opportunistically exploiting alternative foraging opportunities instead.

Optimal exploitation of alternative foraging opportunities requires an accurate estimate of the task's expected marginal payoff conditional on the known information, which—in our hypothesized domain of high-variance, difficult-to-retain payoff data—is comprised of the task's learning type, if known (successful imitation versus de facto innovation); and their level of knowledge on the task. This evolutionarily optimal estimate of the expected marginal payoff—of the student's confidence at the task—is a piecewise function of their level of experience, whose piecewise cutoff point is determined by the optimal point at which tasks learned by de facto innovation are quit. In order for this confidence function to not be always monotonically increasing, it is necessary (as long as Assumption 1 holds) that not all attempted imitation learning is successful: that a positive proportion of tasks are learned instead via de facto innova-

tion.

Moreover, we demonstrate that this confidence function can be non-monotonic in the specifically desired way: general increase with an intermediate period of decrease. This specific non-monotonic pattern, which we have demonstrated for a tractable subfamily of model parametrizations, arises because of the following interplay. Learning via de facto innovation while attempting to imitate a role model is not guaranteed to complete in finite time, because the task may be unlearnable. On the other hand, this risk does not exist when the student learns from authentically imitating a role model, since conditional on the imitation being authentic, the role model must have successfully learned the task beforehand. The student's optimal estimate of the task's expected marginal payoff (confidence) is monotonically increasing in the level of knowledge when it is guaranteed to be learnable in finite time, but eventually decays to zero when it may instead be impossibly difficult. We thus hypothesize that the evolutionarily optimal estimate of the expected marginal payoff can be non-monotonic due to its piecewise definition. The increasing, then decreasing portion of the expected marginal payoff function is conditional on the fact that the task may be unlearnable. The final increasing portion is conditional on having ruled out the risk of unlearnability, because the tasks to which this risk is exclusive—those learned by innovation—should optimally be quit at an intermediate level of knowledge.

In short, we hypothesize that the desired pattern of recurrent non-monotonicity evolved due to a particular interplay between the ecologically rational estimate of task-specific confidence and the ecologically rational strategy of task/role-model turnover. A necessary condition for this interplay is the dichotomy between tasks learned by imitation (for which the risk of unlearnability does not occur) and those learned by innovation (for which it does).

We emphasize that the aforementioned subfamily of model parametrizations was specifically constructed to demonstrate the theoretical plausibility of the desired non-monotonicity in an analytically tractable subset of the family of all parametrizations of our model. We anticipate that the full subset of model parametrizations whose evolutionary optimal estimate of confidence is recurrently non-monotonic in the desired way will be larger.

We are agnostic about the precise combination of adaptive and biological mechanisms by which the ecologically rational strategy (of task-payoff estimation and task/role-model turnover) in an environment of social task-specific learning was achieved. Plausible adaptive mechanisms relevant to this strategy include genetic evolution and contemporary, likely social learning. Given that people often fail to adapt their decision-making to settings of unambiguous individual learning with zero ecological fitness costs of retaining payoff data—such as those of the experiments of Sanchez and Dunning^{203, 204}—we propose that genetic evolution plays at least a partial role in the sense of the ecological rationality hypothesis. On the other hand, cultural evolutionary theory implies that contemporary social learning may also play at least a partial role, especially given the sheer variation of relevant parameters among the myriad environments and groups humans have inhabited and moved between. The biological mechanisms through which ecologically rational strategies of social task-learning are implemented are likely neurological, but may also be partly hormonal. Future research on both the adaptive and the biological mechanisms relevant to strategies of task-payoff estimation, task/role-model turnover, and other aspects of social task-learning would potentially be fruitful.

1.4.1 Implications

Our model proposes to help explain in an interwoven way two related topics: the evolutionary explanation of cognitive biases, and of why people underuse high-variance environmental feedback in the selection of role models. It does so by incorporating—into the general framework of Bayesian decision theory—the cultural-evolutionary-theoretic hypothesis that the primary informational setting of ancestral human learning was the social learning of task-specific knowledge; as well as the insight of the ecological rationality hypothesis that the method by which biological cognitive agents learn from information is constrained in a setting-specific manner, such as by their ancestral environments’ ecological fitness costs of overcommitting attention.

First, our model demonstrates the evolutionarily plausibility of empirically robust cognitive biases regarding confidence, and informs us of potentially useful necessary conditions and sufficient conditions for these patterns to evolve.

1. Task-specific confidence can persistently deviate from the environmental feedback, in a way that conforms to the hard-easy effect. This requires that the ecological fitness cost of retaining payoff data is nonzero, and is guaranteed to occur if the cost is sufficiently high.
2. Task-specific confidence can be recurrently non-monotonic in the desired way: general increase with an intermediate period of decrease. This requires (as long as Assumption 1 holds) that a positive proportion of attempted imitation learning is unknowingly implemented as de facto innovation learning, and is guaranteed to occur in our constructed subfamily of model parametrizations.

In the course of producing these desired conclusions while aiming to maintain model

parsimony, our work has identified a relatively short list of environmental parameters that are potentially key to predicting certain aspects (i.e., task-specific confidence and strategies of task/role-model turnover) of descriptive human learning of a high-variance-payoff task over repeated attempts.

Also, our model augments our understanding of how role-model-selection strategies that persistently fail to meaningfully learn from certain environmental feedback evolved. Cultural evolutionary theory hypothesizes that once some capacity for cultural transmission evolved, natural selection would have favored increasingly effective strategies for cultural learning⁹⁰. In this hypothesis, ancestral humans somehow achieved the threshold level of cultural-learning capacity at which cumulative cultural evolution becomes the primary selection pressure acting on cognition. After crossing this threshold, ancestral humans with a better-than-average capacity for cultural learning would have been favored by natural selection, which would then further amplify cumulative cultural evolution. Thus, gene-culture coevolution caused an autocatalytic cycle of more effective cultural-learning strategies and greater cumulative cultural evolution. A hypothesized example of such an effective cultural-learning strategy is selective social learning: the strategy of learning from preferentially chosen role models who are likely to possess better-than-average knowledge²⁶.

However, empirical studies have uncovered what at first appear to be surprising suboptimalities for the role-model selection strategies that humans have actually evolved. For example, students are substantially inaccurate in assessing the help provided by their teachers^{106,242}. Also, people are persistently vulnerable to maladaptive advice from role models^{47,73,235}, such as that regarding female genital cutting^{110,240}, funerary cannibalism¹³², unfounded shamanistic predictions²¹⁷, membership in an exploitative cult⁶⁷, and medical pseudoscience²⁰⁷. This body of evidence begs a ques-

tion: why did ancestral humans evolve to not meaningfully learn from certain environmental observations relevant to the accurate assessment of role-model quality? One might presume that an informationally rational social learner would base their role-model selection on the payoff data of potential role models, and on the learner’s own payoff data in the process of imitating a given role model.

Our theory contributes to explaining this phenomenon by specializing the ecological-rationality framework (in our setting, by incorporating high ecological fitness costs of retaining environmental observations) to not only the estimation of task-specific payoffs, but also the selection of tasks/role models. Specifically, in our model, these ecological fitness costs can cause role-model-selection strategies (in our model, task/role-model turnover strategies which determine when to quit the status-quo task package for a new one) based on retaining such observations to be informationally inefficient. Classically Bayesian-rational strategies, such as those of role-model selection, are much more likely to be suboptimal when environmental observations occur with high variance. Also, our model proposes explicit mechanisms by which ancestral humans—even in the absence of feasibly retainable environmental feedback—could still have plausibly evolved on-average selective role-model-selection strategies which relied instead on setting-specific sources of information (e.g., the student’s level of knowledge and their speed of learning). By hypothesizing precisely how people’s ostensibly sub-optimal role-model-selection strategies may actually be potentially ecologically rational, our model adds to cultural evolutionary theory’s understanding of its hypothesized on-average selective social learning.

To corroborate the hypothesis that humans achieved on-average selective social learning even for high-variance-payoff tasks, our work highlights the importance of identifying and investigating the relevant mechanisms of selective social learning,

which would need to be robust in the face of high ecological fitness costs of overcommitting attention. One such mechanism, hypothesized by our model, is the potential dependence of task/role-model turnover strategies on setting-specific information, which can inform turnover even in the absence of retained environmental feedback. Another example of such a mechanism is the conformist or reputation-based nature of human role-model-selection strategies^{26,38,89}. To illustrate, descriptive human role-model-selections rely at least partially on granting prestige status to role models based on popularity rather than on the relevant environmental feedback⁹².

These two mechanisms—reliance on setting-specific information and conformist role-model-selection strategies—are not competing explanations for on-average selective social learning in settings of high-variance environmental feedback. In fact, the latter mechanism may require the former, because in order for conformist role-model-selection strategies to facilitate selective social learning in the absence of environmental feedback, the prestige status granted to a popular role model may need to have had incorporated other helpful information at some point in the past. If this information could not feasibly have been environmental feedback, then it must have been setting-specific information in the complement of environmental feedback. Our theory proposes that the student’s level of knowledge and their speed of learning can provide such setting-specific information to achieve an on-average selective strategy of task/role-model choice, even when retaining environmental feedback is unfeasible.

Regardless of whether our model is a good model of ancestral humans’ learning environment, our test for verifying whether a learner is meaningfully incorporating their environmental observations into their decision-making—in the sense of classical Bayesian inference—may be general enough to have various potential applications. To illustrate, public-policy plans are often aimed at least partially at improving societal

well-being. Arguably, the dominant paradigm with which this goal is approached is the assumption that each person's decisions (e.g., the price they are willing to pay or take for an item) reveal an informationally rational aggregate of their private observations relevant to their well-being⁸⁰. Policymakers thus aim to economize on the cost of gathering copious, potentially idiosyncratic information by relying on each person's purported aggregate of their individual observations encapsulated by their decisions. The reliability of this information-gathering strategy is determined by whether each person is actually aggregating their observations in an informationally rational way.

However, as we have seen above, an extensive body of empirical evidence suggests that this assumption of informational rationality may not hold true when the relevant observations occur with high variance. Moreover, we have demonstrated the plausible ecological rationality of empirically robust cognitive biases by constructing an evolutionary model of social learning of task-specific knowledge, hypothesized by cultural evolutionary theory to be the primary mode of ancestral human learning. Our work thus contributes to raising the following research question: in which situations do public-policy plans aimed at improving societal well-being under the assumption of people's informational rationality actually succeed in doing so? It also begs a potentially important follow-up question: can public-policy plans be improved by replacing the assumption of informational rationality with the more empirically tenable assumption of ecological rationality? Domains of high-variance payoff data, such as gambling, may potentially be better served by the latter assumption over the former.

Another preliminary point is that informational rationality may not be an unattainable goal for human cognition. The decision-making of a person who is both trained in statistical methods and has the habit of applying this training to their own observations may be informationally rational. It may thus be fruitful not only to question

the default assumption of people’s informational rationality, but also to explore the potential upside of practical statistics training: such as the habit of keeping track of the mean past payoff data, as implied by our test for informational rationality. This statistical skill can be both a possible remedy to the potentially detrimental misassumption of informational rationality, and a facilitator of improved judgement and role-model selection at the individual level. One potential such benefit is dissuading people from socially learning the practice of repeated gambling on negative-expected-value lotteries.

1.4.2 Model limitations and directions for generalization

Our model is almost certainly an oversimplification of descriptive social task-learning, which in general involves extremely complex social dynamics. We non-exhaustively list several ways in which this is the case. We also note potential remedies, in the form of potential directions for generalization. Thereby generalizing our model may potentially enable it to better represent descriptive social task-learning and thereby better explain the relevant empirical data. We thus propose our model as a barebones representation of social, knowledge-based task learning, on which more sophisticated variants can potentially be built in the future (assuming, of course, that the thrust of the model’s story is essentially correct).

First, our model’s conclusion that the student retains no information from payoff data is oversimplified. Realistically, people can plausibly retain easy-to-remember aspects of their past payoff data, which may include the maximum and minimum payoff values observed so far. People may also temporarily retain a small number of recent payoff data, even when they fail to draw on more distant past data that a Bayesian-

updating belief would incorporate. The realistic assumption that a small number of recent payoff observations may inform decision-making can account for additional empirically documented patterns in descriptive human learning, such as reinforcement learning¹⁶⁵.

Also, our model’s assumption that knowledge affects decision-making through a unidimensional quantification—the level of knowledge b —is an oversimplification. There is no reason to believe that knowledge is unidimensional, an assumption we have used for the sole sake of tractably showing the evolutionary plausibility of recurrently non-monotonic confidence. In fact, given the sheer multifaceted nature of knowledge, we hypothesize that knowledge in general should affect decision-making through a more faithful, multidimensional quantification.

Moreover, our model’s two-dimensional spectrum of task packages—assumed in our model to be comprised of a unidimensional knowledge-based difficulty level and a binary learning type—is an oversimplification. First, as we have noted above, knowledge is likely experienced as a multidimensional quantity, which makes it likely that a unidimensional knowledge-based classification of tasks is an oversimplification. Second, when a student attempts to learn from a role model, their method of learning would in general be placed somewhere on the spectrum between full imitation and full innovation. Third, our two-dimensional spectrum is unlikely to capture all the relevant variations in the task-learning process; idiosyncrasies of the task itself, of how the student learns, of how the teacher imparts (or ostensibly imparts) knowledge, and of the degree to which learning is student-directed as opposed to role-model-directed (for example, whether the student seeks out the role model for a task they already had in mind) may also influence the learning process. In particular, potentially consequential quantities like the speed of learning may vary with respect to characteristics of the

task package that are not captured by this two-dimensional parametrization.

Finally, our model’s assumption that task packages are drawn i.i.d. from a fixed probability distribution is an oversimplification. For one thing, the i.i.d. assumption on our model—added for the sake of tractability—ignores the likely correlations between different task packages due to similarities in either the teachers or the underlying tasks. For another, descriptive selection of tasks/role models is not well-modeled by an i.i.d. draw from a fixed distribution; it is better described as an intrinsically social process that involves dynamically occurring interactions between other students and other potential role models, such as via conformist role-model selection strategies (e.g., prestige status). Such a multi-agent interaction would need to be modeled by a complex game-theoretic model, rather than a comparatively tractable Bayesian decision-theoretic model (which can be solved by dynamic-programming-type methods under quite non-restrictive conditions). Regardless, only a model in the former formulation could veridically represent the relevant social dynamics, such as coordination and punishment.

1.4.3 Empirical tests

We sketch an empirical program to study descriptive human learning in the formulation of our theory. One of the primary goals of such a program would be the eventual corroboration or falsification of the theory itself. However, the program—by pursuing theoretical formulation—may also potentially yield other advances in the psychological sciences’ understanding of descriptive human learning and decision-making, especially since the field is arguably held back by a shortage of theoretical formulation at the moment¹⁶¹.

First, we propose the empirical estimation of the true parameters of various social task-learning environments. Several parameters which we have proposed to be evolutionarily relevant include the proportion of attempted imitation that is successful, the proportion of unlearnable tasks among those that are learned by unsuccessful imitation (de facto innovation learning), the speed of each type of learning, ecological fitness costs of various action choices, and the situation-specific marginal payoff from a task. Empirical studies of how these parameters varied across both ancestral and contemporary human learning environments, as well as studies of whether they can predict the respective evolution of task-specific confidence and strategies of task/role-model turnover, would potentially contribute to a more robust and granular understanding of human cognition. Such studies would also allow us to test whether our model can veridically represent ancestral and contemporary human learning environments.

Estimates of such model parameters in ancestral environments would often be necessarily crude, given the general lack of archaeological and other relevant forms of evidence. As a start, one may feasibly expect ancestral humans who lived in areas where food is complicated to obtain (e.g., tundra)—when compared to those who lived in areas with easy food availability (e.g., rainforests)—to either have a generally lower-valued payoff function, a task difficulty distribution biased towards higher difficulty values, or a greater probability of unlearnable tasks. Empirical studies can then test whether these hypothesized parameter differences in the ancestral environments affect strategies of task-payoff estimation and task/role-model turnover in the ways predicted by our model, such as Proposition 1.1’s prediction that task-specific confidence (conditional on learning not yet having completed) decreases in the proportion of unlearnable tasks. Such efforts, however, may be inevitably limited, due to the

multitude and granular variation of the model parameters, the difficulty of measuring many of them for ancestral environments, and the uncertainty in whether ecologically rational social task-learning strategies were selected via genetic evolution.

More immediately promising would be applying such efforts to investigating the social task-learning of evolutionarily relevant foragers whose lifestyles are hypothesized to be faithful continuations of their ancestors', such as the Hadza people^{130,143}. Such efforts will not be confounded by our current uncertainty in whether the adaptive mechanism by which ecologically rational social task-learning strategies were selected was genetic evolution or contemporary learning. We propose empirical studies of the social task-learning of such peoples as a potentially fruitful first step in testing whether our model (or a sufficient generalization) is a good model of descriptive human learning. If the answer to this question is affirmative, empirical researchers can proceed to study learning environments with granular variations in model parameters, genetic-evolutionary background, and cultural-evolutionary background. Doing so may further corroborate or potentially falsify our model, determine the role of genetic evolution and contemporary learning in the selection of its ecological rational strategies, and investigate the scientific consequences of any such findings.

For instance, suppose that our model is a good model of descriptive human learning, and that the adaptive mechanism by which its ecologically rational strategies were selected was at least partially genetic evolution. Then, our model may provide a way in which otherwise mysterious aspects of ancestral human learning environments can be studied indirectly: via empirical studies (of task-specific confidence and task/role-model turnover) investigating people living today. Specifically, empirical data of these psychological aspects—which are comparatively easy to obtain—can narrow down the feasible region of model parametrizations that can evolutionarily

explain the data of such studies. This would then potentially inform us of characteristics of the respective ancestral human learning environments that would otherwise be difficult to discern. On the other hand, suppose that the adaptive mechanism by which the model's ecologically rational strategies were selected was at least partially contemporary cultural learning. Then, our model may similarly enable certain aspects of a cultural group's social task-learning environment to yield consequences about certain aspects of their decision-making, and vice versa. Such a bridge between different objects of study can increase the number of ways we can study each, and thereby contribute to a more comprehensive literature on human cognition.

It is evident that in all lines of inquiry described above, empirical data from contemporary people's learning (including, but not limited to social task-specific learning) could be crucial. Such data can be obtained from lab studies and field studies of the relevant psychological aspects. A prediction of our theory is that these psychological aspects may be evolutionarily affected by independent variables that are specific to social, knowledge-based learning and not to individual learning: even when in ostensibly unambiguous settings of individual learning with costless environmental feedback. Therefore, it may be potentially beneficial for empirical studies of these psychological aspects—even in domains of individual learning—to keep track of potentially social-learning-specific independent variables like the level of knowledge, the speed of learning, and whether the method of learning is imitation or innovation.

Another prediction is that two psychological aspects in particular—task-specific confidence and task/role-model turnover—are evolutionarily related. We thus propose that they should be studied concurrently. In particular, empirical studies should look for our theory's hypothesized, potentially discernable piecewise cutoff point (a “phase transition”) in the student's task-specific confidence, which should exist and coincide

with the identification of the learning type. They should then investigate precisely when this cutoff point—as well as task/role-model turnover—occurs, which should vary with respect to whether the student’s learning method is authentic imitation or de facto innovation in ways that are elucidated by our model.

Lab studies would do well to incorporate the excellent experimental design of Sanchez and Dunning^{203, 204}, which is effective at studying task-specific confidence over the course of learning a high-variance-outcome task over repeated attempts. To arrive at the setting of our model, the Sanchez–Dunning experimental design could be modified to represent an unambiguous setting of task-specific learning via attempted imitation. Ideally, this modified design would achieve a dichotomy between successful imitation and de facto innovation (e.g., by having some role models teach via the Socratic method, and other role models provide actually helpful knowledge: but not to the point of trivializing task learning), include unlearnable tasks (e.g., by having payoffs of unlearnable tasks occur with full randomness that cannot ever be predicted), grant the option of drawing a new task and/or role model, and—just as in the original experiment—offer an incentive-compatible reward. Such an experimental design could then essentially be a parametrization of our learning environment, albeit an artificial one and not an ancestral one. These artificial model parameters, the genetic and cultural-evolutionary background of the experimental subjects, and other potentially relevant treatment effects (independent variables) can then be varied across studies to test the quantitative predictions of our theory regarding task-specific confidence and task/role-model turnover.

Also, on top of such an artificial model parametrization, empirical researchers could add other hypothesized cultural-evolutionary-theoretic mechanisms that would endow its learning environment with an unambiguously social context. Key examples of such

mechanisms include a nontrivial amount of choice in the selection of new tasks and/or role models, the ability to observe the number of other students that have chosen each task/role model, and the ability to exchange information with other students and role models. The inclusion and veridical representation of such mechanisms could be key to investigating cultural-evolutionary-theoretic dynamics that are not fully captured in a decision-theoretic setting such as that of our model.

In addition, empirical researchers could pursue field research of social task-specific learning, especially pertaining to tasks with high-variance payoffs. In contrast to the lab research proposed above, field research would allow for a more veridical representation of social task-specific learning, at the potential expense of experimental controls and granular variation of the independent variables. Doing such field studies in a manner that comprehensively measures all data relevant to our model would be undoubtedly challenging, given that it may need to keep track of every student and role model's interactions, respective levels of experience, respective speeds of learning, respective payoff data, and—if technologically feasible—informative measurements of knowledge. Even if all such data were collected, there may additionally need to be some degree of nontrivial inference from the data to discern certain model parameters: for example, which packages of tasks and role models were learned via successful imitation rather than de facto innovation. Future advances towards improving and widening the collected data in such field studies would potentially help on these fronts.

In both field studies and lab studies investigating descriptive social learning of high-variance-payoff tasks, empirical researchers would do well to take into account the sheer diversity in potential subjects' psychological profiles and treatment effects, which should ideally be recorded as comprehensively as possible in order to keep track

of all potential independent variables²⁵¹. In fact, consider the following two hypotheses. First, subjects who are most likely to be studied by lab research—individuals of Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies—are in important ways psychological outliers relative to the rest of the human population⁹⁴. Second, much of the genus homo’s two-million-year existence was spent in the non-WEIRD lifestyle of mobile foragers²³¹. A consequence is that a comprehensive understanding of descriptive human learning may require studying the social task-learning of mobile foragers whose lifestyles are faithful continuations of their ancestors’: and studying that of non-WEIRD peoples in general. To their credit, field studies are already doing so extensively^{120,128–130,201,202,210}. It may potentially be fruitful to have more of the relevant lab studies, such as the Sanchez–Dunning experimental design (2018, 2020), to also be targeted at individuals of non-WEIRD societies.

Empirical tests of our model’s assumptions themselves would be potentially valuable for the purpose of assessing whether it is a good model of ancestral learning environments. The program to investigate whether social task-learning comprised the primary selection pressure of ancestral human learning is not new. It is a vibrant line of inquiry that constitutes the center of the debate between cultural evolutionary theory and its competing hypotheses¹⁰, whose resolution has potential implications for other debates: like that regarding the hypothesized evolution of moral, norm-based preferences³⁶. Our model contributes new insights that can add to this program. Most notably, it demonstrates that cultural evolutionary theory can explain otherwise puzzling cognitive biases like recurrently non-monotonic confidence. The fact that descriptive human learning is thereby cognitively biased—even in unambiguous settings of individual learning with costless environmental feedback—grants plausibility to cultural evolutionary theory’s hypothesis that the primary selection pressure on ancestral

human cognition was social, knowledge-based task-learning.

Also, our model identifies several potentially relevant mechanisms in a hypothesized learning environment of social, knowledge-based task-learning: for example, the classification of attempted imitation learning into successful imitation and de facto imitation learning, as well as the risk of an unlearnable task in the case of the latter. In particular, it explicitly posits the predictive importance of ecological fitness costs of overcommitting attention, which determine whether the evolutionarily optimal strategy of selective role-model selection meaningfully learns from the relevant payoff data. Our model's formalization of these parameters can augment empirical assessments of cultural evolutionary theory by informing a potentially fruitful avenue of research: specifically, the estimation of these parameters for various, potentially ancestral learning environments; combined with an investigation of cultural evolutionary theory's relevant predictions and of the degree to which these predictions hold. One such prediction from our model (and suitable generalizations of it) would be that when the ecological fitness cost of retaining payoff data is sufficiently high, the optimal strategy of task/role-model turnover would not retain it, and instead rely on other sources of information that are specific to the hypothesized setting of social, knowledge-based task-learning.

A stronger claim of our theory is that the costly cognitive mechanism by which ancestral humans distinguished successful imitation from de facto innovation was a mental time-measurement experiment, to distinguish their respective learning speeds. Our hypothesized existence of such mental time-measurement experiments is a special case of the generally theorized mental evidence-sampling process preceding a decision¹⁸⁷. Empirical tests of our assumption that the speed of imitation is faster than that of innovation, as well as of our assumption that human learners can and do dif-

ferentiate between the two speeds via mental time-measurement, could help probe the plausibility of our theory.

Other plausible hypotheses for the cognitive mechanism by which the student differentiates between imitation and innovation include a costly-to-observe signal effused by the teacher, or one effused by the accumulated task-specific knowledge at any given point of time. Our model can be suitably modified to use such an alternative hypothesis for this cognitive mechanism. In fact, incorporating such an alternative hypothesis would make the model considerably simpler, since it would not need to consider variation in learning speeds. However, a disadvantage of such an alternative hypothesis is that empirically testing it may be less straightforward, at least without relying on neuroscientific methods. We have thus not pursued these alternatively hypothesized mechanisms in the present paper, although we do not rule their veracity out and hope that they may be feasibly testable in the future.

More generally, it may be plausible that future developments in our neuroscientific knowledge will enable a detailed mechanistic understanding of descriptive human learning. While remarkable empirical advances have been made on this front, our current level of neuroscientific understanding has a long way to go, given the extreme complexity of human cognition and the relative adolescence of the field of neuroscience. However, our sketches of potential empirical studies demonstrate that even at our currently limited level of understanding of descriptive human learning, substantive progress—towards testing our theory and in general—may be plausible. Moreover, evolutionary-theoretic hypotheses like those of our model can inform the design, data collection, and analyses of such empirical studies, and thereby partially compensate for the preliminary nature of the current neuroscientific literature. Given the immediate and far-reaching upside of a comprehensive understanding of descrip-

tive human decision-making, we propose that the eventual benefits of a cumulative program of research working towards this goal (even prior to a full neuroscientific understanding) may outweigh the costs.

Acknowledgements

The author thanks Eric Maskin for suggesting the problem and for giving invaluable guidance throughout the course of this work. The author also thanks Joe Henrich for sharing his immense expertise, invaluable advice on the manuscript, and constant encouragement, all of which have been indispensable for this work. Moreover, the author thanks Chris Bakerlee, Rahul Bhui, Matt Cashman, Charles Efferson, Holly Elmore, Ben Golub, Mark Kisin, Philipp Schoenegger, Suzanne Smith, Alex Stewart, Tomasz Strzalecki, Suren Tavakalov, Wesley Wildman, and Alice Zhu for providing many helpful comments on this paper. Additionally, the author thanks members of the Henrich lab for a substantially helpful feedback session. In particular, he is grateful to Matt Cashman for introducing him to the work of Gerd Gigerenzer, Cammie Curtin for introducing him to the work of Charles Efferson, Patricia Greenfield for providing practical advice on the revision process, Ivan Kroupin for suggesting an emphasis on social learning, Graham Noblit for providing detailed feedback on the manuscript, Rachna Reddy for suggesting an emphasis on concrete ecological fitness costs, and TC Zeng for providing detailed feedback on the abstract and on the topic of selective social learning. Finally, the author thanks Joe Henrich for introducing him to the references Henrich, 2009; Muthukrishna et al., 2018, and Singh, 2018; Barbara Finlay for introducing him to Gladwell, 2019 and Yarkoni, 2020; and Will MacAskill for introducing him to Henrich, 2015.

Funding

This work was supported by the National Science Foundation through the Graduate Research Fellowship Program (grant number DGE1745303), the Centre for Effective Altruism through the Global Priorities Fellowship, and Harvard University through graduate student fellowships.

1.5 Appendix

1.5.1 Proof of Proposition 1.1

Note that

$$g_{\rho, \gamma}(b) = \frac{(1 - \gamma) \left(\log \frac{1}{\gamma} \right) \int_b^{\infty} f(a, b) \eta^a da}{\gamma + (1 - \gamma) \left(\log \frac{1}{\gamma} \right) \int_b^{\infty} \eta^a da} \quad (1.104)$$

$$= \frac{\left(\log \frac{1}{\gamma} \right) \int_b^{\infty} f(a, b) \eta^a da}{\frac{\gamma}{1 - \gamma} + \left(\log \frac{1}{\gamma} \right) \int_b^{\infty} \eta^a da}. \quad (1.105)$$

Recall that $\eta \in (0, 1)$, and that $f(\cdot, b)$ is a continuous, non-negative function satisfying $f(b, b) = 1$. It follows that the integral

$$\mathfrak{w} = \left(\log \frac{1}{\eta} \right) \int_b^{\infty} f(a, b) \eta^a da \quad (1.106)$$

is strictly positive, since we can find a positive-measure subset $[b, b + \varepsilon] \subset [b, \infty)$ on which the integrand

$$\left(\log \frac{1}{\eta} \right) f(a, b) \eta^a \quad (1.107)$$

is lower-bounded by a positive constant close to $f(b, b) \eta^b = \eta^b$. Also, the integral

$$\mathfrak{z} = \left(\log \frac{1}{\eta} \right) \int_b^{\infty} \eta^a da \quad (1.108)$$

is strictly positive.

Check that

$$\frac{\partial}{\partial y} g_{\rho_y}(b) = \frac{\partial}{\partial y} \frac{\mathfrak{w}}{\frac{y}{1-y} + \mathfrak{z}} = \frac{-\mathfrak{w} \frac{1}{(1-y)^2}}{\left(\frac{y}{1-y} + \mathfrak{z}\right)^2} = -\frac{\mathfrak{w}}{(y + \mathfrak{z}(1-y))^2} < 0, \quad (1.109)$$

as desired. In particular, $g_{\rho_y}(b)$ is strictly monotonically decreasing in y , which yields the inequalities (1.46) as a corollary.

1.5.2 Proof of Proposition 1.2

To show part (a), we check that

$$\frac{d}{db} g_{\rho_0}(b) \quad (1.110)$$

is positive. First, we apply a change of variables to obtain

$$\begin{aligned} g_{\rho_0}(b) &= \frac{\left(\log \frac{1}{\eta}\right) \int_b^\infty f(a, b) \eta^a da}{\left(\log \frac{1}{\eta}\right) \int_b^\infty \eta^a da} \\ &= \frac{\left(\log \frac{1}{\eta}\right) \int_b^\infty f(a, b) \eta^a da}{\eta^b} \\ &= \left(\log \frac{1}{\eta}\right) \int_b^\infty f(a, b) \eta^{a-b} da \\ &= \left(\log \frac{1}{\eta}\right) \int_0^\infty f(b+m, b) \eta^m dm. \end{aligned} \quad (1.111)$$

This equality can also be deduced from the memorylessness property of the exponential distribution ρ_0 ,

$$\rho_0(a) = \left(\log \frac{1}{\eta}\right) \eta^a. \quad (1.112)$$

Then, we differentiate the expression (1.111) with respect to b by Leibniz's integral

rule, which yields

$$\begin{aligned}\frac{d}{db}g_{\rho_0}(b) &= \frac{d}{db} \left(\left(\log \frac{1}{\eta} \right) \int_0^\infty f(b+m, b) \eta^m dm \right) \\ &= \left(\log \frac{1}{\eta} \right) \int_0^\infty \left(\frac{\partial}{\partial b} f(b+m, b) \right) \eta^m dm.\end{aligned}\tag{1.113}$$

Recall that $\eta \in (0, 1)$, and that $\frac{\partial}{\partial b} f(b+m, b) > 0$ by Assumption 1. Thus, the expression (1.113) is an integral of a positive and continuous function

$$\left(\log \frac{1}{\eta} \right) \left(\frac{\partial}{\partial b} f(b+m, b) \right) \eta^m\tag{1.114}$$

over $[0, \infty)$. Just as in Appendix 1.5.1, we can find a positive-measure subset of $[0, \infty)$ on which the integrand is lower-bounded by a positive constant. Thus, the integral (1.113) is positive, as desired.

To show part (b), observe that

$$g_{\rho_y}(b) = \int_a f(a, b) d\rho_y^{\text{cond}, a>b}(a),\tag{1.115}$$

where $\rho_y^{\text{cond}, a>b}(a)$ denotes the conditional distribution of ρ_y conditional on $a > b$. Its p.d.f. is given by

$$\rho_y^{\text{cond}, a>b}(a) = \frac{\rho_y(a)}{\int_{a>b} d\rho_y(a)} \quad \text{for } a > b.\tag{1.116}$$

Observe that the conditional distribution $\rho_y^{\text{cond}, a>b}$ places probability

$$\rho_y^{\text{cond}, a>b}(\infty) = \frac{\rho_y^{\text{cond}, a>b}(\infty)}{\int_{a>b} d\rho_y^{\text{cond}, a>b}(a)} = \frac{y}{y + \left(\log \frac{1}{\eta} \right) \eta^b} \rightarrow 1\tag{1.117}$$

on $a = \infty$ as $b \rightarrow \infty$. Equivalently, $\rho_y^{\text{cond}, a>b}$ places probability converging to zero on

the subset of finite difficulty values, (b, ∞) , as $b \rightarrow \infty$. Since $f(\infty, b) = 0$, we can apply the dominated convergence theorem to conclude that

$$\begin{aligned}
0 \leq \lim_{b \rightarrow \infty} g_{\rho_y}(b) &= \lim_{b \rightarrow \infty} \left(\int_{a \in (0, \infty)} f(a, b) d\rho_y^{\text{cond}, a > b}(a) + 0 \cdot \rho_y^{\text{cond}, a > b}(\infty) \right) \\
&\leq \lim_{b \rightarrow \infty} \int_{a \in (0, \infty)} d\rho_y^{\text{cond}, a > b}(a) \\
&= \int_{a \in (0, \infty)} \lim_{b \rightarrow \infty} d\rho_y^{\text{cond}, a > b}(a) = \int_{a \in (0, \infty)} 0 da = 0,
\end{aligned}$$

where we have set $\rho_y^{\text{cond}, a > b}(a) = 0$ for $a \leq b$. Thus, we have the desired equality

$$\lim_{b \rightarrow \infty} g_{\rho_y}(b) = 0. \quad (1.118)$$

To show part (c), we use the quotient rule and Leibniz's integral rule:

$$\begin{aligned}
\frac{d}{db} g_{\rho_y}(b) &= \frac{d}{db} \frac{(1-y) \left(\log \frac{1}{\eta} \right) \int_b^\infty f(a, b) \eta^a da}{y + (1-y)\eta^b} \\
&= \frac{1}{(y + (1-y)\eta^b)^2} \\
&\quad \cdot \left((y + (1-y)\eta^b) (1-y) \left(\log \frac{1}{\eta} \right) \left(-\eta^b + \int_b^\infty \frac{\partial}{\partial b} f(a, b) \eta^a da \right) \right. \\
&\quad \left. - (1-y) \left(\log \frac{1}{\eta} \right) \eta^b (1-y) \left(\log \frac{1}{\eta} \right) \int_b^\infty f(a, b) \eta^a da \right).
\end{aligned} \quad (1.119)$$

Assumption 2 implies that that

$$-\eta^b + \int_b^\infty \frac{\partial}{\partial b} f(a, b) \eta^a da, \quad (1.120)$$

and thereby the entire expression (1.119) for $\frac{d}{db}g_{\rho,y}(b)$, is negative for all sufficiently large b , as desired.

1.5.3 Proof of Lemma 1.3

The proof of this lemma solely uses the fact that the unique solution

$$\begin{bmatrix} \hat{V}_{im}(b_{im}, b_{in}) \\ \hat{V}_{in}(b_{im}, b_{in}) \end{bmatrix} \quad (1.121)$$

to a nondegenerate system of equations

$$\begin{bmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{bmatrix} \begin{bmatrix} \hat{V}_{im}(b_{im}, b_{in}) \\ \hat{V}_{in}(b_{im}, b_{in}) \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{f} \end{bmatrix}. \quad (1.122)$$

is given by

$$\begin{bmatrix} \frac{\mathbf{d}\mathbf{e}-\mathbf{b}\mathbf{f}}{\mathbf{a}\mathbf{d}-\mathbf{b}\mathbf{c}} \\ \frac{\mathbf{a}\mathbf{f}-\mathbf{c}\mathbf{e}}{\mathbf{a}\mathbf{d}-\mathbf{b}\mathbf{c}} \end{bmatrix}. \quad (1.123)$$

Substituting the suitable expressions for the quantities \mathbf{a} , \mathbf{b} , \mathbf{c} , \mathbf{d} , \mathbf{e} , and \mathbf{f} completes our proof. Note that

$$\mathbf{g} = \mathbf{a}\mathbf{d} - \mathbf{b}\mathbf{c}. \quad (1.124)$$

1.5.4 Proof of Proposition 1.4

Choose N large enough that the expected payoff deviation due to the procurement of alternative foraging opportunities in the model parametrization $\mathbf{M}(n)$ is less than $\varepsilon/3$ for all $n \geq N$. By possibly making N larger, the expected payoff deviation due to

time-measurement experiments in the model parametrization $\mathbf{M}(n)$ is also less than $\varepsilon/3$.

Furthermore, by possibly making N even larger, the difference between the expected total payoff $V_n(\pi)$ in the model parametrization $\mathbf{M}(n)$ —henceforward excluding deviations due to side opportunities and time-measurement costs—and that of its approximating continuous learning model $\mathbf{M}(\infty)$, given by $V_\infty(\mathbf{b}(\pi))$, is less than $\varepsilon/3$ for all $n \geq N$. To show this, we may as well assume that the task payoff of each learning period (say, the k th one) of the model parametrization $\mathbf{M}(n)$, given by

$$f(a(k), b(k)) \int_{T(k)}^{T(k+1)} \delta^t dt, \quad (1.125)$$

is obtained as a flow payoff of

$$\delta^t f(a(k), b(k)) dt. \quad (1.126)$$

We then define the function

$$\hat{V}_n(b_{im}, b_{in}) = q\hat{V}_{im,n} + (1-q)\hat{V}_{in,n} \quad (1.127)$$

in terms of the function $\hat{V}_{im,n}, \hat{V}_{in,n} : ((0, \infty) \cup \{\infty\})^2 \rightarrow [0, \infty)$, defined by

$$\hat{V}_{im,n}(b_{im}, b_{in}) = \frac{\mathfrak{d}_n \mathfrak{e}_n - \mathfrak{b}_n \mathfrak{f}_n}{\mathfrak{g}_n} \quad (1.128)$$

and

$$\hat{V}_{in,n}(b_{im}, b_{in}) = \frac{\mathfrak{a}_n \mathfrak{f}_n - \mathfrak{c}_n \mathfrak{e}_n}{\mathfrak{g}_n} \quad (1.129)$$

for

$$\mathfrak{a}_n = 1 - q\delta^{L_{im,\infty}^{-1}(b_{im})} \gamma^{b_{im}} \quad (1.130)$$

$$\mathfrak{b}_n = -(1-q)\delta^{L_{im,\infty}^{-1}(b_{im})}\eta^{b_{im}}, \quad (1.131)$$

$$\mathfrak{c}_n = -q\delta^{L_{im,\infty}^{-1}(b_{im})}\left(p + (1-p)\eta^{b_{im}}\right), \quad (1.132)$$

$$\mathfrak{d}_n = 1 - (1-q)\delta^{L_{im,\infty}^{-1}(b_{im})}\left(p + (1-p)\eta^{b_{im}}\right), \quad (1.133)$$

$$\begin{aligned} \mathfrak{e}_n = & \int_0^{b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(a)} \delta^t f(a, L_{im,n}(t)) dt + \int_{L_{im,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{im}(a) \\ & + \int_{a>b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(b_{im})} \delta^t f(a, L_{im,n}(t)) dt \right) d\mu_{im}(a), \end{aligned} \quad (1.134)$$

$$\begin{aligned} \mathfrak{f}_n = & \int_0^{b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(a)} \delta^t f(a, L_{im,n}(t)) dt + \int_{L_{im,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{in}(a) \\ & + \int_{a>b_{im}} \left(\int_0^{L_{im,\infty}^{-1}(b_{im})} \delta^t f(a, L_{im,n}(t)) dt \right) d\mu_{in}(a), \end{aligned} \quad (1.135)$$

and

$$\mathfrak{g}_n = 1 - \delta^{L_{im,\infty}^{-1}(b_{im})}\left(p + (1-p)\eta^{b_{im}}\right) + q\left(\delta^{L_{im,\infty}^{-1}(b_{im})}\left(p + (1-p)\eta^{b_{im}}\right) - \delta^{L_{im,\infty}^{-1}(b_{im})}\eta^{b_{im}}\right). \quad (1.136)$$

By construction, the functions \hat{V}_n have the property that

$$V_n(\pi) = \hat{V}_n(b, b) \quad (1.137)$$

for a policy π represented by $\mathbf{b} = b$, and

$$V_n(\pi) = \hat{V}_n(b_{im}, b_{in}) \quad (1.138)$$

for a policy π represented by $\mathbf{b} = (b_{im}, b_{in})$.

Under the assumption that \hat{V}_∞ and all functions \hat{V}_n are continuous at $b = 0$, we will complete our proof. We have that $\{\hat{V}_n\}_{n \in \mathbb{N}}$ is a sequence of continuous functions on the compact space

$$\bar{\mathcal{Q}} \cup \{(0, 0)\} = \{(b, b) : b \in [0, \beta]\} \cup \{(b_{im}, b_{in}) : b_{im}, b_{in} \in [\beta, \infty) \cup \{\infty\}\} \quad (1.139)$$

that is monotonically converging to \hat{V}_∞ , which is also continuous. Thus, this convergence is uniform by Dini's theorem. In particular, we have

$$\sup_{\pi \in \Pi} |V_n(\pi) - V_\infty(\mathbf{b}(\pi))| \leq \sup_{(b_{im}, b_{in}) \in \bar{\mathcal{Q}} \cup \{(0, 0)\}} |\hat{V}_n(b_{im}, b_{in}) - \hat{V}_\infty(b_{im}, b_{in})| < \frac{\varepsilon}{3} \quad (1.140)$$

for sufficiently large n as desired, where we have used the fact that the set of all strategies \mathbf{b} of the continuous learning model that represent policies $\pi \in \Pi$ of $\mathcal{M}(n)$ is a subset of $\bar{\mathcal{Q}}$. Our overall theorem statement then follows from the triangle inequality.

It remains to show that \hat{V}_∞ (respectively, all functions \hat{V}_n), which are only defined for $b > 0$, can be continuously extended to $b = 0$. For this, it suffices to show that the constituent functions $\hat{V}_{im, \infty}$ and $\hat{V}_{in, \infty}$ (respectively, $\hat{V}_{im, n}$ and $\hat{V}_{in, n}$) can be continuously extended to $b = 0$. Observe that the numerator and denominator of each constituent function are both equal to zero at $b = 0$, which creates the a priori possible obstruction to continuity. However, by L'Hôpital's rule, if both the numerator and the denominator are differentiable at $b = 0$ and the derivative of the denominator

has nonzero value at $b = 0$, then the limit of the function as $b \rightarrow 0$ is well-defined, as desired.

The derivative of the denominator $\mathfrak{g} = \mathfrak{g}_n$ at zero is computed by the product rule and chain rule:

$$\begin{aligned} \frac{d}{db} \mathfrak{g}(b, b)|_{b=0} &= (1-q) \left(\frac{(\log \frac{1}{\delta}) \delta^{L_{in,\infty}^{-1}(0)}}{\frac{d}{dt} L_{in,\infty}(0)} (p + (1-p)\eta^0) + \delta^{L_{in,\infty}^{-1}(0)} (1-p) \left(\log \frac{1}{\eta} \right) \eta^0 \right) \\ &\quad + q \left(\frac{(\log \frac{1}{\delta}) \delta^{L_{in,\infty}^{-1}(0)}}{\frac{d}{dt} L_{in,\infty}(0)} \eta^0 + \delta^{L_{in,\infty}^{-1}(0)} \left(\log \frac{1}{\eta} \right) \eta^0 \right) \end{aligned} \quad (1.141)$$

$$\begin{aligned} &= (1-q) \left(\frac{(\log \frac{1}{\delta})}{\frac{d}{dt} L_{in,\infty}(0)} (p + (1-p)) + (1-p) \left(\log \frac{1}{\eta} \right) \right) \\ &\quad + q \left(\frac{(\log \frac{1}{\delta})}{\frac{d}{dt} L_{in,\infty}(0)} + \left(\log \frac{1}{\eta} \right) \right) > 0. \end{aligned} \quad (1.142)$$

To conclude via the product rule that the derivatives of the numerators of each of the functions $\hat{V}_{im,\infty}$ and $\hat{V}_{in,\infty}$ (respectively, $\hat{V}_{im,n}$, and $\hat{V}_{in,n}$) is well-defined at $b = 0$, it suffices to check whether the derivatives of \mathfrak{e} (respectively, \mathfrak{e}_n) and \mathfrak{f} (respectively, \mathfrak{f}_n) are well-defined at $b = 0$; this is because

$$\mathfrak{a} = \mathfrak{a}_n, \quad (1.143)$$

$$\mathfrak{b} = \mathfrak{b}_n, \quad (1.144)$$

$$\mathfrak{c} = \mathfrak{c}_n, \quad (1.145)$$

and

$$\mathfrak{d} = \mathfrak{d}_n, \quad (1.146)$$

are clearly differentiable via the chain rule. Indeed, Leibniz's integral rule yields that the derivatives of ϵ_n and f_n are well-defined and given at $b = 0$ by

$$\begin{aligned} \frac{d}{db} \epsilon_n|_{b=0} &= \mu_{im}(0) \left(\int_0^{L_{im,\infty}^{-1}(0)} \delta^\epsilon f(a, L_{im,n}(t)) dt + \int_{L_{im,\infty}^{-1}(0)}^\infty \delta^\epsilon dt \right) \\ &\quad - \mu_{im}(0) \int_0^{L_{im,\infty}^{-1}(0)} \delta^\epsilon f(a, L_{im,n}(t)) dt + \int_{a>0} \frac{\delta^{L_{im,\infty}^{-1}(0)} f(a, 0)}{\frac{d}{dt} L_{im,\infty}(0)} d\mu_{im} \\ &= \left(\log \frac{1}{\eta} \right) \frac{1}{\log \frac{1}{\delta}} + \int_{a>0} \frac{\delta^{L_{im,\infty}^{-1}(0)} f(a, 0)}{\frac{d}{dt} L_{im,\infty}(0)} d\mu_{im} \end{aligned} \quad (1.147)$$

and

$$\begin{aligned} \frac{d}{db} f_n|_{b=0} &= \mu_{in}(0) \left(\int_0^{L_{in,\infty}^{-1}(0)} \delta^\epsilon f(a, L_{in,n}(t)) dt + \int_{L_{in,\infty}^{-1}(0)}^\infty \delta^\epsilon dt \right) \\ &\quad - \mu_{in}(0) \int_0^{L_{in,\infty}^{-1}(0)} \delta^\epsilon f(a, L_{in,n}(t)) dt + \int_{a>0} \frac{\delta^{L_{in,\infty}^{-1}(0)} f(a, 0)}{\frac{d}{dt} L_{in,\infty}(0)} d\mu_{in} \\ &= (1-p) \left(\log \frac{1}{\eta} \right) \frac{1}{\log \frac{1}{\delta}} + \int_{a>0} \frac{\delta^{L_{in,\infty}^{-1}(0)} f(a, 0)}{\frac{d}{dt} L_{in,\infty}(0)} d\mu_{in}. \end{aligned} \quad (1.148)$$

The calculations for ϵ and f are analogous—the only difference being that the function $L_{j,n}(t)$ in the integrand is replaced with $L_{j,\infty}(t)$ —and give the identical answers for the derivative at $b = 0$. The product rule thus yields the derivative of the numerators at $b = 0$, as needed.

1.5.5 Proof of Proposition 1.5

For every strategy $\mathbf{b} = (b_{im}, b_{in})$ such that $b_{im} < \infty$, we construct another strategy \mathbf{b}' that achieves a strictly higher value $V_\infty(\mathbf{b}')$. This shows that a necessary condition for $\mathbf{b} = (b_{im}, b_{in})$ to maximize V_∞ is that $b_{im} = \infty$. Note that the constructed strategy \mathbf{b}'

will not be of the form $\mathbf{b}' = (b'_{im}, b'_{in})$, i.e., will not repeat the same quitting strategy for every drawn task.

Consider the probability measure μ^∞ on the sample space of sequences of tasks drawn i.i.d. from μ (some of which may not be drawn if the student quits finitely many times),

$$\Omega = \mathcal{U}^\infty. \quad (1.149)$$

The distribution is defined as follows. Let \mathcal{F} denote the σ -algebra generated by the algebra

$$\mathcal{F}_0 = \bigcup_{n=1}^{\infty} \mathcal{F}_n, \quad (1.150)$$

where \mathcal{F}_n denotes the collection of events whose occurrence can be determined by the results of the first n draws. The probability distribution μ on \mathcal{U} canonically endows \mathcal{F} with a probability measure μ^∞ , which is used to defined the compute the expected value of the payoff.

Let $V_\infty(\mathbf{b}'', \omega)$ denote the total payoff when the student uses a strategy \mathbf{b}'' and the sequence of task types is $\omega \in \Omega$. Then, the total payoff $V_\infty(\mathbf{b}'')$ is given by

$$V_\infty(\mathbf{b}'') = \int_{\Omega} V_\infty(\mathbf{b}'') d\mu^\infty(\omega). \quad (1.151)$$

We modify $\mathbf{b} = (b_{im}, b_{in})$ to obtain the alternative strategy

$$\mathbf{b}' = (([q, b'_{im}; (1-q)p, b'_{in}], b_{in}), (b_{im}, b_{in}), (b_{im}, b_{in}), \dots), \quad (1.152)$$

where the first factor

$$[q, b'_{im}; (1-q)p, b'_{in}] \quad (1.153)$$

denote the probabilistic quitting strategy of, assuming learning has not completed by then, quitting with probability q at

$$b'_{im} = L_{im,\infty} (2L_{im,\infty}^{-1}(b_{im})) \quad (1.154)$$

and quitting with probability $(1 - q)$ at

$$b'_{in} = L_{im,\infty} (L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(b_{in})). \quad (1.155)$$

The probabilistic strategy \mathbf{b}' can be written as a combination of two deterministic strategies:

$$\mathbf{b}'_{im} = ((b'_{im}, b_{in}), (b_{im}, b_{in}), (b_{im}, b_{in}, \dots)) \quad (1.156)$$

with probability q and

$$\mathbf{b}'_{im} = ((b'_{in}, b_{in}), (b_{im}, b_{in}), (b_{im}, b_{in}, \dots)) \quad (1.157)$$

with probability $1 - q$.

We will show that

$$V_\infty(\mathbf{b}) = \int_{\Omega} V_\infty(\mathbf{b}) d\mu^\infty(\omega) \quad (1.158)$$

is strictly less than

$$V_\infty(\mathbf{b}') = \int_{\Omega} V_\infty(\mathbf{b}') d\mu^\infty(\omega), \quad (1.159)$$

thus showing that \mathbf{b}' strictly outperforms \mathbf{b} .

First, we partition the sample space Ω into subsets

$$\Omega = \Omega_1 \cup \Omega_2, \quad (1.160)$$

defined by

$$\Omega_1 = \{\omega = ((j_1, a_1), \dots) : j_1 = in \text{ or } a_1 \leq b_{im}\} \quad (1.161)$$

$$\Omega_2 = \{\omega = ((j_1, a_1), \dots) : j_1 = im \text{ and } a_1 > b_{im}\}. \quad (1.162)$$

Note that

$$\int_{\Omega_1} V_\infty(\mathbf{b}, \omega) d\mu^\infty(\omega) = \int_{\Omega_1} V_\infty(\mathbf{b}', \omega) d\mu^\infty(\omega). \quad (1.163)$$

Indeed, if $j_1 = im$ and $a_1 \leq b_{im}$ for $\omega \in \Omega_1$, then both \mathbf{b} and \mathbf{b}' learn the first task until completion and stick with it forever; and if $j_1 = in$, the strategies \mathbf{b} and \mathbf{b}' play in the same way for such a task sequence ω .

It thus suffices to show that

$$\int_{\omega \in \Omega_2} V_\infty(\mathbf{b}, \omega) d\mu^\infty < \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}', \omega) d\mu^\infty. \quad (1.164)$$

Partition Ω_2 into subsets

$$\Omega_2 = \Omega_3 \cup \Omega_4 \cup \Omega_5 \quad (1.165)$$

defined by

$$\Omega_3 = \{\omega = ((j_1, a_1), (j_2, a_2), \dots) : j_1 = im, a_1 > b_{im}, \text{ and } j_2 = im\} \quad (1.166)$$

$$\Omega_4 = \{\omega = ((j_1, a_1), (j_2, a_2), \dots) : j_1 = im, a_1 > b_{im}, j_2 = in, \text{ and } a_2 < \infty\} \quad (1.167)$$

and

$$\Omega_5 = \{\omega = ((j_1, a_1), (j_2, a_2), \dots) : j_2 = in, a_2 = \infty, j_2 = in, \text{ and } a_2 = \infty\}. \quad (1.168)$$

It suffices to show that

$$\int_{\omega \in \Omega_3} V_\infty(\mathbf{b}, \omega) d\mu^\infty < q \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{im}, \omega) d\mu^\infty, \quad (1.169)$$

$$\int_{\omega \in \Omega_4} V_\infty(\mathbf{b}, \omega) d\mu^\infty < (1-q)(1-p) \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{in}, \omega) d\mu^\infty, \quad (1.170)$$

and

$$\int_{\omega \in \Omega_5} V_\infty(\mathbf{b}, \omega) d\mu^\infty < (1-q)p \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{in}, \omega) d\mu^\infty, \quad (1.171)$$

since \mathbf{b}' plays as the strategy \mathbf{b}'_{im} with probability q (the proportion of Ω_3 in Ω_2) and as the strategy \mathbf{b}'_{in} with probability $(1-q)p$ (the proportion of Ω_4 and Ω_5 combined in Ω_2).

We first show inequality (1.169). Check that the left-hand side is given by

$$\begin{aligned} \int_{\omega \in \Omega_3} V_\infty(\mathbf{b}, \omega) d\mu^\infty &= \int_{\bar{\omega} \in \Omega'_3} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), \bar{\omega})) \left(\int_{\{(im, a_1): a_1 > b_{im}\}} d\mu \right) d\mu^\infty \\ &= q\eta^{b_{im}} \int_{\bar{\omega} \in \Omega'_3} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), \bar{\omega})) d\mu^\infty, \end{aligned} \quad (1.172)$$

where $\bar{\omega} \in \Omega'_3$ parametrizes the task subsequence of $\omega \in \Omega_3$ given by

$$\bar{\omega} = ((j_2, a_2), (j_3, a_3), \dots), \quad (1.173)$$

$b_{im} + \varepsilon$ is an arbitrary task difficulty level greater than b_{im} ,

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), \bar{\omega})) \quad (1.174)$$

does not depend on the choice of $b_{im} + \varepsilon$, and we have an isomorphism of probability spaces

$$\Omega'_3 \cong \{(im, a) : a > 0\} \times \mathcal{U}^\infty. \quad (1.175)$$

Next, check that the right-hand side can be written as

$$\begin{aligned} q \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{im}, \omega) d\mu^\infty &= q \int_{\{(im, a_1) : a_1 > b_{im}\}} \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}', ((j_1, a_1), \hat{\omega})) d\mu^\infty \right) d\mu \\ &= q\eta^{b_{im}} \int_{\{(im, a) : a > 0\}} \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}', ((j_1, a + b_{im}), \hat{\omega})) d\mu^\infty \right) d\mu, \end{aligned} \quad (1.176)$$

where $\hat{\omega} \in \Omega'_2$ parametrizes the task subsequence of $\omega \in \Omega_2$ given by

$$\hat{\omega} = ((j_2, a_2), (j_3, a_3), \dots), \quad (1.177)$$

and we have an isomorphism of probability spaces

$$\Omega'_2 \cong \mathcal{U}^\infty. \quad (1.178)$$

Using the isomorphisms, we reduce our inequality (1.169) to the following:

$$\begin{aligned} &\int_{((j_2, a), (j_3, a_3), \dots) \in \{(im, a') : a' > 0\} \times \mathcal{U}^\infty} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (j_2, a), (j_3, a_3) \dots)) d\mu^\infty \\ &< \int_{((j_1, a), (j_2, a_2), \dots) \in \{(im, a') : a' > 0\} \times \mathcal{U}^\infty} V_\infty(\mathbf{b}', ((j_1, a + b_{im}), (j_2, a_2) \dots)) d\mu^\infty. \end{aligned} \quad (1.179)$$

There is a clear isomorphism of the probability space of task sequences

$$((j_2, a), (j_3, a_3), \dots) \in \{(im, a') : a' > 0\} \times \mathcal{U}^\infty \quad (1.180)$$

and the probability space

$$((j_1, a), (j_2, a_2), \dots) \in \{(im, a') : a' > 0\} \times \mathcal{U}^\infty. \quad (1.181)$$

It suffices to show that the strict inequality holds for the one-to-one-corresponding integrands in this isomorphism, which we will refer to as the left-hand-side value function

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (j_2, a), (j_3, a_3) \dots)) \quad (1.182)$$

and the right-hand-side value function

$$V_\infty(\mathbf{b}', ((j_1, a + b_{im}), (j_2, a_2) \dots)) \quad (1.183)$$

We need to show that

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (im, a), (j'_3, a'_3), \dots)) < V_\infty(\mathbf{b}', ((im, a + b_{im}), (j'_2, a'_2), (j'_3, a'_3), \dots)) \quad (1.184)$$

Note that the sub-payoff values in the subinterval of time

$$[0, L_{im, \infty}^{-1}(b_{im})] \quad (1.185)$$

for both value functions are identical. This is because the first task is of type $j = im$ and is learned to the point of time $L_{im, \infty}^{-1}(b_{im})$ for both value functions.

Also, conditional on the assumption that the task that is learned at time $t = L_{im,\infty}^{-1}(b_{im})$ (second task and first task, respectively) does not learn to completion—that $a_1 < b_{im}$ and $a_1 + b_{im} < b'_{im}$, respectively—the sub-payoff values in the subinterval of time

$$[2L_{im,\infty}^{-1}(b_{im}), \infty) \quad (1.186)$$

are also identical for both value functions. This is because conditional on this assumption, the aforementioned task is quit at time $t = 2L_{im,\infty}^{-1}(b_{im})$, after which the payoff in the remaining time is the same.

Next, we show that if the task that is learned at time $t = L_{im,\infty}^{-1}(b_{im})$ learns to completion for the left-hand-side value function in that $a_1 < b_{im}$, then it also learns to completion for the right hand-side value function in that $a_1 + b_{im} < b'_{im}$. This is a consequence of the assumption that $L_{im,\infty}(t)$ is convex. It follows that at time $t = L_{im,\infty}^{-1}(b_{im})$, the difference a in knowledge that is required to complete the task learning requires less (or equal) time for the right-hand-side value function, spanning

$$t = L_{im,\infty}^{-1}(b_{im}) \quad \text{to} \quad t = L_{im,\infty}^{-1}(b_{im} + a); \quad (1.187)$$

than the time required to complete the task learning for the left-hand-side value function, spanning

$$t = L_{im,\infty}^{-1}(b_{im}) \quad \text{to} \quad t = L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(a). \quad (1.188)$$

Indeed, our assumption that $L_{im,\infty}(t)$ is convex yields the fact that $L_{im,\infty}^{-1}(b)$ is concave, which yields

$$L_{im,\infty}^{-1}(b_{im} + a) \leq L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(a). \quad (1.189)$$

If learning of this task completes for the left-hand-side, then it also completes for the right-hand-side; consequently, no future tasks are drawn, and the sub-payoff values for the subperiod of time (1.186) are equal. On the other hand, if learning of this task completes for the right-hand-side value function, then no future tasks are drawn for it (but may be drawn for the left-hand-side value function); consequently, the sub-payoff values for the subperiod of time (1.186) automatically satisfy the desired direction of inequality.

Moreover, the respective sub-payoff values in the remaining subperiod of time

$$[L_{im,\infty}^{-1}(b_{im}), 2L_{im,\infty}^{-1}(b_{im})] \quad (1.190)$$

are given by

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{2L_{im,\infty}^{-1}(b_{im})} \delta^t \left\{ \begin{array}{ll} f(a, L_{im,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) & \text{for } t < L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(a) \\ 1 & \text{for } t \geq L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(a) \end{array} \right\} dt \quad (1.191)$$

for the left-hand-side value function and

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{2L_{im,\infty}^{-1}(b_{im})} \delta^t \left\{ \begin{array}{ll} f(b_{im} + a, L_{im,\infty}(t)) & \text{for } t < L_{im,\infty}^{-1}(b_{im} + a) \\ 1 & \text{for } t \geq L_{im,\infty}^{-1}(b_{im} + a) \end{array} \right\} dt \quad (1.192)$$

for the right-hand-side value function. It follows from the inequalities (1.189),

$$\begin{aligned} f(a, L_{im,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) &< f(b_{im} + a, b_{im} + L_{im,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) \\ &\leq f(b_{im} + a, L_{im,\infty}(t)), \end{aligned} \quad (1.193)$$

and

$$f(a, b) \leq 1 \tag{1.194}$$

that the sub-payoff value (1.191) of the left-hand-side value function is strictly less than that (1.192) of the right-hand-side value function.

We have overall shown the inequality of integrands (1.184), which implies the inequality (1.176), and thereby, the inequality (1.169).

The second of our desired inequality (1.170) will be shown analogously. Check that the left-hand side is given by

$$\begin{aligned} & \int_{\omega \in \Omega_4} V_\infty(\mathbf{b}, \omega) d\mu^\infty \\ &= \int_{((in, a_2), \bar{\omega}) \in \{(in, a_2): a_2 \in (0, \infty)\} \times \Omega'_4} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) \left(\int_{\{(im, a_1): a_1 > b_{im}\}} d\mu \right) d\mu^\infty \\ &= q\eta^{b_{im}} \int_{((in, a_2), \bar{\omega}) \in \{(in, a_2): a_2 \in (0, \infty)\} \times \Omega'_4} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) d\mu^\infty \\ &= q\eta^{b_{im}} (1 - q)(1 - p) \int_{a_2 \in (0, \infty)} \left(\log \frac{1}{\eta} \right) \eta^{a_2} \left(\int_{\bar{\omega} \in \Omega'_4} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) d\mu^\infty \right) da_2, \end{aligned} \tag{1.195}$$

where $\bar{\omega} \in \Omega'_4$ parametrizes the task subsequence of $\omega \in \Omega_4$,

$$\bar{\omega} = ((j_3, a_3), (j_4, a_4), \dots), \tag{1.196}$$

$b_{im} + \varepsilon$ is an arbitrary task difficulty level greater than b_{im} ,

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) \tag{1.197}$$

does not depend on the choice of $b_{im} + \varepsilon$, and we have an isomorphism of probability

spaces

$$\Omega'_4 \cong \mathcal{U}^\infty. \quad (1.198)$$

Next, check that the right-hand-side inequality can be written as

$$\begin{aligned} & (1-q)(1-p) \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{in}, \omega) d\mu^\infty \\ &= (1-q)(1-p) \int_{\{(im, a_1): a_1 > b_{im}\}} \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}'_{in}, ((im, a_1), \hat{\omega})) d\mu^\infty \right) d\mu \\ &= (1-q)(1-p) q \eta^{b_{im}} \int_{a \in (0, \infty)} \left(\log \frac{1}{\eta} \right) \eta^a \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}'_{in}, ((im, a + b_{im}), \hat{\omega})) d\mu^\infty \right) da \end{aligned} \quad (1.199)$$

Using the isomorphisms (1.219) and (1.175), we reduce our inequality (1.170) to

$$\begin{aligned} & \int_{(a, (j_3, a_3), \dots) \in (0, \infty) \times \mathcal{U}^\infty} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a), (j_3, a_3) \dots)) d\mu^\infty d\mu_\eta \\ & < \int_{(a, (j_2, a_2), \dots) \in (0, \infty) \times \mathcal{U}^\infty} V_\infty(\mathbf{b}', ((im, a + b_{im}), (j_2, a_2) \dots)) d\mu^\infty d\mu_\eta, \end{aligned} \quad (1.200)$$

where $\mu_\eta = \mu_{im} = \mu_{in}|_{a < \infty}$ denotes the exponential distribution of decay factor η on $(0, \infty)$.

There is a clear isomorphism of the probability space of task sequences

$$(a, (j_3, a_3), \dots) \in (0, \infty) \times \mathcal{U}^\infty \quad (1.201)$$

and the probability space

$$(a, (j_2, a_2), \dots) \in (0, \infty) \times \mathcal{U}^\infty. \quad (1.202)$$

It suffices to show that the strict inequality holds for the one-to-one-corresponding integrands in this isomorphism, which we will refer to as the left-hand-side value function

$$V_{\infty}(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a), (j_3, a_3) \dots)) \quad (1.203)$$

and the right-hand-side value function

$$V_{\infty}(\mathbf{b}', ((im, a + b_{im}), (j_2, a_2) \dots)). \quad (1.204)$$

We need to show that

$$V_{\infty}(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a), (j_3, a_3) \dots)) < V_{\infty}(\mathbf{b}', ((im, a + b_{im}), (j_2, a_2) \dots)). \quad (1.205)$$

Just as before, the sub-payoff-values in the subinterval of time

$$[0, L_{im, \infty}^{-1}(b_{im})] \quad (1.206)$$

for both value functions are identical.

Also, similarly to before, conditional on the assumption that task that is learned at time $t = L_{im, \infty}^{-1}(b_{im})$ (second task and first task, respectively) does not learn to completion—that $a_1 < b_{in}$ and $a_1 + b_{im} < b'_{in}$, respectively—the sub-payoff values in the subinterval of time

$$[L_{im, \infty}^{-1}(b_{im}) + L_{in, \infty}^{-1}(b_{in}), \infty) \quad (1.207)$$

are identical for both value functions.

Next, we show that if the task that is learned at time $t = L_{im, \infty}^{-1}(\infty)$ learns to completion for the left-hand side value function in that $a_1 < b_{in}$, then it also learns to com-

pletion for the right-hand-side value function in that $a_1 + b_{im} < b'_{im}$. This is a consequence of two assumptions: the assumption that $L_{im,\infty}(t)$ is convex (equivalently, that $L_{im,\infty}^{-1}(b)$ is concave) and the assumption that $L_{in,\infty}(t) \leq L_{im,\infty}(t)$ (equivalently, that $L_{im,\infty}^{-1}(b) \leq L_{in,\infty}^{-1}(b)$). It follows that at time $t = L_{im,\infty}^{-1}(b_{im})$, the difference a in knowledge that is required to complete the task learning requires less (or equal) time for the right-hand-side value function, spanning

$$t = L_{im,\infty}^{-1}(b_{im}) \quad \text{to} \quad t = L_{im,\infty}^{-1}(b_{im} + a), \quad (1.208)$$

than the time required to complete the task learning for the left-hand-side value function, spanning

$$t = L_{im,\infty}^{-1}(b_{im}) \quad \text{to} \quad t = L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(a), \quad (1.209)$$

Indeed, our two aforementioned assumptions together yield

$$L_{im,\infty}^{-1}(b_{im} + a) \leq L_{im,\infty}^{-1}(b_{im}) + L_{im,\infty}^{-1}(a) \leq L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(a). \quad (1.210)$$

If learning of this task completes for the left-hand-side, then it also completes for the right-hand-side; consequently, no future tasks are drawn, and the sub-payoff values for the subperiod of time (1.207) are equal. On the other hand, if learning of this task completes for the right-hand-side value function, then no future tasks are drawn for it (but may be drawn for the left-hand-side value function); consequently, the sub-payoff values for the subperiod of time (1.207) automatically satisfy the desired direction of inequality.

Finally, the respective sub-payoff values in the remaining subperiod of time

$$[L_{im,\infty}^{-1}(b_{im}), L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(b_{in})] \quad (1.211)$$

are given by

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{L_{im,\infty}^{-1}(b_{im})+L_{in,\infty}^{-1}(b_{in})} \mathcal{J} \left\{ \begin{array}{ll} f(a, L_{in,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) & \text{for } t < L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(a) \\ 1 & \text{for } t \geq L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(a) \end{array} \right\} dt \quad (1.212)$$

for the left-hand-side value function and

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{L_{im,\infty}^{-1}(b_{im})+L_{in,\infty}^{-1}(b_{in})} \mathcal{J} \left\{ \begin{array}{ll} f(b_{im} + a, L_{im,\infty}(t)) & \text{for } t < L_{im,\infty}^{-1}(b_{im} + a) \\ 1 & \text{for } t \geq L_{im,\infty}^{-1}(b_{im} + a) \end{array} \right\} dt \quad (1.213)$$

for the right-hand-side value function. It follows from the inequalities (1.210),

$$\begin{aligned} f(a, L_{in,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) &\leq f(a, L_{im,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) \\ &< f(b_{im} + a, L_{im,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) \\ &\leq f(a, L_{im,\infty}(t)), \end{aligned} \quad (1.214)$$

and $f(a, b) \leq 1$ that the sub-payoff value (1.212) of the left-hand-side value function is strictly less than that (1.213) of the right-hand-side value function.

Overall, we have shown the inequality of integrands (1.205), which implies the inequality (1.200) and thereby, the inequality (1.170).

It remains to show the inequality (1.171). Check that the left-hand side is given by

$$\int_{\omega \in \Omega_5} V_\infty(\mathbf{b}, \omega) d\mu^\infty$$

$$\begin{aligned}
&= \int_{((in,a_2),\bar{\omega}) \in \{(in,\infty)\} \times \Omega'_5} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) \left(\int_{\{(im,a_1):a_1 > b_{im}\}} d\mu \right) d\mu^\infty \\
&= q\eta^{b_{im}} \int_{((in,a_2),\bar{\omega}) \in \{(in,\infty)\} \times \Omega'_5} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, a_2), \bar{\omega})) d\mu^\infty \\
&= q\eta^{b_{im}}(1-q)p \left(\int_{\bar{\omega} \in \Omega'_5} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, \infty), \bar{\omega})) d\mu^\infty \right) \tag{1.215}
\end{aligned}$$

$$= q\eta^{b_{im}}(1-q)p \int_{\bar{a} \in (0,\infty)} \left(\int_{\bar{\omega} \in \Omega'_5} V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, \infty), \bar{\omega})) d\mu^\infty \right) d\mu_\gamma, \tag{1.216}$$

where $\bar{\omega} \in \Omega'_5$ parametrizes the task subsequence of $\omega \in \Omega_5$,

$$\bar{\omega} = ((j_3, a_3), (j_4, a_4), \dots), \tag{1.217}$$

$b_{im} + \varepsilon$ is an arbitrary task difficulty level greater than b_{im} ,

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, \infty), \bar{\omega})) \tag{1.218}$$

does not depend on the choice of $b_{im} + \varepsilon$, we have an isomorphism of probability spaces

$$\Omega'_5 \cong \mathcal{U}^\infty, \tag{1.219}$$

and \bar{a} , distributed as μ_γ , is a dummy variable. Next, check that the right-hand side of the inequality

$$\begin{aligned}
&(1-q)p \int_{\omega \in \Omega_2} V_\infty(\mathbf{b}'_{in}, \omega) d\mu^\infty \\
&= (1-q)p \int_{\{(im,a_1):a_1 > b_{im}\}} \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}'_{in}, ((im, a_1), \hat{\omega})) d\mu^\infty \right) d\mu \\
&= (1-q)pq\eta^{b_{im}} \int_{a \in (0,\infty)} \left(\int_{\hat{\omega} \in \Omega'_2} V_\infty(\mathbf{b}'_{in}, ((im, a + b_{im}), \hat{\omega})) d\mu^\infty \right) d\mu_\gamma. \tag{1.220}
\end{aligned}$$

There is a clear isomorphism of the probability space of task sequences

$$(a, (j_3, a_3), \dots) \in (0, \infty) \times \mathcal{U}^\infty \quad (1.221)$$

and the probability space

$$(a, (j_2, a_2), \dots) \in (0, \infty) \times \mathcal{U}^\infty. \quad (1.222)$$

It suffices to show that the strict inequality holds for the one-to-one-corresponding integrands in this isomorphism, which we will refer to as the left-hand-side value function

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, \infty), (j_3, a_3) \dots)) \quad (1.223)$$

and the right-hand-side value function

$$V_\infty(\mathbf{b}', ((im, a + b_{im}), (j_2, a_2) \dots)). \quad (1.224)$$

We need to show that

$$V_\infty(\mathbf{b}, ((im, b_{im} + \varepsilon), (in, \infty), (j_3, a_3) \dots)) < V_\infty(\mathbf{b}', ((im, a + b_{im}), (j_2, a_2) \dots)). \quad (1.225)$$

Just as before, the sub-payoff-values in the subinterval of time

$$[0, L_{im, \infty}^{-1}(b_{im})) \quad (1.226)$$

for both value functions are identical.

Also, similarly to before, conditional on the assumption that task that is learned

at time $t = L_{im,\infty}^{-1}(b_{im})$ (second task and first task, respectively) does not learn to completion—that $a_1 < b_{in}$ and $a_1 + b_{im} < b'_{in}$, respectively—the sub-payoff values in the subinterval of time

$$[L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(b_{in}), \infty) \quad (1.227)$$

are identical for both value functions.

Next, we show that if the task that is learned at time $t = L_{im,\infty}^{-1}(\infty)$ learns to completion for the left-hand side value function in that $a_1 < b_{in}$, then it also learns to completion for the right-hand-side value function in that $a_1 + b_{im} < b'_{in}$. In fact, note that learning for this task can never complete for the left-hand-side value function, since the task difficulty is $a = \infty$. Thus, this step is trivially satisfied. consequently, the sub-payoff values for the subperiod of time (1.227) automatically satisfy the desired direction of inequality.

Finally, the respective sub-payoff values in the remaining subperiod of time

$$[L_{im,\infty}^{-1}(b_{im}), L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(b_{in}),] \quad (1.228)$$

are given by

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(b_{in})} \delta^t f(\infty, L_{in,\infty}(t - L_{im,\infty}^{-1}(b_{im}))) dt = 0 \quad (1.229)$$

for the left-hand-side value function and

$$\int_{L_{im,\infty}^{-1}(b_{im})}^{L_{im,\infty}^{-1}(b_{im}) + L_{in,\infty}^{-1}(b_{in})} \delta^t \left\{ \begin{array}{ll} f(b_{im} + a, L_{im,\infty}(t)) & \text{for } t < L_{im,\infty}^{-1}(b_{im} + a) \\ 1 & \text{for } t \geq L_{im,\infty}^{-1}(b_{im} + a) \end{array} \right\} dt \quad (1.230)$$

for the right-hand-side value function. It follows that the sub-payoff value (1.229) of

the left-hand-side value function is strictly less than that (1.230) of the right-hand-side value function.

We have overall shown (1.225). This shows the inequality (1.220), and thereby, the desired inequality (1.171). This completes our proof.

1.5.6 Proof of Proposition 1.6

We use a similar proof strategy as that of the proof of Proposition 1.5. For every strategy $\mathbf{b} = (b_{im}, \infty)$, we construct another strategy \mathbf{b}' , not of the form $\mathbf{b}' = (b'_{im}, b'_{in})$, that achieves a strictly higher value $V_\infty(\mathbf{b}')$. This shows that a necessary condition for $\mathbf{b} = (b_{im}, b_{in})$ to maximize V_∞ is that $b_{in} = \infty$.

The modification \mathbf{b}' is defined by

$$\mathbf{b}' = ((b_{im}, b'), (b_{im}, \infty), (b_{im}, \infty), \dots), \quad (1.231)$$

for a value b' that will be specified later.

We partition the sample space Ω into subsets

$$\Omega = \Omega_6 \cup \Omega_7 \quad (1.232)$$

defined by

$$\Omega_6 = \{\omega = ((j_1, a_1), \dots) : j_1 = im \text{ or } a_1 \leq b'\} \quad (1.233)$$

and

$$\Omega_7 = \{\omega = ((j_1, a_1), \dots) : j_1 = in \text{ and } a_1 > b'\}. \quad (1.234)$$

Note that

$$\int_{\Omega_6} V_\infty(\mathbf{b}, \omega) d\mu^\infty = \int_{\Omega_6} V_\infty(\mathbf{b}', \omega) d\mu^\infty \quad (1.235)$$

Indeed, if $j_1 = in$ and $a_1 \leq b'$ for $\omega \in \Omega_1$, then both \mathbf{b} and \mathbf{b}' learn the first task until completion and stick with it forever; and if $j_1 = im$, the strategies \mathbf{b} and \mathbf{b}' play in the same way for such a task sequence ω .

It thus suffices to show that

$$\int_{\omega \in \Omega_7} V_\infty(\mathbf{b}, \omega) d\mu^\infty < \int_{\omega \in \Omega_7} V_\infty(\mathbf{b}', \omega) d\mu^\infty. \quad (1.236)$$

Observe that for each $\omega \in \Omega_7$, the sub-payoff value of the value function $V_\infty(\mathbf{b}, \omega)$ and that of the value function $V_\infty(\mathbf{b}', \omega)$ for the subperiod of time

$$[0, L_{in, \infty}^{-1}(b')] \quad (1.237)$$

are identical, since both strategies learn the first task during this subperiod.

The key insight is that the sub-payoff value of the value function $V_\infty(\mathbf{b}', \omega)$ in the remaining subperiod of time

$$[L_{in, \infty}^{-1}(b'), \infty) \quad (1.238)$$

is always given by

$$\partial^{L_{in, \infty}^{-1}(b')} V_\infty(\mathbf{b}'', \bar{\omega}), \quad (1.239)$$

where

$$\mathbf{b}'' = ((b_{im}, \infty), (b_{im}, \infty), \dots) \quad (1.240)$$

and

$$\bar{\omega} = ((j_2, a_2), \dots) \quad (1.241)$$

are obtained from \mathbf{b}' and ω , respectively, by truncating the leftmost term. It follows that the integral of this sub-payoff value over Ω_7 is

$$\begin{aligned} \int_{\omega \in \Omega_7} \delta^{L_{in,\infty}^{-1}(b')} V_\infty(\mathbf{b}'', \bar{\omega}) d\mu^\infty &= \delta^{L_{in,\infty}^{-1}(b')} \int_{\{(in,a_1): a_1 > b'\}} \int_{\bar{\omega} \in \mathcal{U}^\infty} V_\infty(\mathbf{b}'', \bar{\omega}) d\mu^\infty d\mu \\ &= \delta^{L_{in,\infty}^{-1}(b')} \left(p + (1-p)\eta^{b'} \right) V_\infty(\mathbf{b}''). \end{aligned} \quad (1.242)$$

In contrast, the integral of the sub-payoff value of the value function $V_\infty(\mathbf{b}, \omega)$ in the subperiod (1.238) is given by

$$\begin{aligned} &\int_{\omega \in \Omega_7} \left(\int_{L_{in,\infty}^{-1}(b')}^\infty f(a_1, L_{in,\infty}(t)) dt \right) d\mu^\infty \\ &= \int_{\{\omega \in \Omega_7: a_1 = \infty\}} \left(\int_{L_{in,\infty}^{-1}(b')}^\infty f(a_1, L_{in,\infty}(t)) dt \right) d\mu^\infty \\ &\quad + \int_{\{\omega \in \Omega_7: a_1 < \infty\}} \left(\int_{L_{in,\infty}^{-1}(b')}^\infty f(a_1, L_{in,\infty}(t)) dt \right) d\mu^\infty \\ &= \int_{\{\omega \in \Omega_7: a_1 < \infty\}} \left(\int_{L_{in,\infty}^{-1}(b')}^\infty f(a_1, L_{in,\infty}(t)) dt \right) d\mu^\infty \\ &\leq \int_{\{\omega \in \Omega_7: a_1 < \infty\}} d\mu^\infty \left(\int_{L_{in,\infty}^{-1}(b')}^\infty 1 dt \right) \\ &= \left((1-p)\eta^{b'} \right) \delta^{L_{in,\infty}^{-1}(b')} \frac{1}{\log \frac{1}{\delta}}. \end{aligned} \quad (1.243)$$

Note that as $b' \rightarrow \infty$, the expression (3.102) divided by $\delta^{L_{in,\infty}^{-1}(b')}$ converges to

$$pV_\infty(\mathbf{b}'') > 0, \quad (1.244)$$

whereas the upper bound (1.243) divided by $\delta^{L_{in,\infty}^{-1}(b')}$ converges to 0. This shows that for b' sufficiently large, the inequality (1.236) holds.

For $\mathbf{b} = b = \infty$, since $V_\infty(b) = V_\infty(b, b)$, we can modify the strategy $(b, b) = (\infty, \infty)$ in the same way as above (for a sufficiently large b') to find a strategy that outperforms \mathbf{b} .

1.5.7 Proof of Corollary 1.7

Let $\hat{V}_\infty^{\bar{p}, \bar{q}}(b_{im}, b_{in})$ denote the expression $\hat{V}_\infty(b_{im}, b_{in})$ when the parameter choices $p = \bar{p}$ and $q = \bar{q}$ are made. Similarly, let $\hat{V}_\infty^{\bar{p}, \bar{q}, \bar{L}_{im, \infty}, \bar{L}_{in, \infty}}(b_{im}, b_{in})$ denote the expression $\hat{V}_\infty(b_{im}, b_{in})$ when the parameter choices $p = \bar{p}$, $q = \bar{q}$, $L_{im, \infty} = \bar{L}_{im, \infty}$ and $L_{in, \infty} = \bar{L}_{in, \infty}$ are made. When parameters $L_{im, \infty}$ and $L_{in, \infty}$ are omitted from the superscript, the meaning is that they are assumed to be the original fixed ones.

For each of part (a) and part (b), we will show a stronger statement than the theorem statement. Specifically, we will show that for any pair of decreasing sequences $\{p_n\}_{n \in \mathbb{N}}$ and $\{q_m\}_{m \in \mathbb{N}}$ converging to zero, there exists N such that for any $n \geq N$, we can find M_n such that the quitting point of innovation-learning tasks b_{in} of any strategy $\mathbf{b} = (\infty, b_{in})$ maximizing $V_\infty^{p_n, q_m}$ is greater than γ for all $m \geq M_n$.

We note that the sequence of continuous functions $\{\hat{V}_\infty^{p_n, 0}\}_{n \in \mathbb{N}}$ pointwise converge to the continuous function $\hat{V}_\infty^{0, 0}$. By part (a) of Lemma 1.8, this sequence of continuous functions in fact monotonically converges (increasing with respect to n) to $\hat{V}_\infty^{0, 0}$. An application of Dini's theorem thus shows that the convergence of $\{\hat{V}_\infty^{p_n, 0}\}_{n \in \mathbb{N}}$ to $\hat{V}_\infty^{0, 0}$ on the compact space $\bar{Q} \cup \{(0, 0)\}$ is uniform.

The proof of Proposition 1.5 implies that the maximum of $\hat{V}_\infty^{0, 0}$ on $\bar{Q} \cup \{(0, 0)\}$ is attained at $(b_{im}, b_{in}) = (b_{\text{arbitrary}}, \infty)$; note that the subscript ‘‘arbitrary’’ means the

choice of that parameter has no effect. Indeed, check that

$$\hat{V}_{\infty}^{0,0}(b_{\text{arbitrary}}, b) = \hat{V}_{\infty}^{\mathcal{P}_{\text{arbitrary}},1,L_{in,\infty},L_{in,\infty}}(b, b'_{\text{arbitrary}}), \quad (1.245)$$

since when $p = 0$, we have the equality of distributions $\mu_{in} = \mu_{in}$; and we have set the learning function of imitation-learning tasks to be $L_{in,\infty}$ as well. The proof of Proposition 1.5 shows that we have

$$\hat{V}_{\infty}^{\mathcal{P}_{\text{arbitrary}},1,L_{in,\infty},L_{in,\infty}}(b, b) < \hat{V}_{\infty}^{\mathcal{P}_{\text{arbitrary}},1,L_{in,\infty},L_{in,\infty}}(\infty, b), \quad (1.246)$$

where we have set $b'_{\text{arbitrary}} = b$. This shows that $(\infty, b'_{\text{arbitrary}})$ maximizes the function $\hat{V}_{\infty}^{\mathcal{P}_{\text{arbitrary}},1,L_{in,\infty},L_{in,\infty}}$; equivalently, $(b_{\text{arbitrary}}, \infty)$ maximizes the function $\hat{V}_{\infty}^{0,0}$.

We will now avoid the use of the term $b_{\text{arbitrary}}$, and instead define

$$\tilde{V}(b) = \hat{V}_{\infty}^{0,0}(b_{\text{arbitrary}}, b). \quad (1.247)$$

Let $\gamma \geq 0$. Consider the positive number

$$\varepsilon = \tilde{V}(\infty) - \max_{b \leq \gamma} \tilde{V}(b). \quad (1.248)$$

By uniform convergence, there exists N such that for all $n \geq N$, we simultaneously have

$$|\hat{V}_{\infty}^{\mathcal{P}_n,0}(b_1, \infty) - \hat{V}_{\infty}^{0,0}(b_1, \infty)| < \frac{\varepsilon}{2} \quad (1.249)$$

and

$$|\hat{V}_{\infty}^{\mathcal{P}_n,0}(b_2, b) - \hat{V}_{\infty}^{0,0}(b_2, b)| < \frac{\varepsilon}{2} \quad (1.250)$$

for all $b \leq \gamma$. Since

$$\hat{V}_\infty^{0,0}(b_1, b) = \tilde{V}(b) < \tilde{V}(\infty) < \hat{V}_\infty^{0,0}(b_2, \infty) \quad (1.251)$$

for every $b \leq \gamma$ with a difference of at least ε , it follows from the triangle inequality that for any $n \geq N$, we have

$$\hat{V}_\infty^{\mathcal{P}_n,0}(b_2, b) < \hat{V}_\infty^{\mathcal{P}_n,0}(b_1, \infty) \quad (1.252)$$

for all $b \leq \gamma$. Note that the choice of b_1 and b_2 has no effect on the values $\hat{V}_\infty^{\mathcal{P}_n,0}(b_2, b)$ and $\hat{V}_\infty^{\mathcal{P}_n,0}(b_1, \infty)$.

Fix $n \geq N$. By part (b) of Lemma 1.8, the continuous functions $\{\hat{V}_\infty^{\mathcal{P}_n, q_m}\}_{m \in \mathbb{N}}$ monotonically converge (decreasing with respect to m) to $\hat{V}_\infty^{\mathcal{P}_n,0}$, which is also continuous. It thus follows from Dini's theorem that the convergence of $\{\hat{V}_\infty^{\mathcal{P}_n, q_m}\}_{m \in \mathbb{N}}$ to $\hat{V}_\infty^{\mathcal{P}_n,0}$ on the compact space $\bar{Q} \cup \{(0,0)\}$ is uniform. Let

$$\varepsilon = \hat{V}_\infty^{\mathcal{P}_n,0}(b_1, \infty) - \sup_{b \leq \gamma} \hat{V}_\infty^{\mathcal{P}_n,0}(b_2, b), \quad (1.253)$$

which is positive by our choice of n . By uniform convergence, there exists M_n such that for all $m \geq M_n$, we simultaneously have the inequality

$$|\hat{V}_\infty^{\mathcal{P}_n, q_m}(\infty, \infty) - \hat{V}_\infty^{\mathcal{P}_n,0}(\infty, \infty)| < \frac{\varepsilon}{2}, \quad (1.254)$$

where we have set $b_1 = \infty$; and the inequality

$$|\hat{V}_\infty^{\mathcal{P}_n, q_m}(b_2, b) - \hat{V}_\infty^{\mathcal{P}_n,0}(b_2, b)| < \frac{\varepsilon}{2} \quad (1.255)$$

for any $(b_2, b) \in \bar{\mathcal{Q}}$. Since

$$\hat{V}_\infty^{\mathcal{P}_n, 0}(b_2, b) < \hat{V}_\infty^{\mathcal{P}_n, 0}(\infty, \infty) \quad (1.256)$$

for all $b \leq \gamma$ with a difference of at least ε , it follows from the triangle inequality that for any $m \geq M_n$, we have

$$\hat{V}_\infty^{\mathcal{P}_n, q_m}(b_2, b) < \hat{V}_\infty^{\mathcal{P}_n, q_m}(\infty, \infty) \quad (1.257)$$

for all $b \leq \gamma$. Setting $b_2 = \infty$ yields part (b), while setting $b_2 = b$ yields part (a).

In this proof, we have applied the proof of Proposition 1.5 to show a necessary lemma that can be described as the following. In the continuous learning model with parameters $q = 1$ and $L_{im, \infty}(t) = L(t)$, the unique strategy to maximize $V_\infty(\mathbf{b})$ is $\mathbf{b} = \infty$, where the single entry denotes that there is no ambiguity in learning types. Strictly speaking, the proof of Proposition 1.5 applied to this continuous learning game only shows that $\mathbf{b} = \infty$ outperforms $\mathbf{b}' = b \in (0, \infty)$, and not necessarily $b \rightarrow 0$. This leaves the possibility that $V_\infty(b)$ is also maximized at $b \rightarrow 0$, with the same function value as $b = \infty$. This implies that V_∞ is decreasing near $b = 0$. However, a quick application of the proof of Proposition 1.5 shows that this possibility does not arise. Specifically, this proof shows that for small $b > 0$, the strategy $\mathbf{b}' = b$ is strictly outperformed by

$$\mathbf{b}'' = (L(2L^{-1}(b)), b, b, \dots), \quad (1.258)$$

which—as a subsequent application of this proof shows—is strictly outperformed by

$$\mathbf{b}''' = (L(2L^{-1}(b)), L(2L^{-1}(b)), b, b, \dots). \quad (1.259)$$

Continuing iteratively, we obtain that \mathbf{b}' is strictly outperformed by the strategy

$$\hat{\mathbf{b}} = L(2L^{-1}(b)). \quad (1.260)$$

Taking b to be small, we see that $\tilde{V}(0) = \lim_{b \rightarrow 0} V(b)$ cannot be decreasing near $b = 0$, and thus it is impossible that the maximum is attained at the two endpoints $b = 0$ and $b = \infty$.

1.5.8 Proof of Lemma 1.8

Recall that

$$\hat{V}_\infty(b_{im}, b_{in}) = q\hat{V}_{im}(b_{im}, b_{in}) + (1-q)\hat{V}_{in}(b_{im}, b_{in}) \quad (1.261)$$

for

$$\hat{V}_{im}(b_{im}, b_{in}) = \frac{\mathfrak{d}\mathfrak{e} - \mathfrak{b}\mathfrak{f}}{\mathfrak{g}} \quad (1.262)$$

and

$$\hat{V}_{in}(b_{im}, b_{in}) = \frac{\mathfrak{a}\mathfrak{f} - \mathfrak{c}\mathfrak{e}}{\mathfrak{g}}. \quad (1.263)$$

First, we show that

$$\frac{\partial}{\partial p} \hat{V}_\infty(b_{im}, b_{in}) \leq 0, \quad (1.264)$$

with equality if and only if $q = 1$. Note that the only one of \mathfrak{a} , \mathfrak{b} , \mathfrak{c} , \mathfrak{d} , \mathfrak{e} , \mathfrak{f} , and \mathfrak{g} that is not constant with respect to p is \mathfrak{f} . We thus have

$$\frac{\partial}{\partial p} \hat{V}_\infty(b_{im}, b_{in}) = \frac{\partial}{\partial p} \hat{V}_{in}(b_{im}, b_{in}) = \frac{-q\mathfrak{b} + (1-q)\mathfrak{a}}{\mathfrak{g}} \left(\frac{\partial}{\partial p} \mathfrak{f} \right). \quad (1.265)$$

Check that

$$\begin{aligned}
-q\mathbf{b} + (1-q)\mathbf{a} &= q(1-q)\delta_{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} + 1-q - q(1-q)\delta_{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} \\
&\geq 1-q - q(1-q)\delta_{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} \\
&\geq 1-q - q(1-q) = (1-q)^2,
\end{aligned} \tag{1.266}$$

which is nonnegative, and positive if and only if $q < 1$. Check also that

$$\mathbf{g} = 1-q - (1-q)\delta_{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right) + q - q\delta_{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} \geq 0. \tag{1.267}$$

with equality if and only if $(b_{in}, b_{in}) = (0, 0)$, since

$$\delta_{L_{in,\infty}^{-1}(b_{in})} \left(p + (1-p)\eta^{b_{in}} \right), \delta_{L_{in,\infty}^{-1}(b_{in})} \eta^{b_{in}} \leq 1. \tag{1.268}$$

Then, check that

$$\begin{aligned}
\frac{\partial}{\partial p} \mathbf{f} &= \frac{\partial}{\partial p} \left(p \cdot 0 + (1-p) \left(\int_0^{b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{\eta}(a) \right. \right. \\
&\quad \left. \left. + \int_{a>b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(b_{in})} \delta^t f(a, L_{in,\infty}(t)) dt \right) d\mu_{\eta}(a) \right) \right) \\
&= - \left(\int_0^{b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^{\infty} \delta^t dt \right) d\mu_{\eta}(a) \right. \\
&\quad \left. + \int_{a>b_{in}} \left(\int_0^{L_{in,\infty}^{-1}(b_{in})} \delta^t f(a, L_{in,\infty}(t)) dt \right) d\mu_{\eta}(a) \right) \leq 0,
\end{aligned} \tag{1.269}$$

with equality if and only if $b_{in} = 0$, which is equivalent to $(b_{in}, b_{in}) = (0, 0)$ in our domain $\bar{\mathcal{Q}} \cup \{(0, 0)\}$. Finally, it follows from the calculation via L'Hôpital's rule in

the proof of Proposition 1.4 that

$$\frac{\frac{\partial}{\partial p} \mathfrak{f}}{\mathfrak{g}} \rightarrow \frac{\log \frac{1}{\eta}}{(\log \frac{1}{\delta}) \frac{d}{db} \mathfrak{g}(b, b)|_{b=0}} > 0 \quad (1.270)$$

as $b \rightarrow 0$ for $(b_{im}, b_{in}) = (b, b)$. The condition $(b_{im}, b_{in}) = (0, 0)$ does not make (1.265) zero.

We thus have shown (1.264), where equality holds if and only if $q = 1$.

Next, suppose that Assumption 1 holds and the imitation-learning knowledge function $L_{im, \infty}(t)$ is convex. We need to show that

$$\frac{\partial}{\partial q} \hat{V}_{\infty}(\infty, b_{in}) > 0. \quad (1.271)$$

Note that ϵ and \mathfrak{f} are constant in q , while \mathfrak{a} , \mathfrak{b} , \mathfrak{c} , \mathfrak{d} , and \mathfrak{g} are not. Check that at $(b_{im}, b_{in}) = (\infty, b_{in})$, we have

$$\mathfrak{a} = 1, \quad (1.272)$$

$$\mathfrak{b} = 0, \quad (1.273)$$

$$\mathfrak{c} = -q \delta^{L_{im, \infty}^{-1}(b_{in})} (p + (1-p)\eta^{b_{in}}), \quad (1.274)$$

and

$$\mathfrak{d} = 1 - (1-q) \delta^{L_{im, \infty}^{-1}(b_{in})} (p + (1-p)\eta^{b_{in}}). \quad (1.275)$$

Let

$$\mathfrak{h} = \delta^{L_{in, \infty}^{-1}(b_{in})} (p + (1-p)\eta^{b_{in}}). \quad (1.276)$$

We next apply the quotient rule to obtain

$$\begin{aligned}
& \frac{\partial}{\partial q} \hat{V}_\infty(\infty, b_{in}) \\
&= \frac{1}{\mathfrak{g}^2} \left(\mathfrak{g} \frac{\partial}{\partial q} (q\mathfrak{d}\mathfrak{e} + (1-q)\mathfrak{f} - (1-q)\mathfrak{c}\mathfrak{e}) - (q\mathfrak{d}\mathfrak{e} + (1-q)\mathfrak{f} - (1-q)\mathfrak{c}\mathfrak{e}) \frac{\partial \mathfrak{g}}{\partial q} \right) \\
&= \frac{(1 - (1-q)\mathfrak{h})(\mathfrak{e} - \mathfrak{f}) - \mathfrak{h}(\mathfrak{f} + q(\mathfrak{e} - \mathfrak{f}))}{(1 - (1-q)\mathfrak{h})^2} \\
&= \frac{\mathfrak{e} - \mathfrak{f} - \mathfrak{e}\mathfrak{h}}{(1 - (1-q)\mathfrak{h})^2}.
\end{aligned}$$

If $b_{in} = \infty$ so that $\mathfrak{h} = 0$, then we have

$$\frac{\partial}{\partial q} \hat{V}_\infty(\infty, b_{in}) = \mathfrak{e} - \mathfrak{f}, \tag{1.277}$$

which is positive; indeed, check that

$$\mathfrak{e} = \int_0^\infty \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_\gamma(a), \tag{1.278}$$

while

$$\mathfrak{f} = (1-p) \int_0^\infty \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_\gamma(a). \tag{1.279}$$

We see that

$$\begin{aligned}
& (1-p) \int_0^\infty \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_\gamma(a) \\
& < \int_0^\infty \left(\int_0^{L_{in,\infty}^{-1}(a)} \delta^t f(a, L_{in,\infty}(t)) dt + \int_{L_{in,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_\gamma(a)
\end{aligned}$$

$$\leq \int_0^\infty \left(\int_0^{L_{im,\infty}^{-1}(a)} \delta^t f(a, L_{im,\infty}(t)) dt + \int_{L_{im,\infty}^{-1}(a)}^\infty \delta^t dt \right) d\mu_\gamma(a), \quad (1.280)$$

as needed, since $L_{in,\infty}(t) \leq L_{im,\infty}(t)$.

Now, suppose that $b_{im} < \infty$, so that $\mathfrak{h} > 0$. Then, we can write

$$\frac{\partial}{\partial q} \hat{V}_\infty(\infty, b_{in}) = \frac{\mathfrak{e} - \mathfrak{f} - \mathfrak{e}\mathfrak{h}}{(1 - (1 - q)\mathfrak{h})^2} = \frac{\hat{V}_{im}(\infty, b_{in}) - \hat{V}_{in}(\infty, b_{in})}{1 - (1 - q)\mathfrak{h}}, \quad (1.281)$$

since

$$\hat{V}_{im}(\infty, b_{in}) - \hat{V}_{in}(\infty, b_{in}) = \frac{\mathfrak{d}\mathfrak{e} - \mathfrak{b}\mathfrak{f} - (\mathfrak{a}\mathfrak{f} - \mathfrak{c}\mathfrak{e})}{\mathfrak{g}} = \frac{\mathfrak{e} - \mathfrak{f} - \mathfrak{e}\mathfrak{h}}{1 - (1 - q)\mathfrak{h}}. \quad (1.282)$$

Thus, we need to show that

$$\hat{V}_{im}(\infty, b_{in}) > \hat{V}_{in}(\infty, b_{in}). \quad (1.283)$$

This inequality is proven by showing that

$$\hat{V}_{im}(\infty, b_{in}) > \hat{V}_{im}(\mathbf{b}) > \hat{V}_{in}(\infty, b_{in}), \quad (1.284)$$

for

$$\mathbf{b} = ((L_{im,\infty}(L_{in,\infty}^{-1}(b_{in})), b_{in}), (\infty, b_{in}), (\infty, b_{in}), \dots). \quad (1.285)$$

The comprising inequality

$$\hat{V}_{im}(\mathbf{b}) < \hat{V}_{im}(\infty, b_{in}) \quad (1.286)$$

follows from the optimality of never quitting tasks of type $j = im$, demonstrated in the

proof of Proposition 1.5. Indeed, conditional on the current task being of type $j = im$, the strategy (∞, b_{in}) is equivalent to never quitting this current task.

Only the comprising inequality

$$\hat{V}_{in}(\infty, b_{in}) < \hat{V}_{im}(\mathbf{b}). \quad (1.287)$$

remains to be shown. The sample space of task sequences for the left-hand-side value function $\hat{V}_{im}(\mathbf{b}, b_{in})$ is

$$(0, \infty) \times \mathcal{U}^\infty \quad (1.288)$$

with the probability measure

$$\mu_{im} \otimes \mu^\infty = \mu_\gamma \otimes \mu^\infty, \quad (1.289)$$

and the sample space of task sequences of the right-hand-side value function $\hat{V}_{in}(b_{im}, b_{in})$ is

$$((0, \infty) \cup \{\infty\}) \times \mathcal{U}^\infty \quad (1.290)$$

with the probability measure

$$\mu_{in} \otimes \mu^\infty, \quad (1.291)$$

where we recall that μ_{in} places probability p on $a = \infty$ and distributes the remaining probability $1 - p$ as the exponential distribution μ_γ . It suffices to show that

$$\begin{aligned} & \int_{a_1 \in (0, \infty)} \left(\int_{(j_2, a_2), \dots \in \mathcal{U}^\infty} V_{in}((\infty, b_{in}), ((in, a_1), (j_2, a_2), \dots)) d\mu^\infty \right) d\mu_\gamma \\ & \leq \int_{a_1 \in (0, \infty)} \left(\int_{((j_2, a_2), \dots) \in \mathcal{U}^\infty} V_{im}(\mathbf{b}, ((im, a_1), (j_2, a_2), \dots)) d\mu^\infty \right) d\mu_\gamma \end{aligned} \quad (1.292)$$

and

$$\begin{aligned} & \int_{a_1 \in \{\infty\}} \left(\int_{(j_2, a_2), \dots \in \mathcal{U}^\infty} V_{in}((\infty, b_{in}), ((in, a_1), (j_2, a_2), \dots)) d\mu^\infty \right) d\chi \\ & < \int_{a_1 \in (0, \infty)} \left(\int_{((j_2, a_2), \dots) \in \mathcal{U}^\infty} V_{im}(\mathbf{b}, ((im, a_1), (j_2, a_2), \dots)) d\mu^\infty \right) d\mu_\gamma \end{aligned} \quad (1.293)$$

for χ the one-point distribution on $\{\infty\}$. Indeed, adding the product of the inequality (1.292) with $(1-p)$ with the product of the inequality (1.293) with p yields the desired inequality (1.287).

The second inequality (1.293) holds immediately because it simplifies to

$$0 < \int_{a_1 \in (0, \infty)} \left(\int_{((j_2, a_2), \dots) \in \mathcal{U}^\infty} V_{im}(\mathbf{b}, ((im, a_1), (j_2, a_2), \dots)) d\mu^\infty \right) d\mu_\gamma. \quad (1.294)$$

The first inequality (1.292) holds because for every sample

$$(a_1, (j_2, a_2), \dots), \quad (1.295)$$

the payoff of the left-hand-side value function

$$V_{in}((\infty, b_{in}), ((in, a_1), (j_2, a_2), \dots)) \quad (1.296)$$

is at most the payoff of the right-hand-side value function

$$V_{im}(\mathbf{b}, ((im, a_1), (j_2, a_2), \dots)). \quad (1.297)$$

This is demonstrated by partitioning $[0, \infty)$ into various subintervals and looking at the respective sub-payoff values corresponding to each subinterval.

In the subinterval

$$[0, L_{in,\infty}^{-1}(b_{in})], \quad (1.298)$$

the sub-payoff value of the left-hand-side value function is at most that of the right-hand-side value function, because the learning of the first task is faster for the latter than the former: $L_{in,\infty}(t) \leq L_{im,\infty}(t)$.

In the subinterval

$$[L_{in,\infty}^{-1}(b_{in}), \infty), \quad (1.299)$$

the sub-payoff value of the left-hand-side value function is equal to that of the right-hand-side value function conditional on the first task being quit at time $t = L_{in,\infty}^{-1}(b_{in})$ for both, i.e., conditional on learning not yet having completed. And conditional on the opposite—that learning of the first task completes for at least one of the value functions by time $t = L_{in,\infty}^{-1}(b_{in})$ —we have the following. If this occurs for the left-hand-side value function, then it also occurs for the right-hand side value function, since $L_{in,\infty}(t) \leq L_{im,\infty}(t)$. Thus, we have the desired inequality for the sub-payoff values corresponding to the subinterval (1.299). If this occurs for the right-hand-side value function, then its sub-payoff value corresponding to the subinterval (1.299) is maximal, so the inequality holds anyway. Thus, we have obtained (1.293), and thereby the desired inequality (1.287).

This concludes our proof of (1.271).

The police admit they don't have enough evidence to convict the pair on the principal charge. They plan to sentence both to two years in prison on a lesser charge. Simultaneously, the police offer each prisoner a Faustian bargain.¹⁸⁸

William Poundstone

2

Cooperation in alternating interactions with memory constraints

Abstract: In repeated social interactions, individuals often employ reciprocal strategies to maintain cooperation. To explore the emergence of reciprocity, many theoretical models assume synchronized decision making. In each round, individuals decide simultaneously whether to cooperate or not. Yet many manifestations of reciprocity in nature are

asynchronous. Individuals provide help at one time and receive help at another. Here, we explore such alternating games in which players take turns. We mathematically characterize all Nash equilibria among memory-one strategies. Moreover, we use evolutionary simulations to explore various model extensions, exploring the effect of discounted games, irregular alternation patterns, and higher memory. In all cases, we observe that mutual cooperation still evolves for a wide range of parameter values. However, compared to simultaneous games, alternating games require different strategies to maintain cooperation in noisy environments. Moreover, none of the respective strategies are evolutionarily stable.

2.1 Introduction

Cooperation can be maintained by direct reciprocity, where individuals help others in repeated interactions^{166,215,233}. Traditionally, researchers capture the logic of direct reciprocity with the repeated prisoner's dilemma^{8,52,68,75,77,97,116,169,174,192,195,209,225,228}. According to that model, two individuals – usually referred to as players – interact with each other over several rounds. In each round, both players can either cooperate or defect. Mutual cooperation yields a better payoff than mutual defection, but each individual has an incentive to defect. Theoretical and experimental work suggests that cooperation can evolve if there are sufficiently many interactions between the individuals⁹⁹. This work has been used to explain a wide variety of behaviors, including why humans are more likely to cooperate in stable groups¹⁵², why certain animal species share food²⁴⁶, and why firms are able to achieve higher market prices when they engage in collusion¹⁷.

A standard assumption that underlies much of this research is that individuals make their decisions simultaneously (or at least in ignorance of the co-player's cur-

rent decision). We refer to this kind of repeated interaction as a simultaneous game; see Figure 2.1a. For many natural manifestations of reciprocity, however, simultaneous cooperative exchanges are unlikely or even impossible, such as when people ask for favors¹⁰⁷, vampire bats donate blood to their conspecifics²⁴⁶, sticklebacks engage in predator inspection¹⁵⁵, or ibis take turns when leading their flock²³⁹. Such interactions are better captured by alternating games, in which players consecutively decide whether to cooperate^{61,148,171,253}. When individuals decide asynchronously, they make their decisions based on different histories. The most recent events one player has in memory differ from the most recent events that the next player takes into account; see Figure 2.1b. Such asymmetries in turn make it more difficult to successfully coordinate on cooperation. As a result, many well-known strategies like Tit-for-Tat or Win-Stay Lose-Shift fail to evolve when players move alternately^{61,171}. Instead, previous computational^{61,171,253} and experimental studies²⁴¹ suggest that individuals need to be more forgiving. However, a full understanding of optimal play in alternating games is lacking, even though optimal behavior in the simultaneous game is by now well-analyzed^{3-5,53,74,96,149,220,221}.

Here, we propose an analytical approach to describe when cooperation evolves in the alternating game. In line with the previous literature, we typically focus on individuals with so-called memory-one strategies²¹⁵. Memory-one strategies depend on each player's most recent move. Our analysis involves two steps. First, we show that successful play in alternating games does not require a sophisticated cognitive apparatus. More specifically, when interacting with a given memory-one opponent, it suffices to respond with a reactive strategy that only depends on the co-player's most recent move. This result is reminiscent of a previous finding of Press and Dyson for the simultaneous game¹⁹⁰. They showed that against a memory-one strategy, there is

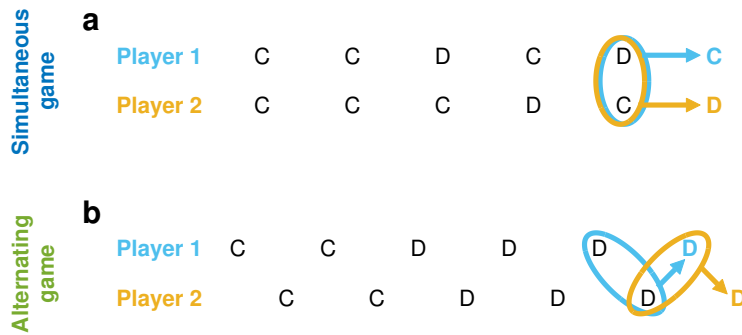


Figure 2.1: Game dynamics for the simultaneous and the alternating game. In both the simultaneous and the alternating game, two players interact repeatedly. In each turn, they decide whether to cooperate (C) or to defect (D). In the simultaneous game (a), they make their decision at the same time (or at least not knowing the other player's decision). In the alternating game (b), one player decides before the other player does. In both cases, we study memory-1 strategies. That is, an individual's next action only depends on each individual's previous action. We illustrate the information each individual takes into account for its last decision with colored ellipses. In the simultaneous game, individuals take into account the same information. In the alternating game, decisions are based on different sets of information.

nothing to gain from having a larger memory than the opponent. Our result for the alternating game goes one step further. Against a memory-one strategy, players can afford to have a strictly lower memory, without any loss to their or their co-player's payoff. As we show, this result crucially depends on the alternating move structure; it is not true when players move simultaneously. In a second step, we show that in order to identify a best response to a given memory-one player, we only need to check the four most extreme reactive strategies: unconditional defection, unconditional cooperation, Tit-for-Tat and Anti-Tit-for-Tat. Using this approach, we identify all Nash equilibria among the memory-one strategies.

In the absence of errors, we find an unexpected equivalence. The very same memory-1 strategies that can be used to enforce cooperation in the simultaneous game also enforce cooperation in the alternating game. However, once we take into account errors, the predictions for the two models diverge. In the simultaneous game, Win-Stay Lose-Shift is evolutionarily stable when the benefit to cost ratio is sufficiently large and when errors are sufficiently rare^{98,137}. In that case, there is a simple rule for how to sustain full cooperation: individuals should repeat their previous action if it yielded a sufficiently large payoff, and switch to the opposite action otherwise. In contrast, in the alternating game, all stable cooperative strategies require players to randomize. After mutual defection, they need to cooperate with some well-defined probability that depends on the game parameters and the error rate. Although the respective strategies in the alternating game are Nash equilibria, we show that none of them is evolutionarily stable. As a result, evolving cooperation rates in the alternating game often tend to be lower than in the simultaneous game, although this difference is smaller than perhaps expected from static stability considerations alone. We summarize our analytical findings in Figure 2.2.

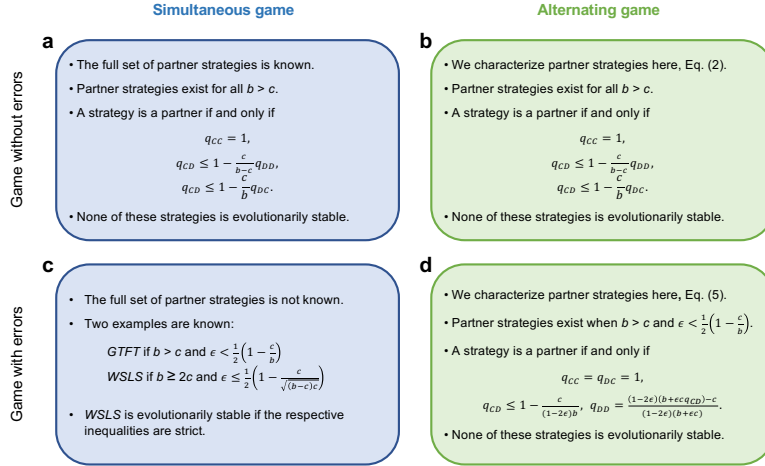


Figure 2.2: A characterization of partners among the memory-1 strategies. Within the class of memory-1 strategies, we provide an overview of the strategies that sustain full cooperation in a Nash equilibrium. The respective strategies are called partner strategies, or partners⁹⁹. a, For the simultaneous game without errors, partners have been first described by Akin^{3,4} (he calls them ‘good strategies’). Akin’s approach has been extended by Stewart and Plotkin²²¹ to describe all memory-1 Nash equilibria of the simultaneous game. In the absence of errors, none of these strategies is evolutionarily stable^{24,136}. Instead, one can always find neutral mutant strategies which act as a stepping stone out of equilibrium⁶⁹. b, For the alternating game without errors, Eq. (2.30) provides a full characterization of all partner strategies. Cooperation is maintained by the same strategies as in the simultaneous game. c, Despite decades of research, the exact set of partner strategies for the simultaneous game with errors is not known. However, there are at least two instances of partner strategies, GTFT^{156,169}, and Win-Stay Lose-Shift, WSLS^{122,170}. For repeated games with errors, evolutionary stability is generally feasible²³. In particular, WSLS is evolutionarily stable if the benefit to cost ratio is sufficiently large and if errors are sufficiently rare¹³⁷. d, For the alternating game with errors, we characterize all partner strategies in Eq. (2.46). None of them is deterministic. As a result, none of them is evolutionarily stable (see Supplementary Information for details).

Our work suggests that in most realistic scenarios, successful play in alternating games requires different kinds of behaviors than predicted by the earlier theory on simultaneous games. In this way, we corroborate earlier experimental work on human cooperation²⁴¹, and provide theoretical methods to further analyze repeated games in the future. Overall, we find that cooperation is still feasible in alternating games. However, the strategies that enforce cooperation can be neutrally invaded, and hence cooperation tends to be more short-lived than in the simultaneous game.

2.2 Results

Model description

In the following, we formulate a simple baseline scenario, which we use to derive our main analytical results (see also Supplementary Note 1). More general scenarios are discussed in a later section, and in full detail in the Supplementary Note 3. We consider interactions between two players, player 1 and player 2. Both players repeatedly decide whether to cooperate (*C*) or defect (*D*). These repeated interactions can take place in two different ways. In the simultaneous game, there is a discrete number of rounds. In each round, both players make their decision at the same time, not knowing their co-player's decision; see Figure 2.1a. In contrast, in the alternating game, the players move consecutively. We consider the strictly alternating game: Player 1 moves first, and then player 2 learns about player 1's decision and moves next; see Figure 2.1b. We note that there are also variants of the alternating game in which the order of moves is random^{148,171}. In particular, one player may by chance make two or more consecutive moves before it is the other player's turn again. The effect of such

irregular alternation patterns will be discussed later.

For the simultaneous game, the possible payoffs in each round can be represented by four parameters. Players receive the reward R in rounds in which they both cooperate; they receive the temptation payoff T and the sucker's payoff S , respectively, if only one player cooperates; and they receive the punishment payoff P in case they both defect. For $T > R > P > S$, we obtain the prisoner's dilemma. In the alternating game, however, it is useful to assume that payoffs can be assigned to each player's individual action¹⁷¹. In that case, the value of one player's cooperation is independent of the co-player's previous or subsequent decision (or equivalently, payoffs are independent of how the two players' decisions are grouped into rounds). As a result, we obtain the donation game²¹⁵. Here, cooperation means to pay a cost $c > 0$ in order to provide a benefit $b > c$ to the co-player. The donation game is a special case of a prisoner's dilemma for which

$$R = b - c, \quad S = -c, \quad T = b, \quad P = 0. \quad (2.1)$$

To compare the alternating game with the simultaneous game, we assume payoffs satisfy (2.1) throughout.

In the baseline scenario, we consider infinitely repeated games, and we study players who make their decisions based on each player's most recent move. In the simultaneous game, the respective strategies are called memory-1 strategies¹⁷⁰; they take into account the outcome of one previous round; see Figure 2.1a. Such strategies can be represented as a 4-tuple, $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$. The entry p_{ij} denotes the probability to cooperate in the next round. This probability depends on the player's action i and the co-player's action j in the previous round. The equivalent strategy class also exists

in alternating games¹⁷¹. In alternating games, however, there is no longer a unique previous round to which both players refer. Instead, the last round that is taken into account depends on the perspective of each player. It consists of the respective last moves of the two players; see Figure 2.1b. An important subset of memory-1 strategies are the so-called reactive strategies. Here, players ignore their own previous action, and only condition their behavior on what the co-player previously did. Reactive strategies are therefore those memory-1 strategies for which $p_{CC} = p_{DC}$ and $p_{CD} = p_{DD}$.

Some well-known examples of memory-1 strategies for the simultaneous game include Always Defect, ALLD = (0, 0, 0, 0), Tit-for-Tat, TFT = (1, 0, 1, 0), and Win-Stay Lose-Shift, WSLS = (1, 0, 0, 1). In the alternating game, a strategy called Firm-but-Fair²¹⁵, defined by FBF = (1, 0, 1, 1) and also referred to as Forgiver²⁵³, has been successful in evolutionary competitions. Out of these examples, ALLD and TFT are reactive, whereas WSLS and FBF are not. We say a strategy is deterministic if each conditional cooperation probability is either zero or one. In particular, all of the above examples are deterministic. Otherwise, we call the strategy stochastic.

Note that our analysis includes the possibility that players sometimes make errors. That is, when a player decides to cooperate, there is some probability ε that the player defects by mistake. Conversely, a player who intends to defect may cooperate with the same probability. We refer to the case of $\varepsilon = 0$ as the game without errors, and to $\varepsilon > 0$ as the game with errors. We note that even a strategy that is deterministic becomes fully stochastic in the game with errors, because in that case a player's effective cooperation probability is always between ε and $1 - \varepsilon$.

Considering memory-1 strategies is useful for two reasons. First, such strategies are straightforward to interpret, and the respective conditional probabilities can be easily inferred from experiments²⁴¹. Second, when both players use memory-1 strate-

gies, their average payoffs are simple to compute (see also Methods). To this end, suppose player 1 uses the strategy \mathbf{p} and player 2 adopts strategy \mathbf{q} . By representing the game as a Markov chain, we can compute the stationary distribution $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$. The entries of this stationary distribution give the probabilities of observing each of the four possible combinations of the players' actions over the course of the game. Based on this stationary distribution, we define player 1's payoff as $\pi(\mathbf{p}, \mathbf{q}) = (v_{CC} + v_{DC})b - (v_{CC} + v_{CD})c$, and similarly for player 2. While the baseline scenario focuses on memory-1 strategies, our results are more general. For example, when we describe which memory-1 strategies are Nash equilibria in the following, co-players are allowed to deviate to strategies with arbitrarily long (but finite) memory. Moreover, similar approaches can also be used to explore the evolutionary dynamics of memory-2 strategies, as we will discuss later.

A recipe for identifying Nash equilibria for alternating games

To predict which memory-1 strategies evolve in the alternating game, we first characterize which of them are Nash equilibria. In the following, we refer to a strategy \mathbf{q} as a Nash equilibrium if $\pi(\mathbf{q}, \mathbf{q}) \geq \pi(\mathbf{p}, \mathbf{q})$ for all alternative memory-1 strategies \mathbf{p} (for stronger results, see Supplementary Note 2). That is, against a co-player who adopts the Nash equilibrium strategy \mathbf{q} , a player has no incentive to choose any different memory-1 strategy. The notion of Nash equilibrium is closely related to evolutionary robustness²²⁰. In a population of size N , a resident strategy \mathbf{q} is called evolutionary robust if no mutant strategy \mathbf{p} has a fixation probability larger than neutral, $1/N$. When selection is sufficiently strong, strategies are evolutionary robust if and only if they are Nash equilibria²²¹.

Verifying that a given strategy \mathbf{q} is a Nash equilibrium is not straightforward. In

principle, this requires us to compare its payoff to the payoff of all possible mutant strategies \mathbf{p} , taken from the uncountable set of all memory-1 strategies. However, for alternating games, it is possible to simplify the task in two steps (see Supplementary Note 2 for details). The first step is to show that it is sufficient to compare \mathbf{q} to all reactive strategies, a strategy set of lower dimension. The intuition for this result is as follows. Even if player 1 starts out with an arbitrary memory-1 strategy \mathbf{p} , it is always possible to find an associated reactive strategy $\tilde{\mathbf{p}}$ that yields the same stationary distribution and the same payoff against \mathbf{q} (Figure 2.3). That is, to find a best response to a strategy that remembers both players' last moves, it is sufficient to explore all strategies that only remember the co-player's last move. In particular, not only is there no advantage of having a strictly larger memory than the opponent, as shown by Press and Dyson for simultaneous games¹⁹⁰. A player can afford to remember strictly less in the alternating game.

The second step is to show that we do not need to consider all reactive strategies to find a best response against \mathbf{q} . Instead, it suffices to consider all deterministic reactive strategies. By combining these two steps, it becomes straightforward to check whether a given memory-1 strategy is a Nash equilibrium. We only need to compare its payoff against itself to the four payoffs that can be achieved by deviating to Always Defect (ALLD), Always Cooperate (ALLC), Tit-for-Tat (TFT), or Anti-Tit-for-Tat (ATFT).

Equilibria in alternating games without errors

Using the above recipe, we first explore which memory-1 strategies can sustain full cooperation in games without errors (see Supplementary Note 2 for all derivations). To this end, we call a memory-1 strategy a partner^{53,96} if (i) it is fully cooperative against itself, and (ii) if it is a Nash equilibrium (such strategies are referred to as



Figure 2.3: In alternating games, individuals can afford to remember less than their opponent. We prove the following result: if two memory-1 players interact, any of the players can switch to a simpler reactive strategy (that only depends on the co-player’s previous action) without changing the resulting payoffs. Here, we illustrate this result for player 1. a, Initially, both players use memory-1 strategies. That is, a player’s cooperation probability depends on the most recent decision of each player. There are four conditional cooperation probabilities. b, The strategies determine how players interact in the alternating game. c, Based on the strategies, we can compute how often we are to observe each pairwise outcome over the course of the game by calculating the game’s stationary distribution. d, Based on the stationary distribution, and on player 1’s memory-1 strategy, we can compute an associated reactive strategy. This reactive strategy only consists of two conditional cooperation probabilities. They determine what to do if the co-player cooperated (or defected) in the previous round. The cooperation probabilities can be calculated as a weighted average of the respective memory-1 strategy’s cooperation probabilities. The resulting reactive strategy for player 1 yields the same outcome distribution against player 2 as the original memory-1 strategy. We note that for this result, the assumption of alternating moves is crucial. In the simultaneous game, the respectively defined reactive strategy does not yield the same outcome distribution against player 2 as the original memory-1 strategy (see Supplementary Information).

‘good’ by Akin³⁻⁵). We find that partners are exactly those memory-1 strategies \mathbf{q} for which the following three conditions hold,

$$q_{CC} = 1, \quad q_{CD} \leq 1 - \frac{c}{b-c}q_{DD}, \quad q_{CD} \leq 1 - \frac{c}{b}q_{DC}. \quad (2.2)$$

The first condition is needed to ensure that the strategy is fully cooperative against itself. The other two conditions restrict how cooperative a player is allowed to be after having been exploited by the co-player. If these last two conditions are violated, the strategy \mathbf{q} can either be invaded by ALLD or ATFT. Together, the three requirements in (2.30) define a three-dimensional polyhedron within the space of all memory-1 strategies (Figure 2.4a). The volume of this polyhedron increases with the benefit to cost ratio b/c . While the polyhedron never contains ALLC, it always contains the conditionally cooperative strategies TFT and GRIM (for these two strategies, we additionally require the respective players to cooperate in the very first round to ensure payoffs are well-defined, see Supplementary Information). Moreover, for $b \geq 2c$, the polyhedron contains WSLS and FBF (independent of the outcome of the first round).

Similarly, we can also identify all Nash equilibria where the players mutually defect. We refer to the respective strategies as defectors. We obtain the following necessary and sufficient condition,

$$q_{DD} = 0, \quad q_{DC} \leq \frac{c}{b}(1-q_{CD}), \quad q_{DC} \leq \frac{c}{b-c}(1-q_{CC}). \quad (2.3)$$

Again, the first equation ensures that two players with the respective strategy end up mutually defecting against each other. The other two conditions ensure that the strategy is comparably unresponsive towards a co-player who tries to initiate cooperation.

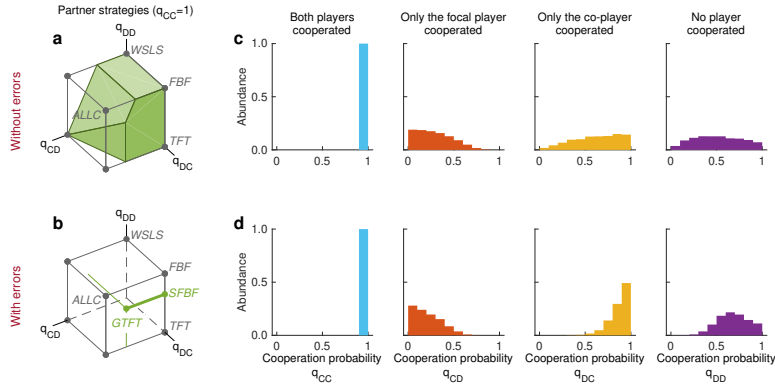


Figure 2.4: Partner strategies in alternating games with and without errors. Partner strategies sustain cooperation in a Nash equilibrium. All such strategies are required to cooperate after mutual cooperation, such that the respective cooperation probability q_{CC} is equal to one. **a**, In the absence of errors, the remaining three cooperation probabilities can be chosen arbitrarily, subject to the constraints in Eq. (2.30). The resulting set of partner strategies takes the shape of a polyhedron. **b**, In the presence of errors, this polyhedron degenerates to a single line segment. This line segment comprises all strategies between Generous Tit-for-Tat (GTFT) and Stochastic Firm-but-Fair (SFBF). **c,d**, We compare these equilibrium results to evolutionary simulations. To this end, we record all strategies that emerge over the course of the simulation. Here, we plot the probability distribution of those strategies that yield at least 80% cooperation against themselves. Without errors, the probability distributions for q_{CD} , q_{DC} , q_{DD} are comparably flat. With errors, players tend to cooperate if they exploited their opponent in the previous round, $q_{DC} \approx 1$. Moreover, they cooperate with some intermediate probability after mutual defection, $q_{DD} \approx 2/3$. Both effects are in line with previous simulation studies^{61,171}, and they confirm the theory. Simulations are run for $b/c = 3$, and $\varepsilon = 0$ or $\varepsilon = 0.02$. For the other parameter values and further details on the simulations, see Methods. Source data are provided as a Source Data file.

Similar to before, the three conditions define a three-dimensional polyhedron (Figure 2.8a). The set of defectors is non-empty for all parameter values, and it always contains the strategy ALLD.

Finally, we identify a third class of Nash equilibria, referred to as equalizers²¹. As in the simultaneous game¹⁹⁰, equalizers are strategies that unilaterally control the co-player's payoff. If one player adopts an equalizer strategy, the co-player's payoff is fixed, independent of the co-player's strategy^{79,95,104,141,147}. In the alternating game, these strategies are characterized by

$$q_{CD} = \frac{b q_{CC} - c(1 + q_{DD})}{b - c}, \quad q_{DC} = \frac{b q_{DD} + c(1 - q_{CC})}{b - c}. \quad (2.4)$$

When both players adopt an equalizer strategy, neither player has anything to gain from deviating; the resulting outcome is a Nash equilibrium.

We also show a converse result: If a memory-1 strategy for the alternating game is a Nash equilibrium, then it either needs to be a partner, a defector, or an equalizer. Remarkably, the same three strategy classes also arise as Nash equilibria of the simultaneous game²²¹. Even the algebraic conditions for being a partner, defector, or equalizer coincide (however, the existing proof for the simultaneous game²²¹ is somewhat more intricate than the proof for the alternating game that we provide in Supplementary Note 4). There is, however, one difference. In the simultaneous game, there is a fourth class of Nash equilibria, referred to as 'alternators'²²¹. Alternators cooperate in one round, only to defect in the next. In the Supplementary Note 2, we show that such patterns of behavior cannot emerge among memory-1 players in the alternating game.

Equilibria in alternating games with errors

Next, we explore how the Nash equilibria change when we introduce errors. In the following, we discuss the case of partner strategies; the analogous results for defectors and equalizers are derived in the Supplementary Note 2. For partner strategies, we find that errors impose additional constraints. First, partners only exist when errors are sufficiently rare, $\varepsilon < \frac{1}{2} \left(1 - \frac{c}{b}\right)$. Second, the respective conditions are now considerably more restrictive,

$$q_{CC} = q_{DC} = 1, \quad q_{CD} \leq 1 - \frac{c}{(1-2\varepsilon)b}, \quad q_{DD} = \frac{(1-2\varepsilon)(b+\varepsilon c q_{CD}) - c}{(1-2\varepsilon)(b+\varepsilon c)}. \quad (2.5)$$

In particular, if the co-player cooperated in the previous round, partners are strictly required to cooperate in the next round, independent of their own previous action (because now $p_{DC} = 1$). If the co-player defected, partners need to cooperate with a well-defined probability, as defined by the last two conditions in (2.46). The last condition guarantees that neither ALLC nor TFT has a selective advantage against \mathbf{q} . In the game without errors, this requirement is satisfied automatically. There, all strategies with $q_{CC} = 1$ yield the full cooperation payoff $b - c$ against each other. In the game with errors, however, such strategies are no longer neutral. Instead, they differ in how quickly they are able to restore cooperation after an error, and to which extent they are able to capitalize on their co-players' mistakes. Noisy environments thus impose additional constraints on self-cooperative strategies to be stable.

As a result of these additional constraints, the three-dimensional polyhedron degenerates to a one-dimensional line segment (Figure 2.4b). On one end of this line segment, there is Generous Tit-for-Tat, which also arises in the simultaneous game^{156,169},

$$\text{GTFT} = \left(1, 1 - \frac{c}{(1-2\varepsilon)b}, 1, 1 - \frac{c}{(1-2\varepsilon)b} \right) \quad (2.6)$$

On the other end of this line segment, we find a strategy that resembles the main characteristics of Firm-but-Fair²¹⁵; we thus refer to this strategy as Stochastic Firm But Fair,

$$\text{SFBF} = \left(1, 0, 1, \frac{(1-2\varepsilon)b-c}{(1-2\varepsilon)(b+\varepsilon c)} \right) \quad (2.7)$$

Behaviors similar to SFBF have been observed in early simulations of alternating games^{61,171}. There, it was found that evolutionary trajectories often lead to strategies that are deterministic, except that they randomize after mutual defection. Our results provide an analytical justification: SFBF is the only such strategy that is a Nash equilibrium.

The above conditions in (2.46) provide a complete characterization of all partner strategies in the alternating game with errors. Despite decades of research, an analogous characterization for the simultaneous game is not yet available (Figure 2.2). However, it is known that particular strategies, most importantly WSLS, can be evolutionarily stable in the presence of noise¹³⁷. That is, in the simultaneous game, cooperation can be sustained with a simple deterministic strategy if $b > 2c$. In contrast, conditions (2.46) imply that no such deterministic strategy is available in the alternating game. Moreover, while the partner strategies characterized by (2.46) are Nash equilibria, we show in the Supplementary Information that they all are vulnerable to neutral invasion by either ALLC or TFT (in fact by all strategies with $q_{CC} = q_{DC} = 1$). These results suggest that cooperation can still evolve in alternating games, but it may be less robust than in the simultaneous game.

Evolutionary dynamics of alternating games

In order to test these equilibrium predictions, we next explore which behaviors emerge when the players' strategies are subject to evolution. To this end, we consider a population of N players. Each member of the population is equipped with a memory-1 strategy. They obtain payoffs by interacting with all other population members. To model the spread of successful strategies, we assume individuals with high payoffs are imitated more often²³² (or equivalently, such individuals produce more offspring²⁴⁹). In addition, new strategies are introduced through random exploration (or equivalently, through mutations). These random strategies are uniformly taken from the space of all memory-1 strategies. We capture the resulting dynamics with computer simulations. For details, see Methods.

First, we explore the evolutionary dynamics for fixed game parameters. We record which strategies the players use over the course of evolution to sustain cooperation. In Figure 2.4, we represent those strategies that yield a cooperation rate against themselves of at least 80%; other threshold values lead to similar conclusions. We call these strategies 'self-cooperative'. By definition, players with these strategies are likely to cooperate after mutual cooperation. Here we are thus interested in how they react when either one or both players defected. Without errors, the respective conditional cooperation probabilities show quite some variation. As a result, the distributions in Figure 2.4c are comparably flat. Overall, players act in such a way that the partner conditions (2.30) are satisfied, but they show no preference for a particular partner strategy. Once we allow for errors, the evolving strategies change (Figure 2.4d). Players tend to always cooperate if the co-player did so in the previous round, with $q_{CC} \approx q_{DC} \approx 1$. Moreover, after mutual defection, they cooperate with some strictly

positive probability. Both patterns are predicted by our equilibrium conditions (2.46). We find a similar match between static theory and evolutionary simulations for defectors, or when we explore evolution in the simultaneous game (Figures 2.7–2.9).

In a next step, we compare the dynamics of the alternating and the simultaneous game across different parameter values. To this end, we systematically vary the benefit of cooperation, the population size, the selection strength, and the mutation rate (Figure 2.5). In games without errors, we observe hardly any difference between the alternating and the repeated game. Both games yield almost identical cooperation rates over time, and these cooperation rates are similarly affected by parameter changes. A difference between the two games only becomes apparent when players need to cope with errors. Here, the simultaneous game leads to systematically higher cooperation rates than the alternating game. This difference is most visible for intermediate benefit-to-cost ratios and intermediate error rates, as one may expect: For small benefits and frequent errors, cooperation evolves in neither game, whereas for large benefits and rare errors, cooperation evolves in both games (Figure 2.10).

Evolutionary results beyond the baseline scenario

Our baseline scenario represents an idealized model of alternating interactions. It assumes (i) the game is infinitely repeated, (ii) players move in a strictly alternating fashion, (iii) games take place in a well-mixed population, and (iv) players use memory-1 strategies. In the following, we use simulations to explore the effect of each of these assumptions in turn. Here, we briefly summarize the respective results. For an exact description of the models, and for a more detailed discussion of the results, we refer to the Supplementary Note 3.

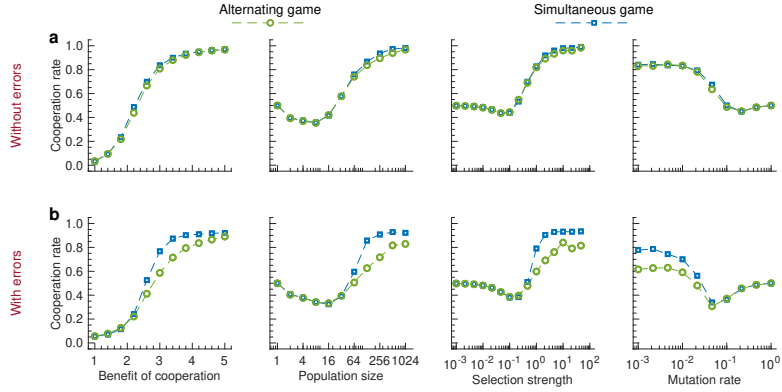


Figure 2.5: Comparing evolution in the alternating and the simultaneous game. To compare the two game versions, we have run additional evolutionary simulations. We systematically vary the benefit of cooperation, the population size, the selection strength, and the mutation rate. In addition, we vary how likely players make errors. Either they make no errors at all ($\varepsilon = 0$), or they make errors at some intermediate rate ($\varepsilon = 0.02$). a, In the absence of errors, there is virtually no difference between the simultaneous and the alternating game. Both games yield the same cooperation rates, and they respond to parameter changes in the same way. For the given baseline parameters, cooperation is favored for large benefits of cooperation, population sizes, and selection strengths. It is disfavored for intermediate and large mutation rates. b, With errors, the cooperation rates in the alternating game are systematically below the simultaneous game. The lower cooperation rates are related to our analytical result that no cooperative memory-1 strategy in the alternating game is evolutionarily stable. In contrast, in the simultaneous game with errors, WSLS can maintain cooperation^{122,170}, it is evolutionarily stable⁹⁸, and it readily evolves in evolutionary simulations (Figure 2.4IPD). As baseline parameters we use a benefit of cooperation $b = 3$, population size $N = 100$, selection strength $\beta = 1$, and the limit of rare mutations $\mu \rightarrow 0$ ^{64,105}. Source data are provided as a Source Data file.

We start by considering games with finitely many rounds. To incorporate a finite game length, we assume that each time both players have made a decision, the game continues with a constant probability δ . Figure 2.6a-c shows the respective evolutionary results for $\delta = 0.96$ (such that games last for 25 rounds on average). We observe similar results as in the infinitely repeated game: The simultaneous game leads to more cooperation (Figure 2.6a); moreover, if players cooperate, their strategies exhibit the main characteristics of WSLS in the simultaneous game, and of SFBF and GTFT in the alternating game (Figure 2.6b). Further simulations suggest that these qualitative results hold when players interact for at least 10 rounds (Figure 2.11). When interactions are shorter, cooperation is unlikely to evolve at all (Figure 2.6c).

In a next step, we explore irregular alternation patterns. To this end, we assume that every time a player has made a decision, with probability s it is the other player who moves next. We refer to s as the game's switching probability. For $s = 1$, we recover the baseline scenario, in which players strictly alternate. For $s = 1/2$, the player to move next is determined randomly. Simulations suggest that in both cases, players again use strategies akin to GTFT and SFBF to sustain cooperation (Figure 2.6e).

However, the robustness of the strategies depends on the switching probability. In particular, mutual cooperation is most likely to evolve when players alternate regularly (Figure 2.6f, Figure 2.12).

To explore the effect of population structure, we follow the framework of Brauchli et al.³¹. Instead of well-mixed populations, players are now arranged on a two-dimensional lattice. They use memory-1 strategies to engage in pairwise interactions with each of their neighbors. For the simultaneous game, we recover the main results of Brauchli et al.³¹: population structure can further enhance cooperation, and it makes it more likely that strategies similar to WSLS evolve (Figure 2.6g-i). For the alternating

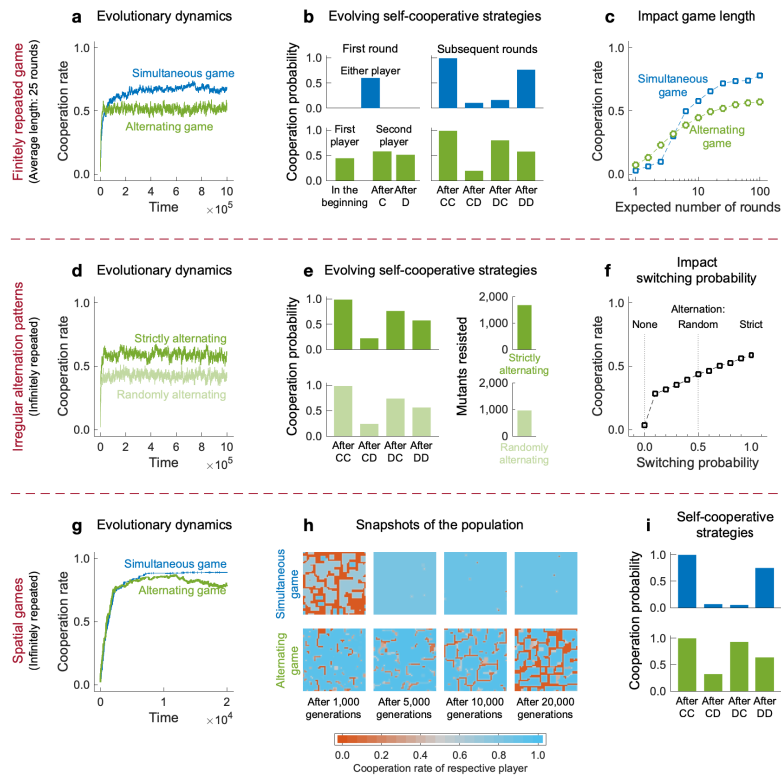


Figure 2.6: Robustness of evolutionary results. We have explored the robustness of our results with various model extensions. Here, we display results for three of them, illustrating the impact of finitely repeated games, of irregular alternating patterns, and of population structure. a–c, The baseline model assumes infinitely repeated games; here we show simulations for games with a finite expected length. If there are sufficiently many rounds, the simultaneous game again leads to more cooperation than the alternating game, and the evolving strategies are largely similar to the ones observed in the baseline model. d–f, The baseline model assumes that players move in a strictly alternating fashion. Instead, here we assume that after each player’s move, the other player moves with some switching probability s . The case $s = 1$ corresponds to strict alternation, whereas $s = 1/2$ represents a case in which the next player to move is completely random. We observe that irregular alternation patterns hardly affect which strategies players use to cooperate. However, it affects the robustness of these strategies. Overall, cooperation is most likely to evolve under strict alternation. g–i, Finally, instead of well-mixed populations, we consider games on a lattice. For the given parameter values, we observe that simultaneous games eventually lead to homogeneous cooperative populations. While this outcome is also possible for alternating games, some simulations also lead to the coexistence of cooperators and defectors (shown here in panel h). The evolving self-cooperative strategies are similar to the strategies that evolve in the baseline model. For a detailed description of these simulations, see Methods and Supplementary Information. Source data for panels a–f, i are provided as a Source Data file.

game, we observe that cooperation remains the most abundant outcome, but spatial structure does not necessarily result in homogeneous populations any longer. Instead, in some simulations we find cooperative and non-cooperative strategies to stably coexist (one particular instance is shown in Figure 2.6h).

Finally, we also analyzed the impact of larger memory. Exploring the dynamics among general memory- k strategies is not straightforward, as the strategy space increases rapidly. For instance, while there are only 16 pure memory-1 strategies, there are 65,536 memory-2 strategies and more than 10^{19} memory-3 strategies⁹⁸. We thus confine ourselves to pure memory-2 strategies in the following. In a first step, we explored which of these strategies are evolutionarily stable, see Figure 2.13a. For the simultaneous game, we find many such strategies, including several strategies with high cooperation rates. In the alternating game, we only find one strategy that is evolutionarily stable for a wide range of parameters, ALLD. Nevertheless, with respect to the evolving cooperation rates, stochastic evolutionary trajectories hardly show any difference between alternating and simultaneous games. The two games differ, however, in terms of the strategies that evolve, and in how robust these strategies are (Figure 2.13b-e).

2.3 Discussion

An overwhelming majority of past research on reciprocity deals with repeated games where individuals simultaneously decide whether to cooperate^{99,215}. In contrast, most natural occurrences of reciprocity require asynchronous acts of giving. Cooperation routinely takes the form of assisting a peer, providing a gift, or taking the lead in a joint endeavor^{107,155,239}. In such examples, simultaneous cooperation can be unfeasi-

ble, undesirable, or unnecessary. Herein, we have thus explored which strategies arise in alternating games where individuals make their decisions in turns. In such games, one individual's cooperation is reciprocated not immediately, but at some point in the future.

To explore the dynamics of cooperation in alternating games, we first describe all Nash equilibria among the memory-1 strategies. Memory-1 strategies are classical tools that have been used to describe the evolutionary dynamics of repeated games for several decades^{122,170,171}. However, most of the early work on memory-1 strategies was restricted to evolutionary simulations. Only with the pioneering work of Press and Dyson¹⁹⁰ and others^{3-5,53,74,96,149,220,221}, better mathematical techniques have become available. Using these techniques, it has become possible to describe all Nash equilibria of the infinitely repeated simultaneous game without errors²²¹. Herein, we make similar progress for the alternating game, both for the case with and without errors (for the simultaneous game with errors, a complete characterization of the Nash equilibria remains an open problem, see Figure 2.2).

Our results suggest that there are both unexpected parallels and important differences between simultaneous and alternating games. The parallels arise when individuals do not make errors. Here, the two models of reciprocity make the same predictions about the feasibility of cooperation. Cooperation evolves in the same environments, and it can be maintained using the same strategies. However, once individuals make mistakes, the predictions of the two models diverge. First, the two models require different kinds of strategies to maintain cooperation. In the simultaneous game, cooperation can be sustained with the deterministic memory-1 strategy Win-Stay Lose-Shift^{122,170}. Individuals with that strategy simply reiterate their previous behavior if it was successful, and they switch their behavior otherwise. In contrast, in the alter-

nating game, no simple deterministic rules for cooperation exist. Although there are still infinitely many memory-1 strategies that can maintain cooperation, all of them require individuals to randomize occasionally. One example of such a strategy for alternating games is Stochastic Firm-But-Fair (SFBF). Individuals with this strategy always reciprocate a co-player's cooperation, never tolerate exploitation, and cooperate with some intermediate probability if both players defected. Similar behaviors have been observed in earlier simulations^{61,171}. Our results provide a theoretical underpinning: SFBF is the unique memory-1 strategy that can sustain cooperation while retaliating against unconditional defectors in the strongest possible way.

The simultaneous game and the alternating game also differ in how stable cooperation is in evolving populations. In the simultaneous game, the evolution of cooperation is hardly affected by errors, provided the error rate is below a certain threshold (Figure 2.5, Figure 2.10). In some instances, errors can even enhance cooperation²⁵⁴. This body of work is based on the insight that evolutionarily stable cooperation is impossible in simultaneous games without errors^{23,24,69,70,136}. For any cooperative resident, it is always possible to find neutral mutant strategies that eventually lead to the demise of cooperation. However, once individuals occasionally commit errors, a strategy like WSLS is no longer neutral with respect to other cooperative strategies; it becomes evolutionarily stable^{23,137}. The situation is different in alternating games. Even in the presence of rare errors, strategies like SFBF remain vulnerable. They can be invaded by unconditional cooperators or by any other strategy that fully reciprocates a co-player's cooperation.

Despite these differences in the stability of their main strategies, evolving cooperation rates in the simultaneous and the alternating game are often surprisingly similar. To interpret these results, we note that when evolution is stochastic and takes

place in finite populations, no strategy persists indefinitely. Even evolutionarily stable strategies are invaded eventually. As a result, the overall abundance of cooperation is not only determined by the stability of any given strategy. Instead, it depends on additional aspects, such as the time it takes cooperative strategies to reappear when they are invaded. The relative importance of these different aspects depends on the details of the considered evolutionary process. To further illustrate these observations, we have run additional simulations for memory-1 players with local mutations²²² (see Supplementary Note 3). Because evolutionary stability considerations are less relevant when mutations are local, we observe that the cooperation rates of the alternating and the simultaneous game become more similar (Figure 2.14).

Cooperation is defined as a behavior where individuals pay a cost in order to increase the payoff or fitness of someone else¹⁶⁶. When individuals interact repeatedly, such cooperative interactions can be maintained by reciprocity. Here we have argued that in many examples, reciprocity arises as a series of asynchronous acts of cooperation. Most often, people do favors not to be rewarded immediately, but to request similar favors in future. Such consecutive acts of cooperation also appear to be at work when vampire bats²⁴⁶, sticklebacks¹⁵⁵, ibis²³⁹, tree swallows¹³⁵, or macaques¹⁵⁹ engage in reciprocity. We have shown that mutual cooperation is still possible in such alternating exchanges. But compared to the predominant model of reciprocity in simultaneous games, cooperation requires different kinds of strategies, and it is more volatile.

2.4 Methods

Calculation of payoffs

When two players with memory-1 strategies interact, their expected payoffs can be computed by representing the game as a Markov chain²¹⁵. To this end, suppose the first player's strategy is $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$, and the second player's strategy is $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$. Depending on the most recent actions of the two players (which can be either CC , CD , DC , or DD), we can compute how likely we are to observe each of the four outcomes in the following round. For the alternating game, we obtain the following transition matrix¹⁷¹,

$$M_A = \begin{pmatrix} p_{CC}q_{CC} & p_{CC}(1-q_{CC}) & (1-p_{CC})q_{CD} & (1-p_{CC})(1-q_{CD}) \\ p_{CD}q_{DC} & p_{CD}(1-q_{DC}) & (1-p_{CD})q_{DD} & (1-p_{CD})(1-q_{DD}) \\ p_{DC}q_{CC} & p_{DC}(1-q_{CC}) & (1-p_{DC})q_{CD} & (1-p_{DC})(1-q_{CD}) \\ p_{DD}q_{DC} & p_{DD}(1-q_{DC}) & (1-p_{DD})q_{DD} & (1-p_{DD})(1-q_{DD}) \end{pmatrix}. \quad (2.8)$$

Based on this transition matrix, we compute how often players observe each of the four outcomes. To this end, we solve the equation for the stationary distribution, $\mathbf{v} = \mathbf{v}M_A$. In most cases, the solution of this equation is unique. Uniqueness is guaranteed, for example, when the players' strategies \mathbf{p} and \mathbf{q} are fully stochastic, or when the error rate is positive. In exceptional cases, however, the transition matrix can allow for two or more stationary distributions. In that case, the outcome of the game is still well-defined, after specifying how players act in the very first round.

Given the stationary distribution $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$, we define the players' payoffs as

$$\begin{aligned} \pi_1 &= (v_{CC} + v_{DC})b - (v_{CC} + v_{CD})c, \\ \pi_2 &= (v_{CC} + v_{CD})b - (v_{CC} + v_{DC})c. \end{aligned} \quad (2.9)$$

This definition implicitly assumes that the game is indefinitely repeated and that future payoffs are not discounted. However, analogous formulas can be given in case there is a constant continuation probability δ , or equivalently, if future payoffs are discounted by δ (see Supplementary Note 3).

We compare our results for the alternating game with the corresponding results for the standard repeated prisoner's dilemma, where players decide simultaneously. Payoffs for the simultaneous game can be calculated in the same way as before. Only the transition matrix needs to be replaced by²¹⁵

$$M_S = \begin{pmatrix} p_{CC}q_{CC} & p_{CC}(1-q_{CC}) & (1-p_{CC})q_{CC} & (1-p_{CC})(1-q_{CC}) \\ p_{CD}q_{DC} & p_{CD}(1-q_{DC}) & (1-p_{CD})q_{DC} & (1-p_{CD})(1-q_{DC}) \\ p_{DC}q_{CD} & p_{DC}(1-q_{CD}) & (1-p_{DC})q_{CD} & (1-p_{DC})(1-q_{CD}) \\ p_{DD}q_{DD} & p_{DD}(1-q_{DD}) & (1-p_{DD})q_{DD} & (1-p_{DD})(1-q_{DD}) \end{pmatrix}. \quad (2.10)$$

Although the two matrices share many similarities, the resulting dynamics can be very different. For example, if the two players use TFT, then the matrix M_S allows for three invariant sets (corresponding to mutual cooperation, mutual defection, and alternating cooperation). However, the respective matrix M_A only allows for the first two invariant sets¹⁷¹. More generally, M_S allows for equilibria where players cooperate in one round but defect in the next round. Such equilibria are impossible for M_A (see Supplementary Note 2).

We sometimes assume players commit errors. We incorporate errors by assuming that with probability ε , a player who intends to cooperate defects by mistake. Analogously, a player who wishes to defect cooperates instead with the same probability. Such errors are straightforward to incorporate into the model. For $\varepsilon > 0$, a player's strategy \mathbf{p} translates into an effective strategy $\mathbf{p}^\varepsilon := (1 - \varepsilon)\mathbf{p} + \varepsilon(1 - \mathbf{p})$. To compute

the payoffs of strategy \mathbf{p} against strategy \mathbf{q} in the presence of errors, we apply the formulas (2.8) – (2.10) to the strategies \mathbf{p}^ε and \mathbf{q}^ε .

Evolutionary dynamics

In the following, we describe the evolutionary process for the baseline scenario. For the various model extensions (Figure 2.6, Figure 2.11 – Figure 2.14), we use appropriately adapted versions of this process, as described in more detail in the Supplementary Note 3. To model how successful strategies spread in well-mixed populations, we use a pairwise comparison process²³². This process considers a population of constant size N . Initially, all population members are unconditional defectors. Each player derives a payoff by interacting with all other population members; for each pairwise interaction, payoffs are given by Eq. (2.9).

To model how strategies with a high payoff spread within a population, we consider a model in discrete time. In each time step, one player is chosen from the population at random. This player is then given an opportunity to revise its strategy. The player can do so in two ways. First, with probability μ (the mutation rate), the player may engage in random strategy exploration. In this case, the player discards its strategy, and samples a new strategy uniformly at random from the set of all memory-1 strategies. Second, with probability $1 - \mu$, the player considers imitating one of its peers. In this case, the player selects a random role model from the population. If the role model's payoff is π_R and the focal player's payoff is π_F , then imitation occurs with a probability given by the Fermi function²²⁴

$$j = \frac{1}{1 + \exp[-\beta(\pi_R - \pi_F)]}. \quad (2.11)$$

If imitation occurs, the focal player discards its previous strategy and adopts the role model's strategy instead. In the formula for the imitation probability, the parameter $\beta \geq 0$ is called the strength of selection. It measures the extent to which players are guided by payoff differences between the players' strategies. For $\beta = 0$, any payoff differences are irrelevant. The focal player adopts the role model's strategy with probability $1/2$. As β becomes larger, payoff differences become increasingly important. In the limiting case $\beta \rightarrow \infty$, imitation only occurs if the role model's payoff at least matches the focal player's payoff.

Overall, the two mechanisms of random strategy exploration and directed strategy imitation give rise to a stochastic process on the space of all population compositions. For positive mutation rates, this process is ergodic. In particular, the average cooperation rate (as a function of the number of time steps) converges, and it is independent of the considered initial population. Herein, we have explored this process with computer simulations. We have recorded which strategies the players adopt over time and how often they cooperate on average. For most of these simulations, we assume that mutations are sufficiently rare²⁴⁸. For those simulations, we require mutant strategies to either fix in the population or to go extinct before the next mutation occurs. Under this regime, the mutant's fixation probability can be computed explicitly¹⁷⁴. This in turn allows us to simulate the evolutionary dynamics more efficiently^{64,105}.

Parameters and specific procedures used for the figures

For the simulations in well-mixed populations, we used the following baseline parameters:

$$\begin{aligned} \text{Benefit of cooperation:} & \quad b=3 \\ \text{Cost of cooperation:} & \quad c=1 \\ \text{Population size:} & \quad N=100 \\ \text{Selection strength:} & \quad \beta=5 \text{ (Figure 2.4, Figures 2.7–2.9) and } \beta=1 \text{ (all other figures)} \\ \text{Error rate:} & \quad \varepsilon=0 \text{ (without errors), or } \varepsilon=0.02 \text{ (with errors)} \\ \text{Mutation rate:} & \quad \mu \rightarrow 0. \end{aligned} \tag{2.12}$$

Changes in these parameters are systematically explored in Figure 2.5 and Figure 2.10. For Figure 2.4, Figure 2.5, and Figure 2.4IPD – Figure 2.12, the respective simulations are run for at least 10^7 time steps each (measured in number of introduced mutant strategies over the course of a simulation). For Figure 2.6, Figure 2.13, and Figure 2.14, simulations are run for a shorter time (as illustrated in the respective panels that illustrate the resulting dynamics). However, here all results are obtained by averaging over 50 – 200 independent simulations.

To report which strategies the players use to sustain cooperation (or defection), we record all strategies that arise during a simulation that have a cooperation rate against themselves of at least 80% (in the case of self-cooperators), or a cooperation rate of less than 20% (in the case of self-defectors). In Figure 2.4, Figure 2.4IPD – Figure 2.8IPD, and Figure 2.11, we show the marginal distributions of all strategies that we have obtained in this way. For these distributions, each strategy is weighted by how long the strategy has been present in the population. In Figure 2.6, Figure

2.13, and Figure 2.14, we represent the self-cooperative strategies by computing the average of the respective marginal distributions. In some cases (Figure 2.6e, Figure 2.13, Figure 2.14), we also report how robust self-cooperative strategies are on average. To this end, we record for each self-cooperative resident strategy how many mutants need to be introduced into the population until a mutant strategy reaches fixation. We consider self-cooperative strategies that resist invasion by many mutant strategies as more robust.

Finally, for the simulations for spatial populations (Figure 2.6g–i), we closely follow the setup of Brauchli et al.³¹. Here, we consider a population of size $N = 2,500$. Players are arranged on a 50×50 lattice with periodic boundary conditions. Players use memory-1 strategies (initially they adopt the strategy ALLD). In every generation, every player interacts in a pairwise game with each of its eight immediate neighbors. After these interactions, all players are independently given an opportunity to update their strategies. With probability $\mu = 0.002$, an updating player chooses a random strategy, uniformly taken from all memory-1 strategies (global mutations). With probability $1 - \mu$, the updating player adopts the strategy of the neighbor with the highest payoff (but only if this neighbor's payoff is better than the focal player's payoff). This elementary process is then repeated for 20,000 generations. Panels Figure 2.6g,i show averages across 50 independent simulations of the process. Panel Figure 2.6(h) illustrates two particular realizations.

Code availability. All simulations were performed with Matlab_R2019b. The respective code is available online¹⁸⁰, at osf.io: <http://dx.doi.org/10.17605/osf.io/v5hgd>.

Data availability. Source data for Figures 2.4, Figure 2.5, Figure 2.6a–f and Figure

2.6i are provided with this paper. Moreover, the raw data generated with the computer simulations, including the data that is necessary to create all figures is available online¹⁸⁰, at osf.io:

<http://dx.doi.org/10.17605/osf.io/v5hgd>.

Acknowledgments. P.S.P. is supported by the National Science Foundation through the Graduate Research Fellowship Program (grant number DGE1745303), by the Centre for Effective Altruism through the Global Priorities Fellowship, and by Harvard University through graduate student fellowships. C.H. acknowledges generous funding by the Max Planck Society and by the ERC Starting Grant E-DIRECT (850529).

Author contributions. P.S.P., M.A.N., and C.H: Designed the research; P.S.P., M.A.N., and C.H: Performed the research; P.S.P., M.A.N., and C.H: Wrote the paper.

Competing interests. The authors declare no conflict of interest.

Supplementary Information

We note that in the following, all references to equations refer to the respective equation in the Supplementary Information document; we do not refer to main text equations herein.

2.5 Supplementary Note 1: Baseline model

Game setup. In the following, we introduce the model in slightly more general terms than in the main text. We consider two players who interact repeatedly. Each turn, they either decide at the same time whether or not to cooperate (simultaneous game), or they decide consecutively, one after the other (alternating game). In the former case, players do not know of the other player's decision when making their own decision. In the latter case, player 1 moves first and player 2 learns the outcome before making its own decision. To ensure payoffs are well-defined in both cases, we assume the payoff of an action can be defined based on that particular action alone (that is, the payoff consequences of one player's cooperation does not depend on the co-player's action). This implies that payoffs take the form of the donation game¹⁷¹. That is, cooperation (C) implies a cost of $c > 0$ to the cooperating player, and it yields a benefit $b > c$ to the co-player. Defection (D) comes with no cost and yields no benefit.

Here we assume the game proceeds indefinitely and future payoffs are not discounted. For such repeated games, we can define the players' payoffs as follows. Let $v_{a_1, a_2}(t)$ denote the probability that the t -th actions of player 1 and player 2 are a_1 and a_2 , respectively, with $a_1, a_2 \in \{C, D\}$. Throughout this paper, we assume that for all

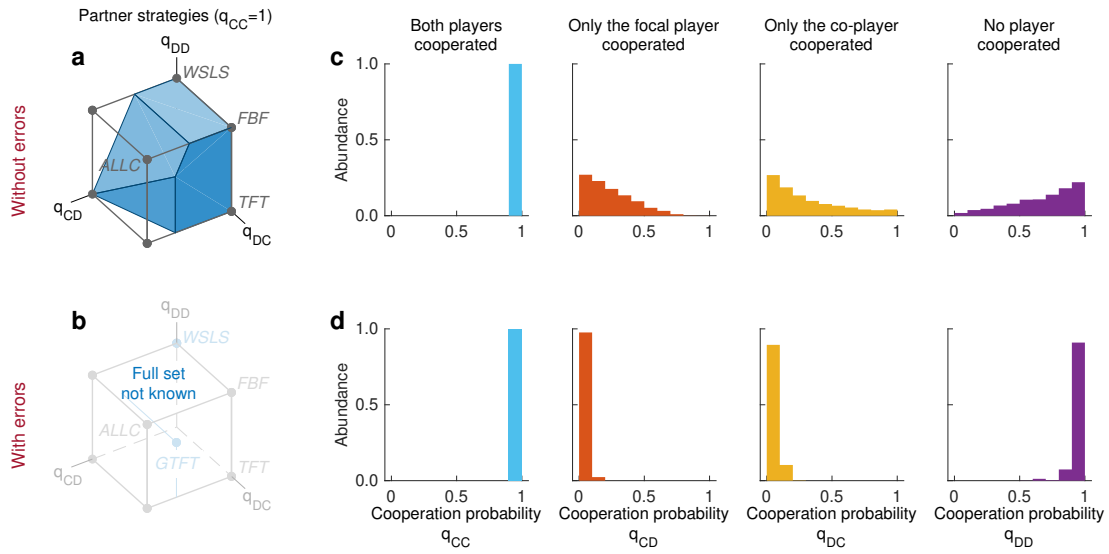


Figure 2.7: Partners in the simultaneous game. Here, we present results analogous to those of Figure 2.4, but illustrating the outcome of simultaneous games. a,c, Without errors, the two models make the same predictions, and also the evolving cooperative strategies are similar. b, With errors, there is no theory yet that characterizes all partner strategies of the simultaneous game. It is only known that particular cooperative strategies are stable under certain conditions. For example, WLSL is a partner strategy if $b > 2c$, provided the error rate is sufficiently small⁹⁸. Another example of a partner strategy is GTFT (defined in the same way as in the alternating game; for reactive strategies, the two games are equivalent¹⁷¹). d, For simultaneous games with errors, our evolutionary simulations confirm that players predominantly maintain cooperation with WLSL. Parameters are the same as in Figure 2.4.

a_1, a_2 the respective limiting averages are well-defined,

$$v_{a_1, a_2} := \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{a_1, a_2}(t). \quad (2.13)$$

These limits are guaranteed to exist, for example, when the two players have finite recall. We collect the four limiting averages defined by Eq. (2.13) in a vector,

$$\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD}). \quad (2.14)$$

Each entry corresponds to the probability to observe the respective outcome at a randomly picked time t . Based on these four probabilities, we define the players' payoffs by

$$\begin{aligned} \pi_1 &= b \cdot (v_{CC} + v_{DC}) - c \cdot (v_{CC} + v_{CD}) \\ \pi_2 &= b \cdot (v_{CC} + v_{CD}) - c \cdot (v_{CC} + v_{DC}). \end{aligned} \quad (2.15)$$

These formulas apply to both the alternating and the simultaneous game (however, the respective limiting averages \mathbf{v} will generally differ, see below).

Memory-1 strategies. We assume players use memory-1 strategies. That is, to decide whether to cooperate in a given round, a player only takes into account each player's most recent decision. Such strategies can be written as a 4-tuple

$$\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD}). \quad (2.16)$$

An entry $p_{a\tilde{a}}$ is the probability the focal player cooperates, given that the focal player's last decision was $a \in \{C, D\}$ and that the opponent's last decision was $\tilde{a} \in \{C, D\}$. Such a strategy is deterministic if all entries are either zero or one; it is semi-stochastic if some but not all entries are between zero and one; and it is fully stochastic if all en-

tries are between zero and one.

When both players use memory-1 strategies, the payoffs according to Eq. (2.15) can be calculated explicitly. To this end, let us consider two players with strategies $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ and $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$, respectively. The game can be represented as a Markov chain, where the states are the possible combinations of the two players' actions at a given point in time, $\{CC, CD, DC, DD\}$. For the alternating game, the Markov chain's transition matrix is

$$M_A(\mathbf{p}, \mathbf{q}) := \begin{pmatrix} p_{CC}q_{CC} & p_{CC}(1-q_{CC}) & (1-p_{CC})q_{CD} & (1-p_{CC})(1-q_{CD}) \\ p_{CD}q_{DC} & p_{CD}(1-q_{DC}) & (1-p_{CD})q_{DD} & (1-p_{CD})(1-q_{DD}) \\ p_{DC}q_{CC} & p_{DC}(1-q_{CC}) & (1-p_{DC})q_{CD} & (1-p_{DC})(1-q_{CD}) \\ p_{DD}q_{DC} & p_{DD}(1-q_{DC}) & (1-p_{DD})q_{DD} & (1-p_{DD})(1-q_{DD}) \end{pmatrix}. \quad (2.17)$$

By the Perron-Frobenius Theorem, the vector \mathbf{v} defined by Eq. (2.14) is an invariant distribution of $M_A(\mathbf{p}, \mathbf{q})$. That is, to compute how often players visit each of the four states, we only need to solve the following linear equation in the unknown $\mathbf{v}(\mathbf{p}, \mathbf{q})$,

$$\mathbf{v}(\mathbf{p}, \mathbf{q}) = \mathbf{v}(\mathbf{p}, \mathbf{q}) \cdot M_A(\mathbf{p}, \mathbf{q}). \quad (2.18)$$

Based on this invariant distribution, one can then compute payoffs based on Eq. (2.15),

$$\begin{aligned} \pi(\mathbf{p}, \mathbf{q}) &= b \cdot (v_{CC}(\mathbf{p}, \mathbf{q}) + v_{DC}(\mathbf{p}, \mathbf{q})) - c \cdot (v_{CC}(\mathbf{p}, \mathbf{q}) + v_{CD}(\mathbf{p}, \mathbf{q})) \\ \pi(\mathbf{q}, \mathbf{p}) &= b \cdot (v_{CC}(\mathbf{p}, \mathbf{q}) + v_{CD}(\mathbf{p}, \mathbf{q})) - c \cdot (v_{CC}(\mathbf{p}, \mathbf{q}) + v_{DC}(\mathbf{p}, \mathbf{q})). \end{aligned} \quad (2.19)$$

For the simultaneous game, payoffs can be computed analogously, but using a differ-

ent transition matrix²¹⁵,

$$M_S(\mathbf{p}, \mathbf{q}) := \begin{pmatrix} p_{CC}q_{CC} & p_{CC}(1-q_{CC}) & (1-p_{CC})q_{CC} & (1-p_{CC})(1-q_{CC}) \\ p_{CD}q_{DC} & p_{CD}(1-q_{DC}) & (1-p_{CD})q_{DC} & (1-p_{CD})(1-q_{DC}) \\ p_{DC}q_{CD} & p_{DC}(1-q_{CD}) & (1-p_{DC})q_{CD} & (1-p_{DC})(1-q_{CD}) \\ p_{DD}q_{DD} & p_{DD}(1-q_{DD}) & (1-p_{DD})q_{DD} & (1-p_{DD})(1-q_{DD}) \end{pmatrix}. \quad (2.20)$$

In cases in which it is clear which game and which strategies \mathbf{p} and \mathbf{q} are considered (or in case the game and the exact strategies do not matter), we will sometimes write \mathbf{v} and M instead of $\mathbf{v}(\mathbf{p}, \mathbf{q})$, $M_A(\mathbf{p}, \mathbf{q})$, and $M_S(\mathbf{p}, \mathbf{q})$.

We note that in degenerate cases, the solution of $\mathbf{v} = \mathbf{v}M$ does not need to be unique. In that case, the correct invariant distribution \mathbf{v} needs to be derived from the players' actions in the very first round. As an example, consider an alternating game in which both players adopt the strategy TFT. The corresponding transition matrix M_A has two absorbing states. The first absorbing state corresponds to indefinite mutual cooperation, and the other corresponds to indefinite mutual defection. Which of these absorbing states is reached (and hence which of the invariant distributions is relevant for the calculation of the players' payoffs) depends on player 1's action in the very first round (when no previous history of actions is yet available). If player 1 cooperates, both players continue to cooperate, and the appropriate invariant distribution is $\mathbf{v} = (1, 0, 0, 0)$. Otherwise, if player 1 defects, the appropriate invariant distribution is $\mathbf{v} = (0, 0, 0, 1)$. We note that if the two TFT players interact in a simultaneous game, the respective transition matrix M_S has a third absorbing state. According to that state, players switch between cooperation and defection. In the alternating game this state is no longer absorbing¹⁷¹, because players now condition their behavior on different past events (as illustrated in Figure 2.1).

Reactive strategies. An important subset of memory-1 strategies are the so-called reactive strategies. While the behavior of a reactive strategy still depends on the opponent's previous decision, it is independent of the player's own previous decision. Such strategies correspond to those 4-tuples $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ for which $p_{CC} = p_{DC}$ and $p_{CD} = p_{DD}$. Slightly abusing notation, we denote reactive strategies as 2-tuples $\mathbf{p} = (p_C, p_D)$. The first entry $p_C := p_{CC} = p_{DC}$ is the player's cooperation probability given that the opponent's last decision was to cooperate. The second entry $p_D := p_{CD} = p_{DD}$ is the player's cooperation probability given that the opponent's last decision was to defect. Examples of reactive strategies include ALLD = (0, 0), ALLC = (1, 1), TFT = (1, 0) and ATFT = (0, 1).

2.6 Supplementary Note 2: Equilibrium analysis for alternating games

In the following, we aim to characterize all symmetric Nash equilibria of the alternating game in the space of memory-1 strategies. In a Nash equilibrium, no player can increase her payoff by unilaterally deviating. To do so, we use an approach that is different from previous approaches for the simultaneous game^{3,4,96,220,221}. Our approach involves two steps. First, we show that for any game between two memory-1 players, one can replace one player's strategy by an appropriately chosen reactive strategy without affecting the resulting payoffs. This step is somewhat reminiscent of a result by Press and Dyson¹⁹⁰. They showed for the simultaneous game that there is no advantage of having a longer memory than the opponent. For alternating games, a stronger result holds. Against a memory-1 opponent, a player can even afford to have a lower memory. It suffices to only remember the opponent's last move and to forget one's own. Second, we show that to find a best response to a given memory-1 strat-

egy, it is sufficient to check it against those reactive strategies that are deterministic. This result implies that one needs to explore only four possible deviations, ALLD, ALLC, TFT, and ATFT, as defined above.

Based on these two results, we show that the alternating game allows for three qualitatively different classes of memory-1 equilibria. According to the first two classes, players either mutually cooperate or mutually defect. We refer to the respective strategies as partners and defectors, respectively. In the last class, players act in such a way that the opponent's payoff is guaranteed to be fixed, irrespective of the opponent's strategy. These strategies have been called equalizers in the context of the simultaneous game^{21,190}.

2.6.1 Sufficiency of reactive strategies

As our first result, we show that when two memory-1 players interact, one player's strategy can be replaced by an appropriate reactive strategy without affecting the game's outcome (all proofs are presented as an appendix in Supplementary Note 4). For simplicity, we show this result for the first player. However, because payoffs are independent of the position of the players¹⁷¹, an analogous result holds for the second player.

Proposition 2.1 (Sufficiency of reactive strategies when both players use memory-1 strategies). Consider two memory-1 players with strategies $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ and $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$, and suppose $\mathbf{v}(\mathbf{p}, \mathbf{q}) = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$ is an invariant distribution of the resulting alternating game. We define a reactive strategy $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$

for player 1 as a solution of

$$\begin{aligned}(v_{CC}+v_{DC})\tilde{p}_C &= v_{CC}p_{CC} + v_{DC}p_{DC} \\ (v_{CD}+v_{DD})\tilde{p}_D &= v_{CD}p_{CD} + v_{DD}p_{DD}.\end{aligned}\tag{2.21}$$

Then $\mathbf{v}(\tilde{\mathbf{p}}, \mathbf{q}) = \mathbf{v}(\mathbf{p}, \mathbf{q})$. We call such a strategy $\tilde{\mathbf{p}}$ a reactive marginalization of \mathbf{p} with respect to \mathbf{q} .

Several remarks are in order.

a Intuition for the result. To gain some intuition for Proposition 2.1, let us assume that the strategies \mathbf{p} and \mathbf{q} are such that player 2 both cooperates and defects with positive probability. In that case, $v_{CC}+v_{DC} > 0$ and $v_{CD}+v_{DD} > 0$, and the reactive marginalization of \mathbf{p} with respect to \mathbf{q} is unique,

$$\begin{aligned}\tilde{p}_C &= \frac{v_{CC}}{v_{CC}+v_{DC}}p_{CC} + \frac{v_{DC}}{v_{CC}+v_{DC}}p_{DC} \\ \tilde{p}_D &= \frac{v_{CD}}{v_{CD}+v_{DD}}p_{CD} + \frac{v_{DD}}{v_{CD}+v_{DD}}p_{DD}.\end{aligned}\tag{2.22}$$

That is, to obtain the value of \tilde{p}_C , we only need to consider how often the first player cooperates in response to the opponent's cooperation on average. To this end, all outcomes in which the first player cooperates are weighted according to how often these outcomes occur in the first place. Figure 3 provides an illustration for two particular examples of strategies \mathbf{p} and \mathbf{q} .

b Proposition 2.1 only applies to alternating games. To illustrate that the statement is not true for simultaneous games, consider the strategies used in Figure 3, with $\mathbf{p} = (0.9, 0.1, 0.5, 0.3)$ and $\mathbf{q} = (0.8, 0.25, 0.75, 0.2)$. By computing the respective transition matrix M_S according to Eq. (2.20), and by solving for

$\mathbf{v} = \mathbf{v}M_S$, we obtain the invariant distribution $\mathbf{v}(\mathbf{p}, \mathbf{q}) \approx (0.23, 0.21, 0.24, 0.32)$. If we use this expression and formula (2.22) to compute the unique reactive marginalization of \mathbf{p} with respect to \mathbf{q} , we obtain $\tilde{\mathbf{p}} \approx (0.698, 0.220)$. However, the respective invariant distribution of a simultaneous game between $\tilde{\mathbf{p}}$ and \mathbf{q} is $\mathbf{v}(\tilde{\mathbf{p}}, \mathbf{q}) \approx (0.22, 0.23, 0.25, 0.30)$, which is different from $\mathbf{v}(\mathbf{p}, \mathbf{q})$. Hence the simultaneous game between the two memory-1 strategies $\tilde{\mathbf{p}}$ and \mathbf{q} induces a dynamics that is different from the simultaneous game between \mathbf{p} and \mathbf{q} .

c Non-uniqueness of reactive marginalizations. In some cases of alternating games, a strategy's reactive marginalization is not unique. This happens, for example, if player 1 uses the strategy $\mathbf{p} = (0, 0, 1, 0)$ and the opponent uses GRIM = $(1, 0, 0, 0)$. The respective transition matrix M_A according to Eq. (2.17) has a unique invariant distribution according to which everyone defects, $\mathbf{v} = (0, 0, 0, 1)$. By Proposition 2.1 it follows that for any reactive strategy $\tilde{\mathbf{p}} = (\tilde{p}_C, 0)$ with $\tilde{p}_C \in [0, 1]$, again \mathbf{v} is an invariant distribution of the game against \mathbf{q} . We note however, that one of these reactive strategies, $\tilde{\mathbf{p}} = (1, 0)$, allows for a second invariant distribution, $\mathbf{v} = (1, 0, 0, 0)$. When this reactive marginalization is chosen, we additionally need to require that player 1 defects in the very first round, such that the correct invariant distribution is selected.

The above Proposition 2.1 suggests that against a given memory-1 opponent, there is no advantage of choosing a memory-1 strategy instead of a reactive strategy: any payoff a player can achieve with a memory-1 strategy can also be achieved with a reactive strategy. This result holds more generally, even if player 1 has access to more complex strategies.

Proposition 2.2 (Sufficiency of reactive strategies when only the second player uses a memory-1 strategy). Consider an alternating game in which the second player uses the memory-1 strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$ whereas the first player uses an arbitrary strategy. Denote by $p_{a_1, a_2}(t)$ the first player's expected probability to cooperate at time t conditional on the players' previous decisions a_1 and a_2 . Suppose the limiting distribution \mathbf{v} according to Eq. (2.14) and the following limits on the right hand side exist, such that we can define \tilde{p}_C and \tilde{p}_D implicitly as a solution of

$$\begin{aligned} (v_{CC} + v_{DC})\tilde{p}_C &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CC}(t)p_{CC}(t) + v_{DC}(t)p_{DC}(t) \\ (v_{CD} + v_{DD})\tilde{p}_D &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CD}(t)p_{CD}(t) + v_{DD}(t)p_{DD}(t). \end{aligned} \tag{2.23}$$

The reactive strategy $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$ so defined satisfies $\mathbf{v}(\tilde{\mathbf{p}}, \mathbf{q}) = \mathbf{v}$.

The requirements imposed by Proposition 2.2 are comparably mild. The existence of the respective limits is guaranteed, for example, if player 1 adopts an arbitrary strategy with finite recall (in some cases, this may again require players to specify their initial actions to make sure the invariant distribution is well-defined). In the following, we say that those strategies that satisfy the conditions in Proposition 2.2 are generic with respect to \mathbf{q} . That is, generic strategies are those for which one can compute how likely a player is to cooperate on average, conditional on the co-player's previous action. In particular, all memory-1 strategies are generic (given their initial actions are defined). In that special case, the respective definition of $\tilde{\mathbf{p}}$ according to Eqs. (2.21) and (2.23) coincide, as one may expect.

We can summarize the results in this section as follows. Given a fixed memory-1 strategy \mathbf{q} for the second player, we can define the following three sets. These sets

describe both players' feasible payoffs, given the first player either adopts a reactive strategy, a memory-1 strategy, or a generic strategy, respectively,

$$\begin{aligned}
\Pi_R(\mathbf{q}) &:= \left\{ \left(\pi(\mathbf{p}, \mathbf{q}), \pi(\mathbf{q}, \mathbf{p}) \right) \in \mathbb{R}^2 \mid \mathbf{p} \text{ is a reactive strategy} \right\} \\
\Pi_M(\mathbf{q}) &:= \left\{ \left(\pi(\mathbf{p}, \mathbf{q}), \pi(\mathbf{q}, \mathbf{p}) \right) \in \mathbb{R}^2 \mid \mathbf{p} \text{ is a memory-1 strategy} \right\} \\
\Pi_G(\mathbf{q}) &:= \left\{ \left(\pi(\mathbf{p}, \mathbf{q}), \pi(\mathbf{q}, \mathbf{p}) \right) \in \mathbb{R}^2 \mid \mathbf{p} \text{ is a generic strategy} \right\}
\end{aligned} \tag{2.24}$$

Then Propositions 2.1 and 2.2 imply the following.

Corollary 2.3. If \mathbf{q} is a memory-1 strategy, then $\Pi_R(\mathbf{q}) = \Pi_M(\mathbf{q}) = \Pi_G(\mathbf{q})$.

That is, against a memory-1 opponent, all payoffs that can either be achieved with a generic strategy, or a memory-1 strategy, can already be achieved with a reactive strategy.

2.6.2 Best responses to memory-1 strategies

In this section, we aim to identify best responses to a given memory-1 strategy. We restrict ourselves to generic best responses. A strategy \mathbf{p} is a generic best response to strategy \mathbf{q} if it is generic, and if

$$\pi(\mathbf{p}, \mathbf{q}) \geq \pi(\mathbf{p}', \mathbf{q}) \quad \text{for all generic strategies } \mathbf{p}'. \tag{2.25}$$

By Proposition 2.2, there is always a generic best response in the space of reactive strategies. The following two results simplify the search for a generic best response even further.

Lemma 2.4. Consider a reactive player with strategy $\mathbf{p} = (p_C, p_D)$ who interacts with a memory-1 opponent with strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$. Then, the payoff of the reactive player is given by

$$\pi(\mathbf{p}, \mathbf{q}) = \frac{bq_{DD} - cq_{DD} \cdot p_C + ((q_{DC} - q_{DD})b - (1 - q_{CD})c) \cdot p_D + c(q_{CC} - q_{CD} - q_{DC} + q_{DD}) \cdot p_C p_D}{1 - q_{CD} + q_{DD} - (q_{CC} - q_{CD}) \cdot p_C - (q_{DD} - q_{DC}) \cdot p_D}. \quad (2.26)$$

In particular, the payoff of the reactive player depends monotonically on each of its inputs p_C and p_D .

The first part of the Lemma gives an explicit formula to compute payoffs. The monotonicity property mentioned in the second part is useful because it allows us to derive the following result.

Proposition 2.5 (Optimality of deterministic reactive strategies). Let \mathbf{q} be some given memory-1 strategy and let $\mathbf{p} \in [0, 1]^2$ be an arbitrary reactive strategy. Then there is a deterministic reactive strategy $\mathbf{p}' \in \{0, 1\}^2$ for which $\pi(\mathbf{p}', \mathbf{q}) \geq \pi(\mathbf{p}, \mathbf{q})$.

In particular, if \mathbf{p} is a best response to \mathbf{q} , then there is at least one deterministic and reactive strategy \mathbf{p}' that yields the same payoff (that is, \mathbf{p}' is also a best response).

By combining Propositions 2.2 and 2.5, we can thus conclude that to find a generic best response to an arbitrary memory-1 strategy, it suffices to consider the four deterministic reactive strategies ALLD = (0, 0), ATFT = (0, 1), TFT = (1, 0), ALLC = (1, 1).

We can use this observation to characterize all generic Nash equilibria among the memory-1 strategies. We say a strategy \mathbf{q} is a generic Nash equilibrium if \mathbf{q} is a generic best reply to itself. By Eq. (2.19), the payoff of a memory-1 strategy \mathbf{q} against itself is

$$\pi(\mathbf{q}, \mathbf{q}) = \frac{(1-q_{CC}+q_{DC})q_{DD}}{(1-q_{CC})(1-q_{CD}+q_{DD})+(1-q_{CC}+q_{DC})q_{DD}} \cdot (b-c). \quad (2.27)$$

Because of Propositions 2.2 and 2.5, we only need to compare this self-payoff to the payoffs of the four deterministic reactive strategies. By Eq. (2.26), the respective payoffs are

$$\begin{aligned} \pi(\text{ALLD}, \mathbf{q}) &= \frac{q_{DD}}{1-q_{CD}+q_{DD}} \cdot b \\ \pi(\text{ATFT}, \mathbf{q}) &= \frac{q_{DC}}{1-q_{CD}+q_{DC}} \cdot b - \frac{1-q_{CD}}{1-q_{CD}+q_{DC}} \cdot c \\ \pi(\text{TFT}, \mathbf{q}) &= \frac{q_{DD}}{1-q_{CC}+q_{DD}} \cdot (b-c) \\ \pi(\text{ALLC}, \mathbf{q}) &= \frac{q_{DC}}{1-q_{CC}+q_{DC}} \cdot b - c, \end{aligned} \quad (2.28)$$

Overall, we obtain the following result.

Theorem 2.6 (Characterization of generic Nash equilibria). Let \mathbf{q} be an arbitrary memory-1 strategy such that the payoffs (2.27) and (2.28) are well-defined. Then \mathbf{q} is a generic Nash equilibrium if and only if

$$\pi(\mathbf{q}, \mathbf{q}) \geq \max \left(\pi(\text{ALLD}, \mathbf{q}), \pi(\text{ATFT}, \mathbf{q}), \pi(\text{TFT}, \mathbf{q}), \pi(\text{ALLC}, \mathbf{q}) \right) \quad (2.29)$$

The assumption on the payoffs (2.27) and (2.28) to be well-defined is not a major restriction. Those cases in which some of the expressions in Eqs. (2.27) and (2.28) cannot be evaluated (for example, when $q_{CC} = 1$ and $q_{DD} = 0$), correspond to those cases in which the invariant distribution \mathbf{v} according to Eq. (2.18) is not unique. In that case, one can resolve the ambiguity by defining an initial cooperation probability for

the very first round. In this way, all relevant payoffs become well-defined, and condition (2.29) remains valid.

2.6.3 Classification of memory-1 Nash equilibria

In the following, we use the general characterization provided in Theorem 2.6 to give a qualitative classification of all generic memory-1 Nash equilibria. To this end, we first describe three distinct behaviors that can be sustained in equilibrium. These three behaviors correspond to players who mutually cooperate, players who mutually defect, and players who unilaterally fix the co-player's payoff to a fixed level. In line with the previous literature on simultaneous games, we refer to the respective equilibrium strategies as partners⁹⁶, defectors⁵³, and equalizers²¹. Then we show that these three classes of behaviors comprise in fact all Nash equilibria of the alternating game.

Partners. We say a strategy is self-cooperative if two players with that strategy obtain the mutual cooperation payoff $b-c$ against each other. For a memory-1 strategy \mathbf{q} to be self-cooperative, Eq. (2.27) implies that q_{CC} needs to be set to one (if $q_{DD} = 0$, the strategy is additionally required to cooperate in the first round). We call a strategy a partner if it is self-cooperative and if it satisfies the Nash condition (2.29). To check whether the Nash condition holds, we note that for any self-cooperative strategy \mathbf{q} , the equality $\pi(\mathbf{q}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q}) = \pi(\text{ALLC}, \mathbf{q}) = b-c$ holds. Thus, we only need to verify the two remaining inequalities, $\pi(\text{ALLD}, \mathbf{q}) \leq b-c$ and $\pi(\text{ATFT}, \mathbf{q}) \leq b-c$. Based on the respective expressions in Eqs. (2.28), we conclude that \mathbf{q} is a partner if

and only if the following three conditions are satisfied,

$$\begin{aligned}
q_{CC} &= 1 \\
(b-c)(1-q_{CD}) &\geq c q_{DD} \\
b(1-q_{CD}) &\geq c q_{DC}.
\end{aligned} \tag{2.30}$$

These conditions define a 3-dimensional subspace of the memory-1 strategies (see Figure 2.4). This subspace is non-empty: since $b > c$, all conditions can be met by choosing sufficiently small cooperation probabilities q_{CD}, q_{DC}, q_{DD} . The subspace of partner strategies increases with b and it decreases with c . That is, the more profitable cooperation is, the easier it becomes to satisfy the conditions for being a partner.

Defectors. We call a strategy self-defective if two players with that strategy end up with the mutual defection payoff when playing against each other. In particular, a self-defective memory-1 strategy \mathbf{q} needs to set q_{DD} to zero (in case payoffs are not well-defined otherwise, it additionally needs to defect in the first round). We say a self-defective strategy is a defector if it additionally satisfies the Nash condition (2.29). Similar to before, two of the four conditions in Eq. (2.29) are automatically met because a self-defective strategy satisfies $\pi(\mathbf{q}, \mathbf{q}) = \pi(\text{ALLD}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q}) = 0$. Thus, we only need to verify $\pi(\text{ATFT}, \mathbf{q}) \leq 0$ and $\pi(\text{ALLC}, \mathbf{q}) \leq 0$. Overall, we obtain the following characterization of defectors,

$$\begin{aligned}
q_{DD} &= 0 \\
b q_{DC} &\leq c(1-q_{CD}) \\
(b-c) q_{DC} &\leq c(1-q_{CC}).
\end{aligned} \tag{2.31}$$

Again, these conditions define a 3-dimensional non-empty subspace of memory-1

strategies (Figure 2.8). The volume of this subspace increases if we either reduce b or increase c .

Equalizers. For simultaneous games, it has been noted that the memory-1 strategies contain a subset of so-called equalizers^{21,190}. If player 2 adopts such a strategy \mathbf{q} , player 1's payoff $\pi(\mathbf{p}, \mathbf{q})$ is a constant, independent of player 1's strategy. Obviously, if both players choose an equalizer strategy, the resulting strategy profile is a Nash equilibrium. In that case, no player can get a different payoff – let alone a larger payoff – by unilaterally deviating.

In the following we aim to identify equalizers in the context of alternating games. That is, we ask which strategies player 2 can use to make sure that player 1's payoff is independent of player 1's strategy. As a minimum requirement, player 2's strategy \mathbf{q} needs to enforce the same payoff upon all co-players with reactive and deterministic strategies, such that

$$\pi(\text{ALLD}, \mathbf{q}) = \pi(\text{ATFT}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q}) = \pi(\text{ALLC}, \mathbf{q}). \quad (2.32)$$

This yields three equations in the four unknown entries of \mathbf{q} . By using Eqs. (2.28) to solve $\pi(\text{ALLD}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q})$ for q_{CD} and $\pi(\text{ALLC}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q})$ for q_{DC} , we obtain

$$\begin{aligned} q_{CD} &= \frac{b q_{CC} - c(1 + q_{DD})}{b - c} \\ q_{DC} &= \frac{b q_{DD} + c(1 - q_{CC})}{b - c}. \end{aligned} \quad (2.33)$$

Given these two relations hold, one can verify that the last relation $\pi(\text{ATFT}, \mathbf{q}) = \pi(\text{TFT}, \mathbf{q})$ holds automatically. Conversely, suppose a memory-1 strategy \mathbf{q} satisfies these two conditions. Because these conditions imply Eq. (2.32) and because

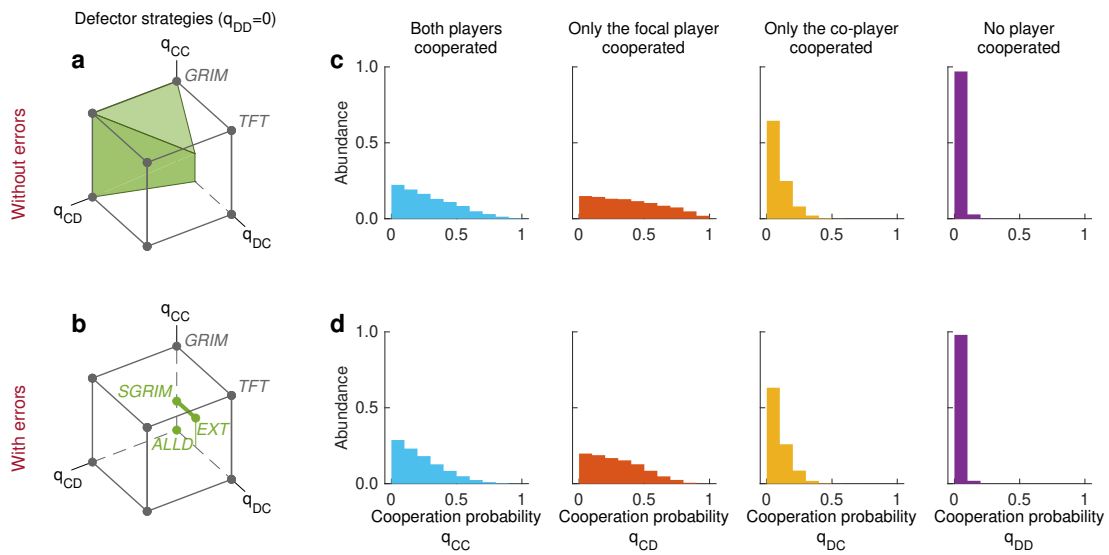


Figure 2.8: Defectors in the alternating prisoner's dilemma. As we have done for partners, we have also characterized all defectors among the memory-1 strategies for the alternating game. a,b, We provide explicit conditions for the case without errors and the case with errors, see Supplementary Note 2. With errors, there are two classes of defector strategies. First, there is the atomic class of unconditional defection (ALLD). Second, there is a line segment that connects a stochastic version of GRIM to the strategy EXT; the latter is a limiting version of the previously described extortionate strategies^{97,190,226,227,250}. c,d, For the evolutionary simulations, we use the same parameters as in Figure 2.4.

the payoffs of a reactive player $\mathbf{p} = (p_C, p_D)$ are monotonic in p_C and in p_D (due to Lemma 2.4), it follows that any reactive strategy obtains the same payoff against \mathbf{q} . Moreover, because generic strategies can be replaced by reactive strategies (Proposition 2.1), it follows that any generic strategy obtains the same payoff against \mathbf{q} . We conclude that equalizers are exactly those strategies that satisfy Eq. (2.33). In particular, equalizers correspond to a 2-dimensional subspace of memory-1 strategies. The following technical result allows us to show that the three above classes of partners, defectors, and equalizers are in fact all generic Nash equilibria within the space of memory-1 strategies.

Lemma 2.7. Consider a memory-1 strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$ that is a generic Nash equilibrium, and let $\tilde{\mathbf{q}} = (\tilde{q}_C, \tilde{q}_D)$ denote its reactive marginalization with respect to itself.

1. If $\tilde{\mathbf{q}}$ is fully stochastic, then \mathbf{q} is an equalizer.
2. If $\tilde{\mathbf{q}}$ is semi-stochastic or deterministic, then \mathbf{q} is either a partner or a defector.

We can summarize the previous results as follows.

Theorem 2.8 (Classification of generic Nash equilibria). A memory-1 strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$ is a generic Nash equilibrium if and only if it is either a partner, a defector, or an equalizer: that is, if and only if it meets the conditions (2.30), (2.31), or (2.33).

Comparing Theorem 2.8 for the alternating game with the respective classification of equilibrium outcomes in the simultaneous game^{4,96,221} yields the following two insights:

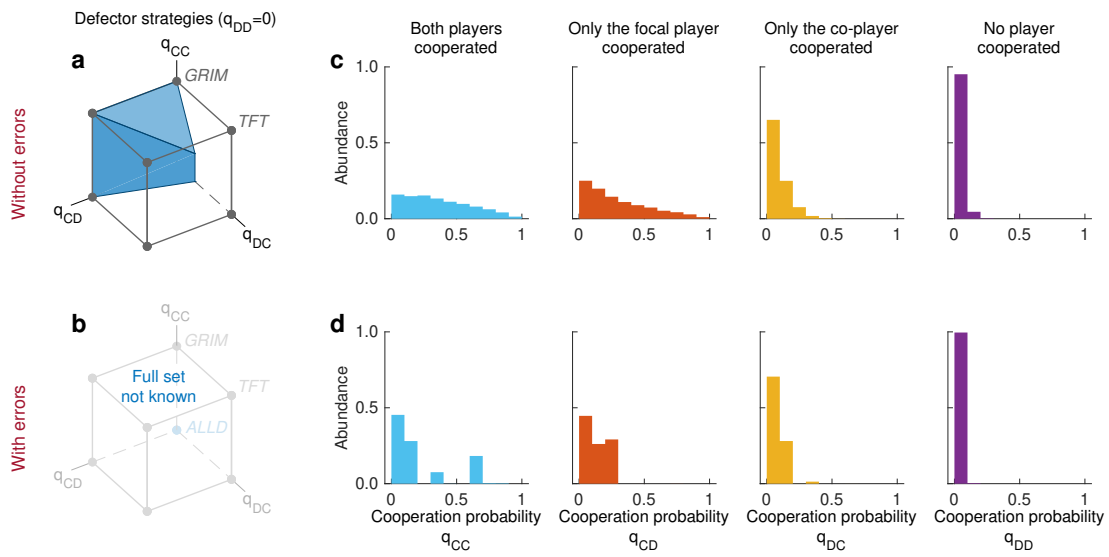


Figure 2.9: Defectors in the simultaneous game. a,b, The figure is analogous to Figure 2.8, but for the case of the simultaneous game instead of the alternating game. For the simultaneous game with errors, there is no complete characterization of defector strategies as of yet. However, it is known that ALLD is a Nash equilibrium for all parameter values, because it simply reiterates the Nash equilibrium of the one-shot game. c,d, For the evolutionary simulations, we use the same parameters as in Figure 2.4.

1. The three classes that we have identified, partners, defectors, and equalizers, also exist in the simultaneous game. In fact, even the respective equilibrium conditions are identical: a memory-1 strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$ is a partner, a defector, or an equalizer in the alternating game if and only if it is a partner, defector, or equalizer in the simultaneous game (assuming that the game parameters b and c are the same).
2. However, the simultaneous game allows for one additional class of equilibrium strategies, called self-alternators²²¹. When two self-alternators interact, they cooperate in turns: one player unilaterally cooperates in one round, and the other player unilaterally cooperates in the next. To be a Nash equilibrium for the simultaneous game, self-alternators need to have the form²²¹

$$q_{CC} \leq \frac{2c}{b+c}, \quad q_{CD}=0, \quad q_{DC}=1, \quad q_{DD} \leq \frac{b-c}{b+c}. \quad (2.34)$$

However, according to Theorem 2.8, strategies that satisfy the conditions in (2.34) do not give rise to a Nash equilibrium in the alternating game.

The intuition is easy to convey with an example. To this end, let us consider the memory-1 strategy $\mathbf{q} = (0, 0, 1, 1/3)$ which satisfies conditions (2.34) for all games with $b > 2c$. In the simultaneous game, two players with strategy \mathbf{q} reliably learn to cooperate in turns irrespective of their first-round behavior. This is illustrated by the following sample path (an asterisk indicates a decision that is partly due to chance). In this path, players reliably alternate from the fourth round onwards,

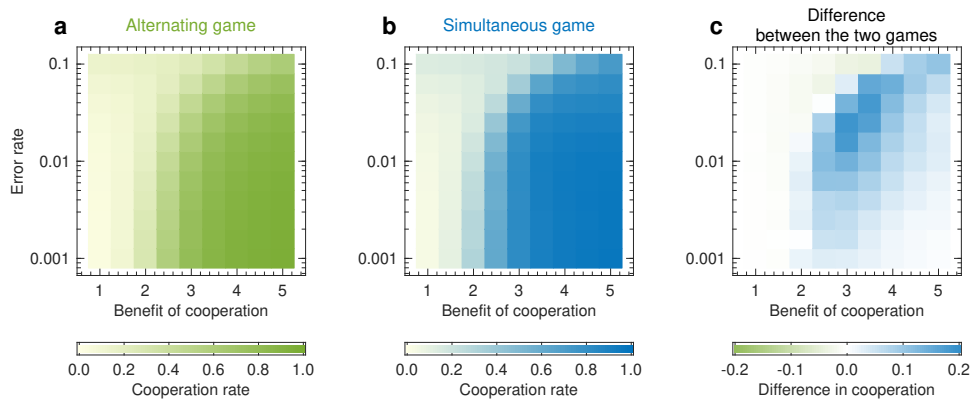


Figure 2.10: Comparing the alternating and the simultaneous game across different error rates and benefit values. We have run further simulations to explore the joint effect of the error rate and the benefit of cooperation, for both the alternating game (a) and the simultaneous game (b). In both cases, we observe high cooperation rates for high benefit values and sufficiently small error rates. c, Here, we plot the difference in cooperation rates between the simultaneous and the alternating game. This difference is small for small benefit values (where defection evolves in both settings). It is also small for large benefits when the error rate is small (for which nearly full cooperation evolves in both settings). In between, for intermediate benefit values and intermediate error rates, the simultaneous game yields systematically more cooperation than the alternating game. Baseline parameters are the same as in Figure 2.5.

Player 1	C	D	D*	C*	D	C	D	C	D	...
Player 2	C	D	D*	D*	C	D	C	D	C	...

In the alternating game, the same first-round behavior gives rise to a more irregular trajectory,

Player 1	C	D	D*	C*	D	D*	C	D	D*	...
Player 2	C	D	D*	C	D	C*	D	D*	C*	...

Here, there are always two consecutive instances of cooperation, which leads both players to defect on their next turn (because $q_{CC} = q_{CD} = 0$). Once both players have defected, it may take several turns for one player to re-start cooperation (because $q_{DD} = 1/3$). As a result, the limiting average payoff in the alternating game is only $(b-c)/3$. If one of the players were to switch to ATFT, the resulting payoff is $\pi(\text{ATFT}, \mathbf{q}) = (b-c)/2 > (b-c)/3$. Hence, \mathbf{q} is unstable. We conclude that strategies of the form (2.34) are no longer a Nash equilibrium in the alternating game because they no longer induce a stable pattern of alternating cooperation.

2.6.4 Alternating games with implementation errors

In the following, we explore how equilibrium behavior is affected by noise. To this end, we assume the players' actions are subject to implementation errors or trembling hand errors,²¹³. That is, each time a player wishes to cooperate, there is some probability ε that the player defects by mistake, with $0 < \varepsilon < 1/2$. Conversely, each time a player wishes to defect, she may cooperate with the same probability. Under this

assumption, a player with strategy \mathbf{p} employs an effective strategy \mathbf{p}^ε , with

$$\mathbf{p}^\varepsilon = \phi^\varepsilon(\mathbf{p}) := (1-\varepsilon)\mathbf{p} + \varepsilon(1-\mathbf{p}) = \varepsilon + (1-2\varepsilon)\mathbf{p}. \quad (2.35)$$

This transformation maps memory-1 strategies $\mathbf{p} \in [0, 1]^4$ to noisy memory-1 strategies $\mathbf{p}^\varepsilon \in [\varepsilon, 1-\varepsilon]^4$. It has two useful properties. First, it is bijective, with the inverse function being defined by

$$(\phi^\varepsilon(\mathbf{p}^\varepsilon))^{-1} = \frac{\mathbf{p}^\varepsilon - \varepsilon}{1-2\varepsilon}. \quad (2.36)$$

Second, the transformation is monotonic: For any previous round's outcome $(a, \tilde{a}) \in \{C, D\}^2$ and for any two memory-1 strategies \mathbf{p} and \mathbf{q} , we have $p_{a\tilde{a}} < q_{a\tilde{a}}$ if and only if $p_{a\tilde{a}}^\varepsilon < q_{a\tilde{a}}^\varepsilon$. That is, if player 1's nominal strategy is more cooperative than player 2's, then the same is true for the respective effective strategies. A few examples of effective strategies are

$$\begin{aligned} \text{ALLD}^\varepsilon &= (\varepsilon, \varepsilon) & \text{ATFT}^\varepsilon &= (\varepsilon, 1-\varepsilon) & \text{GRIM}^\varepsilon &= (1-\varepsilon, \varepsilon, \varepsilon, \varepsilon) \\ \text{ALLC}^\varepsilon &= (1-\varepsilon, 1-\varepsilon) & \text{TFT}^\varepsilon &= (1-\varepsilon, \varepsilon) & \text{FBF}^\varepsilon &= (1-\varepsilon, \varepsilon, 1-\varepsilon, 1-\varepsilon). \end{aligned} \quad (2.37)$$

In particular, even if the nominal strategy is deterministic, the corresponding effective strategy is fully stochastic. For an alternating game with errors between two memory-1 strategies \mathbf{p} and \mathbf{q} we can define the resulting transition matrix, the invariant distribution, and the payoffs based on the respective quantities for the game without errors, given by Eqs. (2.17) – (2.19). This yields

$$M_A^\varepsilon(\mathbf{p}, \mathbf{q}) := M_A(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon), \quad \mathbf{v}^\varepsilon(\mathbf{p}, \mathbf{q}) := \mathbf{v}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon), \quad \pi^\varepsilon(\mathbf{p}, \mathbf{q}) := \pi(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon). \quad (2.38)$$

Since $\varepsilon > 0$, each entry of $M_A(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)$ is positive. Hence, the unique invariant distribution and the resulting payoffs are now well-defined irrespective of the players' actions in the first round.

In the following, we aim to characterize all equilibria among the memory-1 strategies for the alternating game with errors. We follow the same approach as before. In analogy to Propositions 2.1 and 2.2, the following shows that games between a generic player and a memory-1 player can be reduced to a game between a reactive player and a memory-1 player.

Proposition 2.9 (Sufficiency of reactive strategies in games with errors). Consider an alternating game with a positive error rate $0 < \varepsilon < 1/2$, and suppose the second player uses the memory-1 strategy \mathbf{q} .

1. Suppose the first player uses the memory-1 strategy $\mathbf{p} = (p_{CC}, p_{CD}, p_{DC}, p_{DD})$ and the resulting invariant distribution is given by $\mathbf{v}^\varepsilon(\mathbf{p}, \mathbf{q}) = (v_{CC}^\varepsilon, v_{CD}^\varepsilon, v_{DC}^\varepsilon, v_{DD}^\varepsilon)$. Define the reactive marginalization of \mathbf{p} with respect to \mathbf{q} as the unique reactive strategy $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$ for which

$$\begin{aligned}\tilde{p}_C &= \frac{1}{1-2\varepsilon} \left(\frac{v_{CC}^\varepsilon p_{CC}^\varepsilon + v_{DC}^\varepsilon p_{DC}^\varepsilon}{v_{CC}^\varepsilon + v_{DC}^\varepsilon} - \varepsilon \right) \\ \tilde{p}_D &= \frac{1}{1-2\varepsilon} \left(\frac{v_{CD}^\varepsilon p_{CD}^\varepsilon + v_{DD}^\varepsilon p_{DD}^\varepsilon}{v_{CD}^\varepsilon + v_{DD}^\varepsilon} - \varepsilon \right).\end{aligned}\tag{2.39}$$

Then, the reactive marginalization satisfies $\mathbf{v}^\varepsilon(\tilde{\mathbf{p}}, \mathbf{q}) = \mathbf{v}^\varepsilon(\mathbf{p}, \mathbf{q})$.

2. Suppose the first player uses an arbitrary strategy such that $p_{a_1, a_2}^\varepsilon(t) \in [\varepsilon, 1 - \varepsilon]$ is the player's conditional probability to cooperate at time t if the previous outcome is $(a, \tilde{a}) \in \{CC, CD, DC, DD\}$. Let $\mathbf{v}(t) = (v_{CC}(t), v_{CD}(t), v_{DC}(t), v_{DD}(t))$ be the resulting probability distribution for the player's actions at time t . We

assume the following limiting averages to exist,

$$\begin{aligned}
v_{a\tilde{a}} &:= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{a\tilde{a}}(t) \quad \text{for all } a \in \{C, D\}, \tilde{a} \in \{C, D\}. \\
\tilde{p}_{\tilde{a}}^\varepsilon &:= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \frac{v_{C\tilde{a}}(t) \tilde{p}_{C\tilde{a}}^\varepsilon(t) + v_{D\tilde{a}}(t) \tilde{p}_{D\tilde{a}}^\varepsilon(t)}{v_{C\tilde{a}} + v_{D\tilde{a}}} \quad \text{for all } \tilde{a} \in \{C, D\}.
\end{aligned} \tag{2.40}$$

We define the reactive marginalization $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$ of player 1's strategy with respect to \mathbf{q} by

$$\tilde{p}_C = \frac{1}{1-2\varepsilon} (\tilde{p}_C^\varepsilon - \varepsilon) \quad \text{and} \quad \tilde{p}_D = \frac{1}{1-2\varepsilon} (\tilde{p}_D^\varepsilon - \varepsilon). \tag{2.41}$$

Then, this reactive marginalization satisfies $\mathbf{v}^\varepsilon(\tilde{\mathbf{p}}, \mathbf{q}) = \mathbf{v}$.

Both results follow in a straightforward manner from the respective results on alternating games without errors, by applying Propositions 2.1 and 2.2 to the players' effective strategies. In a similar way, we can also generalize Proposition 2.5. To this end, we say a strategy is generic with respect to the opponent strategy and the error rate if the respective limits in Eq. (2.40) exist. In particular, all strategies with finite recall are generic. In addition, we say a strategy \mathbf{p} is a generic best response to \mathbf{q} if it is generic, and if

$$\pi^\varepsilon(\mathbf{p}, \mathbf{q}) \geq \pi^\varepsilon(\mathbf{p}', \mathbf{q}) \quad \text{for all generic strategies } \mathbf{p}'. \tag{2.42}$$

Then we can again show that one can always find a generic best response to a memory-1 strategy among the deterministic reactive strategies.

Proposition 2.10 (Optimality of deterministic reactive strategies in games with er-

rors). Let \mathbf{q} be some given memory-1 strategy and let $\mathbf{p} \in [0, 1]^2$ be an arbitrary reactive strategy. Then, there is a deterministic reactive strategy $\mathbf{p}' \in \{0, 1\}^2$ for which $\pi^\varepsilon(\mathbf{p}', \mathbf{q}) \geq \pi^\varepsilon(\mathbf{p}, \mathbf{q})$.

Similar to before, we say a memory-1 strategy \mathbf{q} is a Nash equilibrium if $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\mathbf{q}', \mathbf{q})$ for any generic deviation strategy \mathbf{q}' . Proposition 2.10 then again allows us to identify memory-1 Nash equilibria more effectively. It suffices to compare the payoff of the strategy \mathbf{q} against itself to the payoffs one could achieve with deterministic reactive strategies. However, now the payoff expressions are more complex. Using Eq. (2.38), we calculate the self-payoff as

$$\pi^\varepsilon(\mathbf{q}, \mathbf{q}) = \frac{(1-q_{CC}^\varepsilon+q_{DC}^\varepsilon)q_{DD}^\varepsilon}{(1-q_{CC}^\varepsilon)(1-q_{CD}^\varepsilon+q_{DD}^\varepsilon)+(1-q_{CC}^\varepsilon+q_{DC}^\varepsilon)q_{DD}^\varepsilon} \cdot (b-c). \quad (2.43)$$

For the possible deviations towards deterministic reactive strategies, we obtain

$$\begin{aligned} \pi^\varepsilon(\text{ALLD}, \mathbf{q}) &= \frac{\varepsilon q_{DC}^\varepsilon + (1-\varepsilon)q_{DD}^\varepsilon}{1-\varepsilon(q_{CC}^\varepsilon-q_{DC}^\varepsilon) - (1-\varepsilon)(q_{CD}^\varepsilon-q_{DD}^\varepsilon)} \cdot b - \varepsilon \cdot c \\ \pi^\varepsilon(\text{ATFT}, \mathbf{q}) &= \frac{(1-\varepsilon)q_{DC}^\varepsilon + \varepsilon q_{DD}^\varepsilon}{1-\varepsilon(q_{CC}^\varepsilon-q_{DD}^\varepsilon) - (1-\varepsilon)(q_{CD}^\varepsilon-q_{DC}^\varepsilon)} \cdot b \\ &\quad - \frac{1-\varepsilon-\varepsilon(1-\varepsilon)q_{CC}^\varepsilon - (1-\varepsilon)^2q_{CD}^\varepsilon + \varepsilon(1-\varepsilon)q_{DC}^\varepsilon + \varepsilon^2q_{DD}^\varepsilon}{1-\varepsilon(q_{CC}^\varepsilon-q_{DD}^\varepsilon) - (1-\varepsilon)(q_{CD}^\varepsilon-q_{DC}^\varepsilon)} \cdot c \\ \pi^\varepsilon(\text{TFT}, \mathbf{q}) &= \frac{\varepsilon q_{DC}^\varepsilon + (1-\varepsilon)q_{DD}^\varepsilon}{1-(1-\varepsilon)(q_{CC}^\varepsilon-q_{DD}^\varepsilon) - \varepsilon(q_{CD}^\varepsilon-q_{DC}^\varepsilon)} \cdot b \\ &\quad - \frac{\varepsilon-\varepsilon(1-\varepsilon)q_{CC}^\varepsilon - \varepsilon^2q_{CD}^\varepsilon + \varepsilon(1-\varepsilon)q_{DC}^\varepsilon + (1-\varepsilon)^2q_{DD}^\varepsilon}{1-(1-\varepsilon)(q_{CC}^\varepsilon-q_{DD}^\varepsilon) - \varepsilon(q_{CD}^\varepsilon-q_{DC}^\varepsilon)} \cdot c \\ \pi^\varepsilon(\text{ALLC}, \mathbf{q}) &= \frac{(1-\varepsilon)q_{DC}^\varepsilon + \varepsilon q_{DD}^\varepsilon}{1-(1-\varepsilon)(q_{CC}^\varepsilon-q_{DC}^\varepsilon) - \varepsilon(q_{CD}^\varepsilon-q_{DD}^\varepsilon)} \cdot b - (1-\varepsilon) \cdot c \end{aligned} \quad (2.44)$$

We define partners, defectors, and equalizers analogously to the case without errors. We call a memory-1 strategy \mathbf{q} self-cooperative if it yields the mutual cooperation payoff against itself as errors become rare, $\varepsilon \rightarrow 0$ (in particular, it must satisfy $q_{CC}=1$). Similarly, \mathbf{q} is self-defective if it yields the self-defection payoff against itself in the limit of rare errors (in particular, $q_{DD} = 0$). Partners are again those Nash equilibria that are self-cooperative, and defectors as those Nash equilibria that are self-defective. As before, a strategy is an equalizer for a given error probability ε if any generic co-player yields the same payoff against that strategy. The following then generalizes the results of Theorems 2.6 and 2.8 to the case of alternating games with errors.

Theorem 2.11 (Classification of Nash equilibria in alternating games with errors).

Consider a memory-1 strategy $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, q_{DD})$ for an alternating game with error probability $0 < \varepsilon < \frac{1}{2} (1 - \frac{c}{b})$. Then, the following are equivalent.

1. The strategy \mathbf{q} is a generic Nash equilibrium.
2. The strategy \mathbf{q} satisfies

$$\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \max(\pi^\varepsilon(\text{ALLD}, \mathbf{q}), \pi^\varepsilon(\text{ATFT}, \mathbf{q}), \pi^\varepsilon(\text{TFT}, \mathbf{q}), \pi^\varepsilon(\text{ALLC}, \mathbf{q})). \quad (2.45)$$

3. The strategy \mathbf{q} is either a partner, a defector, or an equalizer. It is a partner if and only if

$$\begin{aligned} q_{CC} &= q_{DC} = 1, \\ q_{CD} &\leq 1 - \frac{c}{(1-2\varepsilon)b}, \\ q_{DD} &= \frac{(1-2\varepsilon)(b+\varepsilon c q_{CD}) - c}{(1-2\varepsilon)(b+\varepsilon c)}. \end{aligned} \quad (2.46)$$

It is a defector if and only if it is either ALLD or

$$\begin{aligned}
 q_{DD} &= q_{CD} = 0, \\
 q_{DC} &\leq \frac{c}{(1-2\varepsilon)b}, \\
 q_{CC} &= \frac{\varepsilon(1-2\varepsilon)cq_{DC} + c}{(1-2\varepsilon)(b+\varepsilon c)}.
 \end{aligned} \tag{2.47}$$

It is an equalizer if and only if

$$\begin{aligned}
 q_{CD} &= \frac{(1-2\varepsilon)(bq_{CC} - cq_{DD}) - c}{(1-2\varepsilon)(b-c)} \\
 q_{DC} &= \frac{(1-2\varepsilon)(bq_{DD} - cq_{CC}) + c}{(1-2\varepsilon)(b-c)}.
 \end{aligned} \tag{2.48}$$

Several remarks are in order:

a Dimension of the Nash equilibrium classes. Errors lead to a discontinuous reduction in the number of Nash equilibria. For example, for any arbitrarily small (but positive) error probability the set of partners is now one-dimensional instead of three-dimensional. The proof of Theorem 2.11 suggests there are two reasons for this reduction.

First, in games without errors, $p_{CC} = 1$ is sufficient to ensure that the reactive marginalization of \mathbf{q} with respect to itself satisfies $\tilde{q}_C = 1$, see Eq. (2.22). Game outcomes different from mutual cooperation are (almost) never visited, and hence $v_{DC} = 0$. For games with errors, this is no longer true. Independent

of the players' strategies, each game outcome occurs at least when both players choose the respective action by mistake, such that $v_{a\tilde{a}} \geq \varepsilon^2$ for all $(a, \tilde{a}) \in \{CC, CD, DC, DD\}$. To guarantee that the reactive marginalization still satisfies $\tilde{q}_C=1$, we therefore additionally need to require that $q_{DC}=1$.

Second, in alternating games without errors, a partner \mathbf{q} cannot be invaded by the two boundary strategies ALLC and TFT. Instead, all three strategies yield the same payoff $b-c$ against \mathbf{q} . In contrast, in games with errors, the payoffs of the three strategies are generally different. Moreover, by Lemma 2.4 the payoff $\pi^\varepsilon(\mathbf{q}, \mathbf{q})$ is either strictly in between $\pi^\varepsilon(\text{TFT}, \mathbf{q})$ and $\pi^\varepsilon(\text{ALLC}, \mathbf{q})$, or all three payoffs are the same. For \mathbf{q} to be a Nash equilibrium, we thus need to require

$$\pi^\varepsilon(\text{TFT}, \mathbf{q}) = \pi^\varepsilon(\mathbf{q}, \mathbf{q}) = \pi^\varepsilon(\text{ALLC}, \mathbf{q}). \quad (2.49)$$

This equality is equivalent to the last equality in condition (2.46). Finally, the upper bound on q_{CD} ensures that neither a deviation to *ALLD* nor to *ATFT* is profitable.

As a consequence of these observations, we conclude that many of the well-known self-cooperative memory-1 strategies fail to be partners in alternating games with errors. In particular, TFT, WSLS, GRIM and FBF are either no longer self-cooperative, or they are no longer Nash equilibria. Similar considerations also explain why the class of defectors is now one-dimensional.

b Evolutionary stability. While all partner strategies for the alternating game are Nash equilibria by definition, we note that none of them are evolutionarily stable in the sense of Maynard-Smith¹⁴⁶. Instead, by (2.49), mutants who ei-

they adopt TFT or ALLC can invade through neutral drift – in fact, due to the monotonicity property in Lemma 2.4, any mutant strategy \mathbf{q}' with $q_{CC} = q_{DC} = 1$ yields the same payoff against \mathbf{q} as \mathbf{q} does against itself.

Thus, as a corollary of Theorem 2.11, we conclude that in alternating games with errors, there are no evolutionarily stable memory-1 strategies that sustain cooperation. Note that this result is different from previous work suggesting that no strategy in the simultaneous game is stable^{24,69,70,238}. This previous work considers games without errors. Only in that case does it show that for each Nash equilibrium one can identify neutral mutant strategies that eventually lead out of that equilibrium. These arguments do not apply to games with errors, where evolutionary stability is generally feasible²³. For example, for games with $b > 2c$, one can show that WSLS is evolutionarily stable in the simultaneous game, provided the error probability is positive but sufficiently small^{98,137}.

c Existence of the Nash equilibrium classes and comparative statics. Partners do not exist for all parameter values. Because q_{CD} and q_{DD} need to be values in the unit interval, condition (2.46) implies that partners exist if and only if

$$\varepsilon < \frac{1}{2} \left(1 - \frac{c}{b} \right). \quad (2.50)$$

In particular, the conditions for partners to exist are easiest to satisfy if either implementation errors are sufficiently rare, or if the benefit of cooperation is large compared to its costs. The same condition (2.50) also determines whether or not equalizer strategies exist.

In contrast, defectors exist for all parameter values $b > c$ and $0 \leq \varepsilon \leq 1/2$, because ALLD is always a Nash equilibrium. When condition (2.50) does not hold, mutual defection is in fact the only behavior that can be sustained in equilibrium.

d Examples of partner strategies. Provided condition (2.50) holds, the set of partners is given by the line segment connecting the two strategies

$$\mathbf{q}' = \left(1, 0, 1, \frac{(1-2\varepsilon)b-c}{(1-2\varepsilon)(b+\varepsilon c)} \right) \quad (2.51)$$

$$\mathbf{q}'' = \left(1, \frac{(1-2\varepsilon)b-c}{(1-2\varepsilon)b}, 1, \frac{(1-2\varepsilon)b-c}{(1-2\varepsilon)b} \right) \quad (2.52)$$

In particular, we note that while q_{CD} may be chosen to be zero, q_{DD} always needs to be strictly in between zero and one. The first example above, \mathbf{q}' can be considered as a stochastic version of Firm-but-Fair, and hence we refer to it as SFBF. Strategies resembling SFBF have been described previously. For example, Nowak and Sigmund¹⁷¹ observe that the simulations in their Figure 3 converge to the strategy $(1, 0, 1, 2/3)$. Using their parameters $b = 3$, $c = 1$ and $\varepsilon = 0.001$, this is exactly what is predicted by expression (2.51). The second example above, \mathbf{q}'' corresponds to the well-known Generous Tit-for-Tat strategy^{156,169} (GTFT) which has been previously described for the simultaneous game. According to Eqs. (2.51) and (2.52), GTFT is the only partner among the reactive strategies. For reactive strategies, the simultaneous and the alternating game coincide with respect to their dynamics¹⁷¹. In this light, the fact that GTFT also makes an appearance in the alternating game is somewhat less surprising.

In the limit of rare errors, $\varepsilon \rightarrow 0$, the above expressions simplify further. We obtain

$$\mathbf{q}' = \left(1, 0, 1, 1 - \frac{c}{b}\right) \quad \text{and} \quad \mathbf{q}'' = \left(1, 1 - \frac{c}{b}, 1, 1 - \frac{c}{b}\right) \quad (2.53)$$

e Alternative specification of errors. Throughout this section we have assumed that errors originate from players who misimplement their intended actions with a constant probability ε . In this case, a player's effective strategy \mathbf{q}^ε is a linear function of the player's actual strategy \mathbf{q} , as described by Eq. (2.35). However, analogous results can be derived for more general error mappings. Our main results only require that the effect of errors on a player's strategy can be described by a strictly monotonic and bijective transformation $\phi^\varepsilon : [0, 1]^4 \rightarrow [\varepsilon, 1 - \varepsilon]^4$ and that ALLD, ATFT, TFT, ALLC are mapped to the values in Eq. (2.37). Under that more general assumption, condition (2.45) continues to characterize which memory-1 strategies \mathbf{q} are equilibria. Only the exact description of partners, defectors, and equalizers needs to be adapted correspondingly. In addition, one can also extend our results to cases where the error rate depends on the previous outcome, or where it depends on the player's intended action. In this way, one could model cases in which a player who intends to cooperate is more likely to make a mistake than a player who intends to defect.

2.7 Supplementary Note 3: Extensions of the baseline model

The baseline model makes a number of simplifying assumptions: (i) the game is infinitely repeated; (ii) the players move in a strictly alternating fashion; (iii) the simulations only take into account memory-1 strategies; (iv) interactions take place in a well-mixed population; and (v) mutations are global. In the following, we study the impact of each of these assumptions in more detail. In each case, we explore how the respective assumption affects emerging cooperation rates and the strategies that evolve.

2.7.1 Finitely repeated games

Motivation. Our baseline model considers an infinitely repeated game with no discounting of the future. There are two major reasons why the analysis of such games is useful. First, from a mathematical perspective, infinitely repeated games are more convenient to work with because their results tend to be independent of the players' behavior in the early rounds of the game. This in turn allows researchers to consider a simpler strategy space; the players' first-round behavior no longer needs to be specified²¹⁵. Second, such games often serve as a good approximation for games where the number of rounds is large but finite¹⁷². In many cases, results for finitely repeated games resemble the results of infinitely repeated games already for a moderate number of rounds²⁰⁹. To elaborate on the above two points, and to extend the results of the baseline model, in the following we study a model in which players only engage in finitely many interactions.

Game setup. We assume the game proceeds in rounds. In the simultaneous game, the

two players move simultaneously in each round. In the alternating game, one player moves first and the other player moves second. Here, the player who moves first is determined randomly, but kept constant during the game. After each round, the game continues for another round with a constant continuation probability δ . As a result, the expected number of rounds follows a geometric distribution with mean $1/(1-\delta)$. Let $v_{a_1, a_2}(t)$ denote the conditional probability that the two players choose actions a_1 and a_2 in round t , given that round t is reached. Then we can calculate the average probability to observe the respective outcome over the course of the entire game as

$$v_{a_1, a_2} := (1-\delta) \sum_{t=0}^{\infty} \delta^t \cdot v_{a_1, a_2}(t). \quad (2.54)$$

In the limiting case that there is always another round, $\delta \rightarrow 1$, this weighted average converges to the time average (2.13) of the baseline model, provided the limit in Eq. (2.13) exists. As before, we collect these four averages in a vector $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$. Based on this vector, we can compute the players' payoffs in the same way as in the baseline model, using Eq. (2.15).

Memory-1 strategies. To introduce memory-1 strategies for finitely repeated games, we distinguish between the simultaneous and the alternating game. In the simultaneous game, memory-1 strategies take the form⁹⁶

$$\mathbf{p} = (p_{00}; p_{CC}, p_{CD}, p_{DC}, p_{DD}). \quad (2.55)$$

The first entry p_{00} is the player's probability to cooperate in the initial round. The other entries p_{ij} are the respective conditional cooperation probabilities in all subsequent rounds, as defined in the baseline model. In the alternating game, memory-1

strategies take the form

$$\mathbf{p} = (p_{00}; p_{0C}, p_{0D}; p_{CC}, p_{CD}, p_{DC}, p_{DD}). \quad (2.56)$$

Here, p_{00} is the probability to cooperate in the first round if the focal player moves first. The next two probabilities p_{0C} and p_{0D} are the player's probability to cooperate in the first round if the player moves second. In that case, the focal player may condition its decision on the co-player's first round behavior (C or D). The other entries p_{ij} are again the conditional cooperation probabilities that the focal player applies in all subsequent rounds. In particular, we note that while strategies in the baseline model are 4-dimensional, they are now 5-dimensional in the case of simultaneous games, and 7-dimensional in the case of alternating games.

Explicit formulas for the players' payoffs. When two memory-1 players interact, their payoffs can be computed in a similar way as in the baseline model¹⁷². To this end, we consider first the simultaneous game. Suppose the strategies of the two players are

$$\begin{aligned} \mathbf{p} &= (p_{00}; p_{CC}, p_{CD}, p_{DC}, p_{DD}) \\ \mathbf{q} &= (q_{00}; q_{CC}, q_{CD}, q_{DC}, q_{DD}), \end{aligned} \quad (2.57)$$

respectively. Then the outcome distribution in the initial round is

$$\begin{aligned} \mathbf{v}_0 &:= (v_{CC}(0), v_{CD}(0), v_{DC}(0), v_{DD}(0)) \\ &= (p_{00} \cdot q_{00}, p_{00} \cdot (1 - q_{00}), (1 - p_{00}) \cdot q_{00}, (1 - p_{00}) \cdot (1 - q_{00})). \end{aligned} \quad (2.58)$$

Given this initial outcome distribution, we can iteratively compute all subsequent distributions as

$$\mathbf{v}(t) = \mathbf{v}_0 \cdot \mathcal{M}_S^t. \quad (2.59)$$

Here, $M_S = M_S(\mathbf{p}, \mathbf{q})$ is the standard transition matrix for the simultaneous game, as defined by Eq. (2.20). For the average distribution \mathbf{v} according to Eq. (2.54), we therefore obtain

$$\mathbf{v} = (1-\delta) \sum_{t=0}^{\infty} \delta^t \cdot \mathbf{v}(t) = (1-\delta) \mathbf{v}_0 \sum_{t=0}^{\infty} (\delta M_S)^t = (1-\delta) \mathbf{v}_0 (I - \delta M_S)^{-1}. \quad (2.60)$$

Here, I denotes the 4×4 identity matrix, and $(I - \delta M_S)^{-1}$ refers to the respective inverse matrix. Based on \mathbf{v} , we compute player's payoffs using Eq. (2.15). That is,

$$\begin{aligned} \pi_1 &= b \cdot (v_{CC} + v_{DC}) - c \cdot (v_{CC} + v_{CD}) \\ \pi_2 &= b \cdot (v_{CC} + v_{CD}) - c \cdot (v_{CC} + v_{DC}). \end{aligned} \quad (2.61)$$

The payoffs of the alternating game can be computed analogously. However, here we have to distinguish two cases, depending on which of the two players moves first. Suppose the two players' strategies are given by

$$\begin{aligned} \mathbf{p} &= (p_{00}; p_{0C}, p_{0D}; p_{CC}, p_{CD}, p_{DC}, p_{DD}) \\ \mathbf{q} &= (q_{00}; q_{0C}, q_{0D}; q_{CC}, q_{CD}, q_{DC}, q_{DD}). \end{aligned} \quad (2.62)$$

If it is player 1 who moves first, the initial outcome distribution is

$$\mathbf{v}_0^{(1)} = (p_{00} \cdot q_{0C}, p_{00} \cdot (1 - q_{0C}), (1 - p_{00}) \cdot q_{0D}, (1 - p_{00}) \cdot (1 - q_{0D})). \quad (2.63)$$

The respective average distribution over the entire game can then be calculated as

$$\mathbf{v}^{(1)} = (1-\delta) \mathbf{v}_0^{(1)} (I - \delta M_A(\mathbf{p}, \mathbf{q}))^{-1}, \quad (2.64)$$

where $M_A(\mathbf{p}, \mathbf{q})$ is the standard transition matrix (2.17) of the alternating game. Similarly, if it is player 2 who moves first, the initial distribution is

$$\mathbf{v}_0^{(2)} = (q_{00} \cdot p_{0C}, q_{00} \cdot (1 - p_{0C}), (1 - q_{00}) \cdot p_{0D}, (1 - q_{00}) \cdot (1 - p_{0D})). \quad (2.65)$$

The average distribution becomes

$$\mathbf{v}^{(2)} = (1 - \delta) \mathbf{v}_0^{(2)} (I - \delta M_A(\mathbf{q}, \mathbf{p}))^{-1}. \quad (2.66)$$

Because each of the two players is equally likely to move first, average payoffs are now

$$\begin{aligned} \pi_1 &= b \cdot \left(\frac{1}{2} (v_{CC}^{(1)} + v_{DC}^{(1)}) + \frac{1}{2} (v_{CC}^{(2)} + v_{CD}^{(2)}) \right) - c \cdot \left(\frac{1}{2} (v_{CC}^{(1)} + v_{CD}^{(1)}) + \frac{1}{2} (v_{CC}^{(2)} + v_{DC}^{(2)}) \right) \\ \pi_2 &= b \cdot \left(\frac{1}{2} (v_{CC}^{(1)} + v_{CD}^{(1)}) + \frac{1}{2} (v_{CC}^{(2)} + v_{DC}^{(2)}) \right) - c \cdot \left(\frac{1}{2} (v_{CC}^{(1)} + v_{DC}^{(1)}) + \frac{1}{2} (v_{CC}^{(2)} + v_{CD}^{(2)}) \right). \end{aligned} \quad (2.67)$$

Evolutionary dynamics. Given the payoffs (2.61) and (2.67), we can explore the evolutionary dynamics of the finitely repeated game with the same process we used for the infinitely repeated game. That is, again we consider a finite and well-mixed population in which players adopt new strategies by imitation and mutation, as described in the main text.

Figure 2.6a,b shows the corresponding results for a continuation probability of $\delta = 0.96$, such that individuals interact on average for 25 rounds. The figure suggests that the main evolutionary findings are in qualitative agreement with the findings of the baseline model. First, and as in the baseline model, the simultaneous game is slightly more conducive to the evolution of cooperation compared to the alternating

game (Figure 2.6a). Second, the self-cooperative strategies that evolve in the simultaneous game are markedly different from the self-cooperative strategies that evolve in the alternating game (Figure 2.6b). In the simultaneous game, the average self-cooperative strategy shares the main characteristics of win-stay lose-shift¹⁷⁰. Players are most likely to cooperate after mutual cooperation or mutual defection. In the alternating game, evolving strategies rather resemble a mixture of Generous Tit-for-Tat and Stochastic Firm-but-Fair. Here, players are most likely to cooperate if the co-player cooperated in the previous round. In addition, players exhibit a positive probability to cooperate after both players defected. In either game, the payoff derived in the first round only has a modest impact on the player's overall fitness, given the game length. As a result, the values of q_{00}, q_{0C}, q_{0D} are close to $1/2$, as one would expect from traits that are almost neutral.

We explore the dynamics for other continuation probabilities in Figure 2.6c and Figure 2.11. If the continuation probability exceeds some moderate threshold, $\delta \approx 0.8$ (which corresponds to games with five rounds in expectation), the qualitative results are largely comparable to the results of the baseline model. Below this threshold, cooperation is rare in both the simultaneous and the alternating game.

Discussion of the model. We note that by structuring the game into rounds, we implicitly assume that the two players always make the same number of decisions in the alternating game, independent of the realized length of the game. In particular, every time the first player makes a decision, this player can be sure that the second player will have an opportunity to reciprocate. We have made this assumption to make it easier to compare the alternating game to the simultaneous game. By making sure that both players make the same number of decisions, the continuation probability δ has an analogous interpretation in both games. Alternatively, we could have assumed

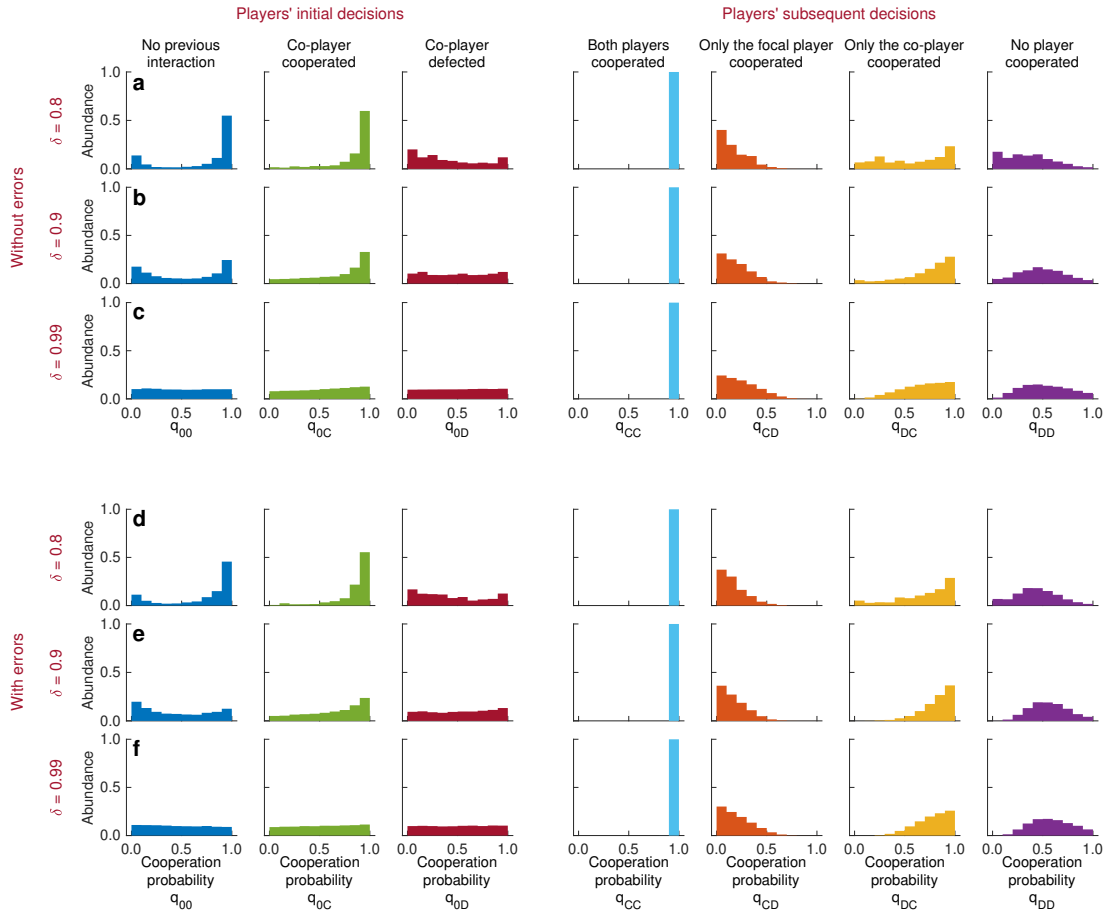


Figure 2.11: Self-cooperative strategies in the finitely repeated alternating game. We explore which strategies the players use to maintain cooperation in finitely repeated alternating games. To this end, we consider two different error scenarios (a-c: $\varepsilon = 0$ and d-f: $\varepsilon = 0.02$), and three expected game lengths (5, 10, or 100 rounds, corresponding to a continuation probability of $\delta = 0.8$, $\delta = 0.9$, and $\delta = 0.99$). In each case, we run simulations and record those strategies that have a self-cooperation rate of at least 80%. Here, we show the distribution of these strategies. We observe the following regularities: (i) The players' first round behavior is only under selection when players interact for a few rounds. For $\delta > 0.9$, the distribution of the respective cooperation probabilities q_{00}, q_{0C}, q_{0D} is comparably flat. (ii) The conditional cooperation probabilities for all subsequent rounds generally resemble the evolving strategies of the baseline model. Apart from the parameters varied explicitly, parameters are the same as in Figure 2.4.

that after each player's decision, the game stops with a certain probability. That alternative scenario is a special case of the model that we consider in the following.

2.7.2 Irregular alternation patterns

Motivation. For both the finitely and the infinitely repeated alternating game, we assumed in the baseline model that players move in a strictly alternating fashion. That is, every time player 1 makes a decision, it is player 2 who makes the next decision. Conversely, every time, player 2 makes a decision, it is player 1 who makes the next decision (provided there is another round). Instead, here we explore what happens when the alternation pattern can be more irregular, such that players may have to make a decision two times in a row before it's the other player's turn to move.

Game setup. To allow for such irregular patterns, we structure the alternating game into a sequence of moves. The player who makes the initial move is determined randomly, with each player having the same chance to move first. After a player has made a move (by deciding whether or not to cooperate), there is a constant probability λ that the game continues. If the game continues, the player who makes the next move is determined randomly. With probability s (the 'switching probability'), it is the other player who makes the next move. With the converse probability $\bar{s} := 1 - s$, it is the same player. In the special case $s = 1$, we recover a setup in which the two players move in a strictly alternating fashion. If $s = 1/2$, the next move is assigned completely randomly, independent of who moved before. Finally, if $s = 0$, there is no alternation at all; the player who moves first is guaranteed to move in all subsequent interactions. The payoffs of each player are defined analogously to the previous cases, by averaging the received benefits and paid costs over all moves of the two players

(for more detail, see further below).

Memory-1 strategies. Similar to the baseline model, we assume that players condition their behavior only on the respective last move of either player (that is, they remember one move per player). In the case of irregular interaction patterns, such strategies take the form of a nine-dimensional vector,

$$\mathbf{p} = (p_{00}; p_{C0}, p_{D0}; p_{0C}, p_{0D}; p_{CC}, p_{CD}, p_{DC}, p_{DD}). \quad (2.68)$$

Here, p_{00} is the player's cooperation probability if no player has moved before. The next two entries, p_{C0} and p_{D0} are the player's cooperation probability if only the focal player has moved before (but not the co-player). The other two entries, p_{0C} and p_{0D} are the player's cooperation probability if only the co-player has moved before (but not the focal player). And finally, the entries p_{ij} with $i, j \in \{C, D\}$ are the usual conditional cooperation probabilities in all subsequent rounds.

Explicit formulas for the players' payoffs. Again assuming that both players use memory-1 strategies, we can compute their payoffs explicitly. To this end, suppose the strategies of player 1 and player 2 are

$$\begin{aligned} \mathbf{p} &= (p_{00}; p_{C0}, p_{D0}; p_{0C}, p_{0D}; p_{CC}, p_{CD}, p_{DC}, p_{DD}) \\ \mathbf{q} &= (q_{00}; q_{C0}, q_{D0}; q_{0C}, q_{0D}; q_{CC}, q_{CD}, q_{DC}, q_{DD}) \end{aligned} \quad (2.69)$$

respectively. We describe the dynamics among the two players by a Markov chain with twelve possible states. The twelve states are (in this order):

$$\begin{aligned} &(1, C, \emptyset), (1, D, \emptyset), (2, \emptyset, C), (2, \emptyset, D), \\ &(1, C, C), (1, C, D), (1, D, C), (1, D, D), (2, C, C), (2, C, D), (2, D, C), (2, D, D). \end{aligned} \quad (2.70)$$

Here, the state (i, a_1, a_2) refers to the case that the previous move was made by player $i \in \{1, 2\}$, and that after this move, the last move by either player is $a_1, a_2 \in \{C, D, \emptyset\}$; the empty set symbol indicates that the respective player did not make a move yet. We obtain the following distribution for the state of the Markov chain after the first move:

$$\mathbf{v}_0 = \left(\frac{p_{00}}{2}, \frac{\bar{p}_{00}}{2}, \frac{q_{00}}{2}, \frac{\bar{q}_{00}}{2}, 0, 0, 0, 0, 0, 0, 0, 0 \right). \quad (2.71)$$

For this initial distribution, we have used the shortcut notation $\bar{p}_{00} := 1 - p_{00}$ and $\bar{q}_{00} := 1 - q_{00}$. The transition matrix of the Markov chain is given by

$$M_I(\mathbf{p}, \mathbf{q}) = \begin{pmatrix} \bar{sp}_{C0} & \bar{sp}_{D0} & 0 & 0 & 0 & 0 & 0 & 0 & sq_{0C} & s\bar{q}_{0C} & 0 & 0 \\ \bar{sp}_{D0} & \bar{sp}_{C0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & sq_{0D} & s\bar{q}_{0D} \\ 0 & 0 & \bar{sq}_{C0} & \bar{sq}_{D0} & sp_{0C} & 0 & \bar{sp}_{0C} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \bar{sq}_{D0} & \bar{sq}_{C0} & sp_{0D} & 0 & \bar{sp}_{0D} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{sp}_{CC} & 0 & \bar{sp}_{DD} & 0 & sq_{CC} & s\bar{q}_{CC} & 0 & 0 \\ 0 & 0 & 0 & 0 & sp_{CD} & 0 & \bar{sp}_{CD} & 0 & sq_{DC} & s\bar{q}_{DC} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{sp}_{DC} & 0 & \bar{sp}_{DC} & 0 & 0 & 0 & sq_{CD} & s\bar{q}_{CD} \\ 0 & 0 & 0 & 0 & sp_{DD} & 0 & \bar{sp}_{DD} & 0 & 0 & 0 & sq_{DD} & s\bar{q}_{DD} \\ 0 & 0 & 0 & 0 & sp_{CC} & 0 & \bar{sp}_{CC} & 0 & \bar{sq}_{CC} & \bar{sq}_{CC} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & sp_{CD} & 0 & \bar{sp}_{CD} & \bar{sq}_{DC} & \bar{sq}_{DC} & 0 & 0 \\ 0 & 0 & 0 & 0 & sp_{DC} & 0 & \bar{sp}_{DC} & 0 & 0 & 0 & \bar{sq}_{CD} & \bar{sq}_{CD} \\ 0 & 0 & 0 & 0 & 0 & sp_{DD} & 0 & \bar{sp}_{DD} & 0 & 0 & \bar{sq}_{DD} & \bar{sq}_{DD} \end{pmatrix}. \quad (2.72)$$

Based on the initial distribution \mathbf{v}_0 and on the transition matrix $M_I(\mathbf{p}, \mathbf{q})$, we can again compute the average distribution to observe any of the twelve states over the course of the game as

$$\mathbf{v} = (1 - \lambda)\mathbf{v}_0(I - \lambda M_I(\mathbf{p}, \mathbf{q}))^{-1}. \quad (2.73)$$

The payoffs of the two players are then given by the formula

$$\begin{aligned}\pi_1 &= 2 \mathbf{v} \cdot (-c, 0, b, 0, -c, -c, 0, 0, b, 0, b, 0)^\top \\ \pi_2 &= 2 \mathbf{v} \cdot (b, 0, -c, 0, b, b, 0, 0, -c, 0, -c, 0)^\top.\end{aligned}\tag{2.74}$$

Here, the factor of two is a normalization constant to take into account that in half of the rounds it's the focal player who moves, whereas in the other half it's the co-player. This constant ensures that two unconditional cooperators obtain an expected payoff of $(1-\varepsilon)(b-c)$, as one would expect.

Evolutionary dynamics. We explore the evolutionary dynamics of the game with irregular alternation patterns using the same process we used for the baseline model. In Figure 2.6d we compare two scenarios. In the first scenario, we assume that players strictly alternate ($s = 1$). In the other scenario, it is randomly determined which player moves next ($s = 1/2$). In both cases, we consider the dynamics of a game that is infinitely repeated ($\lambda \rightarrow 1$), with a moderate error rate ($\varepsilon = 0.02$) and rare mutations ($\mu \rightarrow 1$). We observe that a strictly alternating game leads to higher cooperation rates. To explore this result in more detail, we also recorded which self-cooperative strategies the players are most likely to adopt over the course of time. Surprisingly, both the strictly alternating and the randomly alternating game lead to overall similar strategies (Figure 2.6e). In each case, the average strategy reflects the basic patterns of Stochastic Firm-but-Fair. However, in the strictly alternating case, these self-cooperative strategies tend to be more robust. When players are strictly alternating, it takes on average 1,600 mutants until a resident self-cooperative strategy is successfully invaded. In contrast, for randomly alternating games, this number drops

to 980 mutants. More generally, we observe that cooperation is most likely to evolve the more likely players alternate in a regular manner (Figure 2.6f). In the extreme case that players do not alternate at all ($s = 0$), cooperation does not evolve, as one may expect.

In addition to these simulation results for infinitely repeated games, we have also explored the dynamics when the number of rounds is finite (Figure 2.12). There we show the outcome of two sets of simulations, one in the absence of errors ($\varepsilon = 0$) and one with a moderate error rate ($\varepsilon = 0.02$). In both cases, we vary the expected number of rounds (between 1 and 100), and the switching probability (between 0 and 1). These simulations confirm the previously observed regularities for alternating games: Cooperation is most likely to evolve when (i) errors are rare, (ii) players interact for many rounds, and (iii) players alternate in a regular fashion.

2.7.3 Memory-2 strategies

Motivation. For all evolutionary results so far, we have assumed that players only take into account each player's last action. While it has been argued that memory-1 strategies are good approximations for human behavior in laboratory experiments in simultaneous games⁴⁶, it is natural to ask which of our qualitative results depend on the one-round memory assumption. Exploring the evolutionary dynamics among more complex strategies is not straightforward because the number of available strategies increases super-exponentially in the players' memory capacity.

In the baseline case of an infinitely repeated game, there are 16 pure memory-1 strategies, 65,536 pure memory-2 strategies, and $1.84 \cdot 10^{19}$ pure memory-3 strategies⁹⁸. In the following, we thus confine ourselves to pure memory-2 strategies. Those are all

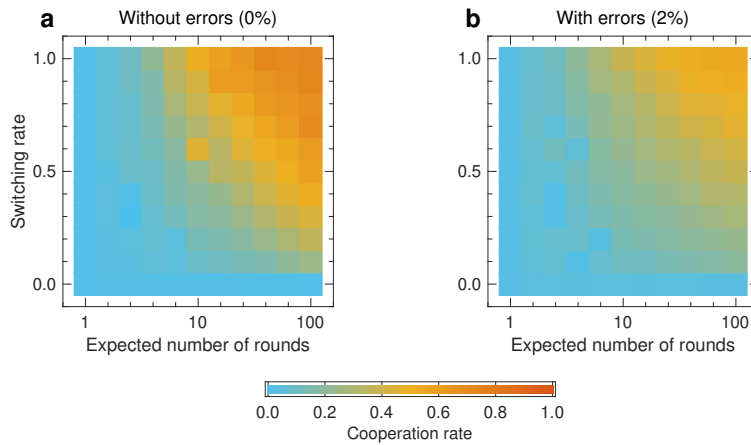


Figure 2.12: Cooperation in the finitely repeated game with irregular alternation patterns. Using the same basic setup as in Figure 2.6, we have explored how likely players are to cooperate in finitely repeated alternating games with irregular alternation patterns. To this end, we vary (i) the expected number of rounds, and (ii) the switching rate that measures how strictly players are to alternate. In addition, we consider two scenarios, depending on whether or not players commit implementation errors (a,b). As already indicated by Figure 2.6, players are most likely to cooperate when there are no errors, when the number of rounds is large, and when players move in a strictly alternating fashion.

strategies that consider the last two moves of each player, and for which the corresponding cooperation probability (in the absence of errors) is either zero or one. Memory-2 strategies. For simplicity, we shall only consider the case of infinitely repeated games here (although the case of finitely games can be treated similarly, as discussed in the previous sections). In infinitely repeated games, memory-2 strategies can be represented by a 16-dimensional vector,

$$\mathbf{p} = \left(p_{CC}^{CC}, p_{CC}^{CD}, p_{CD}^{CC}, p_{CD}^{CD}, p_{CC}^{DC}, p_{CC}^{DD}, p_{CD}^{DC}, p_{CD}^{DD}, p_{DC}^{CC}, p_{DC}^{CD}, p_{DD}^{CC}, p_{DD}^{CD}, p_{DC}^{DC}, p_{DC}^{DD}, p_{DD}^{DC}, p_{DD}^{DD} \right). \quad (2.75)$$

The entries again reflect the player's conditional cooperation probabilities. The upper two indices of an entry represent the last two moves of the focal player (with the very last move coming first and the second-to last move coming second). Analogously, the lower two indices represent the last two moves of the co-player. The space of memory-2 strategies trivially contains the set of all memory-1 strategies as a subset. For example, within the space of memory-2 strategies, Tit-for-Tat takes the form

$$\mathbf{p} = (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0). \quad (2.76)$$

Because each pure memory-2 strategy is a 16-dimensional vector, and each entry is either zero or one, there are indeed $2^{16} = 65,536$ such strategies in total.

Explicit formulas for the players' payoffs. The payoffs of two memory-2 players can again be computed with a Markov chain approach. To this end, suppose the strategies of player 1 and player 2 are \mathbf{p} and \mathbf{q} , respectively, and each of these strategies is of the form (2.75). Then, the respective Markov chain has sixteen possible states, summarizing the last two moves of either player, $\begin{matrix} CC & CC & CD & & DD \\ CC & CD & CC & \dots & DD \end{matrix}$. Slightly abusing

notation, here the upper indices refer to the past two actions of player 1 and the lower two indices refer to the past two actions of player 2. Given this ordering of the states, the transition matrix \mathcal{M}_A of the alternating game takes the following form (The transition matrix for the simultaneous game takes a similar form, and has been derived elsewhere⁹⁸):

$$\begin{pmatrix}
 p_{CC}q_{CC} & 0 & 0 & 0 & p_{CC}\bar{q}_{CC} & 0 & 0 & 0 & \bar{p}_{CC}q_{CC} & 0 & 0 & 0 & \bar{p}_{CC}\bar{q}_{CC} & 0 & 0 & 0 \\
 p_{CC}q_{CD} & 0 & 0 & 0 & p_{CC}\bar{q}_{CD} & 0 & 0 & 0 & \bar{p}_{CC}q_{CD} & 0 & 0 & 0 & \bar{p}_{CC}\bar{q}_{CD} & 0 & 0 & 0 \\
 p_{CD}q_{CC} & 0 & 0 & 0 & p_{CD}\bar{q}_{CC} & 0 & 0 & 0 & \bar{p}_{CD}q_{CC} & 0 & 0 & 0 & \bar{p}_{CD}\bar{q}_{CC} & 0 & 0 & 0 \\
 p_{CD}q_{CD} & 0 & 0 & 0 & p_{CD}\bar{q}_{CD} & 0 & 0 & 0 & \bar{p}_{CD}q_{CD} & 0 & 0 & 0 & \bar{p}_{CD}\bar{q}_{CD} & 0 & 0 & 0 \\
 0 & p_{CC}q_{DC} & 0 & 0 & 0 & p_{CC}\bar{q}_{DC} & 0 & 0 & 0 & \bar{p}_{CC}q_{DC} & 0 & 0 & 0 & \bar{p}_{CC}\bar{q}_{DC} & 0 & 0 \\
 0 & p_{CC}q_{DD} & 0 & 0 & 0 & p_{CC}\bar{q}_{DD} & 0 & 0 & 0 & \bar{p}_{CC}q_{DD} & 0 & 0 & 0 & \bar{p}_{CC}\bar{q}_{DD} & 0 & 0 \\
 0 & p_{CD}q_{DC} & 0 & 0 & 0 & p_{CD}\bar{q}_{DC} & 0 & 0 & 0 & \bar{p}_{CD}q_{DC} & 0 & 0 & 0 & \bar{p}_{CD}\bar{q}_{DC} & 0 & 0 \\
 0 & p_{CD}q_{DD} & 0 & 0 & 0 & p_{CD}\bar{q}_{DD} & 0 & 0 & 0 & \bar{p}_{CD}q_{DD} & 0 & 0 & 0 & \bar{p}_{CD}\bar{q}_{DD} & 0 & 0 \\
 0 & 0 & p_{DC}q_{CC} & 0 & 0 & 0 & p_{DC}\bar{q}_{CC} & 0 & 0 & 0 & \bar{p}_{DC}q_{CC} & 0 & 0 & 0 & \bar{p}_{DC}\bar{q}_{CC} & 0 \\
 0 & 0 & p_{DC}q_{CD} & 0 & 0 & 0 & p_{DC}\bar{q}_{CD} & 0 & 0 & 0 & \bar{p}_{DC}q_{CD} & 0 & 0 & 0 & \bar{p}_{DC}\bar{q}_{CD} & 0 \\
 0 & 0 & p_{DD}q_{CC} & 0 & 0 & 0 & p_{DD}\bar{q}_{CC} & 0 & 0 & 0 & \bar{p}_{DD}q_{CC} & 0 & 0 & 0 & \bar{p}_{DD}\bar{q}_{CC} & 0 \\
 0 & 0 & p_{DD}q_{CD} & 0 & 0 & 0 & p_{DD}\bar{q}_{CD} & 0 & 0 & 0 & \bar{p}_{DD}q_{CD} & 0 & 0 & 0 & \bar{p}_{DD}\bar{q}_{CD} & 0 \\
 0 & 0 & 0 & p_{DC}q_{DC} & 0 & 0 & 0 & p_{DC}\bar{q}_{DC} & 0 & 0 & 0 & \bar{p}_{DC}q_{DC} & 0 & 0 & 0 & \bar{p}_{DC}\bar{q}_{DC} \\
 0 & 0 & 0 & p_{DC}q_{DD} & 0 & 0 & 0 & p_{DC}\bar{q}_{DD} & 0 & 0 & 0 & \bar{p}_{DC}q_{DD} & 0 & 0 & 0 & \bar{p}_{DC}\bar{q}_{DD} \\
 0 & 0 & 0 & p_{DD}q_{DC} & 0 & 0 & 0 & p_{DD}\bar{q}_{DC} & 0 & 0 & 0 & \bar{p}_{DD}q_{DC} & 0 & 0 & 0 & \bar{p}_{DD}\bar{q}_{DC} \\
 0 & 0 & 0 & p_{DD}q_{DD} & 0 & 0 & 0 & p_{DD}\bar{q}_{DD} & 0 & 0 & 0 & \bar{p}_{DD}q_{DD} & 0 & 0 & 0 & \bar{p}_{DD}\bar{q}_{DD}
 \end{pmatrix}. \tag{2.77}$$

Given the transition matrix, we compute the invariant distribution of the respective Markov chain by solving $v = v\mathcal{M}_A$. This invariant distribution is now a 16-dimensional

vector,

$$\mathbf{v} = \left(\begin{array}{cccccccccccccccc} v_{CC}, & v_{CC}, & v_{CD}, & v_{CD}, & v_{CC}, & v_{CC}, & v_{CD}, & v_{CD}, & v_{DC}, & v_{DC}, & v_{DD}, & v_{DD}, & v_{DC}, & v_{DC}, & v_{DD}, & v_{DD} \end{array} \right). \quad (2.78)$$

Using this invariant distribution, we calculate how often each player cooperates on average. To this end, we sum up over all outcomes in which the player cooperates in the last round,

$$\begin{aligned} \rho_1 &= \frac{v_{CC}}{CC} + \frac{v_{CC}}{CD} + \frac{v_{CD}}{CC} + \frac{v_{CD}}{CD} + \frac{v_{CC}}{DC} + \frac{v_{CC}}{DD} + \frac{v_{CD}}{DC} + \frac{v_{CD}}{DD}, \\ \rho_2 &= \frac{v_{CC}}{CC} + \frac{v_{CC}}{CD} + \frac{v_{CD}}{CC} + \frac{v_{CD}}{CD} + \frac{v_{DC}}{CC} + \frac{v_{DC}}{CD} + \frac{v_{DD}}{CC} + \frac{v_{DD}}{CD}. \end{aligned} \quad (2.79)$$

Given these average cooperation rates, the players' payoffs are

$$\pi_1 = b\rho_2 - c\rho_1 \quad \text{and} \quad \pi_2 = b\rho_1 - c\rho_2. \quad (2.80)$$

Stability of pure memory-2 strategies. In Hilbe et al⁹⁸, the authors introduce an algorithm to describe the Nash equilibria of the simultaneous game among all pure memory-2 strategies. In addition, this algorithm outputs the range for the benefit-to-cost ratio b/c for which the respective strategy is an equilibrium. In the following, we briefly recapitulate that algorithm and apply it to the alternating game.

To test whether a given pure strategy \mathbf{p} is stable, we first compute the average probability ρ with which the strategy cooperates against itself, using Eq. (2.79). In particular, the payoff of two players who both use strategy \mathbf{p} is $\pi = b\rho - c\rho$. Now, if one player instead switches to some other pure strategy \mathbf{q} , the payoff of the deviating player is $\tilde{\pi}_{\mathbf{q}} = b\tilde{\rho}_{\mathbf{p}} - c\tilde{\rho}_{\mathbf{q}}$. Here, $\tilde{\rho}_{\mathbf{p}}$ is the average cooperation probability of the \mathbf{p} -player,

and $\tilde{\rho}_{\mathbf{q}}$ is the cooperation probability of the \mathbf{q} -player. For (\mathbf{p}, \mathbf{p}) to be a Nash equilibrium, it needs to be the case that $\pi \geq \tilde{\pi}_{\mathbf{q}}$. This condition simplifies to

$$b \cdot x_{\mathbf{p},\mathbf{q}} \geq c \cdot \gamma_{\mathbf{q},\mathbf{p}}, \quad (2.81)$$

with $x_{\mathbf{p},\mathbf{q}} := \rho - \tilde{\rho}_{\mathbf{p}}$ and $\gamma_{\mathbf{q},\mathbf{p}} := \rho - \tilde{\rho}_{\mathbf{q}}$. Depending on $x_{\mathbf{p},\mathbf{q}}$ and $\gamma_{\mathbf{q},\mathbf{p}}$, there are four possible cases.

1. If $x_{\mathbf{p},\mathbf{q}} > 0$ and $\gamma_{\mathbf{q},\mathbf{p}} > 0$, condition (2.81) is satisfied if and only if $b/c \geq \gamma_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}}$.
2. If $x_{\mathbf{p},\mathbf{q}} < 0$ and $\gamma_{\mathbf{q},\mathbf{p}} \leq 0$, condition (2.81) is satisfied if and only if $b/c \leq \gamma_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}}$.
3. If $x_{\mathbf{p},\mathbf{q}} \leq 0$ and $\gamma_{\mathbf{q},\mathbf{p}} > 0$, condition (2.81) is never satisfied.
4. If $x_{\mathbf{p},\mathbf{q}} \geq 0$ and $\gamma_{\mathbf{q},\mathbf{p}} \leq 0$, condition (2.81) is always satisfied.

Based on these considerations, we define the following three subsets of memory-2 strategies with respect to the focal strategy \mathbf{p} ,

$$\begin{aligned} Q_1(\mathbf{p}) &= \{\mathbf{q} \mid x_{\mathbf{p},\mathbf{q}} > 0 \text{ and } \gamma_{\mathbf{q},\mathbf{p}} > 0\}, \\ Q_2(\mathbf{p}) &= \{\mathbf{q} \mid x_{\mathbf{p},\mathbf{q}} < 0 \text{ and } \gamma_{\mathbf{q},\mathbf{p}} \leq 0\}, \\ Q_3(\mathbf{p}) &= \{\mathbf{q} \mid x_{\mathbf{p},\mathbf{q}} \leq 0 \text{ and } \gamma_{\mathbf{q},\mathbf{p}} > 0\}. \end{aligned} \quad (2.82)$$

Taking into account the four cases described above, the first set $Q_1(\mathbf{p})$ contains all memory-2 strategies against which \mathbf{p} is only stable if b/c is sufficiently large. The second set $Q_2(\mathbf{p})$ contains all memory-2 strategies against which \mathbf{p} is only stable if b/c is sufficiently small. The last set contains the strategies against which \mathbf{p} is never stable,

for no b/c . In particular, we can use these sets to define lower and upper bounds for the benefit-to-cost ratio for \mathbf{p} to be a Nash equilibrium,

$$(b/c)_{LB} = \max \{ y_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}} \mid \mathbf{q} \in Q_1(\mathbf{p}) \} \quad \text{and} \quad (b/c)_{UB} = \min \{ y_{\mathbf{q},\mathbf{p}}/x_{\mathbf{p},\mathbf{q}} \mid \mathbf{q} \in Q_2(\mathbf{p}) \}. \quad (2.83)$$

Using these thresholds, it follows that \mathbf{p} can only be a Nash equilibrium if

$$(b/c)_{LB} \leq b/c \leq (b/c)_{UB} \quad \text{and} \quad Q_3(\mathbf{p}) = \emptyset. \quad (2.84)$$

For a given strategy \mathbf{p} , these conditions can be checked by computing $x_{\mathbf{p},\mathbf{q}}$ and $y_{\mathbf{q},\mathbf{p}}$ for all 2^{16} pure memory-2 strategies \mathbf{q} . In Figure 2.13a, we illustrate the result of this algorithm for both the simultaneous game and the alternating game. For that figure, we call a Nash equilibrium locally robust if it is an equilibrium for a substantial portion of the parameter space; specifically, we require $(b/c)_{UB} - (b/c)_{LB} > 0.2$. The figure then displays all locally robust Nash equilibria for $b/c \leq 5$. We find that in the simultaneous game, there are 34 such equilibria. Out of those, there are several equilibria that yield very little cooperation (colored in red). These equilibrium strategies include, for example, ALLD. On the other hand, for $b/c > 3/2$, there are also several equilibria that yield almost full cooperation. These strategies display so-called all-or-none behavior^{98,185}. Players with these AON_k strategies tend to cooperate if in each of the past k rounds, either both players cooperated or no one did.

For the alternating game, we find that the only locally robust Nash equilibrium among the pure memory-2 strategies is ALLD. There are two additional strategies that are Nash equilibria without being locally robust. These are:

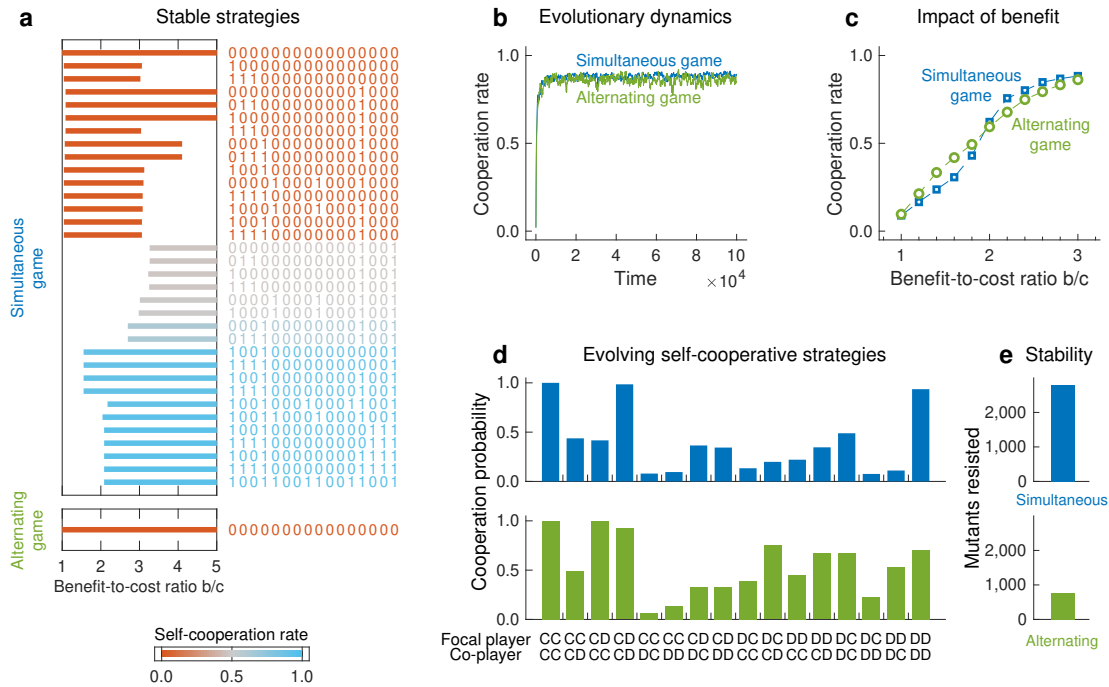


Figure 2.13: Stability and evolutionary dynamics of pure memory-2 strategies. We have also run simulations for (infinitely repeated) simultaneous and alternating games when players have access to all 65,536 memory-2 strategies. These strategies are 16-dimensional vectors that take each player’s last two actions into account. a, In a first step, we have computed which of these strategies are evolutionarily stable for an error rate of $\varepsilon = 0.02$. Here, we display the respective strategies (encoded by the 16 integers on the right hand side), the range of b/c values for which these strategies are stable (indicated by the length of the lines), and the self-cooperation rates of these strategies (indicated by the color of the respective line). In the simultaneous game, there are many evolutionarily stable strategies, including strategies that yield almost full cooperation. In contrast, in the alternating game, ALLD is the only strategy that is evolutionarily stable for a positive range of b/c values. b,c Although only the simultaneous game allows for evolutionarily stable cooperation, simulations suggest that alternating games yield similar average cooperation rates. d, In a next step, we have recorded which strategies the players use to cooperate among themselves (for this simulation we again call a strategy self-cooperative if it yields a cooperation rate of at least 80% against itself). In the simultaneous game, the self-cooperative strategies resemble the previously reported all-or-none strategies⁹⁸. Here, the two players are most likely to cooperate if they both cooperated in the last two rounds, if none of them cooperated in the last two rounds, or if they both cooperated in the last round, but defected in the second-to-last round. In the alternating game, the players’ conditional cooperation probabilities seem more irregular. e, We have also computed how robust self-cooperative strategies are, by recording how many mutant strategies it takes on average to successfully invade into a resident population of self-cooperators. As expected from our evolutionary stability analysis, self-cooperative strategies are more robust in simultaneous games. For details, see Supplementary Note 3.

$$\begin{aligned}
\mathbf{p}_1 &= (0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 1) \\
\mathbf{p}_2 &= (1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0)
\end{aligned} \tag{2.85}$$

Both of these strategies are only stable for a single value of b/c . For the given error rate of $\varepsilon = 0.02$, this value is $b/c = 4.1464$ for the first strategy and $b/c = 1.04166$ for the second. Moreover, none of the two strategies is self-cooperative. The cooperation rate of the first strategy against itself is 64.9%; the self-cooperation rate of the second strategy is 50.0%.

These results show that there is no Nash equilibrium among the pure memory-2 strategies that can sustain full cooperation in the alternating game. Since any evolutionarily stable strategy needs to be a pure strategy, we conclude that evolutionarily stable cooperation in the alternating game is infeasible (as in the case of the memory-1 strategies).

Evolutionary dynamics. In addition to these static results, we have also explored the evolutionary dynamics among pure memory-2 strategies with simulations. For these simulations we consider the case of an infinitely repeated game with a positive error rate ($\varepsilon = 0.02$) in the limit of rare mutations. Figure 2.13b shows the evolving average cooperation rate for both the simultaneous and the alternating game (averaged over 100 independent simulation runs for $b = 3$). Although only the simultaneous game has fully cooperative Nash equilibria, the two scenarios lead to largely comparable overall cooperation rates. We obtain similar results for other benefit values (Figure 2.13c).

To explore this result in more detail, Figure 2.13d analyzes the players' average cooperation probabilities when players adopt a self-cooperative strategy. In the simultaneous game, the player's average cooperation probabilities resemble the typical behavior of AON_2 strategies: players have a high cooperation probability if either

(i) both players mutually cooperated for two rounds, (ii) if both players mutually defected for two rounds, or (iii) if both players cooperated in the last round but defected in the second-to-last round. In all other cases, the conditional cooperation probability is below 50%. In contrast, in the alternating game, players seem to be prepared to cooperate as long as the co-player did not defect more often than the focal player did.

However, Figure 2.13e shows that in line with the equilibrium analysis, the self-cooperative strategies in the simultaneous game tend to be more robust. For this panel, we have recorded for each self-cooperative strategy adopted by the resident population how many mutant strategies it takes on average until the first mutant reaches fixation. We find that in the simultaneous game, it takes on average almost 3,000 mutant strategies to invade a self-cooperative resident. In the alternating game, this number is considerably lower; on average it takes less than 800 random mutant strategies until the resident strategy is successfully invaded. Overall, these results suggest that cooperation in the simultaneous game is generally more robust. However, also in the alternating game, individuals adopt self-cooperative strategies for a substantial amount of time.

2.7.4 Games in spatial populations

Motivation. The simulation results for the baseline model and the previous model extensions are based on the assumption that the population is well-mixed. This assumption has two consequences. First, when playing games, all members of the population are equally likely to interact with everyone else. Second, for the evolutionary updating, all population members are equally likely to act as a role model for any given focal player. The assumption of well-mixed populations has a long tradition

in evolutionary game theory (see, for instance, the text books in Refs. ^{167,215}). However, there is also a rich literature asking how strategies spread in structured populations ^{83,115,168,184}. In the following, we thus explore the strategies that evolve when both interactions and imitation events are local.

To explore the dynamics of spatial games, we closely follow the approach of Brauchli et al ³¹. They study the simultaneous game on a square lattice with periodic boundary conditions. The set of available strategies consists of all (stochastic) memory-1 strategies. With extensive computer simulations, the study shows that spatial games are generally more conducive to the evolution of cooperation. Moreover, evolutionary trajectories are less chaotic, and more likely to result in eventual behavior that is consistent with the strategy Win-Stay Lose-Shift. In the following, we use their setup to (i) repeat their simulations for the simultaneous game, and (ii) extend these simulations to the case of alternating games.

Model setup. For our exploration of games in structured populations, we consider a population with 2,500 individuals placed on a 50×50 square lattice. Individuals use memory-1 strategies to engage in a repeated game with each of their eight immediate neighbors (we use a Moore neighborhood with periodic boundary conditions). We consider two independent scenarios. These scenarios differ in whether the game being played is the simultaneous game or the alternating game. In both cases we use the baseline versions of these games (in which there is no discounting of the future). A player's payoff at any point in time is defined as the player's average payoff against its eight neighbors (taking the sum of the eight pairwise payoffs would yield the same result).

Initially, we assume that all population members adopt the strategy ALLD. In each generation of the simulation, all population members update their strategies. With

probability $1 - \mu$, an individual who is to update its strategy adopts the strategy of the neighbor with the highest payoff. With the converse probability μ , the individual adopts a random memory-1 strategy, by drawing four numbers $(p_{CC}, p_{CD}, p_{DC}, p_{DD})$ from the hypercube $[0, 1]^4$ uniformly at random. This elementary process is then repeated for 20,000 generations. For each generation, we record the strategy that is adopted by each individual, and the average cooperation rate across all interactions taking place in the population.

This overall setup agrees with the setup considered by Brauchli et al³¹, with a few minor exceptions. First, we use a different initial population, ALLD, to better visualize the emergence of cooperation in a population of defectors. In contrast, Brauchli et al³¹ assume that all players initially use the perfectly random strategy $\mathbf{p} = (0.5, 0.5, 0.5, 0.5)$. Second, because we are also interested in the outcome of alternating game, we use one-shot payoffs based on the infinitely repeated donation game, as defined by Eq. (2.19). Brauchli et al³¹ instead use the payoff values of Axelrod⁹, and they consider games that last on average between 100 and 200 rounds. Despite these differences, our simulation results for the simultaneous game are comparable to theirs (as described in more detail below).

Evolutionary dynamics. For our evolutionary simulations, we used parameters that are comparably hostile to cooperation: the benefit of cooperation is smaller than in previous simulations (now $b = 2$ instead of $b = 3$), and errors occur at an appreciable rate, $\varepsilon = 0.02$. Figure 2.6g shows the resulting cooperation dynamics (averaged over 50 independent simulations). Despite the hostile conditions, we observe that spatial games lead to predominantly cooperative populations rather quickly. When we compare the simultaneous game to the alternating game, we observe that the simultaneous game leads to more (and to more robust) cooperation. To explore these results in

more detail, Figure 2.6i shows snapshots of the population at different points in time. In the simultaneous game, we observe that populations almost always converge to a largely homogeneous configuration of cooperative players. In contrast, in the alternating game, different simulation runs can exhibit very different behaviors. In some simulations, we observe a similar dynamics towards almost uniform cooperation as in the simultaneous game. Other simulation runs, however, result in stable mixtures of cooperating and defecting players (this latter case is displayed in the bottom panel of Figure 2.6(h)).

In a next step, we explored which strategies the players use to maintain cooperation in the two scenarios. To this end, we recorded all used strategies that yield a cooperation rate of at least 80% against themselves; then we computed the respective average cooperation probabilities across all these strategies (Figure 2.6(i)). For the simultaneous game, this average strategy exhibits the characteristics of Win-Stay Lose-Shift (as already reported by Brauchli et al³¹). Players are most likely to cooperate after mutual cooperation and mutual defection; after all other outcomes, they tend to defect. For the alternating game, the average strategy reflects some of the characteristics of Stochastic Firm-but-Fair. Here again, players are most likely to cooperate if the opponent's last move was to cooperate.

Overall, our results are in line with the main conclusions of the baseline model: Cooperation in alternating games is slightly less robust, and it requires different kinds of strategies. At the same time, the simulations also highlight the intriguing spatial patterns that can arise in structured populations.

2.7.5 The effect of local mutations

Motivation. For our simulations so far, we assumed that when a mutation occurs, the player's new strategy can be arbitrarily different from the player's present strategy. In that case we speak of 'global mutations'. The assumption of global mutations is fairly common in the evolutionary game theory literature^{84,95,170,221}. However, there is also important work on the effects of local mutations^{105,222}. When mutations are local, they only lead to a slight modification of the players' strategies. Which of the two mutation schemes is more relevant depends on the type of evolution considered. Biological evolution is perhaps better described by local mutations, whereas for cultural processes global mutations may be more reasonable.

Compared to global mutations, local mutations can affect the dynamics in three ways:

1. It can introduce additional (local) equilibria. The corresponding strategies are robust with respect to local mutants, although they can be invaded by strategies further away in the strategy space;
2. It affects how likely any given equilibrium is reached;
3. It affects the robustness of any given equilibrium: When mutations have a sufficiently short range, any mutant strategy has approximately the same payoff as the resident strategy (since payoffs are continuous in the players' strategies). As a result, even a strategy that is evolutionarily stable can be invaded by a strategy that is sufficiently close-by with an approximate probability of $1/N$ (the neutral fixation probability)²⁴⁵. As a result, the concept of evolutionary stability becomes overall less relevant to describe the stochastic evolutionary dynamics in finite populations with local mutations.

To explore these effects in the context of alternating games, we have implemented additional simulations.

Model setup. We use the same basic framework that we used to describe the evolutionary dynamics of the baseline model. However, this time, when a player with strategy \mathbf{p} undergoes a mutation, the new strategy is uniformly chosen among all memory-1 strategies \mathbf{p}' that satisfy

$$|p'_{ij} - p_{ij}| \leq m, \quad \text{for all } i, j \in \{C, D\}. \quad (2.86)$$

We refer to the parameter m as the mutation range; it describes how far apart the mutant strategy can be from the parent strategy. For $m \geq 1$, we recover the case of global mutations. For smaller m , mutations are restricted to generate strategies in a local neighborhood of the parent strategy.

Evolutionary dynamics. In Figure 2.14a,b, we compare the results for global mutations with the corresponding results for local mutations (using $m = 0.05$). We observe that for both the simultaneous and the alternating game, overall cooperation rates under local mutations tend to be lower on average. However, the magnitude of the effect differs: While local mutations strongly reduce cooperation in the simultaneous game, it has a much smaller negative effect on the alternating game.

To analyze this effect in more detail, we again compute an average over all self-cooperative strategies used by the players (Figure 2.14c,d). The resulting average strategies resemble Win-Stay Lose-Shift (in the simultaneous game) and Stochastic Firm-but-Fair (in the alternating game), largely independent of whether mutations are local or global. However, local mutations have a substantial effect on the robustness of these self-cooperative strategies. For local mutations, the number of mutants

it takes to invade a resident self-cooperative strategy is of the order of N (which is 100 in these simulations), as expected. In contrast, under global mutations, resident strategies typically resist $\sim 6,300$ mutant strategies in the simultaneous game, and $\sim 1,600$ mutant strategies in the alternating game.

In Figure 2.14e,f, we systematically explore the effect of different mutation ranges between $m = 0.05$ (local mutations) and $m = 0.95$ (almost global mutations). The alternating game yields slightly more cooperation than the simultaneous game when mutations are local. However, once the mutation range exceeds $m \approx 0.4$, it is the simultaneous game that is more conducive to cooperation.

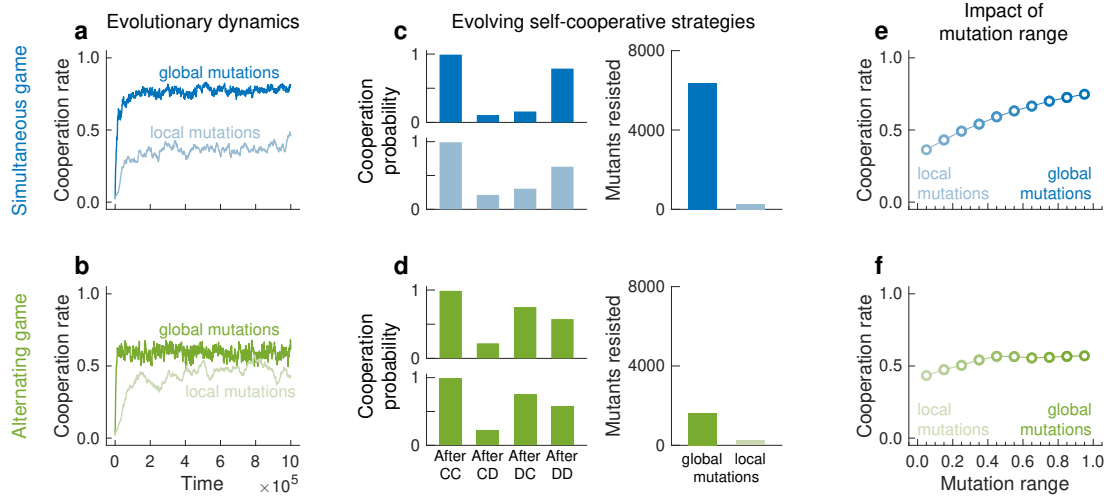


Figure 2.14: Evolutionary dynamics under local mutations. The previous simulations assume that mutations are global: mutant strategies can be arbitrarily far away from the resident strategy. Here we compare this scenario with the case of local mutations, where mutant strategies are required to be in a small neighborhood of the resident strategy. We measure the size of this neighborhood by the mutation range m . The mutation range reflects by how much the mutant's conditional cooperation probabilities are allowed to differ from the resident strategy. Unless noted otherwise, we use $m = 0.05$. a,b, In both the simultaneous and the alternating game, local mutations lead to less cooperation. However, the effect is more notable in the simultaneous game. c,d, Local mutations do not affect which strategies the players use on average to maintain cooperation. However, they affect how robust these strategies are. Under local mutations, all mutants have approximately the same fitness as the resident. As a result, the evolutionary competition is almost neutral; on average, it thus takes an order of N mutants to invade any given resident population (here, the population size is $N = 100$). e,f, We have repeated these simulations for different mutation ranges. The mutation range has a strong effect on cooperation in the simultaneous game (where evolutionarily stable cooperation is possible). It has a comparably weak effect in the alternating game (in which no evolutionarily stable strategy exists that leads to full cooperation).

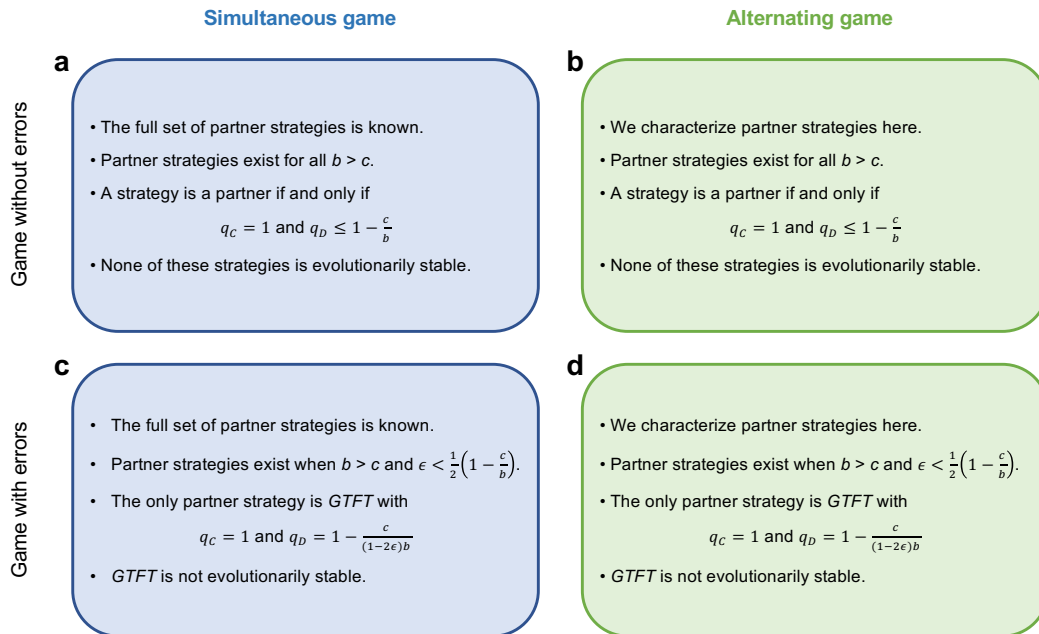


Figure 2.15: A characterization of partners among the reactive strategies. Here, we describe the set of partner strategies within the class of reactive strategies. Reactive strategies are a subset of memory-1 strategies. They consist of two conditional cooperation probabilities, q_C and q_D . The two probabilities describe how a player responds to a co-player's cooperation and defection, respectively. a,b, For reactive strategies, the alternating and the simultaneous game lead to the same payoffs¹⁷¹. As a result, also the partner strategies coincide in each case. c,d, With errors, *GTFT* is the only partner strategy. It can be neutrally invaded by *ALLC*, and hence it is not evolutionarily stable.

2.8 Supplementary Note 4: Proofs of the analytical results

Proof of Proposition 2.1. Let $\mathbf{v}(\mathbf{p}, \mathbf{q})$ be an invariant distribution of the game between strategies \mathbf{p} and \mathbf{q} . Then $\mathbf{v}(\mathbf{p}, \mathbf{q})$ is a solution of the equation $\mathbf{v} = \mathbf{v}\mathcal{M}_A(\mathbf{p}, \mathbf{q})$, with $\mathcal{M}_A(\mathbf{p}, \mathbf{q})$ being the transition matrix of the game as defined by Eq. (2.17). More explicitly, $\mathbf{v}(\mathbf{p}, \mathbf{q})$ solves the following system of linear equations,

$$\begin{aligned}
 v_{CC} &= v_{CC}p_{CC}q_{CC} + v_{CD}p_{CD}q_{DC} + v_{DC}p_{DC}q_{CC} + v_{DD}p_{DD}q_{DC} \\
 v_{CD} &= v_{CC}p_{CC}(1-q_{CC}) + v_{CD}p_{CD}(1-q_{DC}) + v_{DC}p_{DC}(1-q_{CC}) + v_{DD}p_{DD}(1-q_{DD}) \\
 v_{DC} &= v_{CC}(1-p_{CC})q_{CD} + v_{CD}(1-p_{CD})q_{DD} + v_{DC}(1-p_{DC})q_{CD} + v_{DD}(1-p_{DD})q_{DD} \\
 v_{DD} &= v_{CC}(1-p_{CC})(1-q_{CD}) + v_{CD}(1-p_{CD})(1-q_{DD}) + v_{DC}(1-p_{DC})(1-q_{CD}) + v_{DD}(1-p_{DD})(1-q_{DD}).
 \end{aligned} \tag{2.87}$$

By simplifying the right hand's side, we can write Eq. (2.87) as

$$\begin{aligned}
 v_{CC} &= (v_{CC}p_{CC} + v_{DC}p_{DC})q_{CC} + (v_{CD}p_{CD} + v_{DD}p_{DD})q_{DC} \\
 v_{CD} &= (v_{CC}p_{CC} + v_{DC}p_{DC})(1-q_{CC}) + (v_{CD}p_{CD} + v_{DD}p_{DD})(1-q_{DC}) \\
 v_{DC} &= (v_{CC}(1-p_{CC}) + v_{DC}(1-p_{DC}))q_{CD} + (v_{CD}(1-p_{CD}) + v_{DD}(1-p_{DD}))q_{DD} \\
 v_{DD} &= (v_{CC}(1-p_{CC}) + v_{DC}(1-p_{DC}))(1-q_{CD}) + (v_{CD}(1-p_{CD}) + v_{DD}(1-p_{DD}))(1-q_{DD}).
 \end{aligned} \tag{2.88}$$

Now by using assumption (2.21)

$$\begin{aligned}
 (v_{CC}+v_{DC})\tilde{p}_C &= v_{CC}p_{CC} + v_{DC}p_{DC} \\
 (v_{CD}+v_{DD})\tilde{p}_D &= v_{CD}p_{CD} + v_{DD}p_{DD},
 \end{aligned} \tag{2.89}$$

and its equivalent formulation

$$\begin{aligned}
 (v_{CC}+v_{DC})(1-\tilde{p}_C) &= v_{CC}(1-p_{CC}) + v_{DC}(1-p_{DC}) \\
 (v_{CD}+v_{DD})(1-\tilde{p}_D) &= v_{CD}(1-p_{CD}) + v_{DD}(1-p_{DD}),
 \end{aligned} \tag{2.90}$$

we can write Eq. (2.88) as

$$\begin{aligned}
v_{CC} &= (v_{CC}+v_{DC})\tilde{p}_C q_{CC} & + (v_{CD}+v_{DD})\tilde{p}_D q_{DC} \\
v_{CD} &= (v_{CC}+v_{DC})\tilde{p}_C (1-q_{CC}) & + (v_{CD}+v_{DD})\tilde{p}_D (1-q_{DC}) \\
v_{DC} &= (v_{CC}+v_{DC})(1-\tilde{p}_C) q_{CD} & + (v_{CD}+v_{DD})(1-\tilde{p}_D) q_{DD} \\
v_{DD} &= (v_{CC}+v_{DC})(1-\tilde{p}_C)(1-q_{CD}) & + (v_{CD}+v_{DD})(1-\tilde{p}_D)(1-q_{DD}).
\end{aligned} \tag{2.91}$$

This equation can be rewritten as $\mathbf{v} = \mathbf{v}M_A(\tilde{\mathbf{p}}, \mathbf{q})$, where $M_A(\tilde{\mathbf{p}}, \mathbf{q})$ is now the transition matrix of the game between $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$ and \mathbf{q} . If \mathbf{v} solves $\mathbf{v} = \mathbf{v}M_A(\mathbf{p}, \mathbf{q})$, it thus also solves $\mathbf{v} = \mathbf{v}M_A(\tilde{\mathbf{p}}, \mathbf{q})$. \square

Proof of Proposition 2.2. Consider an alternating game in which both players' strategies are fixed and player 2 adopts a memory-1 strategy \mathbf{q} . Let $v_{a_1, a_2}(t)$ denote the probability that the players choose the actions $(a_1, a_2) \in \{CC, CD, DC, DD\}$ at time t in the resulting game. By assumption, the following limiting averages are well-defined,

$$v_{CC} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CC}(t), \quad v_{CD} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CD}(t), \quad v_{DC} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{DC}(t), \quad v_{DD} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{DD}(t). \tag{2.92}$$

We write these four limits as a vector $\mathbf{v} = (v_{CC}, v_{CD}, v_{DC}, v_{DD})$. Moreover, let $p_{a_1, a_2}(t)$ denote the conditional probability that player 1 cooperates at time $t+1$, given the history of the game is such that the players' actions at time t are (a_1, a_2) . Again, by assumption we can define a reactive strategy $\tilde{\mathbf{p}} = (\tilde{p}_C, \tilde{p}_D)$ as an implicit solution of two

equations

$$\begin{aligned}
(v_{CC}+v_{DC})\tilde{p}_C &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CC}(t) p_{CC}(t) + v_{DC}(t) p_{DC}(t) \\
(v_{CD}+v_{DD})\tilde{p}_D &= \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_{CD}(t) p_{CD}(t) + v_{DD}(t) p_{DD}(t).
\end{aligned} \tag{2.93}$$

We need to show that the vector \mathbf{v} satisfies the linear system $\mathbf{v} = \mathbf{v}M_A(\tilde{\mathbf{p}}, \mathbf{q})$, where $M_A(\tilde{\mathbf{p}}, \mathbf{q})$ is the transition matrix defined by Eq. (2.17). We show this for the first equation of the system; all other equations are verified analogously. By the definition of $v_{a_1, a_2}(t)$ and $p_{a_1, a_2}(t)$, we can write $v_{CC}(t+1)$ as follows,

$$v_{CC}(t+1) = v_{CC}(t) p_{CC}(t) q_{CC} + v_{CD}(t) p_{CD}(t) q_{DC} + v_{DC}(t) p_{DC}(t) q_{CC} + v_{DD}(t) p_{DD}(t) q_{DC} \tag{2.94}$$

By summing up this equation for the first T time steps and collecting terms on the right hand side, we obtain

$$\sum_{t=1}^T v_{CC}(t+1) = \left(\sum_{t=1}^T v_{CC}(t) p_{CC}(t) + v_{DC}(t) p_{DC}(t) \right) q_{CC} + \left(\sum_{t=1}^T v_{CD}(t) p_{CD}(t) + v_{DD}(t) p_{DD}(t) \right) q_{DC} \tag{2.95}$$

Dividing both sides by T , taking the limit $T \rightarrow \infty$, and replacing the limits by the respective expressions in Eqs. (2.92) and (2.93), we obtain

$$v_{CC} = (v_{CC}+v_{DC})\tilde{p}_C q_{CC} + (v_{CD}+v_{DD})\tilde{p}_D q_{DC}. \tag{2.96}$$

This is exactly the first equation of the linear system $\mathbf{v} = \mathbf{v}M_A(\tilde{\mathbf{p}}, \mathbf{q})$. \square

Proof of Lemma 2.4. The payoff equation (2.26) follows immediately from the formula

in Eq. (2.19) when using the strategies \mathbf{p} and \mathbf{q} as input. To show the monotonicity property in p_C , we keep \mathbf{q} and p_D fixed and consider the function $p_C \mapsto f_C(p_C) := \pi(\mathbf{p}, \mathbf{q})$. By Eq. (2.26), this function can be written as

$$f_C(p_C) = \frac{a_1 + a_2 p_C}{a_3 + a_4 p_C}, \quad (2.97)$$

where a_1, a_2, a_3, a_4 are constants that are independent of p_C . Calculating the derivative yields

$$\frac{\partial f_C}{\partial p_C} = \frac{a_2 a_3 - a_1 a_4}{(a_3 + a_4 p_C)^2}. \quad (2.98)$$

In particular, the sign of the derivative is independent of p_C . That is, $f_C(p_C) = \pi(\mathbf{p}, \mathbf{q})$ is either strictly increasing (if $a_2 a_3 > a_1 a_4$), strictly decreasing (if $a_2 a_3 < a_1 a_4$), or constant in p_C (if $a_2 a_3 = a_1 a_4$). A similar argument shows that also the map $p_D \mapsto f_D(p_D) := \pi(\mathbf{p}, \mathbf{q})$ is monotonic in p_D . \square

Proof of Proposition 2.5. The proof is by iterated application of Lemma 2.4. Let \mathbf{q} and \mathbf{p} be arbitrary but fixed. We iteratively define $\mathbf{p}^0 = (p_C^0, p_D^0) := (p_C, p_D)$,

$$\mathbf{p}^1 = \begin{cases} (1, p_D^0) & \text{if } \pi((1, p_D^0), \mathbf{q}) \geq \pi(\mathbf{p}, \mathbf{q}) \\ (0, p_D^0) & \text{otherwise,} \end{cases} \quad (2.99)$$

and

$$\mathbf{p}^2 = \begin{cases} (p_C^1, 1) & \text{if } \pi((p_C^1, 1), \mathbf{q}) \geq \pi(\mathbf{p}, \mathbf{q}) \\ (p_C^1, 0) & \text{otherwise} \end{cases} \quad (2.100)$$

Because we only change one component in each step, it follows by the monotonicity

property in Lemma 2.4 that $\pi(\mathbf{p}^2, \mathbf{q}) \geq \pi(\mathbf{p}^1, \mathbf{q}) \geq \pi(\mathbf{p}^0, \mathbf{q})$. Moreover, $\mathbf{p}^2 \in \{0, 1\}^2$ by construction. Therefore, defining $\mathbf{p}' := \mathbf{p}^2$ yields the desired result. \square

Proof of Lemma 2.7.

1. Suppose \mathbf{q} is a generic Nash equilibrium and $\tilde{\mathbf{q}} = (\tilde{q}_C, \tilde{q}_D) \in (0, 1)^2$ is its reactive marginalization with respect to itself. In particular, $\tilde{\mathbf{q}}$ is a generic best response to \mathbf{q} , since \mathbf{q} is a best response to itself. It follows that the map $\tilde{q}_C \mapsto \pi(\tilde{\mathbf{q}}, \mathbf{q})$ needs to be constant (otherwise Lemma 2.4 implies that it is either strictly increasing or decreasing, which both contradicts the best reply property). As a consequence, both respective boundary strategies $\tilde{\mathbf{q}}' := (0, \tilde{q}_D)$ and $\tilde{\mathbf{q}}'' := (1, \tilde{q}_D)$ are also generic best responses to \mathbf{q} . With the same argument, one can now show that the maps $\tilde{q}_D \mapsto \pi(\tilde{\mathbf{q}}', \mathbf{q})$ and $\tilde{q}_D \mapsto \pi(\tilde{\mathbf{q}}'', \mathbf{q})$ are also constant. Therefore all respective boundary strategies – ALLD, ATFT, TFT, ALLC – are generic best responses to \mathbf{q} . In particular, all four boundary strategies yield the same payoff against \mathbf{q} . That is, we have shown Eq. (2.32). Because Eq. (2.32) implies Eq. (2.33), we conclude that \mathbf{q} is an equalizer.
2. Using Eqs. (2.18) and (2.22), we can compute the reactive marginalization $\tilde{\mathbf{q}}$ of \mathbf{q} with respect to itself explicitly. This yields

$$\tilde{\mathbf{q}} = (\tilde{q}_C, \tilde{q}_D) = \left(\frac{q_{DC}}{1 - q_{CC} + q_{DC}}, \frac{q_{DD}}{1 - q_{CD} + q_{DD}} \right) \quad (2.101)$$

By assumption, this reactive marginalization is either semi-stochastic or deterministic, and therefore either $\tilde{q}_C \in \{0, 1\}$, $\tilde{q}_D \in \{0, 1\}$, or both. This gives rise to

four possible cases,

$$\begin{aligned}
\tilde{q}_C \in \{0, 1\} &\Leftrightarrow \mathbf{q} = (1, q_{CD}, q_{DC}, q_{DD}) \quad \text{or} \quad \mathbf{q} = (q_{CC}, q_{CD}, 0, q_{DD}) \\
\tilde{q}_D \in \{0, 1\} &\Leftrightarrow \mathbf{q} = (q_{CC}, 1, q_{DC}, q_{DD}) \quad \text{or} \quad \mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, 0).
\end{aligned}
\tag{2.102}$$

The first and the last case, $\mathbf{q} = (1, q_{CD}, q_{DC}, q_{DD})$ and $\mathbf{q} = (q_{CC}, q_{CD}, q_{DC}, 0)$ correspond to the self-cooperating and self-defecting players, respectively. They give rise to a generic Nash equilibrium if and only if the respective conditions for being a partner, or for being a defector are satisfied, as given by Eqs. (2.30) and (2.31). In the following, we discuss the remaining two cases. For those we can assume without loss of generality that $q_{CC} < 1$ and $q_{DD} > 0$.

First, suppose that $\mathbf{q} = (q_{CC}, q_{CD}, 0, q_{DD})$. Then by Eq. (2.27), the payoff of \mathbf{q} against itself is

$$\pi(\mathbf{q}, \mathbf{q}) = \frac{q_{DD}}{1 - q_{CD} + 2q_{DD}} (b - c).
\tag{2.103}$$

If a player deviates to ALLD instead, its payoff according to Eq. (2.28) becomes

$$\pi(\text{ALLD}, \mathbf{q}) = \frac{q_{DD}}{1 - q_{CD} + q_{DD}} \cdot b.
\tag{2.104}$$

In particular, $\pi(\text{ALLD}, \mathbf{q}) > \pi(\mathbf{q}, \mathbf{q})$.

Second, suppose $\mathbf{q} = (q_{CC}, 1, q_{DC}, q_{DD})$. Again we use Eq. (2.27) to compute the payoff of \mathbf{q} against itself, which yields

$$\pi(\mathbf{q}, \mathbf{q}) = \frac{1 - q_{CC} + q_{DC}}{2(1 - q_{CC}) + q_{DC}} (b - c).
\tag{2.105}$$

However, if a player deviates to ALLD, its payoff becomes $\pi(\text{ALLD}, \mathbf{q}) = b > \pi(\mathbf{q}, \mathbf{q})$.

□

Proof of Proposition 2.9.

Both results follow from a straightforward application of our earlier results for the case without errors. For (1), we note that

$$\begin{aligned}
\mathbf{v}^\varepsilon(\tilde{\mathbf{p}}, \mathbf{q}) &\stackrel{\text{Eq. (2.38)}}{=} \mathbf{v}(\tilde{\mathbf{p}}^\varepsilon, \mathbf{q}^\varepsilon) \\
&\stackrel{\text{Eq. (2.35)}}{=} \mathbf{v}\left(\left(\frac{v_{CC}^\varepsilon(\mathbf{p}, \mathbf{q})p_{CC}^\varepsilon + v_{DC}^\varepsilon(\mathbf{p}, \mathbf{q})p_{DC}^\varepsilon}{v_{CC}^\varepsilon(\mathbf{p}, \mathbf{q}) + v_{DC}^\varepsilon(\mathbf{p}, \mathbf{q})}, \frac{v_{CD}^\varepsilon(\mathbf{p}, \mathbf{q})p_{CD}^\varepsilon + v_{DD}^\varepsilon(\mathbf{p}, \mathbf{q})p_{DD}^\varepsilon}{v_{CD}^\varepsilon(\mathbf{p}, \mathbf{q}) + v_{DD}^\varepsilon(\mathbf{p}, \mathbf{q})}\right), \mathbf{q}^\varepsilon\right) \\
&\stackrel{\text{Eq. (2.38)}}{=} \mathbf{v}\left(\left(\frac{v_{CC}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)p_{CC}^\varepsilon + v_{DC}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)p_{DC}^\varepsilon}{v_{CC}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon) + v_{DC}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)}, \frac{v_{CD}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)p_{CD}^\varepsilon + v_{DD}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)p_{DD}^\varepsilon}{v_{CD}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon) + v_{DD}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon)}\right), \mathbf{q}^\varepsilon\right) \\
&\stackrel{\text{Prop. 2.1}}{=} \mathbf{v}(\mathbf{p}^\varepsilon, \mathbf{q}^\varepsilon) \\
&\stackrel{\text{Eq. (2.38)}}{=} \mathbf{v}^\varepsilon(\mathbf{p}, \mathbf{q}).
\end{aligned} \tag{2.106}$$

For (2), we note that by their definition in Eq. (2.41), \tilde{p}_C^ε and \tilde{p}_D^ε satisfy the conditions (2.23) in Proposition 2.2. Therefore we can conclude

$$\mathbf{v}^\varepsilon(\tilde{\mathbf{p}}, \mathbf{q}) \stackrel{\text{Eq. (2.38)}}{=} \mathbf{v}(\tilde{\mathbf{p}}^\varepsilon, \mathbf{q}^\varepsilon) \stackrel{\text{Proposition 2.2}}{=} \mathbf{v}. \tag{2.107}$$

□

Proof of Proposition 2.10.

Because the error transformation $\mathbf{p} \mapsto \mathbf{p}^\varepsilon = \phi^\varepsilon(\mathbf{p})$ is strictly monotonically increasing and defined on each component separately, it follows from Lemma 2.4 that also the maps

$$p_C \rightarrow \pi^\varepsilon(\mathbf{p}, \mathbf{q}) = \pi\left(\phi^\varepsilon((p_C, p_D)), \mathbf{q}\right) \quad \text{and} \quad p_D \rightarrow \pi^\varepsilon(\mathbf{p}, \mathbf{q}) = \pi\left(\phi^\varepsilon((p_C, p_D)), \mathbf{q}\right) \quad (2.108)$$

are either strictly monotonically increasing, decreasing, or constant. The result then follows with the same argument as in the proof of Proposition 2.5, by replacing $\pi(\mathbf{p}, \mathbf{q})$ with $\pi^\varepsilon(\mathbf{p}, \mathbf{q})$. \square

Proof of Theorem 2.11.

(1) \Rightarrow (2). Because \mathbf{q} is a generic Nash equilibrium, $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\mathbf{p}, \mathbf{q})$ for all generic strategies \mathbf{p} , which includes all reactive strategies.

(2) \Rightarrow (3). By Proposition 2.9, we can compute the payoff $\pi^\varepsilon(\mathbf{q}, \mathbf{q})$ of a memory-1 strategy \mathbf{q} against itself by computing the payoff of its reactive marginalization $\tilde{\mathbf{q}}$ against itself, $\pi^\varepsilon(\tilde{\mathbf{q}}, \mathbf{q})$. To this end, we use Eq. (2.39) to compute the entries of $\tilde{\mathbf{q}} = (\tilde{q}_C, \tilde{q}_D)$, yielding

$$\begin{aligned} \tilde{q}_C &= \frac{\varepsilon q_{CC} + (1-\varepsilon)q_{DC}}{1 - (1-2\varepsilon)(q_{CC} - q_{DC})}, \\ \tilde{q}_D &= \frac{\varepsilon q_{CD} + (1-\varepsilon)q_{DD}}{1 - (1-2\varepsilon)(q_{CD} - q_{DD})}. \end{aligned} \quad (2.109)$$

Similar to Lemma 2.7, we distinguish two cases, depending on whether or not this reactive marginalization is fully stochastic, that is whether or not $\tilde{\mathbf{q}} \in (0, 1)^2$.

- (i) $\tilde{\mathbf{q}}$ is fully stochastic. Because for reactive strategies $\mathbf{p} = (p_C, p_D)$ the maps $p_C \rightarrow \pi^\varepsilon(\mathbf{p}, \mathbf{q})$ and $p_D \rightarrow \pi^\varepsilon(\mathbf{p}, \mathbf{q})$ are either strictly monotonically increasing, decreasing, or constant (previous proof), it follows from $\tilde{\mathbf{q}} \in (0, 1)^2$ and the assumption (2.45) that $\pi^\varepsilon(\mathbf{p}, \mathbf{q})$ is constant for all reactive strategies \mathbf{p} . That is, $\phi^\varepsilon(\mathbf{q})$ needs to satisfy Eq. (2.33). By applying the back-transformation (2.36), it follows that \mathbf{q} needs to satisfy the conditions in Eq. (2.48).
- (ii) $\tilde{\mathbf{q}}$ is semi-stochastic or deterministic. In this case, either $\tilde{q}_C \in \{0, 1\}$ or $\tilde{q}_D \in \{0, 1\}$. By Eq. (2.109) this implies

$$\begin{aligned} \tilde{q}_C \in \{0, 1\} &\Leftrightarrow \mathbf{q} = (1, q_{CD}, 1, q_{DD}) \quad \text{or} \quad \mathbf{q} = (0, q_{CD}, 0, q_{DD}) \\ \tilde{q}_D \in \{0, 1\} &\Leftrightarrow \mathbf{q} = (q_{CC}, 1, q_{DC}, 1) \quad \text{or} \quad \mathbf{q} = (q_{CC}, 0, q_{DC}, 0). \end{aligned} \tag{2.110}$$

We discuss each of these four cases in turn:

- $\mathbf{q} = (1, q_{CD}, 1, q_{DD})$. In this case, we can use the payoff formulas (2.43) and (2.44) to verify that $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{TFT}, \mathbf{q})$ if and only if

$$q_{DD} \leq \frac{(1-2\varepsilon)(b+\varepsilon c q_{CD}) - c}{(1-2\varepsilon)(b+\varepsilon c)}. \tag{2.111}$$

On the other hand, an analogous computation shows that $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ALLC}, \mathbf{q})$ if and only if the inequality in Eq. (2.111) is reversed.

Together these two requirements imply the last condition in the characterization of partners (2.46). Finally, for strategies that satisfy

$q_{CC} = q_{DC} = 1$ and the last condition in (2.46), both additional requirements

$\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ALLD}, \mathbf{q})$ and $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ATFT}, \mathbf{q})$ are

met if and only if

$$q_{CD} \leq 1 - \frac{c}{(1-2\varepsilon)b}. \quad (2.112)$$

Overall, we conclude that such a strategy $\mathbf{q} = (1, q_{CD}, 1, q_{DD})$ is robust with respect to deviations towards the four deterministic reactive strategies if and only if the additional conditions in (2.46) hold. In particular, such strategies are generic Nash equilibria, because robustness against all deterministic reactive strategies implies robustness against all generic strategies by Propositions 2.9 and 2.10.

- $\mathbf{q} = (0, q_{CD}, 0, q_{DD})$. Using the payoff formulas (2.43) and (2.44) one can show that $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) < \pi^\varepsilon(\text{ALLD}, \mathbf{q})$ unless $q_{CD} = q_{DD} = 0$, that is $\mathbf{q} = \text{ALLD}$. We discuss the case of $\mathbf{q} = \text{ALLD}$ further below.
- $\mathbf{q} = (q_{CC}, 1, q_{DC}, 1)$. It is easy to show that $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) < \pi^\varepsilon(\text{ALLD}, \mathbf{q})$ for all such \mathbf{q} .
- $\mathbf{q} = (q_{CC}, 0, q_{DC}, 0)$. If \mathbf{q} is ALLD, then it is a Nash equilibrium (because defection is an equilibrium of the one-shot game). In the following let us thus assume that $q_{CC} > 0$ or $q_{DC} > 0$. In this case we obtain $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{TFT}, \mathbf{q})$ if and only if

$$q_{CC} \leq \frac{\varepsilon(1-2\varepsilon)c q_{DC} + c}{(1-2\varepsilon)(b+\varepsilon c)}. \quad (2.113)$$

On the other hand, the requirement $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ALLD}, \mathbf{q})$ is met if and only if the inequality in (2.113) is reversed. Together these two requirements imply the last condition in (2.47). Given this last condition and $q_{DD} = q_{CD} = 0$, it follows that $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ALLC}, \mathbf{q})$

and $\pi^\varepsilon(\mathbf{q}, \mathbf{q}) \geq \pi^\varepsilon(\text{ATFT}, \mathbf{q})$ if and only if

$$q_{DC} \leq \frac{c}{(1-2\varepsilon)b}. \quad (2.114)$$

We conclude that for strategies \mathbf{q} that satisfy all conditions in (2.47) there are no profitable deviations among the deterministic reactive strategies. Because of Propositions 2.9 and 2.10, this implies that there are no profitable deviations among the generic strategies, and hence \mathbf{q} is a generic Nash equilibrium.

(3) \Rightarrow (1) Follows immediately because partners and defectors are generic Nash equilibria by definition. Equalizers are generic Nash equilibria because any deviating player yields exactly the same payoff against an equalizer as the equalizer strategy yields against itself.

□

Trade has been going on for as long as humans have needed or wanted something that others had and they did not.³⁷

Mark Cartwright

3

A theory of specialization, exchange, and innovation in human groups

Abstract: To study how key aspects of cumulative cultural evolution interact, we model the coevolution of occupational specialization, exchange, and innovation. This allows for the study of which group conditions favor the innovation of occupational diversity in the long run. For example, our model predicts that human groups will innovate more occupational spe-

cializations in the long run when the population size is large, egalitarian norms enforce equal sharing, there is low noncomplementarity of specializations in the environment, junior people are moderately overconfident, and groups with redundant specializations begin to trade. Existing evidence from small-scale societies, large-scale societies, and the archaeological record provides preliminary corroboration of nine out of the model's ten predictions, and we propose future empirical tests for the remaining tenth prediction. Contemporary human societies' propensity for complex specialization and trade may thus not be a modern outlier, but an adaptation rooted in our ancestral past. This theory can help explain multiple evolutionary puzzles of human cognition, such as the prevalence of egalitarian sharing norms in evolutionarily relevant societies, the evolution of overconfidence, and cooperation in large groups of non-kin.

3.1 Introduction

Unlike other mammals, humans have long been characterized by flexible occupational specializations, extensive exchanges of goods/services, and cumulative cultural innovation. Examples include the cultural evolution of knowledge pertaining to fire-making, cooking, stone tools, adhesives, projectiles, ornamentation and novel food sources.

However, specialization, exchange, and innovation have all varied in their breadth and intensity across time and space for hundreds of thousands of years^{90,101,154}. Here, we develop an analytical model for the coemergence and coevolution of these key features of our species. We do so by integrating the fitness-maximizing decisions of individuals with the intergenerational transmission of specialized knowledge.

The theory of culture-gene coevolution offers an interconnected framework capable of rigorously explaining how innovation, specialization, and exchange have emerged and varied throughout the two-million-year history of our genus^{15,91,160,162}. Under

suitable initial conditions, cumulative cultural evolution can generate an autocatalytically expanding body of adaptive knowledge—tools, techniques, and practices—capable of driving the genetic-evolutionary expansion of brains to acquire, store, and organize this knowledge. Then, as brain expansion became too evolutionarily costly, cultural evolution began shaping the distribution of accumulated knowledge across individuals²¹⁷ and over the course of a lifetime²¹¹. Knowledge was first partitioned into a sex-based division of labor, and then into broader specializations, including those related to midwifery, medicinal plants, shamanism, storytelling, fishing, littoral gathering, big-game hunting, honey collecting, and making specialized tools such as watercraft. Parallel to this process, cumulative cultural evolution also generated social norms, a category of knowledge that has particularly important implications⁹³. Social norms influence the dynamics by which the group maintains and innovates occupational specializations as a collective brain. This occurs via the effect of social norms on cooperation, apprenticeship, exchange, and other relevant interactions.

We develop a theoretical framework for this overall process of cumulative cultural specialization. Mathematically, our model is inspired by the international-trade model of Krugman¹²⁵, lauded for identifying the interrelatedness of scale exchange economies and specialization in an analytically tractable manner. We adapt and modify Krugman’s model so that these scale exchange economies are characterized by cumulative cultural specialization. First, we incorporate into the model several mechanisms central to the process of cumulative cultural specialization, such as apprenticeship, specialized expertise, and social norms. Second, we modify Krugman’s static, unigenerational model into a dynamic, multigenerational process. The number of specializations accumulated in the group evolves over generations to an equilibrium number, which varies predictably with respect to factors like population size, egalitarianism, cogni-

tive biases, environmental parameters, and trade. We then deduce from our analytical model a number of predictions about cumulative cultural specialization. For example, our model predicts that human groups will innovate more specializations in the long run when they have a larger population size, when the diversity of an individual's resources results in higher fitness returns, and when groups with redundant specializations begin to trade.

In total, we deduce ten predictions, nine of which find preliminary empirical support in small-scale societies, large-scale societies, and the archaeological record. To illustrate, our model predicts that societies with larger population size will, over time, give rise to an increased variety of specialized activities. This prediction finds support in various sources of evidence, such as the Tasmanian archaeological record^{88,111}, ancient DNA evidence elucidating the emergence of modern-level toolkit complexity¹⁸⁹, research examining early-European-contact-era toolkit intricacy across Oceanic islands¹¹⁹, and an analysis of the toolkit sophistication in 40 nonindustrial agricultural and pastoralist societies⁴¹. Our model also predicts that larger societies will, as a result of their increased variety of specialized activities, tend to achieve higher average fitness. On this front, the contemporary studies of Redding and Venables¹⁹⁴ as well as of Head and Mayer⁸⁶ have shown that contemporary countries' income levels rise in correlation with a metric of "market potential"—a proxy for average group fitness—and that this metric itself exhibits a positive relationship with country size.

Moreover, our model predicts that (all else equal) egalitarian societies innovate more occupational specializations than their non-egalitarian counterparts. This counterintuitive prediction is corroborated by our analyses of the ethnographic data provided by the Ethnographic Atlas, abbreviated as the EA dataset^{12,22,76,118,121,158}; and the Western North American Indians, abbreviated as the WNAI dataset^{112-114,118}.

Another counterintuitive prediction of our model is that the average fitness of a non-egalitarian society is an inverse-U-shaped function of the confidence level of its juniors. This function is maximized not at the point of rational calibration, but at an intermediate level of overconfidence. A non-egalitarian group comprised of moderately overconfident innovators enjoys an equally optimal average fitness as that robustly enjoyed by egalitarian groups. This counterintuitive prediction finds support in the contemporary-society literature. Specifically, it is corroborated by the finding of Cieřlik et al.⁴⁰ that countries' Gross Domestic Product (GDP), a proxy for average group fitness, is predicted by an inverse-U-shaped function of their level of entrepreneurial overconfidence; and that this function is maximized at an intermediate level of overconfidence. Thus, our model may contribute to explaining why the degree to which people are overconfident about their task abilities varies substantially around the world¹⁶⁴.

Our model yields a number of counterintuitive, but empirically corroborated predictions about societal specialization and trade. This strengthens the case that specialization and trade can be productively studied via the theory of cumulative cultural evolution. In it, human cognition influences and is influenced by a causal interplay between specialized knowledge, apprenticeship, cooperative production, social norms, cognitive biases, the pursuit of individual fitness, and the environment.

3.2 The model

Our model represents human groups as amalgamations of two overlapping generations: seniors and juniors. These group members assort themselves into specialized guilds (e.g., hunting, fishing, gathering), each of which cooperatively produce the cor-

responding specialized good or service. While we call these cooperative production units “guilds,” they may not precisely map on to the cooperative units in traditional cultural groups that have historically been called guilds under various definitions. Our notion of “guild” is an abstraction that aims to capture the notion of cooperative production in an analytically tractable manner. As is standard for tractability, we approximate the number of potentially discoverable guilds with a continuum; let $i \in [0, \infty)$ index the set of potentially discoverable specializations.

There are two types of guilds: traditional guilds and emergent guilds. A traditional guild is led by experienced seniors who have the option of taking on junior apprentices. Seniors stipulate what share of the guild’s production juniors receive. An emerging guild is founded by juniors who have decided not to apprentice with any traditional guilds. Seniors offer apprenticeships, and juniors either decide whether to join or innovate a new guild. We consider new guilds to involve innovations.

After all juniors make their decisions, guild members produce their specialized goods or services. To produce an amount x of its good/service, an emerging guild is assumed to need $\ell = \alpha + \beta x$ members. Here, $\alpha > 0$ denotes the fixed cost and $\beta > 0$ denotes the per-unit cost. On the other hand, each traditional guild benefits from an intergenerationally accumulated body of knowledge that has been optimized for its specialization. So, a traditional guild only needs $\ell = \hat{\alpha} + \hat{\beta}x$ members to produce x , for a lower fixed cost $0 < \hat{\alpha} < \alpha$ and a lower per-unit cost $0 < \hat{\beta} < \beta$. This incorporates into the model, admittedly in an abstracted form, the effect of cumulative cultural evolution among guilds.

After production, each guild divides the fruits of its labors and expertise among its members in one of two ways. In egalitarian groups, every guild shares its production equally among all its members¹⁹. Here, we assume guilds simply apply egalitarian

social norms already common in their communities, such as those found among mobile foragers^{63,127}. In non-egalitarian groups, apprentices receive shares determined by seniors, who determine allocations so as to maximize their fitness, paying apprenticeships based on their initial promise. Here, we assume that some normative mechanism enforces promises within communities, but otherwise allocations are determined to maximize seniors' payoffs. Another aspect of the social norm we consider pertains to groupwide cognitive biases (or lack thereof). Specifically, we allow variation in the degree to which juniors are overconfident, underconfident, or rationally calibrated about their prospects of innovating versus apprenticing.

Finally, all group members partake in a system of exchange. Goods are exchanged back-and-forth until no mutually beneficial exchange is possible anymore. Each individual is assumed to maximize fitness as defined by the utility function

$$U(c) = \int_{i \in [0, \infty)} c(i)^\theta di \quad \text{for } 0 < \theta < 1. \quad (3.1)$$

Originally developed by Dixit and Stiglitz⁵⁰, this utility function represents a preference for diverse bundles of goods c when all else is equal. This diversity benefit can occur due to decreasing marginal utility of goods and due to complementarity. The degree of the diversity benefit is represented by the exponent parameter $0 < \theta < 1$, which represents the noncomplementarity of specializations. A decrease in θ constitutes an increase in the fitness returns to the diversity of an individual's resources. For example, consider an environment where available sources of food required a specialized guild's cooperation to forage. If each of these food sources provided all of the necessary nutrients and thus were interchangeable with one another, then the degree of specialization noncomplementarity θ would be high. But if the food sources pro-

vided different subsets of nutrients and thus were ideally consumed as a bundle, then the degree of specialization noncomplementarity θ would be low.

In its original economic formulation, the utility function U represents an individual's preference for which bundle of goods to prefer over another given bundle of goods. In economics, no assumptions are generally made about the causal determination of the preference. In our culture-gene coevolutionary formulation, we propose that the preference underlying the utility function U is caused by some combination of an individual's adaptive preference—the degree to which their preference is self-perceived as successful and adaptive in the local environment—the individual's cultural fitness—the degree to which their preference is outwardly perceived as successful and thereby likely to be emulated—and the individual's genetic fitness—the degree to which their preference is genetically evolved to prefer certain bundles of goods over others for robust benefits to survival and reproduction. Specifying how these causal factors precisely combine is unnecessary for predicting the outcome of our model, which only requires that their combination robustly causes surviving group members to adhere to the ecologically optimal utility function U .

The final outcome of the current time step is uniquely determined by the following two assumptions.

Assumption 1: Each individual exchanges to maximize fitness.

Thus, after an individual's net exchange, their terminal bundle of goods c^* is assumed to solve the fitness maximization program

$$\max_c U(c) \quad \text{subject to} \quad \int_{i \in [0, \infty)} p(i)c(i)di = W. \quad (3.2)$$

Here, W denotes the trade value of the goods with which the individual starts the exchanging phase and $p(i)$ denotes the price of good i (relative to other goods, so that the choice of unit is arbitrary).

Assumption 2: Each guild aims to achieve its optimal size by either taking on co-founders or apprentices.

To illustrate for egalitarian groups, consider the decision of seniors in traditional guilds on how many apprentices to take on. They are assumed to solve the optimal guild-size program for traditional guilds that are forced to share equally:

$$\text{maximize } \frac{px(p)}{\hat{\alpha} + \hat{\beta}x(p)} \quad \text{subject to } p > 0 \quad (3.3)$$

The optimal guild size for the maximization problem (3.3) is

$$\ell^* = \frac{\hat{\alpha}}{1 - \theta}. \quad (3.4)$$

Similarly, when seniors of traditional guilds in non-egalitarian groups decide how many apprentices to take on, they are assumed to solve the optimal guild-size program for traditional guilds that promise apprentices the exact share of goods that would make them indifferent between apprenticing and innovating:

$$\text{maximize } \frac{px(p) - w^{\text{inn}}(\hat{\alpha} + \hat{\beta}x(p) - \ell^{\text{trad}}(i))}{\ell^{\text{trad}}(i)} \quad \text{subject to } p > 0. \quad (3.5)$$

Here, $\ell^{\text{trad}}(i)$ denotes the number of guild i 's seniors who have survived from the previous time step, and w^{inn} denotes the trade value of an equal share of an emerging

guild's produced goods.

The optimal guild size for the maximization problem (3.5) is

$$\ell^* = \hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1}. \quad (3.6)$$

In summary, each individual's net exchange maximizes their fitness, and in anticipation of this, each guild achieves (either via taking on apprentices or co-founders) the precise size that optimizes the trade value of each current guild member's share of produced goods. The resulting allocation of goods is called the equilibrium outcome of the time step.

The time steps occur sequentially. At the end of each time step, the senior individuals pass away. Some fortunate junior individuals survive to become the senior individuals of the next time step. A proportion $q_{\text{trad}} \in [0, 1]$ of traditional builds and a proportion $q_{\text{emer}} \in [0, 1]$ of emergent guilds lose all of their members and thereby disappear; these specializations are lost would need to be rediscovered in the future. Guilds with any survivors remaining all become traditional guilds. The next time step is then carried out just like the previous time step, and the overall process is repeated a countably infinite number of times.

The above model represents a single group as a closed system. Another version of our model represents a two-group closed system, in which members of each of the two groups have the option of exchanging with the other group. Such inter-group exchanges may be subject to an inefficiency. We assume that $\tau \geq 1$ units of a good must be sent by one group for the other group to receive one unit of that good. The parameter τ denotes the trade barrier, where $\tau = 1$ represents a setting where no goods are lost after passing the trade barrier, and $\tau > 1$ represents a setting where a pro-

portion of goods is lost after passing it. The notion of a trade barrier can represent various forms of trade inefficiency, such as a linguistic trade barrier between neighboring ethnolinguistic groups or subcultures, as well as geographic distance and obstacles.

See Figure 1 for a diagram representing the model step-by-step.

3.3 Results

For the sake of presentation, we will initially suppress three features of the model. First, we will assume—unless otherwise stated—that no specializations are forgotten at the end of each time step due to all members passing away. This assumption becomes tenable when guilds are comprised of many members and the probability of each member’s death in each time step is low and sufficiently uncorrelated with other members’ probability of death. The assumption becomes less tenable, however, when guilds are comprised of few members, when the probability of each member’s death in each time step is high, and when members’ death events are correlated¹⁵. We introduce the possibility of specializations being forgotten (due to all guild members dying) in Subsection 3.3.4.

Second, we assume that the population size of the group converges to a limit of L as the number of time steps approaches infinity. For brevity of exposition, we will treat the limit of the population size L as an exogenous parameter throughout the main text, in that we do not assume that other mechanisms of the model affect L . However, endogenous models of population growth are more realistic^{123,199}. To illustrate, if the group’s average fitness at time step t represents the absolute average number of surviving offspring—not relative to a carrying capacity—then this average fitness value would affect the change in population size at that time step. Our model

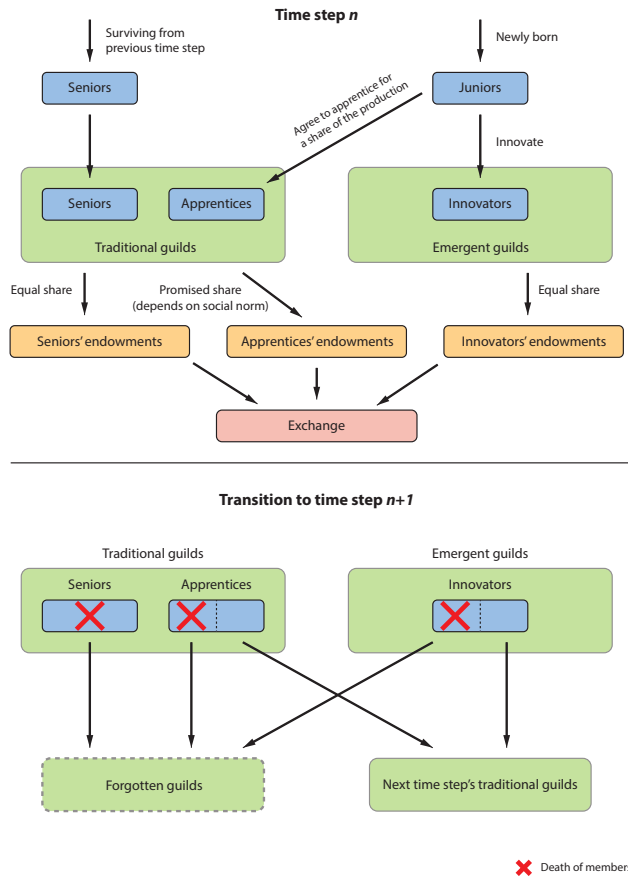


Figure 3.1. Life cycle diagram of the model. It is comprised of two overlapping generations: seniors and juniors. Guilds can be traditional, led by experienced seniors; or emergent, innovated by juniors who have decided not to apprentice with any traditional guilds. On the other hand, a junior can join a traditional guild as an apprentice after being promised a share of its future production. After production and sharing within each guild, individuals exchange goods to maximize their fitness. Afterwards, the transition to the next time step occurs, in which all seniors and some juniors die, the surviving juniors become the next time step's seniors, and their specializations survive onto the next time step in the form of traditional guilds. The specializations of the guilds in which all members die are forgotten.

can also be modified so that the population size $L(t)$ at each time step t is causally affected by not only the average fitness, but also by other parameters and mechanisms of the model. However, we presently do not assume or specify the functional form of this complex effect, and propose that it be investigated empirically. In the meantime, we abstract all endogenous demographic effects into the long-run limit of population size L , and deduce predictions based on a general, unspecified value of L .

Third, we will assume that descriptive parameters are shared across individuals, across emerging guilds, and across traditional guilds. We assume all individuals share the same fitness-utility function U . We also assume that all discoverable guilds share the same cost structure in the emerging phase, as well as share the same cost structure in the traditional phase. In principle, we can modify our model so that different types of individuals have different fitness-utility functions. We can also modify it so that different types are predisposed to work in different types of guilds, each with a different cost structure in the emerging phase, traditional phase, or both. Such modifications would allow our model to precisely represent—albeit potentially via overfitting—societies with heterogeneous guild sizes (among emerging guilds and among traditional guilds) as well as heterogeneous amounts of goods exchanged and consumed by each individual. In the present paper, we focus on the simple variant of the model without these heterogeneities, with the goal of deducing parsimonious, potentially generalizable causal relationships in an analytically tractable manner at the potential cost of model precision.

In the following, we derive from our model ten causal predictions that can help explain various evolutionary puzzles surrounding the emergence, development, and adaptive strategies of ancestral human societies, as well as analogous empirical puzzles pertaining to contemporary human societies. For brevity of exposition, we present

some mathematical predictions in the SI rather than in the main text. We note that our list of predictions is not meant to be exhaustive. It is very possible that other mathematical predictions we did not deduce or specify in the present paper may turn out to be explanatorily valuable and empirically corroborated.

3.3.1 Number of specializations

In the limit of the time step $t \rightarrow \infty$, the proportion of guilds that are traditional rather than innovated approaches 1. It follows that as $t \rightarrow \infty$, the limiting number of traditional guilds—which is equal to the limiting number of specialized guilds overall—is given by the population size divided by the equilibrium size of traditional guilds ℓ^* ,

$$n = \frac{L}{\ell^*}. \quad (3.7)$$

Moreover, the Dixit–Stiglitz utility function U has the nice property that the traditional guild size ℓ^* does not depend on the population size L . It can be shown that the traditional guild size ℓ^* of egalitarian groups is given by

$$\ell^* = \frac{\hat{\alpha}}{1 - \theta}, \quad (3.8)$$

a formula that is unaffected by the population size L . Similarly, the traditional guild size ℓ^* of non-egalitarian groups is given by

$$\ell^* = \hat{\alpha} + \frac{\alpha\theta}{1 - \theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta} - 1}, \quad (3.9)$$

which is also unaffected by the population size L .

While the equilibrium size of traditional guilds ℓ^* does not depend on the population size L , it does depend on the allocation of payoffs made by senior to juniors. Here, we explore two possible allocations by comparing the long-run number of specializations in egalitarian groups versus in non-egalitarian groups.

For non-egalitarian groups, where seniors are maximizing their fitness, the number n of specializations in the long run is given by

$$n = \frac{L}{\ell^*} = \frac{L}{\hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}}\right)^{\frac{1}{1-\theta}-1}} \quad (3.10)$$

In egalitarian groups, where seniors are constrained by social norms, the number n of specializations in the long run is given by

$$n = \frac{L}{\ell^*} = \frac{L(1-\theta)}{\hat{\alpha}} \quad (3.11)$$

Observe that the long-run number of specializations n scales linearly in the population size L . This is because as we have stated before, the denominator ℓ^* of the formula $n = \frac{L}{\ell^*}$ does not have any dependence on L ; it is essentially a positive constant. We can thus deduce the following corollary.

Prediction 1: Groups with a larger population size L have a higher number n of innovated specializations in the long run.

Moreover, observe that the traditional guild size ℓ^* of egalitarian groups (3.8) and that of non-egalitarian groups (3.9) increase with θ , the parameter presenting the non-complementarity of specializations in the environment. Intuitively, guilds choose to

be at the size at which the benefit from efficient economies-of-scale and the benefit from uncovering new specializations are at equilibrium. The latter benefit becomes larger relative to the former when a diverse bundle of specialized goods yields large benefits to fitness-utility. This explains why having non-complementary specializations causes the equilibrium size of traditional guilds (in both egalitarian and non-egalitarian groups) to increase; and thereby causes the long-run number of specializations in the group, $n = L/\ell^*$, to decrease. We thus have the following result.

Prediction 2: The long-run number n of specializations is higher when the degree θ of specialization noncomplementarity in the environment is low.

Predictions 1 and 2 are illustrated in the plots of Figure 3.2, which detail the comparative-static effects of several model parameters on the long-run number of innovations.

3.3.2 Group members' average fitness-utility in the long run

Egalitarian norms foster greater innovation and specialization. To see this, consider that as $t \rightarrow \infty$, group members converge towards achieving an equidistributed bundle c^* of goods that is comprised of an equal amount of each specialized good produced by a traditional guild. In other words, the limit of the average group member's equilibrium bundle c^* can be computed in the following way. First, all group members first produce in traditional guilds of size $\ell = \ell^*$. Second, all group members achieve an equidistributed bundle of goods c^* during the exchanging phase. The group members' average utility is then given by $U(c^*)$ for this equidistributed bundle c^* of goods.

Note that the above calculation of the limiting average bundle can be done for any

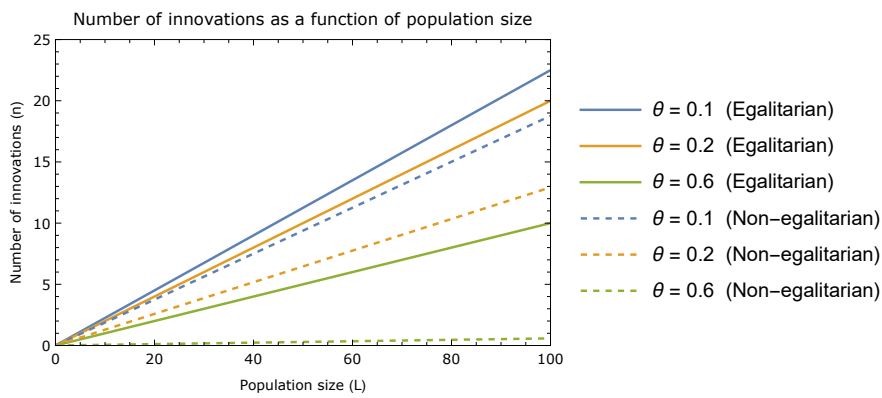


Figure 3.2. The number of innovations (n) as a function of population size (L). The cost parameters are $\alpha = 10$, $\beta = 10$, $\hat{\alpha} = 4$, and $\hat{\beta} = 2$, while the degree of specialization noncomplementarity θ is varied among $\theta = 0.1, 0.2$, and 0.6 .

choice of guild size ℓ , meaning that we can study the average limiting fitness $U(c^*)$ as a function of an arbitrary guild size ℓ . Studying this, we find that the average limiting fitness $U(c^*)$ is an increasing function for $\ell < \frac{\hat{\alpha}}{1-\theta}$, a decreasing function for $\ell > \frac{\hat{\alpha}}{1-\theta}$, and is maximized at $\ell = \frac{\hat{\alpha}}{1-\theta}$.

But recall that $\ell^* = \frac{\hat{\alpha}}{1-\theta}$ is the equilibrium size of traditional guilds in egalitarian groups. This yields a counterintuitive result that egalitarian groups robustly innovate the optimal number of specializations in the long run. In non-egalitarian groups, the equilibrium size

$$\ell^* = \hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} > \frac{\hat{\alpha}}{1-\theta} \quad (3.12)$$

of traditional guilds is too large compared to the optimal level. Intuitively, this is because non-egalitarian social norms permit seniors to underpay their apprentices. In equilibrium, apprentices are underpaid to the precise point where juniors would be indifferent between apprenticing and innovating. In order to maximize their self-interest, seniors in non-egalitarian groups are incentivized to take on many underpaid apprentices, which allows them to optimally take advantage of economies-of-scale. This is in contrast to egalitarian groups, where the comparatively large equal share that would need to be paid to each apprentice incentivizes seniors to make do with only a few apprentices. All else equal, more juniors are left to innovate (due to less apprenticing opportunities) in egalitarian groups, resulting in more groupwide innovation in the long run.

Prediction 3: Egalitarian groups robustly innovate the optimal number of specializations n in the long run. Non-egalitarian groups innovate a suboptimally low number of specializations n in the long run..

The contrast between the optimal innovation of egalitarian groups and the suboptimally low innovation of non-egalitarian groups is qualitatively illustrated in Figure 3.3(a).

3.3.3 The effect of groupwide overconfidence on innovation

Non-egalitarian groups are subject to a permanent penalty to the equilibrium amount of innovated specialization, which is suboptimally low compared to that of egalitarian groups. But counterintuitively, a social norm of overconfidence can offset non-egalitarian groups' long-term penalty to groupwide innovation and thereby, to group fitness.

Specifically, let us modify our model so that the group's juniors believe that the fixed cost and the per-unit cost of an emerging guild are not necessarily equal to their true respective values. In this interpretation, $\alpha_{\text{perceived}}$ and $\beta_{\text{perceived}}$ denote the perceived fixed cost and per-unit cost, in the perception of juniors. We distinguish this from $\alpha = \alpha_{\text{true}}$ and $\beta = \beta_{\text{true}}$, the true values of these respective parameters.

The size ℓ^* of traditional guilds in non-egalitarian groups changes when changing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$. It changes from

$$\hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} \quad (3.13)$$

to

$$\ell^* = \hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1}. \quad (3.14)$$

On the other hand, the size of traditional guilds in egalitarian groups is situation-

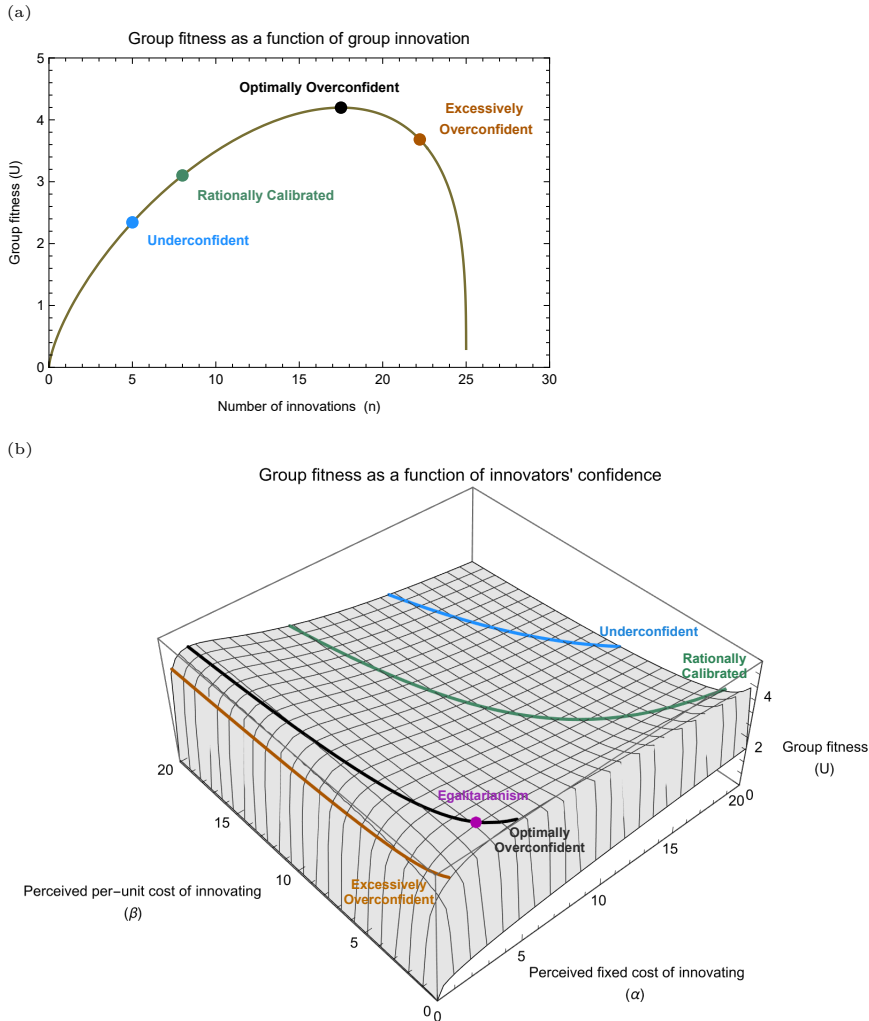


Figure 3.3. Plots pertaining to the effects of overconfidence on group innovation and fitness. For a non-egalitarian group with environmental parameters $\theta = 0.3$, $\hat{\alpha} = 4$, and $\hat{\beta} = 2$, we present (a) group members' average long-run fitness as a function of the group's long-run number of innovations and (b) as a function of the confidence norm of the group's junior innovators. The fitness of the group is maximized at an intermediate level of overconfidence (dark green curve). Rational confidence (green curve) cause the group to innovate a suboptimally low number of specializations, and this is made even more suboptimally low by a social norm of underconfidence (blue curve). Excessive overconfidence (orange curve) begins to cause decreasing group fitness, due to an overpayment of the per-guild fixed cost. In contrast, the fitness of an egalitarian group (purple point) is robustly maximal, because the group robustly innovates the optimal number of specializations regardless of small changes in confidence norms.

ally determined. Apprenticing is strictly better than innovating for juniors if and only if

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta < \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.15)$$

If this inequality holds, then the equilibrium size ℓ^* is unchanged; it is given by

$$\ell^* = \frac{\hat{\alpha}}{1-\theta}. \quad (3.16)$$

If this inequality holds in the other direction, however, traditional guilds must pay strictly higher than an equal share to incentivize juniors to apprentice. They must pay the precise share at which juniors are indifferent between apprenticing and innovating, just like the traditional guilds of non-egalitarian groups. In this case, the equilibrium size ℓ^* of traditional guilds in egalitarian groups is given by (3.14), just like that of non-egalitarian groups.

Recall that in the limit, the proportion of guilds that are traditional rather than innovated increases to 1. Thus, the group's limiting average fitness is a function of the equilibrium size ℓ^* of traditional guilds.

In non-egalitarian groups, the equilibrium size ℓ^* of traditional guilds granularly changes with respect to $\alpha_{\text{perceived}}$ and $\beta_{\text{perceived}}$. This means that the group's long-term limiting average fitness changes as well. As either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ increases, the equilibrium size ℓ^* of traditional guilds increases, because apprentices are less expensive to hire. Thus, the amount of innovated specialization that happens in the long run decreases, making it more suboptimally low than before.

On the other hand, as either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ decreases, the equilibrium size ℓ^* of traditional guilds decreases, because apprentices are more expensive to hire. Thus, the amount of innovated specialization that happens in the long run increases.

The region of increase can be divided into two parts:

$$\hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} > \frac{\hat{\alpha}}{1-\theta} \quad (3.17)$$

and

$$\hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} < \frac{\hat{\alpha}}{1-\theta}. \quad (3.18)$$

In the first region (3.17), decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ results in the group's long-term average fitness increasing, and thereby becoming closer to the optimal value. In the second region (3.18), decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ results in the group's long-term average fitness decreasing, and thereby becoming further from the optimal value. In between these two regions, the optimal long-term average fitness is achieved.

Prediction 4: Among non-egalitarian groups, the amount of long-run innovated specialization (and thereby, group fitness) is an inverse-U-shaped function of innovators' overconfidence. The optimal level of overconfidence occurs when juniors are perfectly indifferent between apprenticing and innovating.

This can be seen in Figure 3.3, which illustrates that a non-egalitarian group with the optimal level of overconfidence yields the same average group fitness, all else equal, as the rationally calibrated egalitarian group.

The group-fitness-optimality of egalitarianism is robust, in that the equilibrium size ℓ^* of traditional guilds in egalitarian groups does not easily change from the optimal value. Suppose that either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ increases. The region of increase can

be divided into two parts:

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta < \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.19)$$

and

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta > \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.20)$$

In the first region (3.19), we have seen that changing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ does not change the equilibrium guild size ℓ^* . This is mathematically equivalent to the fact that the optimality of egalitarianism is robust to changes in the real fixed cost α_{real} and real per-unit cost β_{real} when assuming the juniors are rational, since the long-run formula for the number of specializations n does not change whether α and β denote the perceived cost values or the real cost values. Overall, the fact that egalitarianism achieves the optimal amount of innovated specialization is robust to a wide range of confidence parameter values.

Prediction 5: Egalitarian groups are more likely to have their number of long-run innovated specialization (and thereby, group fitness) be unaffected by changes in innovators' confidence norms, in the fixed cost of emerging guilds, and in the per-unit cost of emerging guilds than their non-egalitarian counterparts.

On the other hand, decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ to the point of reaching the second region (3.20) does change the equilibrium guild size ℓ^* . This is because in this region, a traditional guild's equal share is insufficient to convince juniors to apprentice rather than innovate. Thus, traditional guilds must pay juniors the precise share at which they are indifferent between apprenticing and innovating. This leads

the equilibrium size ℓ^* of traditional guilds to suboptimally increase to (3.14), same as the unconditional equilibrium size of traditional guilds in non-egalitarian groups. This makes the long-term degree of specialized innovation to be suboptimally high, where the suboptimality is due to an excessive loss of productivity from the fixed cost requisite for each specialized guild.

3.3.4 When innovated specializations are lost

Many ancestral humans likely died prematurely due to factors like disease, starvation, environmental accidents, predation, and warfare. In between time steps, some junior individuals pass away, and the remaining ones comprise the next time step's senior generation. The guilds with at least some members surviving become traditional guilds, indicating that their specialized methods of production are successfully passed down intergenerationally. The guilds with no members surviving do not become traditional guilds, meaning they must be innovated again in the future.

In a finite guild with k junior members and a probability r of each junior's death, the guild would disappear with a probability of r^k . But recall that our model is an infinite, analytically tractable approximation of a discrete reality where finitely many guilds each have finitely many members. We do not specify an approximation process by which a sequence of increasingly fine-grained discrete models—one of which is the proposed model of the real system—becomes our infinite, analytically tractable model, such as the approximation process described in Park¹⁷⁸ for knowledge-based learning strategies.

However, let us make the realistic assumption that in this approximation process, the probability of a traditional guild's death q_{trad} and the probability of a probability

of an emerging guild's death q_{emer} are well-defined. Recall that emerging guilds are comprised entirely of juniors, and are larger than traditional guilds in equilibrium. Thus, the number of juniors in an emerging guild will tend to be higher than that in a traditional guild. Mathematically, we can without loss of generality suppose that the random variable of the former number first-order stochastically dominates that of the latter number. It then follows that the limiting probability of an emerging guild's death, q_{emer} , is less than that of a traditional guild's death, q_{trad} . We do not make any additional assumptions or specifications of q_{emer} and q_{trad} , and propose that they be empirically studied.

Regardless, we can still compute the long-run number of specializations as a function of the unspecified probability values q_{emer} and q_{trad} . Let n_{emer} and n_{trad} denote the long-run number of emerging guilds and that of traditional guilds. Assuming that the system is in equilibrium, the values of n_{emer} and n_{trad} are uniquely determined by the system of equations

$$q_{\text{emer}}n_{\text{emer}} + q_{\text{trad}}n_{\text{trad}} = n_{\text{emer}} \quad (3.21)$$

and

$$\ell_{\text{emer}}^*n_{\text{emer}} + \ell_{\text{trad}}^*n_{\text{trad}} = L. \quad (3.22)$$

Here, the first equation (3.21) denotes the condition that the totality of the forgotten guilds in a given time step are replaced in equilibrium by innovating the same number of emerging guilds in the following time step. It is equivalent to

$$n_{\text{emer}} = \frac{q_{\text{trad}}n_{\text{trad}}}{1 - q_{\text{emer}}} \quad (3.23)$$

The second equation (3.22) denotes the condition that the limit of the total number of people across all guilds must be L . Substituting in (3.23), we obtain

$$n_{\text{trad}} = \frac{L}{\ell_{\text{trad}}^* + \ell_{\text{emer}}^* \frac{q_{\text{trad}}}{1 - q_{\text{emer}}}}. \quad (3.24)$$

Notice that the expression (3.24) is increasing in q_{emer} . It is also increasing in q_{trad} . This means that if a higher proportion of either emerging guilds or traditional guilds are forgotten due to all member dying, then the limiting proportion of traditional guilds decreases. Since $\ell_{\text{trad}}^* < \ell_{\text{emer}}^*$, it follows that the limiting number of guilds also decreases. Stated in another way, since more of the guilds are emerging guilds in the long run, and emerging guilds take up more members per guild, the long-run number of guilds—of specializations—decreases. We thus have the following prediction:

Prediction 6: Groups with a higher rate of premature deaths sustain a smaller number of specializations in the long run.

Premature deaths cause a higher proportion of a group's guilds to stay at the initially inefficient level of production rather than the efficient level that comes with experience and tradition. This inefficiency results in a lower fitness value for the average group member: a lower group fitness. This is qualitatively illustrated in Figure 3.4(c).

3.3.5 Inter-group trade

Finally, we discuss the predictions of our two-group model. Our contributions for the two-group case include incorporating multigenerational dynamics and cumulative cultural evolution into the model, demonstrating that the original economic model's

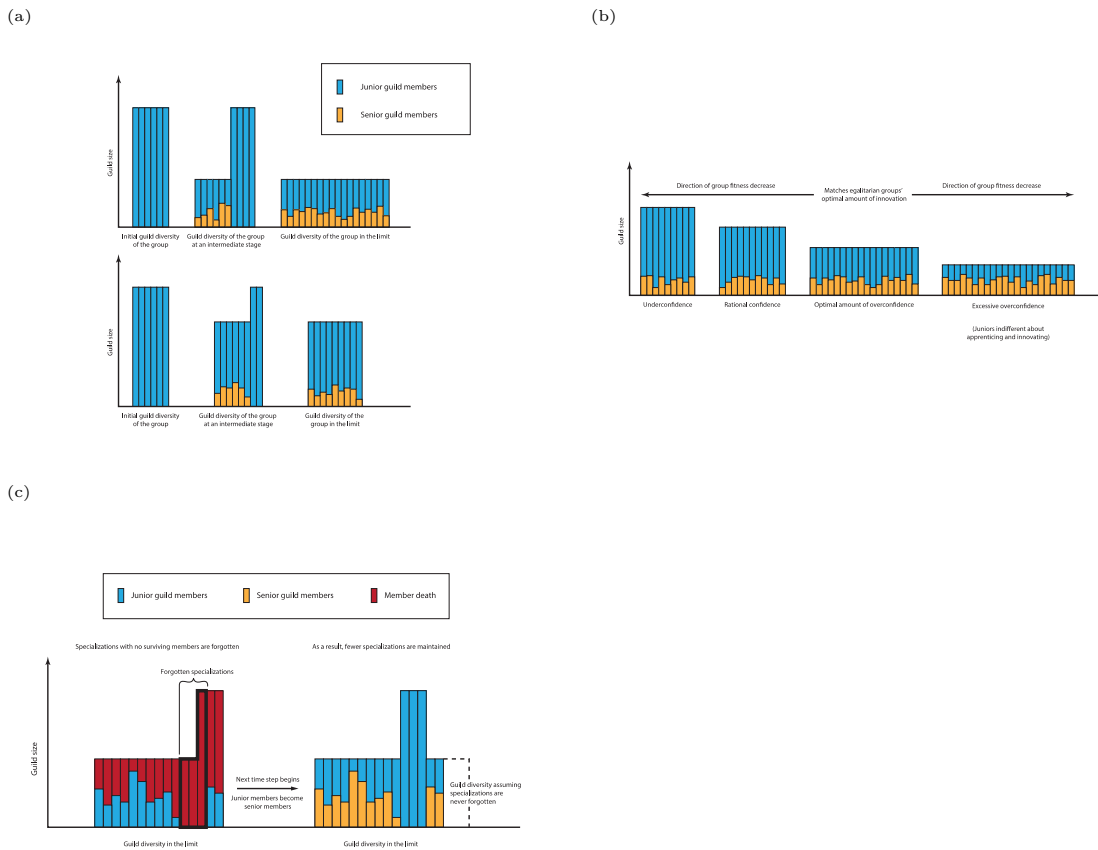


Figure 3.4. Qualitative visualizations of the single-group model's dynamics. The visualization (a) contrasts the optimal innovation of an egalitarian group (top) with the suboptimally low innovation, all else equal, of a non-egalitarian group (bottom). This is due to the suboptimally large size of traditional guilds in the latter, and the suboptimally high number of juniors that apprentice rather than innovate as a result. The visualization (b) shows that a non-egalitarian group's total number of innovated specializations changes based on the level of juniors' overconfidence in innovating versus apprenticing. There exists an optimal level of overconfidence that allows it to match the optimal number of innovated specializations, all else equal, of egalitarian groups. The visualization (c) shows that if there is a positive probability that all members of a guild die in a given time step, then this reduces the long-run number of innovated specializations.

trade predictions are robust to this incorporation, and connecting the model to the trade of ancestral societies in addition to that of contemporary ones.

Our first result is that two groups that engage in trade will specialize in mutually exclusive niches, whereas their specializations may overlap if they do not engage in trade. In other words, phenotypic specialization at the group level will tend to increase as groups come into contact with each other for trade. We note that this result predicts the opposite of what is predicted by a solely memetic evolutionary model of phenotype, in which a group becomes more similar to another group when the two make contact. In fact, trade contact can have a differentiating effect, rather than a homogenizing effect, on societies' culturally transmitted phenotypes⁹¹.

Prediction 7: Trade increases specialization.

One interesting aspect of this effect any nonzero amount of trade contact is sufficient to incentivize the two groups to specialize in non-overlapping niches. The two groups' specializations can overlap only if they engage in no trade. This all-or-nothing effect is qualitatively illustrated in Figure 5(a). The all-or-nothing nature of this effect is an artifact of the model assumptions, and may not be realistic. However, our model shows the possibility that the differentiating effect of trade contact on societies' areas of specializations may have a discrete effect akin to an activation function, rather than a continuous effect akin to a linear or similar function.

Our second result is that both groups gain from trade. More generally, a decrease in the trade barrier causes a strict increase in the fitness values of both groups' members. This occurs because easier trade with the other group enables an individual to take advantage of a wider variety of exchangeable goods, which is more preferable

and adaptive due to our assumption that a diversity of resources results in increasing returns to fitness. The mutually beneficial effect of decreasing trade barriers may help explain the ubiquity of mechanisms to reduce various types of trade barriers in contemporary and plausibly ancient societies³⁰, such as exogamy, shared rituals, currency, merchants, trade roads, and free trade agreements.

Prediction 8: Groups benefit from trade. A decrease in the trade barrier causes an increase in the fitness of both groups' members.

An example of this effect is illustrated in Figure 3.5(b), which plots the average group fitness of two groups (sharing certain environmental parameters) that trade across a barrier of varying degree τ .

Our third result pertains to the trade, all else equal, between a group with a smaller population L_1 , denoted by Group 1; and a group with a larger population L_2 , denoted by Group 2. First, the smaller group exchanges a higher proportion of its produced goods to the larger group than the other way around. This can be intuitively be seen by considering the case of no trade barrier $\tau = 1$. In this case, the two groups trade as if they comprised one overarching group. The trade of one overarching group would be such that every member gets an equidistributed bundle of goods. This would mean that every individual would trade $L_1/(L_1+L_2)$ of their starting goods to Group 1 members and $L_2/(L_1 + L_2)$ of their starting goods to Group 2 members, regardless of the group in which they reside. In other words, the less populous Group 1 would export a higher proportion of its starting goods to the more populous Group 2.

Moreover, the difference in export proportions becomes even higher in the remaining case of a positive trade barrier $\tau > 1$, because the less populous group must pay

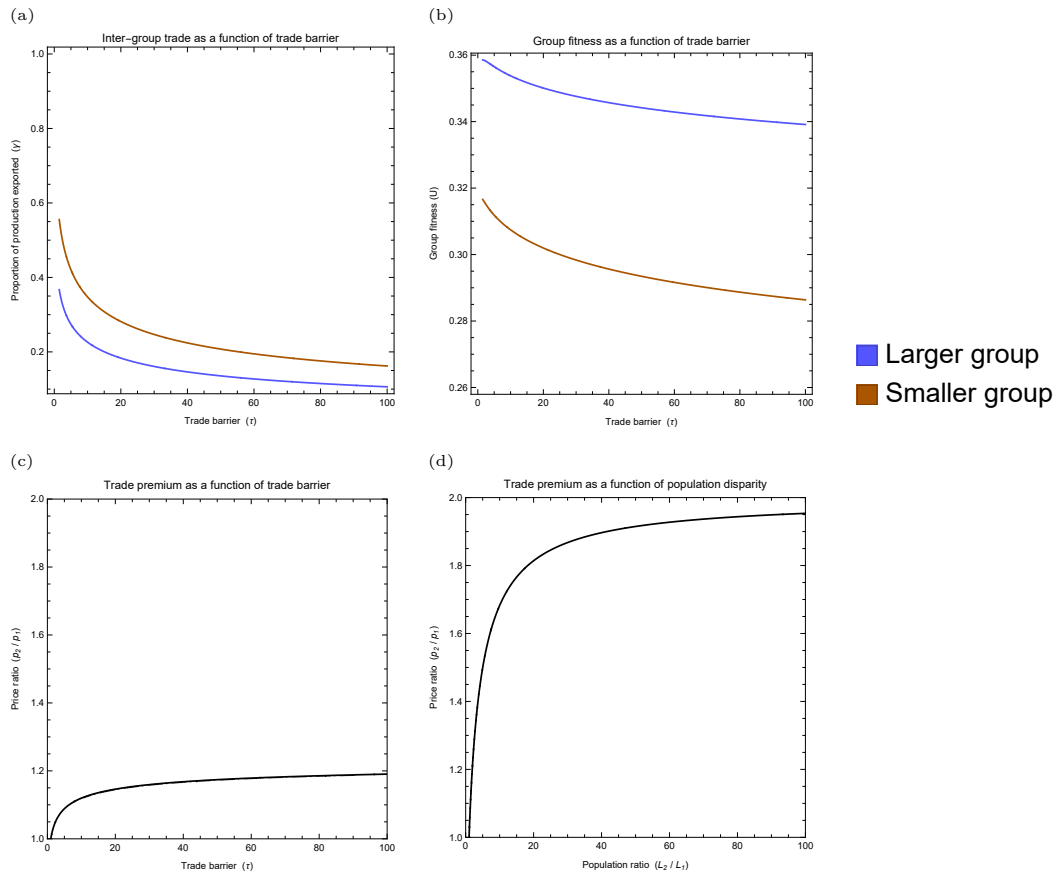


Figure 3.5. Plots pertaining to the trade between Group 1 and Group 2. They are assumed to share the environmental parameters $\theta = 0.3$, $\alpha = 10$, $\beta = 10$, $\hat{\alpha} = 4$, and $\hat{\beta} = 2$. Our plots comprise (a) the proportion of produced goods that each group exports as a function of the trade barrier size, (b) average member fitness of each group as a function of the trade barrier size, (c) the trade premium (defined as the relative price of the more populous Group 2's good to that of the less populous Group 1's good) as a function of the trade barrier size, and (d) the trade premium as a function of the population ratio between the two groups. The base parameters are $\tau = 10$ and $L_2/L_1 = 1.5$. In each plot, exactly one of the parameters is varied from its base value.

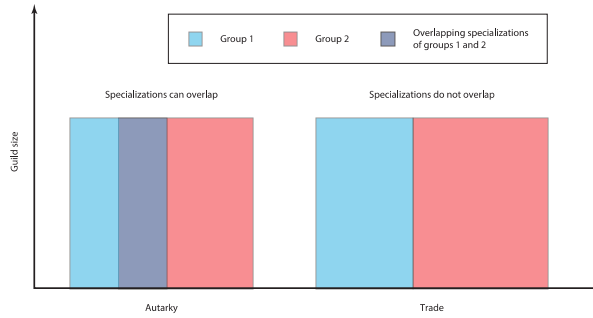
a higher price than normal to trade with the more populous group. Intuitively, this is due to the fact that the opportunity cost is higher for the larger group's goods. Specifically, an individual in the larger group can exchange their initial goods for a large variety of goods produced by their fellow group members, instead of trading it for the other group's goods. On the other hand, an individual in the smaller group can exchange their initial goods only for a small variety of goods produced by their fellow group members, which is a worse deal. Thus, individuals of the larger group can ask for a higher price relative to those of the smaller group. It follows that the less populous Group 1 exports a proportion that is even higher than $L_2/(L_1 + L_2)$ of its starting goods to the more populous Group 2 in the presence of a trade barrier $\tau > 1$.

Prediction 9: Consider the trade between a larger group and a smaller group. All else equal, the smaller group exports a higher proportion of its produced goods to the larger group.

An example of this effect is illustrated in Figure 3.5(a), which plots the average proportion of goods exported when two groups (sharing certain environmental parameters) trade across a barrier of varying degree τ . Indeed, the plot shows that the smaller group exports a higher proportion of its produced goods than the larger group. This effect is also shown in a visualization presented in Figure 3.6(b).

Our final prediction pertains to the trade premium between a larger group and the smaller group. The trade premium is defined as the price ratio (i.e., exchange rate) between a good produced by the larger group and a good produced by the smaller group. Members of the smaller group need to pay a trade premium in order to trade with members of the larger group. Intuitively, this is because the option of trading

(a)



(b)

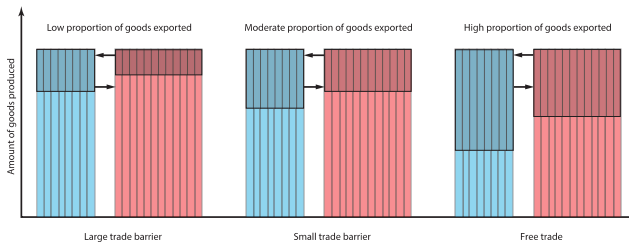


Figure 3.6. Qualitative visualizations of the two-group model's trade dynamics. The visualization (a) denotes the phenomenon in which two groups that do not trade can innovate overlapping specializations, but two groups that engage in trade intentionally specialize away from each other. The visualization (b) illustrates how much of each group's goods are exchanged across groups. The smaller group exports a higher proportion of its produced goods than the larger group. Moreover, increasing the trade barrier results in the increase of both groups' proportion of goods exported.

with other in-group members is more favorable to members of the larger group than to those of the smaller group. The difference in opportunity cost forces members of the smaller group to pay a premium to sufficiently incentivize members of the larger group to trade with them.

Prediction 10: Consider the trade between a larger group and a smaller group. If the trade barrier is nontrivial, then the smaller group must pay a premium to trade with the larger group. The trade premium is increasing with respect to the trade barrier, and increasing with respect to the population ratio of the two groups.

An example of this effect is illustrated in Figure 3.5(c), which plots the trade premium when two groups (sharing certain environmental parameters) have a varying trade barrier; and in Figure 3.5(d), which plots the trade premium when the two groups have a varying population ratio.

Due to the complicated formulas involved, we have limited this section to a verbal discussion of four predictions. A more detailed, quantitative discussion—as well as additional predictions—can be found in the SI.

3.3.6 Empirical corroboration

Above, we have deduced from our model ten predictions. Preliminary empirical corroboration for nine of the ten predictions is summarized in Tables 3.1 and 3.2, and discussed in further detail within the SI. The former pertains to predictions from the single-group model and the latter, those from the two-group model.

Table 3.1: Testable predictions for an exchange economy comprised of a single group.

Prediction	Empirical Evidence
Prediction 1: Groups with a larger population size have a higher number of innovated specializations in the long run.	EA ethnic groups with a larger population tend to have a greater number of specialized activities (see Subsection 3.6.1 of the SI). This prediction is also corroborated by the Tasmanian archaeological record ^{88,111} , by ancient DNA evidence pertaining to the emergence times of modern-level toolkit complexity ¹⁸⁹ , by a study of early-European-contact-era toolkit complexity in the islands of Oceania ¹¹⁹ , and by a study of the toolkit complexity in 40 nonindustrial farming and pastoralist groups ⁴¹ . Moreover, contemporary economic studies by Redding and Venables ¹⁹⁴ and by Head and Mayer ⁸⁶ demonstrate that income increases with respect to a measure of “market potential,” which increases with respect to country size.
Prediction 2: The long-run number of specializations is higher when the degree of specialization noncomplementarity in the environment is low.	Ethnic groups of the WNAI and of the EA that inhabit ecologically diverse environments tend to have a greater number of specialized activities (see Subsection 3.6.2 of the SI). Ecological diversity is used as a proxy for the noncomplementarity of specializations.
Prediction 3: Egalitarian groups robustly innovate the optimal number of specializations in the long run. Non-egalitarian groups innovate a suboptimally low number of specializations in the long run.	Ethnic groups of the WNAI and of the EA with more Subsection sharing norms tend to have a greater number of specialized activities (see Section 3.6.3 of the SI).
Prediction 4: Among non-egalitarian groups, the amount of long-run innovated specialization (and thereby, group fitness) is an inverse-U-shaped function of innovators’ overconfidence. The optimal level of overconfidence occurs when juniors are perfectly indifferent between apprenticing and innovating.	A study by Cieřlik et al. ⁴⁰ found that countries’ Gross Domestic Product (GDP) measure—a proxy for both the amount of innovated specialization and for group fitness—is predicted by an inverse-U-shaped function of their level of entrepreneurial overconfidence. In particular, an intermediate level of entrepreneurial overconfidence maximizes GDP.
Prediction 5: Egalitarian groups are more likely to have their number of long-run innovated specialization (and thereby, group fitness) be unaffected by changes in innovators’ confidence norms, in the fixed cost of emerging guilds, and in the per-unit cost of emerging guilds than their non-egalitarian counterparts.	Remains to be tested.
Prediction 6: Groups with a higher rate of premature deaths sustain a smaller number of specializations in the long run.	Archaeological data suggest that technological innovations were lost as a result of rapid population decline among Northern Europeans during the Late Glacial Period ^{49,198} .

Table 3.2: Testable predictions for an exchange economy comprised of two groups.

Prediction	Empirical Evidence
Prediction 7: Trade increases specialization.	Groups engaged in trade specialize in differentiated types of goods, a phenomenon observed in both ancestral societies ^{200,219} and contemporary societies ^{44,138,196} . Contemporary trade data also suggest that groups specialize more when trade between them opens up. To illustrate, countries of the European Union increasingly specialized in differentiated types of goods following their 1957 economic integration ^{6,33} . Moreover, as the regional economies comprising the U.S. became more integrated between the 1800s and the early 1900s, they increasingly specialized in differentiated types of goods ¹¹⁷ .
Prediction 8: Groups benefit from trade. A decrease in the trade barrier causes an increase in the fitness of both groups' members.	Contemporary data strongly corroborate this hypothesis. For example, a study by the OECD ¹⁷⁵ has found that countries that have a relatively open trade policy have experienced twice the average annual economic growth compared to those that have a relatively closed trade policy. Also, a study by Frankel and Romer ⁶⁰ has shown that an increase in the geographical barrier to countries' trade decreases their average income levels.
Prediction 9: Consider the trade between a larger group and a smaller group. All else equal, the smaller group exports a higher proportion of its produced goods to the larger group.	In contemporary international trade, smaller countries tend to have a higher trade-to-GDP ratio than larger countries ¹⁷⁶ .
Prediction 10: Consider the trade between a larger group and a smaller group. If the trade barrier is nontrivial, then the smaller group must pay a premium to trade with the larger group. The trade premium is increasing with respect to the trade barrier, and increasing with respect to the population ratio of the two groups.	Smaller countries tend to face higher aggregate prices for tradeable goods than larger countries ¹⁴² . Similarly, oral-tradition data from the evolutionarily relevant forager societies of Borneo suggest that these forager societies traded at a disadvantage with larger agriculturalist societies ²¹² .

3.4 Discussion

3.4.1 Growing recognition of paleo-complexity

Our model of human specialization and trade yields a number of predictions: some intuitive, and others less so. These predictions pertain not just to specialization and trade, but also to a wide range of factors in human cognition, such as social norms (egalitarianism vs. non-egalitarianism) and cognitive biases (overconfidence vs. rational confidence). The empirical corroboration of these predictions is currently preliminary. However, it nevertheless bolsters the growing body of evidence that Pleistocene human societies' propensity for complex specialization and trade is not a modern outlier, but an adaptation that emerged in the Holocene human groups which comprised most of human evolutionary history^{30,32,133}. This adaptation was crucially shaped by ancestral humans' selection pressure towards the optimal social accumulation of specialized knowledge.

Prediction 3 of our theory states that egalitarian groups enjoy a higher average group fitness than non-egalitarian groups, due to innovating a larger and more optimal number of specializations in the long run. This may help explain the consistent emergence of egalitarian social norms in evolutionarily relevant forager societies. We have found preliminary corroboration of this prediction in our analyses of the EA and the WNAI ethnographic datasets. This adds to the body of multidisciplinary evidence behind the dominant paradigm that egalitarianism has consistently emerged throughout most of human evolutionary history^{19,63,127}. Egalitarianism was adaptive despite its consequent free-rider problem²⁰, because among ancestral humans whose primary selection pressure pertained to the optimal accumulation of specialized knowledge at

the group level, egalitarianism robustly helped achieve this.

Prediction 4 of our model states that the penalty to long-run innovation imposed by non-egalitarianism can be offset by an intermediate increase in juniors' overconfidence (in their prospects from innovating versus apprenticing). This prediction finds preliminary corroboration in the study of Cieřlik et al.⁴⁰ of how countries' GDP varies with their average level of entrepreneurial overconfidence. The number of specializations that an egalitarian group innovates, in contrast, is predicted to be more robust to changes in confidence norms. This is because egalitarian groups innovate the optimal number of specializations already for a wide range of model parameters.

The evolution of overconfidence, and of cognitive biases in general, may seem evolutionarily puzzling at first glance. How did natural selection select for cognitive biases when they systematically cause errors in judgement? In fact, cognitive biases like overconfidence—which seem like systematic flaws in how a person learns from their individual observations—may in fact be behavioral byproducts of a strategy optimized for the social learning of specialized knowledge¹⁷⁸. Our theory adds to the plausibility of this hypothesis by formally deducing that even if overconfidence is maladaptive at the individual level, it can be robustly adaptive at the group level due to its optimizing effects on the social accumulation of specialized knowledge.

The phenomenon of human cooperation in large groups of non-kin is also an evolutionary puzzle. Direct reciprocity has traditionally been proposed as an evolutionary explanation of non-kin cooperation, for which the theorem that cooperation can be evolutionarily sustained in the repeated prisoner's dilemma—the most commonly used model of direct reciprocity—is often asserted as evidence²³³. However, reciprocal cooperation is much more difficult to sustain when the repeated prisoner's dilemma is modified to be more biologically realistic¹⁸¹. Moreover, data from social animal

species—including but not limited to humans—generally do not identify a high frequency of reciprocity opportunities to be as strong of a predictor of cooperative behavior as predicted by the theory of direct reciprocity^{78,145,243}. This overall suggests that direct reciprocity, by itself, may not be sufficient to evolutionarily explain the sheer degree to which humans cooperate.

Cultural group selection is a more viable evolutionary explanation of human ultrasocial cooperation^{25,87,197}. This differs from genetic group selection, which is hampered as an explanation by the negligibility of genetic differences between competing human groups³⁹. Indeed, the accumulated norms, specialized knowledge, and other culturally transmissible forms of phenotype display far more variation between groups than genetically transmissible forms of phenotype¹⁴. Our theory helps explain why inter-group cultural variation is larger than inter-group genetic variation by specifying a source of cultural variation other than random mutations. In our theory, different groups endogenously specialize away from each other when making trade contact, a pattern that consistently emerges from their members' pursuit of their individual fitness. Intentional specialization rooted in individual incentives is more likely to create and maintain sizable phenotypic variation than random phenotypic changes like genetic mutations, especially when the gene flow between groups is substantial. Cultural group selection can act on this sizable phenotypic variation to robustly select for group-cooperative phenotypes, despite their large negative effect on individual fitness in excess of what can be explained away by direct reciprocity.

Cultural group selection is also a viable explanation for how group phenotypes that increase group fitness by facilitating a more adaptive collective brain can robustly persist⁹⁰. Potential examples of such selection events suggested by our theory include the hypothesized spread of norms that facilitate the maintenance of societies and

exchange networks with a high population size, the hypothesized spread of egalitarian sharing norms throughout most ancestral human societies, and the hypothesized spread of social norms like exogamy and shared rituals that reduce various barriers to trade. Our model thus bolsters the hypothesis that the human propensity for complex specialization and trade is not a modern outlier. It is an ancestral adaptation with a crucial role in the theory of cultural group selection.

3.4.2 Limitations of the model

We constructed our model with the goal of tractably deducing predictions about how innovation, specialization, and trade interact. For this analytic tractability, we made a number of oversimplifying model assumptions, including the linearity of the production function, symmetry between different guilds and their specialized goods, as well as the homogeneity of members' preferences (and thereby, the existence of a representative member). Such assumptions are standard in complex fields of social science such as international trade, where constructing mechanistically accurate models is computationally impractical and in practice, generalizable predictions come at the expense of precision. We provide a non-exhaustive list several of these oversimplifying model assumptions and discuss how they may be made more realistic in future variants of the model.

Cumulative cultural evolution is abstracted away

In our model, cumulative cultural evolution among guilds was modeled as a one-shot event, in which guilds in the “emerging” phase enter the “traditional” phase and thereby obtain a one-time increase in production efficiency. In reality, cumulative

cultural evolution is a process in which the practices that more successful or more ostensibly successful than status-quo practices are innovated, adopted, and disseminated. In particular, a traditional guild's gain in production efficiency may occur as a continual, stochastically determined sequence rather than as a one-time deterministic event. We made the oversimplifying assumption of a one-shot efficiency increase based on experience with the goal of a model that can simultaneously achieve the goals of analytically tractability and the investigation of how cumulative cultural evolution interacts with other model mechanisms.

The social norm is not an explicit evolutionary outcome

In our model, the social norm of whether the guilds in a group share their production equally with apprentices (egalitarian) or without constraint (non-egalitarian) is assumed to be an exogenous factor. The emergence, survival, and spread of such a social norm is not explicitly modeled in a cultural-evolutionary manner. We note, however, that future variants of our model that aim to explicitly model these cultural-evolutionary dynamics may use the average group fitness for this purpose.

Preference is not an explicit evolutionary outcome

In our model, group members' innate types are solely characterized by their preference U . Members are born and die in our model, which may lead a standard evolutionary model to incorporate the effect of natural selection: such as on members' preference. Also, human preferences in reality are significantly dependent on the local environment, and on the preferences of associates. However, our economics-based model abstracts away the effects of natural selection and of adaptive learning on members'

preferences. A mechanistically realistic model would allow members' preferences U to undergo variations—via either genetic mutation or adaptive learning—as well as to differentially survive and propagate (via genetic- or cultural-evolutionary channels). However, recall that we have abstracted away this process by assuming that in equilibrium, members' genetic evolution and adaptive learning will cause surviving members to have their preference coincide with the ecologically optimal utility function U . Our approach has the advantage of epistemic modesty on how the fitness-utility function U is determined in reality, but has the disadvantage of lacking realism of how such a preference may be mechanistically shaped by various factors spanning genetics and the environment.

3.4.3 Future Directions

Effect of overconfidence on societal innovation

Prediction 5 of our model states that egalitarian groups' number of innovated specializations is more robust to changes in confidence levels. Specifically, an increased social norm overconfidence will tend to increase societal innovation in non-egalitarian groups. While this effect can also occur in egalitarian groups, it will tend to occur in fewer situations.

We have not yet found empirical corroboration of the prediction that egalitarian groups' innovation are more robust to changes in confidence (or falsification, for that matter). While we have been able to find datasets on groups that vary in their degree of egalitarianism, as well as datasets on groups that vary in confidence norms, it remains to compile a dataset on groups that sizably vary in both. We propose the future collection of such a dataset in order to either corroborate or falsify the at-large

prediction.

Further tests of our prediction that overconfidence tends to generally increase societal innovation—and thereby, group fitness—will also require a dataset on societies that show sufficient variation in their degrees of overconfidence. While the finding of Cieřlik et al.⁴⁰ that countries' entrepreneurial overconfidence affects their GDP values in the way predicted by our model is a valuable start, robust and non-confounded measurements of overconfidence or underconfidence norms at the society level is a nontrivial but doable task¹⁶³. We propose the future collection of data on social norms pertaining to the degree of overconfidence or underconfidence in various cultures around the world, so that the effect of these confidence norms on societal innovation may be further studied, ideally in a robust and non-confounded way.

Effect of changing cost parameters on societal innovation

We similarly propose empirical tests of another mathematical interpretation of Prediction 4. Specifically, this is the prediction that egalitarian groups' number of innovated specializations is more robust to changes in emerging guilds' real (rather than perceived) cost values: the fixed cost and the per-unit cost. Nontrivial tests of this prediction for real cost values (rather than perceived cost values, as a parametrization of innovators' overconfidence or underconfidence) would necessitate domain-specific knowledge of the cost parameters ahead of time, such as how many labor hours are needed for the production of a given amount of a certain good or service. In general, corroborating further or falsifying any of our model's predictions via additional data would be scientifically productive.

Incorporating resource curse

Our model’s prediction that non-egalitarian groups experience a penalty to average group fitness due to innovating too few specializations may seem counterintuitive. Small-scale mobile foraging groups—the egalitarian societies that exist today—are less specialized than modern non-egalitarian societies. However, this is not a controlled comparison, since the former is characterized by a much lower population size than the latter. Our theory predicts that societies with larger populations tend to innovate more specializations, a prediction corroborated by various archaeological data which suggest that technological complexity increases with population size^{41,88,111,119,189}. The increasing effect of population size on innovation can more-than-offset the decreasing effect due to non-egalitarian social norms. The question of why non-egalitarianism consistently emerges in modern societies (and agricultural societies in general) may be studied in future work, by incorporating into our model the resource curse and other additional causal mechanisms.

Acknowledgements

We would like to thank Mona Xue for her help in digitally illustrating the visualizations in Figures 3.1, 3.4, and 3.6. We would also like to thank Christian Hilbe, Michael Muthukrishna, and Micah Pietraho for helpful comments on the draft.

3.5 Supplementary Information: The model

Suppose first that there is a single group, whose population is modeled by a volume. It is comprised of numerous members, each of whom has infinitesimal volume in the population. Let $Z = [0, L]$ model the population of these members, where $L > 0$ denotes the total volume of all members in the population. In other words, the quantity L represents the normalized approximation of the group's population size. For conciseness, we call L the population size. Members are divided into two overlapping generations: seniors and juniors. Juniors who survive to the next time step become seniors, whose experience improves the productivity of their specialization.

3.5.1 Production

The group's members coordinate on a number of specializations (e.g., hunting, fishing, gathering), each of which manifests as a guild which cooperatively produces the corresponding specialized good (e.g., meat, fish, gathered food items). Each specialized guild contains a finite number of members. As a result, since the group's population size is modeled by a continuum, the set of specializations must also be modeled by a continuum. Let the continuum $I = [0, \infty)$ denote the set of all possible specializations. The index $i \in I$ will be used to mean, depending on the context, either the given specialized guild or the corresponding specialized good that is produced by the guild.

We assume that specialized guilds come in two types. The first type is a recently emerging guild. We assume that recently emerging guilds are inefficient, due to their lack of experience. The second type is a traditional guild, whose methods of production have been passed down from generation to generation. We assume that the tradi-

tional guilds, compared to their recently innovated counterparts, produce with higher efficiency due to their experiential learning. Traditional guilds are led by their senior leaders, who take on junior apprentices. This is the mechanism of cultural learning. The remaining juniors coordinate to create and produce in new guilds. This is the mechanism of innovation.

Suppose a specialized guild i was recently innovated. We assume that the total amount of good i produced by this recently emerging guild is equal to the value $x(i)$ which solves the equation

$$\ell(i) = \alpha + \beta x(i). \quad (3.25)$$

Observe that the production function (3.26) is characterized by increasing returns. Specifically, there is a fixed cost, denoted by α . The fixed cost denotes the amount of time that members of the specialized guild i have to collectively spend. For example, the specialized guilds of mobile forager bands may have needed to spend a fixed amount of time and resources to survey the current area, to create the necessary tools, and to invest in other overhead costs. The per-unit cost is denoted by β . In the above example, the per-unit cost denotes the average amount of goods created by one guild member in one unit of time after the phase of surveying the area and developing the cooperative foraging strategy.

On the other hand, suppose a specialized guild i is a traditional guild. We assume that the total amount of good i produced by this traditional guild is equal to the value $x(i)$ which solves the equation

$$\ell(i) = \hat{\alpha} + \hat{\beta} x(i). \quad (3.26)$$

Here, we assume that the fixed cost $\hat{\alpha}$ and the per-unit cost $\hat{\beta}$ are assumed to be lower

than their counterparts for recently emerging guilds: that $\hat{\alpha} \leq \alpha$ and $\hat{\beta} \leq \beta$ (with at least one of the inequalities assumed to hold strictly).

After every specialized guild produce its respective good, the good is distributed among the guild members. The ratio of the distribution is determined by the group's social norm. In non-egalitarian groups, senior leaders have the right to give each junior apprentice a smaller proportion of their guild's produced good. Thus, they will in equilibrium pay each junior apprentice the proportion that would make the apprentice indifferent between apprenticing at a generic traditional guild and founding a generic emerging guild. In other words, the exchange value of the proportion allocated to an apprentice of a traditional guild must be equal to that of the proportion allocated to a founder of an emerging guild.

However, in egalitarian groups, senior leaders do not have the right to give each junior apprentice any amount less than they accord each of themselves. Consequently, in equilibrium, every traditional guild shares equally, even among both senior leaders and apprentices.

3.5.2 Preference

A bundle of goods $c \in [0, \infty)^I$ is defined by the collection of nonnegative numbers $c(i)$, each of which denotes the amount of the specialized good i corresponding to guild i .

An agent's fitness from obtaining the bundle of goods $c \in [0, \infty)^I$ is

$$U(c) = \int_{i \in I} c(i)^\theta di \quad \text{for } \theta > 0. \quad (3.27)$$

It is a utility function on the space

$$\mathcal{C} = \{c \in \mathcal{L}^2(I) : c(i) \geq 0 \text{ for all } i \in I\} \quad (3.28)$$

of potentially consumed information-good bundles. Here, $\mathcal{L}^2(S)$ denotes the normed space of square-integrable, Lebesgue-measurable functions $S \rightarrow \mathbb{R}$, which comes with the caveat that two Lebesgue-measurable functions f_1, f_2 are considered equivalent if they only differ on a measure-zero subset.

In the above, “fitness” denotes some combination of cultural fitness and genetic fitness, the exact ratio of which we are agnostic about. The combination is likely weighted towards the former, given the compelling body of evidence that cultural evolution is the primary adaptive mechanism acting on human phenotypes. In equilibrium, every agent’s decision-making maximizes their fitness function. Thus, we use the terms “fitness function” and “utility function” interchangeably.

When $\theta < 1$, the fitness function U has the property of increasing returns to specialization. Specifically, consider subsets $S_1 \subsetneq S_2 \subsetneq I$ of positive measures $n_1 < n_2$; without loss of generality, we can let $S_1 = [0, n_1]$ and $S_2 = [0, n_2]$. The property of increasing returns to specialization is that for any $C \in (0, \infty)$,

$$\int_0^{n_1} C^\theta di < \int_0^{n_2} \left(\frac{Cn_1}{n_2}\right)^\theta di. \quad (3.29)$$

In other words, the fitness function U strictly prefers spreading out a fixed number of units for consumption within the larger subset S_2 over doing so within the smaller subset S_1 .

On the other hand, when $\theta > 1$, the fitness function U has the property of increas-

ing returns to homogeneity. The definition of this property is given by replacing the inequality (3.29) with the reverse inequality

$$\int_0^{n_1} C^\theta di > \int_0^{n_2} \left(\frac{Cn_1}{n_2} \right)^\theta di. \quad (3.30)$$

Finally, when $\theta = 1$, the fitness function U is indifferent between increasing returns to specialization and increasing returns to homogeneity. It prioritizes the summed amount of all goods, without distinguishing between distinct goods.

Throughout this paper, we will assume that $\theta < 1$ unless otherwise stated. This assumption represents the case in which the returns to specialization are increasing: the case where niche specialization and mutually beneficial trade occurs.

For tractability, we have assumed symmetry among the possible goods $i \in I$ in our fitness utility function U . Also, we have implicitly assumed that the mechanisms causing increasing returns to specialization and those causing increasing returns to homogeneity can be aggregated into a unidimensional parameter $\theta > 0$. We note that these are significant oversimplifications, and that our model can be generalized to incorporate realistic variability in these and other aspects.

3.5.3 Exchange

Without loss of generality, suppose that the subset of traditional guilds is $[0, n_{\text{trad}})$ and the subset of emerging guilds is $[n_{\text{trad}}, n)$.

If $i \in [n_{\text{trad}}, n)$ denotes an emerging guild, say comprising a population size of $\ell(i)$ larger than the fixed cost α , it produces

$$x(i) = \frac{\ell(i) - \alpha}{\beta} \quad (3.31)$$

units of goods i . On the other hand, if the producers of a traditional guild $i \in [0, n_{\text{trad}})$, say comprising a population size of $\ell(i)$ larger than the fixed cost $\hat{\alpha}$, it produces

$$x(i) = \frac{\ell(i) - \hat{\alpha}}{\hat{\beta}} \quad (3.32)$$

units of good i . We require that $\ell(i)$ is a Lebesgue-measurable function that satisfies

$$\int_{i \in I} \ell(i) = L. \quad (3.33)$$

Simultaneously, the specialized guild producing each good i offers a relative price $p(i) : [0, n] \rightarrow [0, \infty)$ for their product. The function $p(i)$ is assumed to be contained in the space $\mathcal{L}^2([0, n])$ of square-integrable functions. Then, the members maximize their utility by voluntarily exchanging their produced goods with each other according to the relative price vector $p(i)$. The total value that a producer of good i has to exchange is given by the wage function

$$w(i) = \frac{p(i)x(i)}{\ell(i)}. \quad (3.34)$$

Assume that $w(i) = W$ is the total exchange value of the produced goods that an agent starts with. After the agent's net exchange, their bundle of goods is assumed to solve the fitness-maximization program:

$$\max_{c(i) \in \mathcal{C}} U(c) \quad (3.35)$$

with respect to the constraint

$$\int_{i \in [0, n]} p(i)c(i)di = W. \quad (3.36)$$

The solution $c = c^*$ of the fitness-maximization program (3.35) must satisfy the continuum version of the Lagrange-multiplier necessary condition. The necessary condition is that the Gâteaux derivative of the Lagrangian functional

$$\mathcal{L}[c, \lambda] = \int_{i \in [0, n]} \left(c(i)^\theta - \lambda p(i)c(i) \right) di \quad (3.37)$$

with respect to each variable $c(i)$ equals zero at $c = c^*$. For each variable $c(i)$, the Gâteaux derivative with respect to $c(i)$ is given by

$$\frac{\partial \mathcal{L}[c, \lambda]}{\partial c(i)} = \theta c(i)^{\theta-1} - \lambda p(i). \quad (3.38)$$

Setting this equal to zero, we obtain the necessary condition

$$\theta c(i)^{\theta-1} = \lambda p(i). \quad (3.39)$$

The collection of necessary conditions for each variable $c(i)$ uniquely determines the fitness-maximizing bundle of goods $c = c^*$ that the individual achieves during the exchanging phase. The explicit formula for c^* further yields an explicit formula $x(p)$ for every guild's optimal amount of goods to produce when their goods are priced at p . This is done as follows.

First, the necessary condition (3.39) is rewritten as

$$c(i) = \left(\frac{\lambda p(i)}{\theta} \right)^{-\frac{1}{1-\theta}}. \quad (3.40)$$

This determines the relative amount consumed of each good i ,

$$\frac{c(i)}{c(j)} = \left(\frac{p(j)}{p(i)} \right)^{\frac{1}{1-\theta}}. \quad (3.41)$$

Rearrange (3.41) to get

$$c(i) = c(j) \left(\frac{p(j)}{p(i)} \right)^{\frac{1}{1-\theta}}. \quad (3.42)$$

Then, multiply by $p(i)$ and integrate to get

$$\int_0^n p(i)c(i)di = c(j)p(j)^{\frac{1}{1-\theta}} \int_0^n p(i)^{1-\frac{1}{1-\theta}} dt. \quad (3.43)$$

Note that the left-hand side of (3.43) is precisely W . So, we get

$$p(j)^{-\frac{1}{1-\theta}} \left(\int_0^n p(i)^{-\frac{\theta}{1-\theta}} di \right)^{-1} W = c(j). \quad (3.44)$$

Take the values of the relative price function $p(\cdot)$ and the values of the output function $c(\cdot)$ to be fixed, except for the values $p(j)$ and $x(j)$. This determines the starting exchange value W_a of agent a by

$$W_a = \int_{i \in [0,j) \cup (j,n]} p(i)c_a(i)di. \quad (3.45)$$

The expression (3.44) we obtained before can then be written as

$$c_a(j) = p(j)^{-\frac{1}{1-\theta}} \left(\int_0^n p(i)^{-\frac{\theta}{1-\theta}} di \right)^{-1} W_a. \quad (3.46)$$

Take the integral of (3.46) over all members a , which yields

$$C^{\text{total}}(j) = p(j)^{-\frac{1}{1-\theta}} \left(\int_0^n p(i)^{-\frac{\theta}{1-\theta}} di \right)^{-1} W^{\text{total}}, \quad (3.47)$$

This gives a formula for the optimal amount of goods $x_j(p)$ to produce, assuming a relative price value of p .

$$x_j(p) = p^{-\frac{1}{1-\theta}} \left(\int_0^n p(i)^{-\frac{\theta}{1-\theta}} di \right)^{-1} W^{\text{total}}, \quad (3.48)$$

Size of emerging guilds

Suppose i is an emerging guild, say with number of members

$$\ell = \alpha + \beta x. \quad (3.49)$$

Their optimal guild size x is determined by the optimal guild-size program:

$$\text{maximize } \frac{px_j(p)}{\alpha + \beta x_j(p)} \text{ subject to } p > 0. \quad (3.50)$$

The maximum is obtained at the point where the derivative is zero:

$$\frac{\partial}{\partial p} \frac{px_j(p)}{\alpha + \beta x_j(p)} = \frac{(\alpha + \beta x_j(p)) \left(x_j(p) + p \frac{\partial x_j}{\partial p}(p) \right) - px_j(p) \beta \frac{\partial x_j}{\partial p}(p)}{(\alpha + \beta x_j(p))^2} = 0. \quad (3.51)$$

This is simplified to

$$(\alpha + \beta x_j(p)) \left(\frac{x_j(p)}{\frac{\partial x_j}{\partial p}(p)} + p \right) - \beta p x_j(p) = 0. \quad (3.52)$$

The function $x_j(p)$ is given by $p^{-\frac{1}{1-\theta}}$ times a constant. Thus, we have

$$\frac{x_j(p)}{\frac{\partial x_j}{\partial p}(p)} = -p(1 - \theta). \quad (3.53)$$

Substituting this into (3.52), we get

$$\frac{\partial}{\partial p} \frac{p x_j(p)}{\alpha + \beta x_j(p)} = \frac{\alpha \theta - \beta(1 - \theta)x_j(p)}{(\alpha + \beta x_j(p))^2}. \quad (3.54)$$

This derivative is zero if and only if

$$x_j = \frac{\alpha \theta}{\beta(1 - \theta)}. \quad (3.55)$$

This is the optimal size of emerging guilds, the size that maximizes the exchange value accorded to each member.

Note. We remark that the above logic relies on the assumption that $\theta < 1$. What happens when $\theta \geq 1$ instead? In that case, the derivative (3.54) is positive whenever p and $x_j(p)$ are positive. This is because the numerator

$$\alpha \theta - \beta(1 - \theta)x_j(p) \quad (3.56)$$

is positive. Thus, a larger guild size will always yield more exchange value per member than a smaller guild. This intuitively makes sense. With either neutral or decreas-

ing returns to specialization, it is in the individuals' interest to coordinate on a small number of populous guilds. This has the benefit of mitigating the inefficiency posed by the fixed cost and, for the case of strictly decreasing returns to specialization, of achieving the consequently increasing returns to homogeneity. Since we are interested in studying groups of people who exchange specialized goods for their mutual benefit, we will return to the case of $\theta < 1$ (increasing returns to specialization, decreasing returns to homogeneity) for the remainder of this paper.

Size of traditional guilds

By the same argument as above, the optimal size of traditional guilds in egalitarian groups is given by

$$\ell_j = \hat{\alpha} + \hat{\beta}x_j = \frac{\hat{\alpha}}{1 - \theta}. \quad (3.57)$$

This is because the optimal amount to produce per member is given by

$$x_j = \frac{\hat{\alpha}\theta}{\hat{\beta}(1 - \theta)}. \quad (3.58)$$

However, the optimal size of traditional guilds in non-egalitarian groups must be computed differently. The total exchange value allocated to each innovator is

$$w^{\text{inn}} = \frac{\text{total value produced}}{\text{number of producers}} = \frac{p^{\text{inn}}x^{\text{inn}}}{\ell^{\text{inn}}} = \frac{p^{\text{inn}}\theta}{\beta}. \quad (3.59)$$

The seniors need to promise at least this much to each junior apprentice. As a result, the seniors maximize the optimal guild-size program:

$$\text{maximize } \frac{px_j(p) - w^{\text{inn}}(\hat{\alpha} + \hat{\beta}x_j(p) - \ell^{\text{trad}}(j))}{\ell^{\text{trad}}(j)} \text{ subject to } p > 0. \quad (3.60)$$

The maximizing price p must solve the first-order condition

$$x_j(p) + p \frac{\partial x_j}{\partial p}(p) - w^{\text{inn}} \hat{\beta} \frac{\partial x_j}{\partial p}(p) = 0 \quad (3.61)$$

Substitute in the expression from before,

$$\frac{x_j(p)}{\frac{\partial x_j}{\partial p}(p)} = -p(1 - \theta), \quad (3.62)$$

to get the maximizing price:

$$p^{\text{trad}} = \frac{w^{\text{inn}} \hat{\beta}}{\theta} = \frac{p^{\text{inn}} \hat{\beta}}{\beta}. \quad (3.63)$$

Given the inequality $\hat{\beta} \leq \beta$, traditional guilds' goods have an equal or lower price

$$p^{\text{trad}} = \frac{p^{\text{inn}} \hat{\beta}}{\beta} \quad (3.64)$$

than that of the emerging guilds' goods, p^{inn} .

From the optimal size of emerging guilds,

$$x^{\text{inn}} = \frac{\alpha \theta}{\beta(1 - \theta)} \iff \ell^{\text{inn}} = \alpha + \beta x^{\text{inn}} = \frac{\alpha}{1 - \theta}, \quad (3.65)$$

we can compute the optimal size of traditional guilds in non-egalitarian groups:

$$x^{\text{trad}} = \frac{\alpha \theta}{\beta(1 - \theta)} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}} \iff \ell^{\text{trad}} = \hat{\alpha} + \hat{\beta} x^{\text{trad}} = \hat{\alpha} + \frac{\alpha \theta}{1 - \theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1}. \quad (3.66)$$

Juniors in egalitarian groups prefer apprenticing to innovating

In non-egalitarian groups, seniors pay their apprentices the precise proportion of the produced goods that would make a junior person indifferent between apprenticing and innovating. What about in egalitarian groups? We demonstrate below that in egalitarian groups, juniors strictly prefer apprenticing to innovating. Thus, the seniors will open up the optimal number of apprentice slots that each pay an equal share, juniors will rush and compete to fill them, and the remaining juniors will cooperate to innovate new guilds.

The trade value of a new guild's equal share is given by

$$w^{\text{inn}} = \frac{p^{\text{inn}} x^{\text{inn}}}{\ell^{\text{inn}}} = \frac{p^{\text{inn}} \left(\frac{\alpha \theta}{\beta(1-\theta)} \right)}{\frac{\alpha}{1-\theta}} = \frac{\theta p^{\text{inn}}}{\beta}. \quad (3.67)$$

Similarly, the trade value of an old guild's equal share is given by

$$w^{\text{trad}} = \frac{p^{\text{trad}} x^{\text{trad}}}{\ell^{\text{trad}}} = \frac{p^{\text{trad}} \left(\frac{\hat{\alpha} \theta}{\hat{\beta}(1-\theta)} \right)}{\frac{\hat{\alpha}}{1-\theta}} = \frac{\theta p^{\text{trad}}}{\hat{\beta}}. \quad (3.68)$$

Apprenticing is strictly better than innovating if and only if

$$\frac{\theta p^{\text{inn}}}{\beta} < \frac{\theta p^{\text{trad}}}{\hat{\beta}}, \quad (3.69)$$

which can be simplified to

$$\hat{\beta} p^{\text{inn}} < \beta p^{\text{trad}}. \quad (3.70)$$

Recall from the formula (3.48) that the optimal amount $x_j(p)$ to produce of a good

priced at p is given by

$$x_j(p) = p^{-\frac{1}{1-\theta}} \cdot \text{constant}. \quad (3.71)$$

Thus, it follows that

$$p^{\text{inn}} = p^{\text{trad}} \left(\frac{\hat{\alpha}\hat{\beta}}{\alpha\beta} \right)^{1-\theta}. \quad (3.72)$$

Substituting (3.72) into (3.70), we obtain that apprenticing is strictly better than innovating if and only if

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta < \alpha^{1-\theta}\beta^\theta. \quad (3.73)$$

This is always true, since $\hat{\alpha} \leq \alpha$ and $\hat{\beta} \leq \beta$, with one of the two inequalities holding strictly.

Thus, the exchange value of the goods a junior person would gain from apprenticing is strictly higher than that from innovating. Juniors will first fill up the apprentice slots, and only after these slots are all filled up will they choose to innovate.

Juniors who are fortunate enough to apprentice earn more than juniors who are left out and forced to innovate. This is because junior apprentices, unlike junior innovators, enjoy the strictly better benefit of equally sharing with seniors of traditional guilds, whose production is more efficient than emerging guilds. Counterintuitively, the cause of this economic inequality among juniors is the group's egalitarian social norm (equal sharing within each guild, even between seniors and juniors).

3.5.4 Multigenerational dynamics

The discussion above pertains to the group's outcome (of coordination, production, and exchange) in a single generation of the model. In the following discussion, we will model how the group's outcome changes over multiple generations.

Let $L(t)$ denote the population of the group at the t th generation, for $t \in \mathbb{N}$.

Suppose that the group, at each time step $t \in \mathbb{N}$, is comprised of an senior generation $L_{\text{senior}}(t)$ and a junior generation $L_{\text{junior}}(t)$. When the time step t is incremented, the following occurs:

1. Members of both the senior generation and the junior generation give birth to the new junior generation $L_{\text{junior}}(t+1)$.
2. All of the previous senior generation $L_{\text{senior}}(t)$, and some proportion $\varphi(t) \in [0, 1)$ of the previous junior generation $L_{\text{junior}}(t)$ dies off.
3. The remaining proportion $1 - \varphi(t)$ becomes the next senior generation $L_{\text{senior}}(t+1)$.

We assume that as $t \rightarrow \infty$, the population size $L(t)$ converges to a limit \bar{L} . Both endogenous and endogenous mechanisms will factor into the limiting population size \bar{L} in complex and interrelated ways.

We also assume that the number of surviving seniors in each guild is smaller than the equilibrium size of traditional guilds ℓ^{trad} . The benefit of this assumption is that the proportion of guilds that are traditional converges to a well-defined limit of 1, of a full proportion. This is helpful for exposition, although an analogous analysis can be straightforwardly done without assuming that the number of surviving seniors in each guild is small.

Egalitarianism is robustly optimal in the long run

Consider a group comprised solely of guilds with size ℓ . The representative group member's fitness is given by the function

$$u(\ell) = \int_0^n c(i)^\theta di = n c^\theta = \frac{L}{\ell} \left(\frac{(\ell - \hat{\alpha})/\hat{\beta}}{L} \right)^\theta = \frac{(\ell - \hat{\alpha})^\theta L^{1-\theta}}{\ell \hat{\beta}^\theta}. \quad (3.74)$$

Observe that $u(\ell)$ is a non-monotonic function that is maximized at $\ell = \frac{\hat{\alpha}}{1-\theta}$. Specifically, it is increasing in the region $\ell < \frac{\hat{\alpha}}{1-\theta}$ and decreasing in the region $\ell > \frac{\hat{\alpha}}{1-\theta}$.

In the limit, the proportion of guilds that are traditional rather than innovated increases to 1. Thus, the group's limiting average fitness is a function of the equilibrium size ℓ^{trad} of traditional guilds.

Recall that in egalitarian groups, the limiting number of specialized guilds is given by L/ℓ^{trad} for

$$\ell^{\text{trad}} = \frac{\hat{\alpha}}{1-\theta}. \quad (3.75)$$

Thus, the limiting average group fitness is $u\left(\frac{\hat{\alpha}}{1-\theta}\right)$, the maximum possible value.

Colloquially, egalitarianism achieves the long-term niche diversity that maximizes average fitness. A social norm of equal sharing within each guild leads group members to innovate the optimal number of specialized guilds in the long run, which allows the group to capture the maximum possible long-term benefits from innovated specialization.

In contrast, a non-egalitarian group innovates a suboptimally low number of specialized guilds in the long run. Recall that the equilibrium size of traditional guilds in

non-egalitarian groups is

$$\ell^{\text{trad}} = \hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1}. \quad (3.76)$$

Since we have assumed that $\hat{\alpha} \leq \alpha$ and $\hat{\beta} \leq \beta$ with at least one of these two inequalities holding strictly, we see that the limiting guild size (3.76) is larger than the optimal value, $\frac{\hat{\alpha}}{1-\hat{\theta}}$. Equivalently, the limiting number of guilds in the group is smaller than the optimal value. The degree of suboptimality increases with how much $\hat{\alpha}$ is smaller than α , as well as with how much $\hat{\beta}$ is smaller than β .

Colloquially, a non-egalitarian social norm that allows seniors to pay their apprentices a lower share causes the group to innovate a suboptimally low number of specializations in the long run. If the parameters of the group and its environment are such that an emerging guild is nearly as efficient as a traditional guild, then the degree of suboptimality is low. However, if the parameters are such that an emerging guild is much less efficient than a traditional guild, then the degree of suboptimality is high.

Moderately overconfident juniors can help offset the innovation penalty posed by non-egalitarianism

Suppose we modify our model so that the group's juniors believe that the fixed cost and the per-unit cost of an emerging guild are not necessarily equal to their true respective values. In this interpretation, $\alpha_{\text{perceived}}$ and $\beta_{\text{perceived}}$ denote the perceived fixed cost and per-unit cost, in the perception of juniors. We distinguish this from $\alpha = \alpha_{\text{true}}$ and $\beta = \beta_{\text{true}}$, the true values of these respective parameters.

The size ℓ^{trad} of traditional guilds in non-egalitarian groups changes when changing

either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$. It changes from

$$\hat{\alpha} + \frac{\alpha\theta}{1-\theta} \left(\frac{\beta}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} \quad (3.77)$$

to

$$\ell^* = \hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1}. \quad (3.78)$$

On the other hand, the size of traditional guilds in egalitarian groups is situationally determined. A similar argument to the one presented in (3.73) yields that apprenticing is strictly better than innovating for juniors if and only if

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta < \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.79)$$

If this inequality holds, then the equilibrium size ℓ^* is unchanged; it is given by

$$\ell^* = \frac{\hat{\alpha}}{1-\theta}. \quad (3.80)$$

If this inequality holds in the other direction, however, traditional guilds must pay strictly higher than an equal share to incentivize juniors to apprentice. They must pay the precise share at which juniors are indifferent between apprenticing and innovating, just like the traditional guilds of non-egalitarian groups. In this case, the equilibrium size ℓ^* of traditional guilds in egalitarian groups is given by (3.78), just like that of non-egalitarian groups.

Recall that in the limit, the proportion of guilds that are traditional rather than innovated increases to 1. Thus, the group's limiting average fitness is a function of the equilibrium size ℓ^* of traditional guilds.

In non-egalitarian groups, the equilibrium size ℓ^* of traditional guilds granularly changes with respect to $\alpha_{\text{perceived}}$ and $\beta_{\text{perceived}}$. This means that the group's long-term limiting average fitness changes as well. As either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ increases, the equilibrium size ℓ^* of traditional guilds increases, because apprentices are less expensive to hire. Thus, the amount of innovated specialization that happens in the long run decreases, making it more suboptimally low than before.

On the other hand, as either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ decreases, the equilibrium size ℓ^* of traditional guilds decreases, because apprentices are more expensive to hire. Thus, the amount of innovated specialization that happens in the long run increases. The region of increase can be divided into two parts:

$$\hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} > \frac{\hat{\alpha}}{1-\theta} \quad (3.81)$$

and

$$\hat{\alpha} + \frac{\alpha_{\text{perceived}}\theta}{1-\theta} \left(\frac{\beta_{\text{perceived}}}{\hat{\beta}} \right)^{\frac{1}{1-\theta}-1} < \frac{\hat{\alpha}}{1-\theta}. \quad (3.82)$$

In the first region (3.81), decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ results in the group's long-term average fitness increasing, and thereby becoming closer to the optimal value. In the second region (3.82), decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ results in the group's long-term average fitness decreasing, and thereby becoming further from the optimal value. In between these two regions, the optimal long-term average fitness is achieved.

In egalitarian groups, however, the equilibrium size ℓ^{trad} of traditional guilds does not always change from the optimal value. Suppose that either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$

increases. The region of increase can be divided into two parts:

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta < \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.83)$$

and

$$\hat{\alpha}^{1-\theta}\hat{\beta}^\theta > \alpha_{\text{perceived}}^{1-\theta}\beta_{\text{perceived}}^\theta. \quad (3.84)$$

In the first region (3.83), we have seen that decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ does not change the equilibrium guild size ℓ_{trad} . In other words, the fact that egalitarianism achieves the optimal amount of innovated specialization is robust to juniors's irrationalities.

Decreasing either $\alpha_{\text{perceived}}$ or $\beta_{\text{perceived}}$ to the point of reaching the second region (3.84) does change the equilibrium guild size ℓ_{trad} . This is because in this region, a traditional guild's equal share is insufficient to convince juniors to apprentice rather than innovate. Thus, traditional guilds must pay juniors the precise share at which they are indifferent between apprenticing and innovating. This leads the equilibrium size ℓ_{trad} of traditional guilds to suboptimally increase to (3.78), same as the unconditional equilibrium size of traditional guilds in non-egalitarian groups. This makes the long-term degree of specialized innovation to be suboptimally high, where the suboptimality is due to an excessive loss of productivity from the fixed cost requisite for each specialized guild.

3.5.5 Inter-group trade: A two-group model

To investigate the properties of inter-group trade, we consider a model of two groups which have the option of trading with each other. The two groups may potentially

have a trade barrier. This notion can geographic barriers to trade. It can also represent a linguistic trade barrier between neighboring ethnolinguistic groups or sub-cultures, because such a trade barrier would make trade more difficult. To illustrate, trade patterns within Switzerland and external counterparts show that linguistic differences causally decreases trade volume⁵⁵. This finding was also corroborated by the study of Melitz¹⁵³, who moreover found that literacy and linguistic diversity at home cause foreign trade to increase. Additionally, a study of 67 countries found that greater cultural differences between trade partners tend to decrease trade²²⁹.

Denote the two groups by Group 1 and Group 2, say of populations L_1 and L_2 , respectively. Suppose that for $k \in \{1, 2\}$, the traditional guilds of Group j produce according to the productivity function

$$\ell(i) = \hat{\alpha}_k + \hat{\beta}_k x(i) \tag{3.85}$$

and the emerging guilds of Group j produce according to the productivity function

$$\ell(i) = \alpha_k + \beta_k x(i). \tag{3.86}$$

We also allow each group to have a social norm that can vary in whether guilds are egalitarian, the degree of overconfidence or underconfidence of juniors, and even potentially other dimensions of variation. We let ℓ_k^{trad} and ℓ_k^{inn} denote the equilibrium sizes of traditional guilds and emerging guilds in Group j . We have implicitly assumed that the social norm of Group j is such that the equilibrium size ℓ_k^{trad} of traditional guilds and the equilibrium size ℓ_k^{inn} of emerging guilds are well-defined quantities.

Moreover, we suppose that the trade barrier between the two groups is represented

by an iceberg transport cost, meaning that one unit of a good reduces to $1/\tau$ when transported from one group from another. Here, we assume that $\tau \geq 1$, indicating that the trade barrier impedes the exchange of goods between groups.

For $k \in \{1, 2\}$, let p_k^{trad} denote the unit price of a traditional guild's good produced in Group k , and p_k^{inn} denote that of an emerging guild's good produced in Group k . In the limit, the proportion of traditional guilds in both groups increases to 1. Thus, only the price values $p_k = p_k^{\text{trad}}$ of traditional guilds' goods matter in the limit. Members of Group 1 can obtain goods produced in Group 1 at a price of p_1 , and goods produced in Group 2 as a price of τp_2 . Members of Group 2 can obtain goods produced in Group 2 at a price of p_2 , and goods produced in Group 1 as a price of τp_1 . Similarly, only the size $\ell_k = \ell_k^{\text{trad}}$ of traditional guilds matter in the limit.

Recall the formula (3.41) for the ratio between the amounts consumed of two goods,

$$\frac{c(i)}{c(j)} = \left(\frac{p(j)}{p(i)} \right)^{\frac{1}{1-\theta}}. \quad (3.87)$$

It follows that

$$\frac{c_{1,1}}{c_{1,2}} = \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}} \quad (3.88)$$

and

$$\frac{c_{2,1}}{c_{2,2}} = \left(\frac{p_2}{\tau p_1} \right)^{\frac{1}{1-\theta}}, \quad (3.89)$$

where $c_{k,m}$ denotes the amount consumed of an good produced by group m by an agent of group k .

In the limit, there are $n_1 = L_1/\ell_1$ types of goods in Group 1 and $n_2 = L_2/\ell_2$ types of

goods in Group 2. It follows that among the total value

$$n_1 p_1 \frac{\ell_1 - \hat{\alpha}_1}{\hat{\beta}_1} \quad (3.90)$$

of goods produced by Group 1, a proportion

$$\gamma_1 = \frac{n_2 \tau p_2 c_{1,2}}{n_1 p_1 c_{1,1} + n_2 \tau p_2 c_{1,2}} \quad (3.91)$$

is spent on the consumption of Group 2's goods. Substituting in (3.88), we get

$$\gamma_1 = \frac{n_2 \tau p_2}{n_2 \tau p_2 + n_1 p_1 \left(\frac{\tau p_2}{p_1}\right)^{\frac{1}{1-\theta}}}. \quad (3.92)$$

We thus obtain that the wage amount

$$\frac{n_2 \tau p_2}{n_2 \tau p_2 + n_1 p_1 \left(\frac{\tau p_2}{p_1}\right)^{\frac{1}{1-\theta}}} n_1 p_1 \frac{\ell_1 - \hat{\alpha}_1}{\hat{\beta}_1} \quad (3.93)$$

is spent on the consumption of Group 2's goods.

By an analogous argument, the total wage of Group 2 is

$$n_2 p_2 \frac{\ell_2 - \hat{\alpha}_2}{\hat{\beta}_2}, \quad (3.94)$$

and of it, the proportion

$$\gamma_2 = \frac{n_1 \tau p_1 c_{2,1}}{n_2 p_2 c_{2,2} + n_1 \tau p_1 c_{2,1}} \quad (3.95)$$

is spent on the consumption of Group 1's goods. Substituting in (3.89), we get

$$\gamma_2 = \frac{n_1 \tau p_1}{n_1 \tau p_1 + n_2 p_2 \left(\frac{\tau p_1}{p_2} \right)^{\frac{1}{1-\theta}}} n_2 p_2 \frac{\ell_2 - \hat{\alpha}_2}{\hat{\beta}_2} \quad (3.96)$$

We thus obtain that the wage amount

$$\frac{n_1 \tau p_1}{n_1 \tau p_1 + n_2 p_2 \left(\frac{\tau p_1}{p_2} \right)^{\frac{1}{1-\theta}}} n_2 p_2 \frac{\ell_2 - \hat{\alpha}_2}{\hat{\beta}_2} \quad (3.97)$$

is spent on the consumption of Group 1's goods.

In equilibrium, every exchange between the two groups must be between equal values. It follows that the value imported and the value exported must coincide: that (3.93) equals (3.97). This yields the condition

$$\begin{aligned} & \frac{(L_2/\ell_2)\tau p_2}{(L_2/\ell_2)\tau p_2 + (L_1/\ell_1)p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}}} (L_1/\ell_1)p_1 \frac{\ell_1 - \hat{\alpha}_1}{\hat{\beta}_1} \\ &= \frac{(L_1/\ell_1)\tau p_1}{(L_1/\ell_1)\tau p_1 + (L_2/\ell_2)p_2 \left(\frac{\tau p_1}{p_2} \right)^{\frac{1}{1-\theta}}} (L_2/\ell_2)p_2 \frac{\ell_2 - \hat{\alpha}_2}{\hat{\beta}_2}. \end{aligned} \quad (3.98)$$

Simplifying, we obtain

$$\begin{aligned} & \frac{1}{(L_2/\ell_2)\tau p_2 + (L_1/\ell_1)p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}}} \frac{\ell_1 - \hat{\alpha}_1}{\hat{\beta}_1} \\ &= \frac{1}{(L_1/\ell_1)\tau p_1 + (L_2/\ell_2)p_2 \left(\frac{\tau p_1}{p_2} \right)^{\frac{1}{1-\theta}}} \frac{\ell_2 - \hat{\alpha}_2}{\hat{\beta}_2}. \end{aligned} \quad (3.99)$$

Note that when p_2 increases relative to p_1 , the left-hand side decreases while the right-hand side increases. Thus, there is precisely one ratio p_2/p_1 that solves the above

equality condition.

Comparative statics of the proportion of goods exported

The first comparative-statics result we will prove pertains to the proportion of each group's goods that is exported to the other group. We will show that γ_1 and γ_2 are both strictly decreasing in τ . We compute

$$\begin{aligned} \frac{\partial}{\partial \tau} \gamma_1 &= \frac{n_2 p_2 \left(n_2 \tau p_2 + n_1 p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}} \right) - n_2 \tau p_2 \left(n_2 p_2 + \frac{1}{1-\theta} n_1 p_1 \tau^{\frac{1}{1-\theta}-1} \left(\frac{p_2}{p_1} \right)^{\frac{1}{1-\theta}} \right)}{\left(n_2 \tau p_2 + n_1 p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}} \right)^2} \\ &= \frac{-\left(\frac{1}{1-\theta} - 1 \right) n_2 p_2 n_1 p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}}}{\left(n_2 \tau p_2 + n_1 p_1 \left(\frac{\tau p_2}{p_1} \right)^{\frac{1}{1-\theta}} \right)^2} < 0. \end{aligned} \quad (3.100)$$

A similar computation shows that $\frac{\partial}{\partial \tau} \gamma_2 < 0$, as well.

Comparative statics as the population ratio varies

The second comparative-statics result we will prove pertains to two groups of different population sizes and equal values for the other parameters. Suppose that the two groups share the same fixed cost for traditional guilds $\hat{\alpha} = \hat{\alpha}_1 = \hat{\alpha}_2$ and the same per-unit cost for traditional guilds $\hat{\beta} = \hat{\beta}_1 = \hat{\beta}_2$. Moreover, we suppose that the two groups' respective social norms are such that they share the same equilibrium size of traditional guilds, $\ell = \ell_1 = \ell_2$. However, we suppose that the two groups' population sizes, L_1 and L_2 , are not necessarily the same.

Without loss of generality, suppose that $p_2 \geq p_1$. Then, let $y = p_2/p_1$. Similarly, we let $z = n_2/n_1 = L_2/L_1$. We will let the population ratio z vary, and observe its

comparative-statics effect on the price ratio y .

We can simplify the equality (3.99) to

$$\frac{1}{z + (\tau y)^{\frac{1}{1-\theta}-1}} = \frac{y}{1 + z \left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}-1}}, \quad (3.101)$$

and further to

$$0 = \left(1 + z \left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}-1}\right) - y \left(z + (\tau y)^{\frac{1}{1-\theta}-1}\right). \quad (3.102)$$

This can be rearranged to

$$zy^{1-\frac{1}{1-\theta}} \left(\tau^{\frac{1}{1-\theta}-1} - y^{\frac{1}{1-\theta}}\right) = y^{\frac{1}{1-\theta}} \tau^{\frac{1}{1-\theta}-1} - 1 \quad (3.103)$$

of equation (3.102). Since $y \geq 1$ and $\tau \geq 1$, the right-hand side of (3.103) is nonnegative, and zero if and only if $y = \tau = 1$. Thus, the same applies to the left-hand side of (3.103): we have

$$\tau^{\frac{1}{1-\theta}-1} \geq y^{\frac{1}{1-\theta}}. \quad (3.104)$$

with equality if and only if $\tau = y = 1$.

We can now show via implicit differentiation that y is strictly increasing with respect to z . Differentiate (3.102) with respect to z to get

$$0 = \frac{1}{\tau y} \left(\left(\left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}} - \tau \right) y^2 - z \left(\tau + \left(\frac{1}{1-\theta} - 1\right) \left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}} \right) y \frac{\partial y}{\partial z} - \frac{1}{1-\theta} (\tau y)^{\frac{1}{1-\theta}} \frac{\partial y}{\partial z} \right). \quad (3.105)$$

Solving for $\frac{\partial y}{\partial z}$, we obtain

$$\frac{\partial y}{\partial z} = \frac{\left(\left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}} - \tau\right)y^2}{z\left(\tau + \left(\frac{1}{1-\theta} - 1\right)\left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}}\right)y + \frac{1}{1-\theta}(\tau y)^{\frac{1}{1-\theta}}}. \quad (3.106)$$

It follows from (3.104) that

$$\left(\frac{\tau}{y}\right)^{\frac{1}{1-\theta}} - \tau \geq 0, \quad (3.107)$$

with equality if and only if $\tau = y = 1$. Thus, the same applies to the expression for $\frac{\partial y}{\partial z}$, given by (3.106). We have $\frac{\partial y}{\partial z} \geq 0$, with equality if and only if $\tau = y = 1$.

Moreover, we check the limit of (3.102) as $z \rightarrow \infty$. This yields the equality

$$0 = \lim_{z \rightarrow \infty} \left(zy^{1-\frac{1}{1-\theta}} \left(\tau^{\frac{1}{1-\theta}-1} - y^{\frac{1}{1-\theta}}\right) - \left(y^{\frac{1}{1-\theta}} \tau^{\frac{1}{1-\theta}-1} - 1\right)\right). \quad (3.108)$$

It follows that

$$0 = \lim_{z \rightarrow \infty} \left(y^{\frac{1}{1-\theta}} \tau^{\frac{1}{1-\theta}-1} - 1\right). \quad (3.109)$$

Solving for the limit, we obtain

$$\lim_{z \rightarrow \infty} = \tau^\theta, \quad (3.110)$$

the threshold value. We thus see that the domain $z \in [1, \infty)$ is mapped bijectively to the range $y \in [1, \tau^\theta)$ in a monotonically increasing way.

Colloquially, we have shown the following. First, the larger group produces goods of a higher relative price than the smaller group. Second, the ratio y between the two groups' relative prices is strictly increasing in the ratio z of their populations. Finally, the ratio y between the two groups' relative prices converges to the finite value τ^θ as the ratio z of their populations is taken to infinity.

The ratio y between the two groups' relative prices quantifies how much more exchange value the good of an agent of the more populous group has. It is intuitive that the maximum possible value for this ratio, τ^θ , is increasing in τ . An increase in the trade barrier τ would significantly reduce the effective market for the less populous group's goods. However, it would only slightly reduce the effective market for the more populous group's goods.

A slightly less obvious fact is that the maximum possible ratio, τ^θ , decreases when love-of-variety increases (when θ decreases), although this can also be intuitively explained. An increase in the love-of-variety increases the willingness of the more populous group to—despite the trade barrier—trade for the goods of the less populous group.

Comparative statics as the trade barrier varies

The third comparative-statics result we will prove again pertains to two groups of different population sizes and equal values for the other parameters. The result is about how the price ratio of a foreign good to a domestic good varies with respect to the trade barrier τ .

Consider

$$\tau y = \frac{\tau p_2}{p_1}, \tag{3.111}$$

the effective price ratio of a foreign good to a domestic good in the less populous Group 1. Since y is at least weakly increasing in τ , it follows that (3.111) is strictly increasing in τ . In other words, for members of the less populous Group 1, the effective price ratio of a foreign good to a domestic good is higher when the trade barrier is higher.

Next, consider

$$\frac{\tau}{y} = \frac{\tau p_1}{p_2}, \quad (3.112)$$

the effective price ratio of a foreign good to a domestic good in the less populous Group 1. We can show that like (3.111), the ratio (3.112) is also strictly increasing in τ . In other words, for members of the more populous Group 2, the effective price ratio (3.112) of a foreign good to a domestic good is higher when the trade barrier is higher. To demonstrate this, substitute $g = \frac{\tau}{y}$ into (3.102), which yields the equality

$$0 = \left(1 + z g^{\frac{1}{1-\theta}-1}\right) - \frac{\tau}{g} \left(z + \left(\frac{\tau^2}{g}\right)^{\frac{1}{1-\theta}-1}\right). \quad (3.113)$$

Differentiating with respect to τ , we obtain

$$0 = - \left(\left(\frac{2}{1-\theta} - 1 \right) \left(\frac{\tau^2}{g} \right)^{\frac{1}{1-\theta}-1} g \right) + \tau z \frac{\partial g}{\partial \tau} + \left(\frac{1}{1-\theta} - 1 \right) z g^{\frac{1}{1-\theta}} \frac{\partial g}{\partial \tau} + g \left(\frac{\frac{1}{1-\theta} \left(\frac{\tau^2}{g} \right)^{\frac{1}{1-\theta}} \frac{\partial g}{\partial \tau}}{\tau} - z \right). \quad (3.114)$$

We can then solve for $\frac{\partial g}{\partial \tau}$,

$$\begin{aligned} \frac{\partial g}{\partial \tau} &= \frac{1}{\tau \left(\tau^2 z + \frac{1}{1-\theta} \left(\frac{\tau^2}{g} \right)^{\frac{1}{1-\theta}} g \right) + \left(\frac{1}{1-\theta} - 1 \right) \tau z g^{\frac{1}{1-\theta}}} \\ &\cdot g \left(\tau^2 z + \left(\frac{2}{1-\theta} - 1 \right) \left(\frac{\tau^2}{g} \right)^{\frac{1}{1-\theta}} g \right). \end{aligned} \quad (3.115)$$

Since this expression is positive, we have shown that $\frac{\partial g}{\partial \tau} > 0$.

Increasing the trade barrier increases the base price of the more populous group's

goods compared to that of the less populous group's goods. However, this effect is outweighed by the increase in the trade barrier itself. From the perspective of the less populous group's members, the effective price of the more populous group's goods is monotonically increasing with respect to the size of the trade barrier.

Trade barrier decreases group members' utility

Finally, we show that the members of both the more populous Group 2 and the less populous Group 1 have utility strictly decreasing in the trade barrier τ . This follows from the fact that the price of foreign goods in terms of domestic goods, given respectively by (3.111) and (3.112), are both strictly increasing in τ .

Indeed, without loss of generality, we can assume that the equilibrium we have computed occurs sequentially via the following exchanges. First, all members produce their respective goods. This yields the starting allocation. Second, all members of Group 1 exchange within the group to obtain equidistributed goods, as do all members of Group 2. Because this step homogenizes members' allocations within each group, we can without loss of generality consider each group to be comprised of a representative agent. Finally, the representative agent of Group 1 and that of Group 2 exchanges fractions of their respective allocations.

We can without loss of generality assume that the exchanged items are two discrete units: a collection of one unit of each Group 1 good, and a collection of one unit of each Group 2 good. For the representative agent of Group 1, an increase in τ results in a strict increase in the price of the latter in terms of the former, (3.112). This results in a strict decrease in the maximum utility achievable by the starting allocation, since the utility function U is concave and increasing. An analogous argument with (3.111) shows the same result for the representative agent of Group 2. Thus, we have

shown that the representative members' utility values strictly decrease in the trade barrier τ .

3.6 Supplementary Information: Empirical results

The present paper introduces a theoretical framework that aims to understand the process of cultural evolution through the examination of various features of human behavior, including social learning, innovation, self-confidence, sharing behavior, and inter-group trade. By integrating these aspects, the model generates a range of potentially testable predictions. These predictions are theoretically elaborated upon in Section 3.3 and summarized in Tables 3.1 and 3.2.

In this section, we present an empirical analysis of several of these predictions using data from the Ethnographic Atlas, abbreviated as the EA dataset^{12,22,76,118,121,158}; and the Western North American Indians, abbreviated as the WNAI dataset^{112–114,118}. Specifically, the analysis corroborates the following predictions.

Prediction 1: Groups with a larger population size L have a higher number n of innovated specializations in the long run.

Prediction 2: The long-run number n of specializations is higher when the degree θ of specialization noncomplementarity in the environment is low.

Prediction 3: Egalitarian groups robustly innovate the optimal number of specializations n in the long run. Non-egalitarian groups innovate a suboptimally low number of specializations n in the long run..

3.6.1 Group population and specialization

This subsection empirically explores relationship between group's population and the number of cultural niches entailed by the Prediction 1 of the theoretical model. As predicted by the theory, our analysis establishes that groups with larger population on the local level tend to be characterized by a higher number of activities with specialization.

To explore the relationship empirically the average population of local communities, as captured by the corresponding variable in the EA* is mapped into the measure that reflects the number of specialized activities across ethnic groups in EA. The measure is based on the presence of occupational specialization across nine potential activities (e.g., metal working, weaving, building, etc).[†] The relationship is explored using an Ordinary Least Squares (OLS) linear-log model, while accounting for: (i) geographic controls (e.g., elevation, terrain ruggedness, distance to coast or river, climate), (ii) continental fixed-effects, and (iii) several ethnographic characteristics (e.g., dependence on hunting vs gathering).

The baseline relationship between the size of local population and the number of specialized activities is depicted in Figure 3.7 in a form of binned scattered plot of residuals with a fitted line. The baseline analysis establishes a statistically significant relationship of magnitude 0.34 in terms of standardized β coefficient[‡], which is significant at 0.1% level. The empirical relationship is further explored in Table 3.3,

*The population size of the local community measure is based on variable 31 of EA. Coded as the average value of each population size interval of variable 31.

[†]The number of specialized activities is based on variables 57 through 65 of EA, and is constructed as a sum of indicators of whether the specialization in each activity exists in the ethnic group.

[‡] β coefficient of magnitude d implies that an explained variable y changes by d standard deviations if explanatory variable x is increased by one standard deviation.

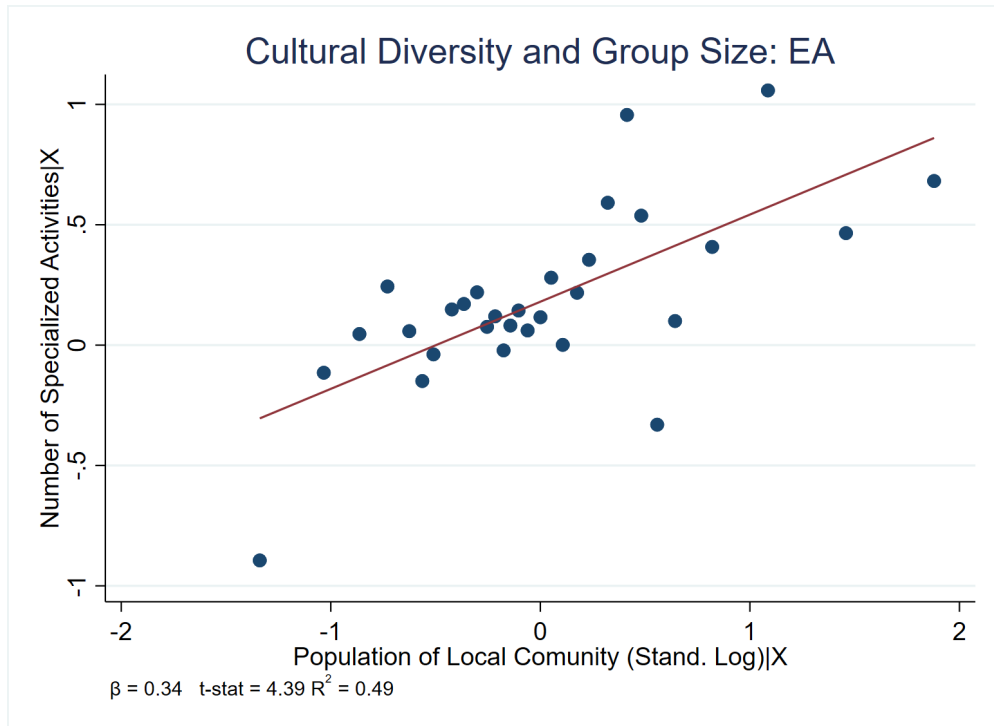


Figure 3.7. The residualized and binned scatterplot of cultural diversity and group size among EA ethnic groups, plotted with the accompanying OLS linear-log regression. Cultural diversity, denoting the number of specialized activities, is based on variables 57 through 65 of the EA dataset. Group size is based on variable 31. The OLS linear-log regression accounts for geographic controls, continental fixed-effects, and ethnographic characteristics. The relationship is statistically significant at the 0.1% level, with a magnitude of 0.34 in terms of the standardized β coefficient.

Table 3.3: Cultural diversity and group size among EA ethnic groups.

	Number of Specialized Activities					
	All Sample					non-WEIRD
	(1)	(2)	(3)	(4)	(5)	(6)
Local Population (Log)	0.65*** (0.07)	0.52*** (0.07)	0.64*** (0.07)	0.41*** (0.07)	0.36*** (0.07)	0.34*** (0.08)
Continental FE	No	Yes	No	No	Yes	Yes
Geographic Controls	No	No	Yes	No	Yes	Yes
Ethnographic Controls	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.29	0.40	0.32	0.40	0.46	0.46
Observations	508	508	508	508	508	475

Notes: Using the OLS regression, this table establishes a positive significant effect of the population of local community on the number of specialized activities, across the ethnic groups in EA. Geographic controls are absolute latitude, mean elevation, terrain ruggedness, agricultural suitability, distance to coast or river, temperature and precipitation annual mean and standard variation. Ethnographic controls include dependence on hunting, gathering and agriculture, year of groups ethnographic observation. Both dependent and independent variables has not been normalized to have zero mean in a standard deviation of 1. Spatial correlation robust standard error⁴² estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

which summarizes the result under the different combination of controls across different samples.

3.6.2 Cultural diversity and ecological variety

This subsection aims to empirically assess the validity of Prediction 2. As predicted by the theory, groups that inhabit an environment with low specialization noncomplementarity would develop a higher number of specializations.

To evaluate this prediction, we investigate the relationship between the number of specialized activities and the ecological diversity—defined by the number of plants and animal species present—of the environments inhabited by ethnic groups observed in EA and WNAI. The empirical test is based on the assumption that an ecologically diverse environment allows for a wider range of unique strategies for plant and animal species, and that this correlates with a wider range of unique strategies for human specializations.

First, the analysis focuses on the WNAI dataset, which contains the data on both the presence of specialization across 10 activities (e.g., weaving, boat-construction, leather-working, etc)[§] and ecological diversity, as captured by the number of plants and animal species present around the ethnic group.[¶] The relationship is explored via linear OLS model, while accounting for: (i) geographic controls (e.g., elevation, distance to coast or river, climate), (ii) regional fixed-effects, and (iii) several ethnographic characteristics (e.g., dependence on hunting and agriculture).

[§]The number of specialized activities is based on variables 218, 221, 224, 227, 230, 233, 236, 241, 246, 249 of WNAI dataset and is constructed as a sum of indicators of whether the specialization in each activity exists in the ethnic group.

[¶]The measure of ecological variety is based on the binary variables 11 – 140 of WNAI dataset, and is constructed as the sum of the indicators for the presence of each plant or animal species.

The baseline relationship between the size of local population and the number of specialized activities is depicted in Figure 3.8 in a form of binned scattered plot of residuals with a fitted line. The baseline analysis establishes a statistically significant relationship of magnitude 0.27 in terms of standardized β coefficient, which is significant at 0.5% level. The empirical relationship is further explored in Table 3.4, which summarizes the result under the different combination of controls across different samples.

In the second part of the analysis, we examine the relationship between the number of specialized activities and ecological diversity among ethnic groups in EA. As the dataset does not include a ready-to-use measure of ecological diversity, we develop a novel metric for this purpose. Our measure encompasses various aspects of the local ecology, including plants, animals, terrain, and climate, and is constructed through the following steps:

1. A 200km-radius zone around the centroid of each ethnic group in EA is constructed;
2. With each zone a number of potentially cultivable crops⁵⁹, number of mammals⁵⁷, range of elevation and terrain ruggedness⁸², and variation in mean annual temperature and annual precipitation⁵⁸ are calculated;
3. These 6 dimension are standardized to have a zero mean and a standard deviation of one;
4. A first principal component of these 6 dimensions is constructed to represent ecological variety.

Relationship between the ecological variety and the number of specialized

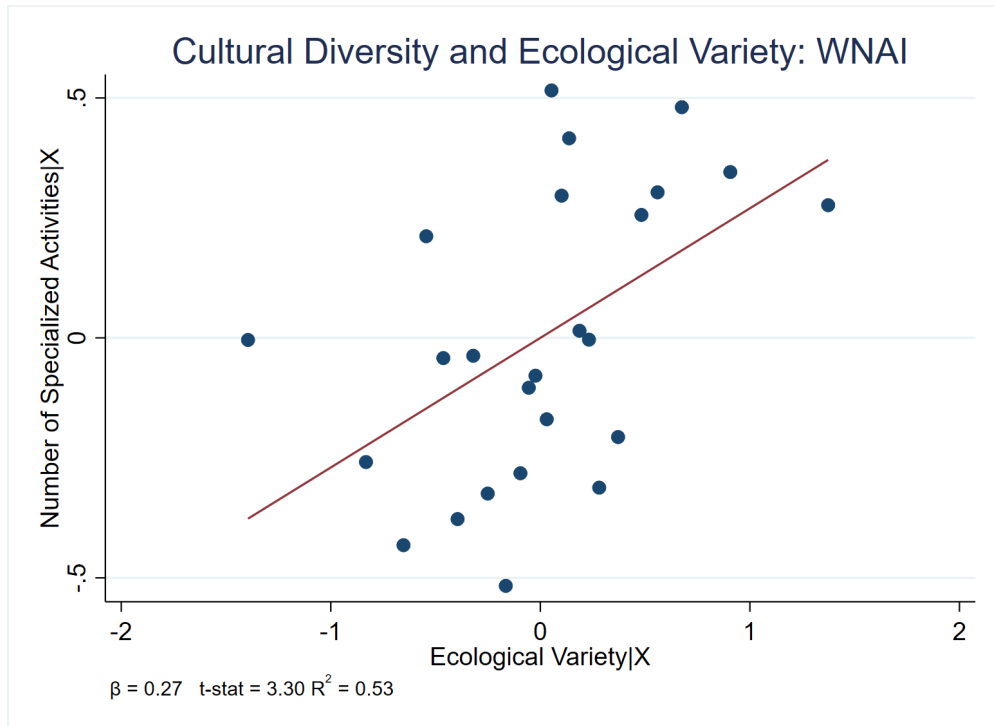


Figure 3.8. The residualized and binned scatterplot of cultural diversity and ecological variety among EA ethnic groups, plotted with the accompanying OLS regression. Cultural diversity, denoting the number of specialized activities, is based on variables 57 through 65 of the EA dataset. Our measure of ecological variety encompasses the number of potentially cultivable crops⁵⁹, the number of mammals⁵⁷, range of elevation and terrain ruggedness⁸², and variation in mean annual temperature and annual precipitation⁵⁸. The relationship is statistically significant at the 0.1% level, with a magnitude 0.13 in terms of the standardized β coefficient.

Table 3.4: Cultural diversity and ecological variety among WNAI ethnic groups

	Number of Specialized Activities					
	All Sample					Foragers
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Variety	0.56*** (0.10)	0.52*** (0.13)	0.60*** (0.10)	0.25*** (0.09)	0.27*** (0.10)	0.34*** (0.12)
Region FE	No	Yes	No	No	Yes	Yes
Ethnographic Controls	No	No	Yes	No	Yes	Yes
Geographic Controls	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.31	0.37	0.33	0.42	0.48	0.58
Observations	172	172	172	172	172	129

Notes: Using the OLS regression, this table establishes a positive significant effect of the ecological diversity on the number of specialized activities, across the ethnic groups in WNAI. Geographic controls are latitude/ longitude polynomial of the second order, mean elevation, distance to coast or river, temperature and precipitation mean. Ethnographic controls include dependence on hunting and agriculture. Regional FE are based on 6 general regions in the dataset: Mexico, Northwestern, South-Central and Southwestern USA, Western Canada and Subarctic America. Both dependent and independent variables has not been normalized to have zero mean in a standard deviation of 1. Spatial correlation robust standard error⁴² estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

activities is explored using a linear OLS model, while accounting for other geographic and ethnographic controls, as well as continental fixed effects (i.e., in the same manner as above). The baseline result is presented in Figure 3.9. It establishes a statistically significant relationship of magnitude 0.13 in terms of standardized β coefficient, which is significant at 0.1% level. The relationship is further explored in Table 3.5, demonstrating the robustness of the results to the combination of controls.

3.6.3 Egalitarianism and cultural diversity

In this subsection, we empirically examine one of the more intriguing predictions of the theory: the relationship between the presence of egalitarian social norms and the degree of cultural diversity. As previously mentioned in Subsection 3.3.2, the model posits that groups with egalitarian norms and institutions will tend to have a greater number of specialized cultural niches over time, all else being equal.

To test the prediction, we first turn to the WNAI dataset. The measure that captures the number of specialized activities, defined above, is related to the indicator of the presence of food sharing norm in the society.^{||} The distributions of societies with and without the social norm of food sharing over the number of specialized activities are depicted in Figure 3.10. Figure 3.10 suggests that more egalitarian societies (i.e., ones that practise the sharing of food and chattels) are far more likely to have a higher number of activities with specialization, while the societies with less egalitarian norms more commonly possess a lower number of specialized activities. This observation is supported by a statistical test that compares the means of two distributions. The observed difference between two groups in terms of the number of specialized

^{||}The measure is based on the variable 253 of WNAI that is coded as 0 if there is no redistribution of chattels and/or food and 1 otherwise.

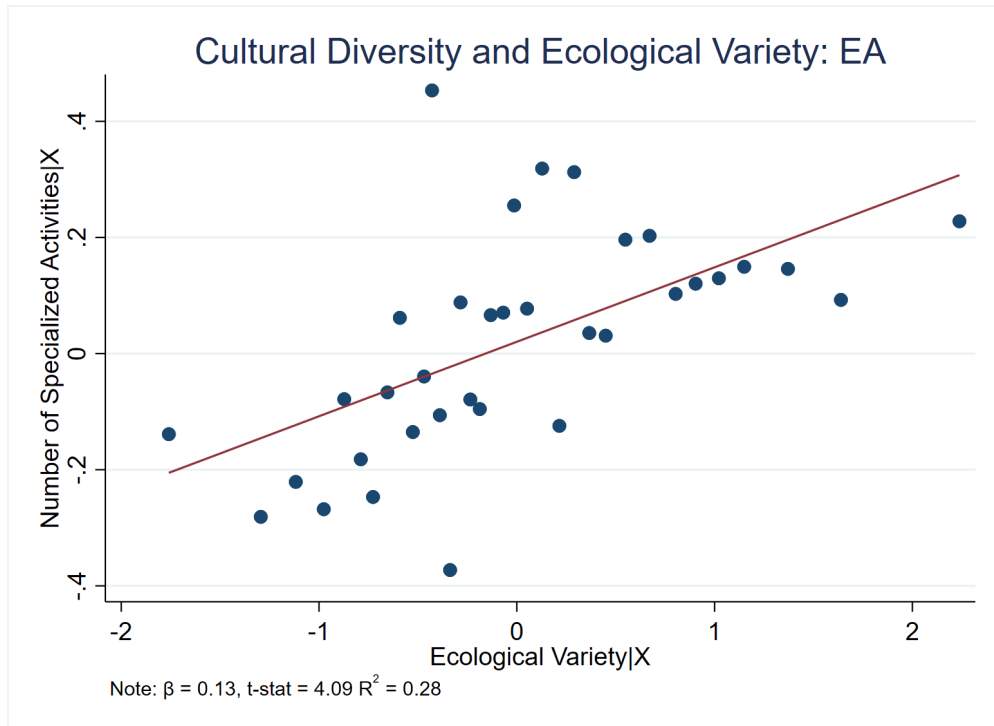


Figure 3.9. The residualized and binned scatterplot of cultural diversity and ecological variety among WANI ethnic groups, plotted with the accompanying OLS regression. Cultural diversity, denoting the number of specialized activities, is based on variables 218, 221, 224, 227, 230, 233, 236, 241, 246, 249 of the WNAI dataset. Ecological variety is based on the binary variables 11–140. The OLS regression accounts for geographic controls, regional fixed-effects, and several ethnographic characteristics. The relationship is statistically significant at the 0.5% level, with a magnitude of 0.27 in terms of the standardized β coefficient.

Table 3.5: Cultural diversity and ecological variety among EA ethnic groups

	Number of Specialized Activities					
	All Sample					non-WEIRD
	(1)	(2)	(3)	(4)	(5)	(6)
Ecological Variety	0.13*** (0.04)	0.15*** (0.03)	0.17*** (0.04)	0.14*** (0.03)	0.13*** (0.03)	0.13*** (0.03)
Region FE	No	Yes	No	No	Yes	Yes
Geographic Controls	No	No	Yes	No	Yes	Yes
Ethnographic Controls	No	No	No	Yes	Yes	Yes
Adjusted- R^2	0.02	0.19	0.03	0.22	0.27	0.27
Observations	1058	1058	1058	1058	1058	1012

Notes: Using the OLS regression, this table establishes a positive significant effect of the ecological variety on the number of specialized activities, across the ethnic groups in EA. Geographic controls are absolute latitude, mean elevation, terrain ruggedness, agricultural suitability, distance to coast or river, temperature and precipitation annual mean and standard variation. Ethnographic controls include dependence on hunting, gathering and agriculture, year of groups ethnographic observation. Both dependent and independent variables has not been normalized to have zero mean in a standard deviation of 1. Spatial correlation robust standard error⁴² estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

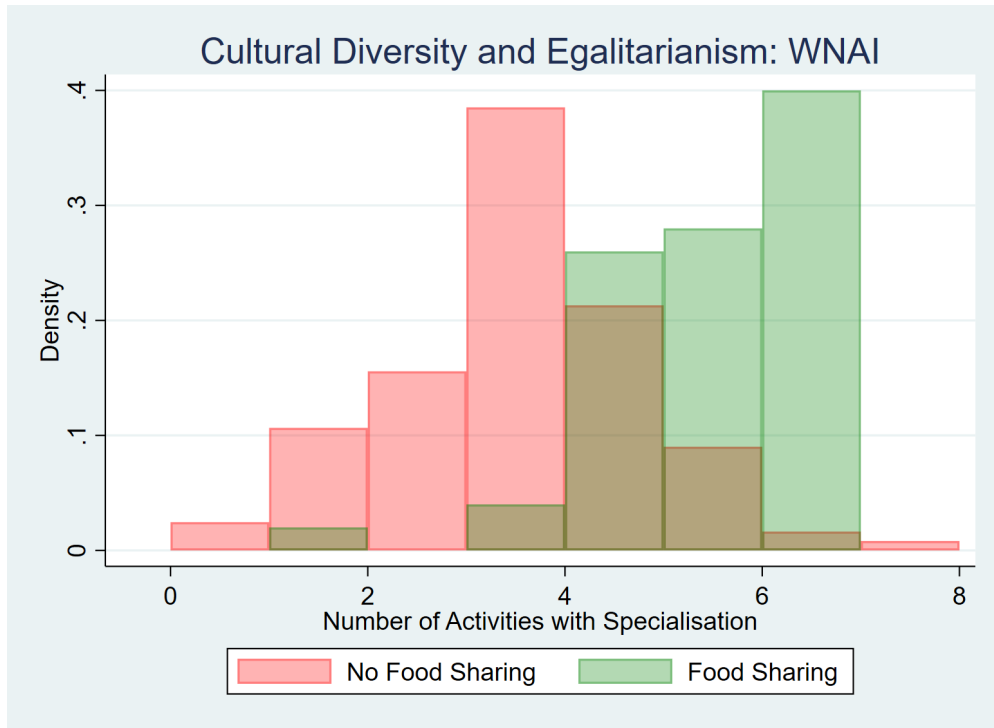


Figure 3.10. Histogram of cultural diversity and egalitarianism among WNAI ethnic groups. Cultural diversity, denoting the number of specialized activities, is based on variables 218, 221, 224, 227, 230, 233, 236, 241, 246, 249 of the WNAI dataset, while the measure of egalitarianism is based on variable 253 (redistribution of food and chattels within a society).

activities is 2.1, which allows us to reject the null-hypothesis of mean with equality more than 99.9% confidence.

Since in this case the outcome variable is an ordered categorical variable while the explanatory variable is binary, the relationship could be explored in a greater detail using the ordered probit regression model. Estimation of the empirical model establishes a positive relationship between the presences of food sharing social norm on the number of specialized activities, across a number of specifications (Table 3.6). The results are statistically significant at less than 1% level throughout. Although this is reassuring, the interpretation of the coefficients is not straightforward. In order to better understand the implications of these parameters, Figure 3.11 presents the average marginal effects of food sharing for each number of specialized activities for the specification in column (5) of Table 3.6. The figure measures the number of activities on the horizontal axis and the average marginal effect of food sharing with its 95% confidence interval on the vertical axis. As can be seen there, the average marginal effect of food sharing is negative for low number of specialized niches and increases until it becomes positive for high values. This implies that having a food sharing norm in the society decreases the probability of observing low number of specialized activities and increases the probability of observing high values. Thus, the presence of egalitarian social norm shifts the probability distribution of the number of specialized activities rightwards.

Additionally, we perform a similar empirical analysis using the ethnic group-level data from EA. The number of specialized activities, as constructed in Subsection 3.6.1 is mapped into the measure of wealth equality across the societies.** The

**The measure of wealth equality intensity is constructed based on the variable 66 of EA. The measure is coded as binary equal to 0 if the presence of class stratification based on wealth is observed in the society and coded 1 otherwise.

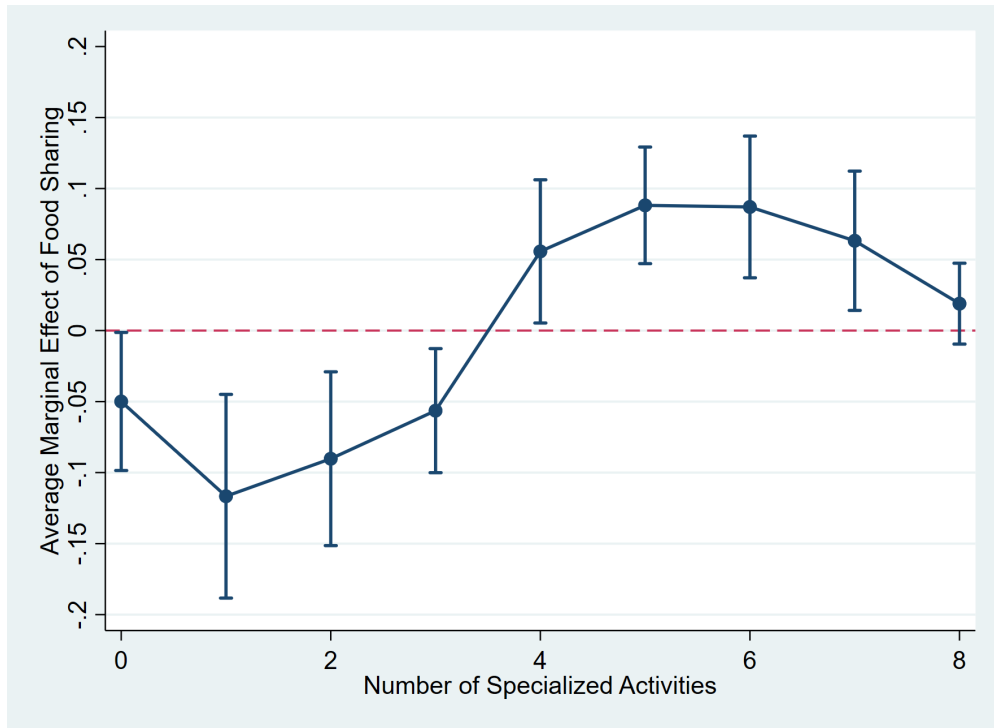


Figure 3.11. Average marginal effect of the presence of food sharing on the number of specialized activities, among WNAI ethnic groups. The figure shows that the average marginal effect of food sharing on the number of specialized activities in a society is negative for low values and positive for high values. This indicates that the presence of a food sharing norm decreases the likelihood of observing low numbers of specialized activities and increases the likelihood of observing high numbers. This suggests that egalitarian social norms shift the distribution of specialized activities towards higher values.

Table 3.6: Cultural diversity and egalitarianism: Ordered probit (WNAI)

	Number of Specialized Activities					
	All Sample					Foragers
	(1)	(2)	(3)	(4)	(5)	(6)
Presence of Food Sharing	1.67*** (0.22)	1.59*** (0.23)	1.78*** (0.24)	1.22*** (0.30)	1.29*** (0.32)	1.73*** (0.39)
Region FE	No	Yes	No	No	Yes	Yes
Ethnographic Controls	No	No	Yes	No	Yes	Yes
Geographic Controls	No	No	No	Yes	Yes	Yes
Pseudo- R^2	0.11	0.15	0.12	0.20	0.22	0.29
Observations	172	172	172	172	172	129

Notes: Using ordered probit regression, this table establishes a positive significant effect of the presences of food sharing social norm on the number of specialized activities, across the ethnic groups in WNAI. Geographic controls are latitude/ longitude polynomial of the second order, mean elevation, distance to coast or river, temperature and precipitation mean. Ethnographic controls include dependence on hunting and agriculture. Regional FE are based on 6 general regions in the dataset: Mexico, Northwestern, South-Central and Southwestern USA, Western Canada and Subarctic America. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

distributions of societies with high and low levels of wealth equality over the number of specialized activities are depicted in Figure 3.12. Although the observed distinction in the distributions is not as stark as in the case of WNAI dataset, with a difference in observed mean number of specialized activities being equal to 0.21 between the two groups, the simple t-test, nevertheless, rejects the null-hypothesis of mean equality with more than 99% confidence.

The effect of wealth equality on the cultural diversity is further analyzed via the ordered probit regression model. Results of Table 3.7 establish a positive and statistically significant relationship between the presence of wealth equality in the society and the observed number of activities with specialization. The result is robust to combination of geographic and ethnographic controls, as well as continental fixed effects. The interpretation of the result for a full specification in column (5) of Table 3.7 in terms of the average marginal effects is depicted in Figure 3.13. As can be seen there, the average marginal effect of wealth equality is negative for zero specialized niches and positive for positive values. This implies that having a wealth equality in the society significantly reduces chances of having no specialized cultural niches, while increasing the probability of having larger number of activities with specialization.

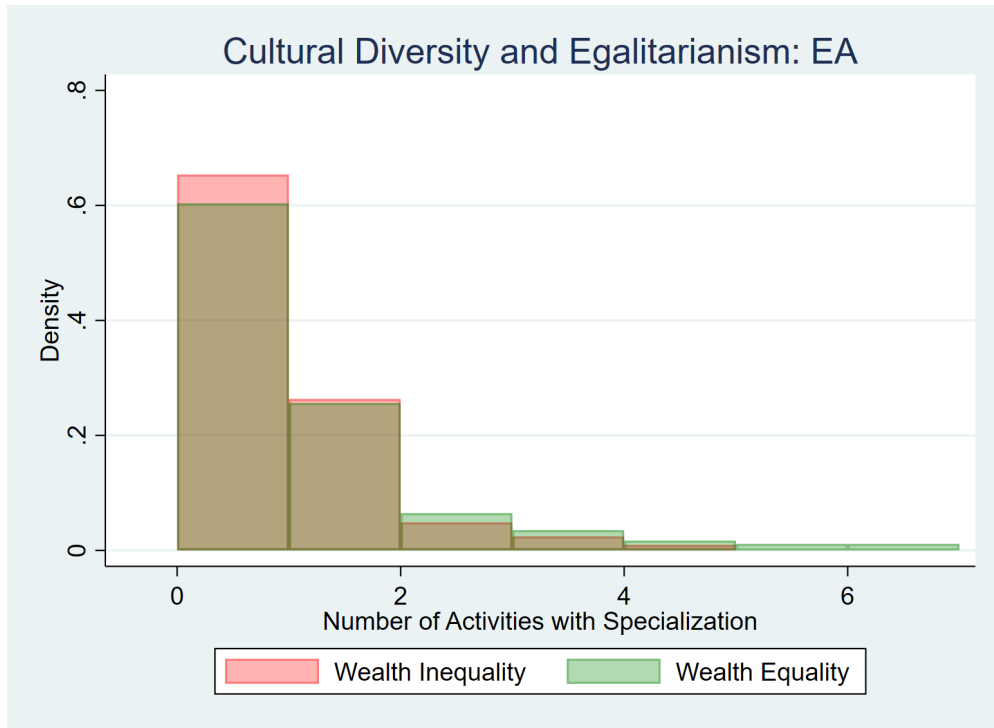


Figure 3.12. Histogram of cultural diversity and egalitarianism among EA ethnic groups. Cultural diversity, denoting the number of specialized activities, is based on variables 57 through 65 of the EA dataset. The measure of egalitarianism is based on class differentiation based on wealth, which is part of variable 66.

Table 3.7: Cultural diversity and egalitarianism among EA ethnic groups: Ordered probit

	Number of Specialized Activities					
	All Sample					non-WEIRD
	(1)	(2)	(3)	(4)	(5)	(6)
Presence of Wealth Equality	0.32*** (0.10)	0.27** (0.12)	0.27*** (0.11)	0.34*** (0.12)	0.35*** (0.12)	0.35*** (0.13)
Continental FE	No	Yes	No	No	Yes	Yes
Geographic Controls	No	No	Yes	No	Yes	Yes
Ethnographic Controls	No	No	No	Yes	Yes	Yes
Pseudo- R^2	0.06	0.21	0.09	0.21	0.26	0.26
Observations	843	843	843	843	843	808

Notes: Using the ordered probit regression, this table establishes a positive significant effect of the presence of equal wealth distribution on the number of specialized activities, across the ethnic groups in EA. Geographic controls are absolute latitude, mean elevation, terrain ruggedness, agricultural suitability, distance to coast or river, temperature and precipitation annual mean and standard variation. Ethnographic controls include dependence on hunting, gathering and agriculture, year of groups ethnographic observation. Heteroskedasticity robust standard error estimates are reported in parentheses; *** denotes statistical significance at the 1% level, ** at the 5% level, and * at the 10% level, all for two-sided hypothesis tests.

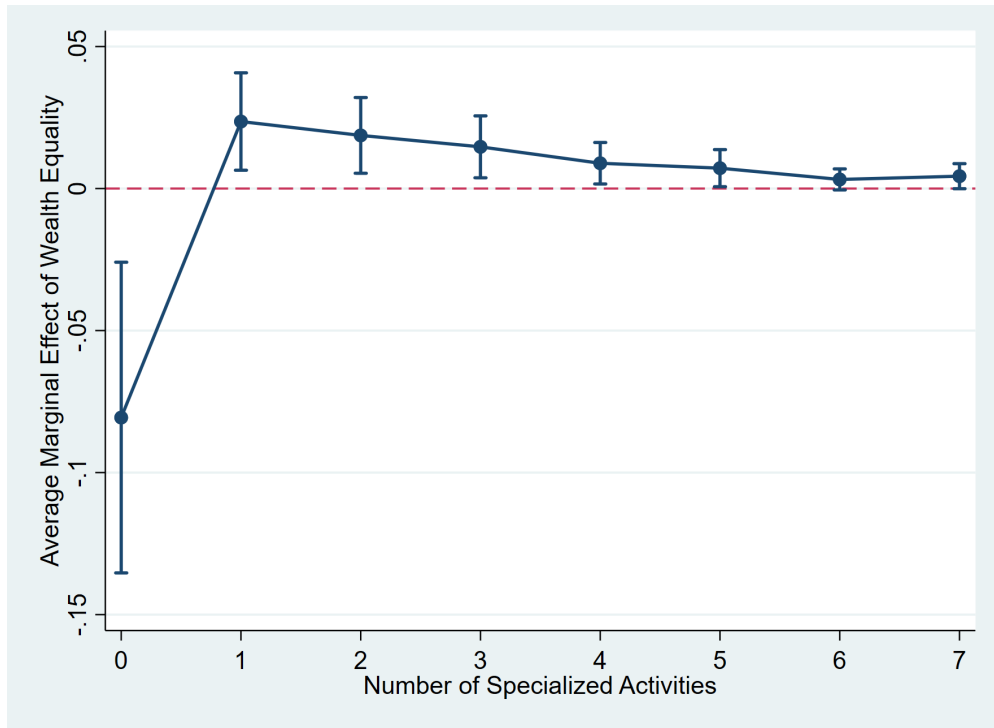


Figure 3.13. Average marginal effect of the intensity of wealth equality on the number of specialized activities, among EA ethnic groups. The figure shows that the average marginal effect of wealth equality intensity on the number of specialized activities in a society is negative at zero specialized activities. The average marginal effect is positive at positive numbers of specialized activities. This overall indicates that wealth equality in a society reduces the probability of having no specialized niches and increases the probability of having a higher number of specialized activities.

References

- [1] S. Agarwal, J. C. Driscoll, X. Gabaix, and D. Laibson. Learning in the credit card market. 2008.
- [2] G. A. Akerlof and R. J. Shiller. *Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism*. Princeton University Press, Princeton, NJ, 2009. ISBN 978-0-691-14233-3.
- [3] E. Akin. What you gotta know to play good in the iterated prisoner’s dilemma. *Games*, 6:175–190, 2015.
- [4] E. Akin. The iterated prisoner’s dilemma: Good strategies and their dynamics. In I. Assani, editor, *Ergodic Theory, Advances in Dynamics*, pages 77–107. de Gruyter, Berlin, 2016.
- [5] E. Akin. Good strategies for the iterated prisoner’s dilemma: Smale vs. Markov. *Journal of Dynamics and Games*, 4:217–253, 2017.
- [6] M. Amiti. Specialization patterns in europe. *Weltwirtschaftliches Archiv*, 135 (4):573–593, 1999. ISSN 0043-2636.
- [7] N. Augenblick and M. Rabin. Belief movement, uncertainty reduction, and rational updating. *Quarterly Journal of Economics*, 136(2):933–985, 2021. doi: 10.1093/qje/qjaa043.
- [8] R. Axelrod. *The evolution of cooperation*. Basic Books, New York, NY, 1984.
- [9] R. Axelrod and W. D. Hamilton. The evolution of cooperation. *Science*, 211: 1390–1396, 1981.
- [10] A. Baimel, M. Juda, S. Birch, and J. Henrich. Machiavellian strategist or cultural learner? Mentalizing and learning over development in a resource-sharing game. *Evolutionary Human Sciences*, 3, 2021. ISSN 2513-843X. doi: 10.1017/ehs.2021.11.

- [11] B. M. Barber and T. Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics*, 116(1):261–292, 2001. ISSN 0033-5533. doi: 10.1162/003355301556400.
- [12] H. Barry, III. Ethnographic atlas xxviii. *Ethnology*, 19(2):245–263, 1980. ISSN 0014-e.
- [13] G. Becker, M. Degroot, and J. Marschak. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232, 1964. ISSN 00057940. doi: 10.1002/bs.3830090304.
- [14] A. V. Bell, P. J. Richerson, and R. McElreath. Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proceedings of the National Academy of Sciences - PNAS*, 106(42):17671–17674, 2009. ISSN 0027-8424.
- [15] Y. Ben-Oren, O. Kolodny, and N. Creanza. Cultural specialization as a double-edged sword: division into specialized guilds might promote cultural complexity at the cost of higher susceptibility to cultural loss. *bioRxiv*, 2022.
- [16] D. J. Benjamin. Errors in probabilistic reasoning and judgment biases. In B. D. Bernheim, S. DellaVigna, and D. Laibson, editors, *Handbook of Behavioral Economics—Foundations and Applications 2*, pages 69–186. North-Holland, Amsterdam, 2019. doi: 10.1016/bs.hesbe.2018.11.002.
- [17] D. Bernheim and M. D. Whinston. Multimarket contact and collusive behavior. *The RAND Journal of Economics*, 21:1–26, 1990.
- [18] N. Bird-David. Beyond ‘The hunting and gathering mode of subsistence’: Culture-sensitive observations on the Nayaka and other modern hunter-gatherers. *Man*, 27(1):19–44, 1992. ISSN 00251496. doi: 10.2307/2803593.
- [19] C. Boehm. *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Harvard University Press, Cambridge, Mass., 1999. ISBN 0674390318.
- [20] C. Boehm. Bullies: Redefining the Human Free-Rider Problem. In *Darwin’s Bridge: Uniting the Humanities and Sciences*. Oxford University Press, Aug. 2016. ISBN 978-0-19-023121-7. doi: 10.1093/acprof:oso/9780190231217.003.0002.
- [21] M. C. Boerlijst, M. A. Nowak, and K. Sigmund. Equal pay for all prisoners. *American Mathematical Monthly*, 104:303–307, 1997.
- [22] D. Bondarenko, A. Kazankov, D. Khaltourina, and A. Korotayev. Ethnographic atlas xxxi: Peoples of easternmost europe. *Ethnology*, 44(3):261–289, 2005. ISSN 0014-1828.

- [23] R. Boyd. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *Journal of Theoretical Biology*, 136:47–56, 1989.
- [24] R. Boyd and J. Lorberbaum. No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature*, 327:58–59, 1987.
- [25] R. Boyd and P. J. Richerson. Cultural transmission and the evolution of cooperative behavior. *Human Ecology: An Interdisciplinary Journal*, 10(3):325–351, 1982. ISSN 0300-7839.
- [26] R. Boyd and P. J. Richerson. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago, 1985. ISBN 0-226-06931-1.
- [27] R. Boyd and P. J. Richerson. An evolutionary model of social learning: The effects of spatial and temporal variation. In T. Zentall and B. Galef, editors, *Social Learning: Psychological and Biological Perspectives*, pages 29–48. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 1988. ISBN 978-1-317-76688-9. doi: 10.4324/9781315801889.
- [28] R. Boyd and P. J. Richerson. Why does culture increase human adaptability? *Ethology and Sociobiology*, 16(2):125–143, 1995. ISSN 0162-3095. doi: 10.1016/0162-3095(94)00073-G.
- [29] R. Boyd and P. J. Richerson. Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533):3281–3288, 2009. ISSN 0962-8436.
- [30] R. Boyd and P. J. Richerson. Large-scale cooperation in small-scale foraging societies. *Evolutionary Anthropology*, 31(4):175–198, 2022. ISSN 1060-1538.
- [31] K. Brauchli, T. Killingback, and M. Doebeli. Evolution of cooperation in spatially structured populations. *Journal of Theoretical Biology*, 200:405–417, 1999.
- [32] A. S. Brooks, J. E. Yellen, R. Potts, A. K. Behrensmeier, A. L. Deino, D. E. Leslie, S. H. Ambrose, J. R. Ferguson, F. d'Errico, A. M. Zipkin, S. Whittaker, J. Post, E. G. Veatch, K. Foecke, and J. B. Clark. Long-distance stone transport and pigment use in the earliest Middle Stone Age. *Science*, 360(6384):90–94, 2018. ISSN 0036-8075.
- [33] M. Brulhart. Trading places: Industrial specialization in the european union. *Journal of Common Market Studies*, 36(3):319–346, 1998. ISSN 0021-9886.
- [34] K. A. Burson, R. P. Larrick, and J. Klayman. Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, 90(1):60–77, 2006. ISSN 0022-3514. doi: 10.1037/0022-3514.90.1.60.

- [35] R. Byrne and A. Whiten. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford Science Publications. Clarendon Press; Oxford University Press, Oxford; New York, 1988. ISBN 0-19-852179-0.
- [36] V. Capraro and M. Perc. Mathematical foundations of moral preferences. *Journal of the Royal Society interface*, 18(175):20200880–20200880, 2021. ISSN 1742-5662. doi: 10.1098/rsif.2020.0880.
- [37] M. Cartwright. Trade in the ancient world, Feb 2019. URL <https://www.worldhistory.org/collection/39/trade-in-the-ancient-world/>.
- [38] L. L. Cavalli-Sforza and M. W. Feldman. *Cultural Transmission and Evolution: A Quantitative Approach*. Monographs in population biology ; 16. Princeton University Press, Princeton, N.J., 1981. ISBN 0-691-08280-4.
- [39] J.-K. Choi and S. Bowles. Coevolution of parochial altruism and war. *Science (American Association for the Advancement of Science)*, 318(5850):636–640, 2007. ISSN 0036-8075.
- [40] J. Cieřlik, E. Kaciak, and A. van Stel. Country-level determinants and consequences of overconfidence in the ambitious entrepreneurship segment. *International small business journal*, 36(5):473–499, 2018. ISSN 0266-2426.
- [41] M. Collard, A. Ruttle, B. Buchanan, and M. J. O’Brien. Population size and cultural evolution in nonindustrial food-producing societies. *PLOS ONE*, 8(9): e72628–e72628, 2013. ISSN 1932-6203.
- [42] T. G. Conley. Gmm estimation with cross sectional dependence. *Journal of econometrics*, 92(1):1–45, 1999.
- [43] A. Corner and U. Hahn. Normative theories of argumentation: Are some norms better than others? *Synthese*, 190(16):3579–3610, 2012. ISSN 0039-7857. doi: 10.1007/s11229-012-0211-y.
- [44] A. Costinot and D. Donaldson. Ricardo’s theory of comparative advantage: old idea, new evidence. *American Economic Review*, 102(3):453–58, 2012.
- [45] P. Dal Bo and G. R. Frechette. The evolution of cooperation in infinitely repeated games: Experimental evidence. *American Economic Review*, 101(1): 411–429, 2011. ISSN 0002-8282.
- [46] P. Dal Bo and G. R. Fr’echette. Strategy choice in the infinitely repeated prisoner’s dilemma. *American Economic Review*, 109:3929–3952, 2019.

- [47] G. de Francesco. *The Power of the Charlatan*. Harvard College Library history of science project, 00655. Yale University Press, New Haven, CT, 1939.
- [48] S. DellaVigna and U. Malmendier. Paying not to go to the gym. *American Economic Review*, 96(3):694–719, 2006. ISSN 0002-8282. doi: 10.1257/aer.96.3.694.
- [49] S. Dev and F. Riede. Quantitative functional analysis of late glacial projectile points from northern europe. *Lithics*, 33:40–55, 01 2012.
- [50] A. K. Dixit and J. E. Stiglitz. Monopolistic competition and optimum product diversity. *American Economic Review*, 67(3):297–308, 1977. ISSN 0002-8282.
- [51] N. F. Dixon. *On the Psychology of Military Incompetence*. Cape, London, 1976. ISBN 0-224-01161-8.
- [52] M. Doebeli and C. Hauert. Models of cooperation based on the prisoner’s dilemma and the snowdrift game. *Ecology Letters*, 8:748–766, 2005.
- [53] K. Donahue, O. Hauser, M. Nowak, and C. Hilbe. Evolving cooperation in multichannel games. *Nature Communications*, 11:3885, 2020.
- [54] J. L. Doob. Application of the theory of martingales. In *Actes du Colloque International Le Calcul des Probabilités et ses applications (Lyon, 28 Juin – 3 Juillet, 1948)*, pages 23–27. Paris CNRS, Paris, 1949.
- [55] P. H. Egger and A. Lassmann. The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. *The Economic Journal*, 125(584):699–745, 2015.
- [56] J. Ehrlinger, K. Johnson, M. Banner, D. Dunning, and J. Kruger. Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1):98–121, 2008. ISSN 0749-5978. doi: 10.1016/j.obhdp.2007.05.002.
- [57] S. Faurby, M. Davis, R. Ø. Pedersen, S. D. Schowanek, A. Antonelli, and J.-C. Svenning. Phylacine 1.2: the phylogenetic atlas of mammal macroecology. *Ecology*, 99(11):2626, 2018.
- [58] S. E. Fick and R. J. Hijmans. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, 37(12):4302–4315, 2017.
- [59] G. Fischer, F. O. Nachtergaele, H. van Velthuizen, F. Chiozza, G. Francheschini, M. Henry, D. Muchoney, and S. Tramberend. *Global agro-ecological zones (gaez v4)-model documentation*. 2021.

- [60] J. Frankel and D. Romer. Does trade cause growth? *American Economic Review*, 89(3):379–399, 1999. ISSN 0002-8282.
- [61] M. R. Frean. The prisoner’s dilemma without synchrony. *Proceedings of the Royal Society B*, 257:75–79, 1994.
- [62] D. A. Freedman. On the Asymptotic Behavior of Bayes’ Estimates in the Discrete Case. *Annals of Mathematical Statistics*, 34(4):1386–1403, 1963. ISSN 0003-4851.
- [63] D. Fry, C. Keith, and P. Söderberg. Social complexity, inequality and war before farming: Congruence of comparative forager and archaeological data, pages 303–320. McDonald Institute for Archaeological Research, 2020.
- [64] D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *Journal of Economic Theory*, 131:251–262, 2006.
- [65] D. Fudenberg and E. Maskin. Evolution and cooperation in noisy repeated games. *American Economic Review*, 80(2):274–279, 1990. ISSN 0002-8282.
- [66] D. Fudenberg, D. G. Rand, and A. Dreber. Slow to anger and fast to forgive: Cooperation in an uncertain world. *American Economic Review*, 102(2):720–749, 2012. ISSN 0002-8282.
- [67] M. Galanter. *Cults: Faith, Healing, and Coercion*. Oxford University Press, New York, 1989. ISBN 0-19-505631-0.
- [68] J. Garc’ia and A. Traulsen. The structure of mutations and the evolution of cooperation. *PLOS ONE*, 7:e35287, 2012.
- [69] J. Garc’ia and M. van Veelen. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory*, 161:161–189, 2016.
- [70] J. Garc’ia and M. van Veelen. No strategy can win in the repeated prisoner’s dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI*, 5:102, 2018.
- [71] G. Gigerenzer. *Adaptive Thinking: Rationality in the Real World*. Evolution and Cognition. Oxford University Press, Oxford; New York, 2000. ISBN 0-19-513622-5.
- [72] G. Gigerenzer and P. M. Todd. *Simple Heuristics that Make Us Smart*. Oxford University Press, New York, 1999. ISBN 0-19-512156-2.

- [73] M. Gladwell. *Talking to Strangers: What We Should Know about the People We Don't Know*. Allen Lane, London, 2019. ISBN 978-0-241-35157-4.
- [74] N. Glynatsi and V. Knight. Using a theory of mind to find best responses to memory-one strategies. *Scientific Reports*, 10:1–9, 2020.
- [75] N. Glynatsi and V. Knight. A bibliometric study of research topics, collaboration and centrality in the field of the Iterated Prisoner's Dilemma. *Humanities and Social Sciences Communications*, 8:45, 2021.
- [76] J. Gray. A corrected ethnographic atlas. *World Cultures*, 10(1):24–85, 1999.
- [77] J. Grujic, C. Gracia-Lazaro, M. Milinski, D. Semmann, A. Traulsen, J. A. Cuesta, Y. Moreno, and A. Sanchez. A comparative analysis of spatial prisoner's dilemma experiments: Conditional cooperation and payoff irrelevance. *Scientific reports*, 4(1):4615–4615, 2014. ISSN 2045-2322.
- [78] P. Hammerstein. Why is reciprocity so rare in social animals? A protestant appeal. In *Genetic and cultural evolution of cooperation.*, Dahlem workshop report., pages 83–93. MIT Press, Cambridge, MA, US, 2003. ISBN 0-262-08326-4 (Hardcover).
- [79] D. Hao, Z. Rong, and T. Zhou. Extortion under uncertainty: Zero-determinant strategies in noisy games. *Physical Review E*, 91:052803, 2015.
- [80] A. C. Harberger. Three basic postulates for applied welfare economics: An interpretive essay. *Journal of Economic Literature*, 9(3):785–797, 1971. ISSN 0022-0515.
- [81] M. P. Haselhuhn, D. G. Pope, M. E. Schweitzer, and P. Fishman. The impact of personal experience on behavior: Evidence from video-rental fines. *Management Science*, 58(1):52–61, 2012. ISSN 0025-1909. doi: 10.1287/mnsc.1110.1367.
- [82] D. A. Hastings, P. K. Dunbar, G. M. Elphinstone, M. Bootz, H. Murakami, H. Maruyama, H. Masaharu, P. Holland, J. Payne, N. A. Bryant, T. L. Logan, J.-P. Muller, G. Schreier, and J. S. MacDonald. The global land one-kilometer base elevation (globe) digital elevation model, version 1.0, 1999. Digital data base on the World Wide Web (URL: <http://www.ngdc.noaa.gov/mgg/topo/globe.html>) and CD-ROMs.
- [83] C. Hauert and M. Doebeli. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature*, 428:643–646, 2004.
- [84] C. Hauert and H. G. Schuster. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society B*, 264:513–519, 1997.

- [85] D. E. Haun, A. Zeringue, A. Leach, and A. Foley. Assessing the competence of specimen-processing personnel. *Laboratory Medicine*, 31(11):633–637, 2000. ISSN 0007-5027. doi: 10.1309/8Y66-NCN2-J8NH-U66R.
- [86] K. Head and T. Mayer. Gravity, market potential and economic development. *Journal of economic geography*, 11(2):281–294, 2011. ISSN 1468-2702.
- [87] J. Henrich. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of economic behavior & organization*, 53(1):3–35, 2004. ISSN 0167-2681.
- [88] J. Henrich. Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses—the tasmanian case. *American antiquity*, 69(2):197–214, 2004. ISSN 0002-7316.
- [89] J. Henrich. The evolution of costly displays, cooperation and religion: credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, 30(4):244–260, 2009. ISSN 1090-5138. doi: 10.1016/j.evolhumbehav.2009.03.005.
- [90] J. Henrich. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press, Princeton, 2015. ISBN 978-0-691-17843-1.
- [91] J. Henrich and R. Boyd. Division of labor, economic specialization, and the evolution of social stratification. *Current Anthropology*, 49(4):715–724, 2008. ISSN 0011-3204.
- [92] J. Henrich and F. J. Gil-White. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3):165–196, 2001. ISSN 1090-5138. doi: 10.1016/S1090-5138(00)00071-4.
- [93] J. Henrich and M. Muthukrishna. The origins and psychology of human cooperation. *Annual Review of Psychology*, 72(1):207–240, Jan 2021. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev-psych-081920-042106. URL <https://www.annualreviews.org/doi/10.1146/annurev-psych-081920-042106>.
- [94] J. Henrich, S. J. Heine, and A. Norenzayan. The WEIRDest people in the world? *Behavioral and Brain Sciences*, 33(2-3):61–83, 2010. ISSN 0140-525X. doi: 10.1017/S0140525X0999152X.
- [95] C. Hilbe, M. A. Nowak, and K. Sigmund. The evolution of extortion in iterated prisoner’s dilemma games. *Proceedings of the National Academy of Sciences USA*, 110:6913–6918, 2013.

- [96] C. Hilbe, A. Traulsen, and K. Sigmund. Partners or rivals? strategies for the iterated prisoner’s dilemma. *Games and Economic Behavior*, 92:41–52, 2015.
- [97] C. Hilbe, K. Hagel, and M. Milinski. Asymmetric power boosts extortion in an economic experiment. *PLOS ONE*, 11:e0163867, 2016.
- [98] C. Hilbe, L. A. Martinez-Vaquero, K. Chatterjee, and M. A. Nowak. Memory- n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences USA*, 114:4715–4720, 2017.
- [99] C. Hilbe, K. Chatterjee, and M. A. Nowak. Partners and rivals in direct reciprocity. *Nature Human Behaviour*, 2:469–477, 2018.
- [100] M. Hoffman and S. V. Burks. Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics*, 11(1): 315–348, 2020. doi: 10.3982/QE834.
- [101] P. L. Hooper, K. Demps, M. Gurven, D. Gerkey, and H. S. Kaplan. Skills, division of labour and economies of scale among amazonian hunters and south indian honey collectors. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1683):20150008, Dec 2015. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2015.0008. URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2015.0008>.
- [102] P. L. Hooper, K. Demps, M. Gurven, D. Gerkey, and H. S. Kaplan. Skills, division of labour and economies of scale among Amazonian hunters and South Indian honey collectors. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 370(1683):20150008, 2015. ISSN 09628436. doi: 10.1098/rstb.2015.0008.
- [103] N. Humphrey. The social function of intellect. In P. P. G. Bateson and R. A. Hinde, editors, *Growing Points in Ethology*. Cambridge University Press, Cambridge, UK, 1976. ISBN 0-521-21287-1.
- [104] G. Ichinose and N. Masuda. Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology*, 438:61–77, 2018.
- [105] L. A. Imhof and M. A. Nowak. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B*, 277:463–468, 2010.
- [106] M. Insler, A. F. McQuoid, A. Rahman, and K. A. Smith. Fear and loathing in the classroom: Why does teacher quality matter? IZA Discussion Paper No. 14036, 2021. URL <https://ssrn.com/abstract=3767273>.

- [107] M. O. Jackson, T. Rodriguez-Barraquer, and X. Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, pages 1857–1976, 2012.
- [108] R. A. Jansen, A. N. Rafferty, and T. L. Griffiths. A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour*, 5:756–763, 2021. ISSN 2397-3374. doi: 10.1038/s41562-021-01057-0.
- [109] D. D. P. Johnson. *Overconfidence and War: The Havoc and Glory of Positive Illusions*. Harvard University Press, Cambridge, MA, 2004. ISBN 0-674-01576-2.
- [110] H. Jones, N. Diop, I. Askew, and I. Kabore. Female genital cutting practices in Burkina Faso and Mali and their negative health outcomes. *Studies in Family Planning*, 30(3):219–230, 1999. ISSN 0039-3665. doi: 10.1111/j.1728-4465.1999.00219.x.
- [111] R. Jones. The demography of hunters and farmers in tasmania. In D. J. Mulvaney and J. Golson, editors, *Aboriginal Man and Environment in Australia*, pages 271–287. Australian National University Press, Canberra, 1971.
- [112] J. G. Jorgensen. *Western Indians: Comparative Environments, Languages, and Cultures of 172 Western American Indian tribes*. W. H. Freeman, San Francisco, 1980. ISBN 0716711044.
- [113] J. G. Jorgensen. An empirical procedure for defining and sampling culture bearing units in continuous geographic areas. *World Cultures*, 10(2):139–143, 1999.
- [114] J. G. Jorgensen. Codebook for Western Indians data. *World Cultures*, 10(2): 144–293, 1999.
- [115] T. Killingback and M. Doebeli. Spatial evolutionary game theory: Hawks and doves revisited. *Proceedings of the Royal Society B*, 263:1135–1144, 1996.
- [116] T. Killingback and M. Doebeli. The continuous Prisoner’s Dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *American Naturalist*, 160:1–23, 2002.
- [117] S. Kim. Expansion of markets and the geographic distribution of economic activities: The trends in u.s. regional manufacturing structure, 1860–1987. *Quarterly Journal of Economics*, 110(4):881–908, 1995. ISSN 0033-5533.
- [118] K. R. Kirby, R. D. Gray, S. J. Greenhill, F. M. Jordan, S. Gomes-Ng, H.-J. Bibiko, D. E. Blasi, C. A. Botero, C. Bowern, C. R. Ember, D. Leehr, B. S. Low, J. McCarter, W. Divale, and M. C. Gavin. D-place: A global database of

- cultural, linguistic and environmental diversity. *PLOS ONE*, 11(7):e0158391–e0158391, 2016. ISSN 1932-6203. Accessed from Jan. 15, 2023.
- [119] M. A. Kline and R. Boyd. Population size predicts technological complexity in oceania. *Proceedings of the Royal Society. B, Biological sciences*, 277(1693): 2559–2564, 2010. ISSN 0962-8452.
- [120] M. A. Kline, R. Boyd, and J. Henrich. Teaching and the life history of cultural transmission in Fijian villages. *Human Nature*, 24(4):351–374, 2013. ISSN 1045-6767. doi: 10.1007/s12110-013-9180-1.
- [121] A. Korotayev, A. Kazankov, S. Borinskaya, D. Khaltourina, and D. Bondarenko. Ethnographic atlas xxx: Peoples of siberia. *Ethnology*, 43(1):83–92, 2004. ISSN 0014-1828.
- [122] D. P. Kraines and V. Y. Kraines. Pavlov and the prisoner’s dilemma. *Theory and Decision*, 26:47–79, 1989.
- [123] M. Kremer. Population growth and technological change: One million b.c. to 1990. *The Quarterly journal of economics*, 108(3):681–716, 1993. ISSN 0033-5533.
- [124] J. Kruger and D. Dunning. Unskilled and unaware of It: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134, 1999. ISSN 0022-3514. doi: 10.1037/0022-3514.77.6.1121.
- [125] P. R. Krugman. Increasing returns, monopolistic competition, and international trade. *Journal of International Economics*, 9(4):469–479, 1979.
- [126] K. N. Laland. *Darwin’s Unfinished Symphony: How Culture Made the Human Mind*. Princeton University Press, Princeton, 2017. ISBN 978-0-691-15118-2.
- [127] R. B. Lee. Hunter-gatherers and human evolution: New light on old debates. *Annual Review of Anthropology*, 47(1):513–531, 2018. ISSN 0084-6570.
- [128] S. Lew-Levy and A. H. Boyette. Evidence for the adaptive learning function of work and work-themed play among Aka forager and Ngandu farmer children from the Congo Basin. *Human Nature*, 29(2):157–185, 2018. ISSN 1045-6767. doi: 10.1007/s12110-018-9314-6.
- [129] S. Lew-Levy, R. Reckin, N. Lavi, J. Cristóbal-Azkarate, and K. Ellis-Davies. How do hunter-gatherer children learn subsistence skills? A meta-ethnographic review. *Human Nature*, 28:367–394, 2017. doi: 10.1007/s12110-017-9302-2.

- [130] S. Lew-Levy, E. J. Ringen, A. N. Crittenden, I. A. Mabulla, T. Broesch, and M. A. Kline. The life history of learning subsistence skills among Hadza and BaYaka foragers from Tanzania and the Republic of Congo. *Human Nature*, 32(1):16–47, 2021. doi: 10.1007/s12110-021-09386-9.
- [131] S. Lichtenstein and B. Fischhoff. Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20(2):159 – 183, 1977. ISSN 0030-5073. doi: 10.1016/0030-5073(77)90001-0.
- [132] S. Lindenbaum. Kuru, prions, and human affairs: Thinking about epidemics. *Annual Review of Anthropology*, 30(1):363–385, 2001. ISSN 0084-6570. doi: 10.1146/annurev.anthro.30.1.363.
- [133] M. Lipson, E. A. Sawchuk, J. C. Thompson, J. Oppenheimer, C. A. Tryon, K. L. Ranhorn, K. M. De Luna, K. A. Sirak, I. Olalde, S. H. Ambrose, J. W. Arthur, K. J. W. Arthur, G. Ayodo, A. Bertacchi, J. I. Cerezo-Roman, B. J. Culleton, M. C. Curtis, J. Davis, A. O. Gidna, A. Hanson, P. Kaliba, M. Kantonogo, A. Kwekason, M. F. Laird, J. Lewis, A. Z. P. Mabulla, F. Mapemba, A. Morris, G. Mudenda, R. Mwafulirwa, D. Mwangomba, E. Ndiema, C. Ogola, F. Schilt, P. R. Willoughby, D. K. Wright, A. Zipkin, R. Pinhasi, D. J. Kennett, F. K. Manthi, N. Rohland, N. Patterson, D. Reich, and M. E. Prendergast. Ancient dna and deep population structure in sub-saharan african foragers. *Nature*, 603(7900):290–296, 2022. ISSN 0028-0836.
- [134] M. Lisi, G. Mongillo, G. Milne, T. Dekker, and A. Gorea. Discrete confidence levels revealed by sequential decisions. *Nature Human Behaviour*, 5(2):273–280, 2021. ISSN 2397-3374. doi: 10.1038/s41562-020-00953-1.
- [135] M. P. Lombardo. Mutual restraint in tree swallows: A test of the tit for tat model of reciprocity. *Science*, 227:1363–1365, 1985.
- [136] J. Lorberbaum. No strategy is evolutionary stable in the repeated Prisoner’s Dilemma. *Journal of Theoretical Biology*, 168:117–130, 1994.
- [137] J. P. Lorberbaum, D. E. Bohning, A. Shastri, and L. E. Sine. Are there really no evolutionarily stable strategies in the iterated prisoner’s dilemma? *Journal of Theoretical Biology*, 214:155–169, 2002.
- [138] G. D. A. MacDougall. British and american exports: A study suggested by the theory of comparative costs. part i. *The Economic journal (London)*, 61(244): 697–724, 1951. ISSN 0013-0133.
- [139] D. MacKenzie. Computer-related accidental death: An empirical exploration. *Science and Public Policy*, 21(4):233–248, 1994. ISSN 0302-3427. doi: 10.1093/spp/21.4.233.

- [140] U. Malmendier and G. Tate. CEO overconfidence and corporate investment. *Journal of Finance*, 60(6):2661–2700, 2005. ISSN 0022-1082. doi: 10.1111/j.1540-6261.2005.00813.x.
- [141] A. Mamiya and G. Ichinose. Zero-determinant strategies under observation errors in repeated games. *Physical Review E*, 102:032115, 2020.
- [142] A. Mantovani, G. Rossini, and P. Zanghieri. Country size and the price of tradeables: is there any relationship beyond wishful thinking? (443), 2002. doi: 10.6092/unibo/amsacta/4855. URL <http://hdl.handle.net/10419/159284>.
- [143] F. Marlowe. *The Hadza: Hunter-gatherers of Tanzania. Origins of human behavior and culture ; 3.* University of California Press, Berkeley, 2010. ISBN 9780520253414.
- [144] C. F. Martin, R. Bhui, P. Bossaerts, T. Matsuzawa, and C. Camerer. Chimpanzee choice rates in competitive games match equilibrium game theory predictions. *Scientific Reports*, 4(1):5182–5182, 2014. ISSN 2045-2322.
- [145] S. Mathew, R. Boyd, and M. Van Veelen. Human Cooperation among Kin and Close Associates May Require Enforcement of Norms by Third Parties. In *Cultural Evolution: Society, Technology, Language, and Religion*. The MIT Press, Nov. 2013. ISBN 978-0-262-01975-0. doi: 10.7551/mitpress/9780262019750.003.0003. URL <https://doi.org/10.7551/mitpress/9780262019750.003.0003>.
- [146] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, 1982.
- [147] A. McAvoy and C. Hauert. Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences*, 113:3573–3578, 2016.
- [148] A. McAvoy and C. Hauert. Autocratic strategies for alternating games. *Theoretical Population Biology*, 113:13–22, 2016.
- [149] A. McAvoy and M. A. Nowak. Reactive learning strategies for iterated games. *Proceedings of the Royal Society A*, 475:20180819, 2019.
- [150] R. McKay and C. Efferson. The subtleties of error management. *Evolution and Human Behavior*, 31(5):309–319, 2010. ISSN 1090-5138. doi: 10.1016/j.evolhumbehav.2010.04.005.
- [151] J. McNamara and A. Houston. The application of statistical decision theory to animal behaviour. *Journal of Theoretical Biology*, 85(4):673–690, 1980. ISSN 0022-5193. doi: 10.1016/0022-5193(80)90265-9.

- [152] A. P. Melis and D. Semmann. How is human cooperation different? *Philosophical Transactions of the Royal Society B*, 365:2663–2674, 2010.
- [153] J. Melitz. Language and foreign trade. *European Economic Review*, 52(4): 667–699, 2008.
- [154] A. B. Migliano and L. Vinicius. The origins of human cumulative culture: from the foraging niche to collective intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1843):20200317, Jan 2022. ISSN 0962-8436, 1471-2970. doi: 10.1098/rstb.2020.0317. URL <https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0317>.
- [155] M. Milinski. Tit For Tat in sticklebacks and the evolution of cooperation. *Nature*, 325:433–435, 1987.
- [156] P. Molander. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution*, 29:611–618, 1985.
- [157] D. A. Moore and P. J. Healy. The trouble with overconfidence. *Psychological Review*, 115(2):502–517, 2008. ISSN 0033-295X. doi: 10.1037/0033-295X.115.2.502.
- [158] G. P. Murdock. *Ethnographic Atlas, Installments i-xxvii*. *Ethnology*, 1-10, 1962-1971.
- [159] Y. Muroyama. Mutual reciprocity of grooming in female japanese macaques (*macaca fuscata*). *Behaviour*, 119:161–170, 1991.
- [160] M. Muthukrishna and J. Henrich. Innovation in the collective brain. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 371(1690):20150192, 2016. ISSN 0962-8436. doi: 10.1098/rstb.2015.0192.
- [161] M. Muthukrishna and J. Henrich. A problem in theory. *Nature Human Behaviour*, 3(3):221–229, 2019. ISSN 2397-3374. doi: 10.1038/s41562-018-0522-1.
- [162] M. Muthukrishna, M. Doebeli, M. Chudek, and J. Henrich. The cultural brain hypothesis: How culture drives brain expansion, sociality, and life history. *PLOS Computational Biology*, 14(11):e1006504, Nov 2018. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1006504. URL <http://dx.plos.org/10.1371/journal.pcbi.1006504>.
- [163] M. Muthukrishna, M. Doebeli, M. Chudek, and J. Henrich. The Cultural Brain Hypothesis: How culture drives brain expansion, sociality, and life history. *PLoS Computational Biology*, 14(11):e1006504–e1006504, 2018. ISSN 1553-734X. doi: 10.1371/journal.pcbi.1006504.

- [164] M. Muthukrishna, J. Henrich, W. Toyokawa, T. Hamamura, T. Kameda, and S. J. Heine. Overconfidence is universal? elicitation of genuine overconfidence (ego) procedure reveals systematic differences across domain, task knowledge, and incentives in four populations. *PLOS ONE*, 13(8):e0202288–e0202288, 2018. ISSN 1932-6203.
- [165] H. H. Nax and M. Perc. Directional learning and the provisioning of public goods. *Scientific Reports*, 5(1):8010–8010, 2015. ISSN 2045-2322. doi: 10.1038/srep08010.
- [166] M. A. Nowak. Five rules for the evolution of cooperation. *Science*, 314:1560–1563, 2006.
- [167] M. A. Nowak. *Evolutionary dynamics*. Harvard University Press, Cambridge, MA, 2006.
- [168] M. A. Nowak and R. M. May. Evolutionary games and spatial chaos. *Nature*, 359:826–829, 1992.
- [169] M. A. Nowak and K. Sigmund. Tit for tat in heterogeneous populations. *Nature*, 355:250–253, 1992.
- [170] M. A. Nowak and K. Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner’s dilemma game. *Nature*, 364:56–58, 1993.
- [171] M. A. Nowak and K. Sigmund. The alternating prisoner’s dilemma. *Journal of Theoretical Biology*, 168:219–226, 1994.
- [172] M. A. Nowak and K. Sigmund. Invasion dynamics of the finitely repeated prisoner’s dilemma. *Games and Economic Behavior*, 11:364–390, 1995.
- [173] M. A. Nowak and K. Sigmund. Evolution of indirect reciprocity by image scoring. *Nature*, 393(6685):573–577, 1998. ISSN 0028-0836.
- [174] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428:646–650, 2004.
- [175] OECD. *Open Markets Matter: The Benefits of Trade and Investment Liberalisation*. OECD Publishing, Paris, 1998. ISBN 1-280-03309-6. p. 10, 39.
- [176] OECD. *OECD Science, Technology and Industry Scoreboard, 2011*. OECD Science, Technology and Industry Scoreboard. 2011. Paris, 2011. ISBN 1-283-29113-4. p. 177.
- [177] OpenAI. OpenAI charter, September 2020. URL <https://openai.com/charter/>.

- [178] P. S. Park. The evolution of cognitive biases in human learning. *Journal of Theoretical Biology*, 541:111031–111031, 2022. ISSN 0022-5193.
- [179] P. S. Park and C. Hilbe. The alternating prisoner’s dilemma with cognitive costs. In progress, 2022.
- [180] P. S. Park, M. A. Nowak, and C. Hilbe. Cooperation in alternating interactions with memory constraints – source code and data, 2022. URL <http://dx.doi.org/10.17605/osf.io/v5hgd>. OSF.
- [181] P. S. Park, M. A. Nowak, and C. Hilbe. Cooperation in alternating interactions with memory constraints. *Nature Communications*, 13(1):737–737, 2022. ISSN 2041-1723.
- [182] P. S. Park, V. Savitskiy, and J. Henrich. A theory of specialization, exchange, and innovation in human groups. *Cultural Evolution Society Conference*, 2022.
- [183] Y. J. Park and L. Santos-Pinto. Overconfidence in tournaments: Evidence from the field. *Theory and Decision*, 69(1):143–166, 2010. ISSN 0040-5833. doi: 10.1007/s11238-010-9200-0.
- [184] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L. M. Floría, and Y. Moreno. Evolutionary dynamics of group interactions on structured populations: A review. *Journal of The Royal Society Interface*, 10:20120997, 2013.
- [185] F. L. Pinheiro, V. V. Vasconcelos, F. C. Santos, and J. M. Pacheco. Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Computational Biology*, 10:e1003945, 2014.
- [186] S. Pinker. The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107(Supplement 2): 8993–8999, 2010. ISSN 0027-8424. doi: 10.1073/pnas.0914630107.
- [187] T. J. Pleskac and J. R. Busemeyer. Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3): 864–901, 2010. ISSN 0033-295X. doi: 10.1037/a0019737.
- [188] W. Poundstone. *Prisoner’s dilemma*. Anchor Books, New York, 1st anchor books ed. edition, 1993. ISBN 038541580X. p. 118.
- [189] A. Powell, S. Shennan, and M. G. Thomas. Late pleistocene demography and the appearance of modern human behavior. *Science*, 324(5932):1298–1301, 2009. ISSN 0036-8075.
- [190] W. H. Press and F. D. Dyson. Iterated prisoner’s dilemma contains strategies that dominate any evolutionary opponent. *PNAS*, 109:10409–10413, 2012.

- [191] E. Proto, A. Rustichini, and A. Sofianos. Intelligence, personality, and gains from cooperation in repeated interactions. *The Journal of political economy*, 127(3):1351–1390, 2019. ISSN 0022-3808.
- [192] A. Rapoport and A. M. Chammah. *Prisoner’s Dilemma*. University of Michigan Press, Ann Arbor, 1965.
- [193] S. M. Reader, Y. Hager, and K. N. Laland. The evolution of primate general and cultural intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1017–1027, 2011. ISSN 0962-8436. doi: 10.1098/rstb.2010.0342.
- [194] S. Redding and A. J. Venables. Economic geography and international inequality. *Journal of International Economics*, 62(1):53–82, 2004. ISSN 0022-1996.
- [195] J. G. Reiter, C. Hilbe, D. G. Rand, K. Chatterjee, and M. A. Nowak. Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness. *Nature Communications*, 9:555, 2018.
- [196] D. Ricardo. *On the principles of political economy*. J. Murray London, 1821.
- [197] P. Richerson, R. Baldini, A. V. Bell, K. Demps, K. Frost, V. Hillis, S. Mathew, E. K. Newton, N. Naar, L. Newson, C. Ross, P. E. Smaldino, T. M. Waring, and M. Zefferman. Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences*, 39:e30–e30, 2016. ISSN 0140-525X.
- [198] F. Riede. The laacher see-eruption (12,920 bp) and material culture change at the end of the allerød in northern europe. *Journal of archaeological science*, 35(3):591–599, 2008. ISSN 0305-4403.
- [199] P. M. Romer. Endogenous technological change. *Journal of Political Economy*, 98(5):S71–S102, 1990. ISSN 0022-3808.
- [200] M. J. Rowlands. The archaeological interpretation of prehistoric metalworking. *World archaeology*, 3(2):210–224, 1971. ISSN 0043-8243.
- [201] G. D. Salali, N. Chaudhary, J. Thompson, O. M. Grace, X. M. van der Burgt, M. Dyble, A. E. Page, D. Smith, J. Lewis, R. Mace, L. Vinicius, and A. B. Migliano. Knowledge-sharing networks in hunter-gatherers and the evolution of cumulative culture. *Current Biology*, 26(18):2516–2521, 2016. ISSN 0960-9822. doi: 10.1016/j.cub.2016.07.015.
- [202] G. D. Salali, N. Chaudhary, J. Bouer, J. Thompson, L. Vinicius, and A. B. Migliano. Development of social learning and play in BaYaka hunter-gatherers

- of Congo. *Scientific Reports*, 9(1):11080–10, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-47515-8.
- [203] C. Sanchez and D. Dunning. Overconfidence among beginners: Is a little learning a dangerous thing? *Journal of Personality and Social Psychology*, 114(1): 10–28, 2018. ISSN 0022-3514. doi: 10.1037/pspa0000102.
- [204] C. Sanchez and D. Dunning. Decision fluency and overconfidence among beginners. *Decision*, 7(3):225–237, 2020. ISSN 2325-9965. doi: 10.1037/dec0000122.
- [205] L. J. Savage. *The Foundations of Statistics*. Dover Publications, New York, second revised ed. edition, 1972. ISBN 0-486-62349-1.
- [206] J. A. Scheinkman and W. Xiong. Overconfidence and speculative bubbles. *Journal of Political Economy*, 111(6):1183–1220, 2003. ISSN 1537-534X. doi: 10.1086/378531.
- [207] W. Scheirer. A pandemic of bad science. *Bulletin of the Atomic Scientists*, 76(4):175–184, 2020. ISSN 0096-3402. doi: 10.1080/00963402.2020.1778361.
- [208] E. Schlosser. *Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety*. The Penguin Press, New York, 2013. ISBN 978-1-59420-227-8.
- [209] L. Schmid, K. Chatterjee, C. Hilbe, and M. Nowak. A unified framework of direct and indirect reciprocity. *Nature Human Behaviour*, 5:1292–1302, 2021.
- [210] E. Schniter, M. Gurven, H. S. Kaplan, N. T. Wilcox, and P. L. Hooper. Skill ontogeny among Tsimane forager-horticulturalists. *American Journal of Physical Anthropology*, 158(1):3–18, 2015. ISSN 0002-9483. doi: 10.1002/ajpa.22757.
- [211] E. Schniter, N. T. Wilcox, B. A. Beheim, H. S. Kaplan, and M. Gurven. Information transmission and the oral tradition: Evidence of a late-life service niche for tsimane amerindians. *Evolution and Human Behavior*, 39(1):94–105, Jan 2018. ISSN 10905138. doi: 10.1016/j.evolhumbehav.2017.10.006. URL <https://linkinghub.elsevier.com/retrieve/pii/S1090513817301411>.
- [212] B. Sellato. *Nomades et sédentarisation à Bornéo : histoire économique et sociale*. *Etudes insulindiennes-Archipel 9*. Editions de l'école des hautes études en sciences sociales, Paris], 1989. ISBN 271320917X. p. 205.
- [213] R. Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4:25–55, 1975.

- [214] D. Shariatmada. Daniel Kahneman: 'What would I eliminate if I had a magic wand? Overconfidence', Jul 2015. URL <https://www.theguardian.com/books/2015/jul/18/daniel-kahneman-books-interview>.
- [215] K. Sigmund. *The Calculus of Selfishness*. Princeton Univ. Press, Princeton, NJ, 2010.
- [216] K. Sigmund and M. A. Nowak. Evolution of indirect reciprocity. *Nature*, 437(7063):1291–1298, 2005. ISSN 0028-0836.
- [217] M. Singh. The cultural evolution of shamanism. *Behavioral and Brain Sciences*, 41:1–83, 2018. ISSN 0140-525X. doi: 10.1017/S0140525X17001893.
- [218] M. Singh and Z. H. Garfield. Evidence for third-party mediation but not punishment in mentawai justice. *Nature Human Behaviour*, 6(7):930–940, 2022. ISSN 2397-3374.
- [219] M. T. Stark. Ceramic production and community specialization: A kalinga ethnoarchaeological study. *World Archaeology*, 23(1):64–78, 1991. ISSN 0043-8243.
- [220] A. J. Stewart and J. B. Plotkin. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA*, 110:15348–15353, 2013.
- [221] A. J. Stewart and J. B. Plotkin. Collapse of cooperation in evolving games. *Proceedings of the National Academy of Sciences USA*, 111:17558–17563, 2014.
- [222] A. J. Stewart and J. B. Plotkin. The evolvability of cooperation under local and non-local mutations. *Games*, 6:231–250, 2015.
- [223] S. E. Street, A. F. Navarrete, S. M. Reader, and K. N. Laland. Coevolution of cultural intelligence, extended life history, sociality, and brain size in primates. *Proceedings of the National Academy of Sciences*, 114(30):7908–7914, 2017. ISSN 0027-8424. doi: 10.1073/pnas.1620734114.
- [224] G. Szabó and C. Tóke. Evolutionary Prisoner's Dilemma game on a square lattice. *Physical Review E*, 58:69–73, 1998.
- [225] G. Szabó, T. Antal, P. Szabó, and M. Droz. Spatial evolutionary prisoner's dilemma game with three strategies and external constraints. *Physical Review E*, 62:1095–1103, 2000.
- [226] A. Szolnoki and M. Perc. Evolution of extortion in structured populations. *Physical Review E*, 89:022804, 2014.

- [227] A. Szolnoki and M. Perc. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific Reports*, 4:5496, 2014.
- [228] A. Szolnoki, M. Perc, and G. Szab'o. Phase diagrams for three-strategy evolutionary prisoner's dilemma games on regular graphs. *Physical Review E*, 80: 056104, 2009.
- [229] B. Tadesse and R. White. Does cultural distance hinder trade in goods? a comparative study of nine oecd member nations. *Open economies review*, 21(2): 237–261, 2010.
- [230] C. E. Tarnita, E. O. Wilson, and M. A. Nowak. The evolution of eusociality. *Nature*, 466(7310):1057–1062, 2010. ISSN 0028-0836.
- [231] C. Townsend. *Egalitarianism, Evolution of*. The Wiley Blackwell International Encyclopedia of Anthropology, 2018.
- [232] A. Traulsen, J. M. Pacheco, and M. A. Nowak. Pairwise comparison and selection temperature in evolutionary game dynamics. *Journal of Theoretical Biology*, 246:522–529, 2007.
- [233] R. L. Trivers. The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46(1):35–57, 1971. ISSN 0033-5770.
- [234] A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974. ISSN 1095-9203. doi: 10.1126/science.185.4157.1124.
- [235] J. E. Uscinski, C. Klofstad, and M. D. Atkinson. What drives conspiratorial beliefs? The role of informational cues and predispositions. *Political Research Quarterly*, 69(1):57–71, 2016. ISSN 1065-9129. doi: 10.1177/1065912915621621.
- [236] T. J. Valone. Are animals capable of Bayesian updating? An empirical review. *Oikos*, 112(2):252–259, 2006. ISSN 0030-1299. doi: 10.1111/j.0030-1299.2006.13465.x.
- [237] C. P. van Schaik and J. M. Burkart. Social learning and evolution: The cultural intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1008–1016, 2011. ISSN 0962-8436. doi: 10.1098/rstb.2010.0304.
- [238] M. van Veelen, J. Garc'ia, D. G. Rand, and M. A. Nowak. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA*, 109:9929–9934, 2012.

- [239] B. Voelkl et al. Matching times of leading and following suggest cooperation through direct reciprocity during V-formation flight in ibis. *Proceedings of the National Academy of Sciences USA*, 112:2115–2120, 2015.
- [240] N. Wagner. Female genital cutting and long-term health consequences – Nationally representative estimates across 13 countries. *The Journal of Development Studies*, 51(3):226–246, 2015. ISSN 0022-0388. doi: 10.1080/00220388.2014.976620.
- [241] C. Wedekind and M. Milinski. Human cooperation in the simultaneous and the alternating prisoner’s dilemma: Pavlov versus generous tit-for-tat. *Proceedings of the National Academy of Sciences USA*, 93:2686–2689, 1996.
- [242] B. A. Weinberg, M. Hashimoto, and B. M. Fleisher. Evaluating teaching in higher education. *Journal of Economic Education*, 40(3):227–261, 2009. ISSN 0022-0485. doi: 10.3200/JECE.40.3.227-261.
- [243] S. West, A. Griffin, and A. Gardner. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, 20(2):415–432, 2007. ISSN 1010-061X.
- [244] A. Whiten and C. P. van Schaik. The evolution of animal ‘cultures’ and social intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480):603–620, 2007. ISSN 0962-8436. doi: 10.1098/rstb.2006.1998.
- [245] G. Wild and A. Traulsen. The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247:382–390, 2007.
- [246] G. S. Wilkinson. Reciprocal food-sharing in the vampire bat. *Nature*, 308:181–184, 1984.
- [247] D. S. Wilson and E. Sober. Reintroducing group selection to the human behavioral sciences. *Behavioral and Brain Sciences*, 17(4):585–608, 1994. ISSN 0140-525X.
- [248] B. Wu, C. S. Gokhale, L. Wang, and A. Traulsen. How small are small mutation rates? *Journal of Mathematical Biology*, 64:803–827, 2012.
- [249] B. Wu, B. Bauer, T. Galla, and A. Traulsen. Fitness-based models and pairwise comparison models of evolutionary games are typically different—even in unstructured populations. *New Journal of Physics*, 17:023043, 2015.
- [250] Z.-X. Wu and Z. Rong. Boosting cooperation by involving extortion in spatial prisoner’s dilemma games. *Physical Review E*, 90:062102, 2014.

- [251] T. Yarkoni. The generalizability crisis. *Behavioral and Brain Sciences*, pages 1–37, 2020. doi: 10.1017/S0140525X20001685.
- [252] A. Zador and Y. LeCun. Don't fear the terminator. *Scientific American*, Sept. 2019. URL <https://blogs.scientificamerican.com/observations/dont-fear-the-terminator/>.
- [253] B. M. Zagorsky, J. G. Reiter, K. Chatterjee, and M. A. Nowak. Forgiver triumphs in alternating prisoner's dilemma. *PLOS ONE*, 8:e80814, 2013.
- [254] H. Zhang. Errors can increase cooperation in finite populations. *Games and Economic Behavior*, 107:203–219, 2018.