The evolution of social learning

DISSERTATION

zur Erlangung des akademischen Grades

Dr. rer. nat.
im Fach Biologie

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von
Dipl.-Biol. Benjamin Vinh Bossan

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Stefan Hecht PhD

Gutachter:
1. Peter Hammerstein
2. Hanspeter Herzel
3. Robert Boyd

eingereicht am: 07.05.2013
Tag der mündlichen Prüfung: 21.10.2013

Abstract

Humans differ most from other animals in that their lives are shaped by many cultural practices. Having cultural traits allowed human populations to grow considerably in a short time and to conquer almost all terrestrial habitats on Earth. Cultural traits are not inborn but are instead transmitted between humans through social learning – no individual could build a fully functional kayak without learning from others. Concluding that cultural evolution is thus a separate process from genetic evolution would, however, be rash. The latter has endowed humans with the possibility to learn from others in the first place and prepared learning to make it especially adaptive. To find out what makes humans unique, cultural and genetic evolution, therefore, have to be studied in concert. Although nobody doubts that evolution gave rise to social learning and that the resulting cultural practices serve an adaptive purpose, theoretical works have shown that simple forms of social learning do not improve human adaptedness. This finding contradicts the observations and thus implies that the understanding of social learning is incomplete. Several authors have proposed solutions to this paradox but we find the paradox to be more resilient than is believed. We propose new forms of social learning that could solve it, albeit only under very narrow circumstances. Furthermore, we argue for a new perspective on social learning and, consequently, for a different framework that allows for more realistic learning models. We suggest that the study of the evolutionary origin of social learning should be given equal weight as the study of the evolutionary origin of cooperation and illustrate this by elaborating on the impact of social learning on modern societies and market behaviors in general, and on financial crises specifically.

Zusammenfassung

Menschen unterscheiden sich von anderen Tieren insbesondere dadurch, dass ihr Alltag durch vielfältige kulturelle Praktiken bestimmt wird. Diese erlaubten es dem Menschen, fast alle terrestrischen Habitate auf der Erde in hoher Dichte zu besiedeln. Kulturelle Merkmale werden nicht genetisch vererbt, sondern durch soziales Lernen zwischen Menschen übertragen – niemand könnte ohne den vorhandenen Wissensbeitrag anderer ein funktionstüchtiges Kajak bauen. Daraus zu schließen, kulturelle und genetische Evolution seien komplett getrennt zu behandeln, wäre allerdings vorschnell. Genetische Evolution hat es überhaupt erst erlaubt, von anderen in adaptiver Weise zu lernen. Kulturelle und genetische Evolution müssen zusammen betrachtet werden, um die Einzigartigkeit des Menschen zu verstehen. Der offensichtlich vorhandene adaptive Nutzen sozialen Lernens konnte in theoretischen Arbeiten allerdings nicht repliziert werden. Das deutet darauf hin, dass das Verständnis über die Funktionsweise sozialen Lernens noch unvollständig ist. Zwar wurden mögliche Lösungen für dieses Paradox vorgeschlagen, aber unser Modell zeigt, dass sich der Widerspruch hartnäckiger hält als geglaubt. Wir analysieren zwar neue soziale Lernstrategien, die den Widerspruch lösen können, doch erfolgt das nur unter sehr beschränkten Bedingungen. Außerdem treten wir für eine neue Sicht auf soziales Lernen ein und damit einhergehend für einen Modellierungsansatz, der Lernformen in realistischerer Weise berücksichtigt. Die Untersuchung des evolutionären Ursprungs sozialen Lernens sollte den gleichen Stellenwert haben wie jene des evolutionären Ursprungs kooperativen Verhaltens. Dass dies sinnvoll wäre, belegen wir, indem wir zeigen, welchen Einfluss soziales Lernen sogar auf moderne Gesellschaften und Volkswirtschaften hat und wie es beispielsweise hilft, Finanzkrisen besser zu verstehen.

Contents

1 Introduction to models of social learning

This chapter has two goals, frist, to introduce the reader to some basic approaches of how to model social learning, second, to present some of the most important findings from social learning theory, most notably Rogers' paradox [23, 146]. These goals can be reached in one step by presenting the model that led to this important finding, while also explaining potential modifications of the model. Instead of summarizing the important literature on social learning upfront, we discuss it when appropriate during our journey. We present the basics of our own model, which we will further develop over the course of the following chapters, and explain why we made certain choices that differ from existing approaches. Before we begin, however, we first need to establish the context of this work.

1.1 Why study social learning

1.1.1 From the calculus of planetary motions ...

> Look round our world; behold the chain of love
> Combining all below and all above.[1]

For much of mankind's history, we stared up at the stars and saw with endless awe their constant movements on the firmament. We marveled at what enigmatic motions act the souls of the eternal beings who above us circle in eternity. When people came to comprehend the implications of the works of Newton it was like a thunder waking up a sleeping man from his deceitful dreams. For Isaac Newton showed that same mathematic laws that govern earthly matters also govern movements of celestial bodies. It is impossible to underrate the gravity of this; these realms once thought as separated were combined at last.

> Whether with reason or with instinct blest,
> Know, all enjoy that power which suits them best;
> To bliss alike by that direction tend,
> And find the means proportion'd to their end.[2]

[1] Alexander Pope, 1734, An Essay on Man, Epistle III, 7–8.
[2] Ibid., 79–82.

One century later Condorcet attempted to envision mathematic laws describing social matters and behaviors. Although accomplishments of social sciences when compared to physics have been modest, aspirations to achieve this great ideal to this day persist. However, what could be the science that will decipher what makes humans spin, that will achieve to mark the laws of culture and decision making?

> God, in the nature of each being, founds
> Its proper bliss, and sets its proper bounds[3]

The organisms populating Earth have one characteristic that distinguishes them from all inanimate objects. And that distinction is their almost perfect adaptation to the world surrounding them and their ability to navigate it. This wondrous observation led the theologian Paley to insist that this must be the work of a divine Designer. (see [39]). If, wandering in a forest, one were to find such a perfect apparatus as a watch, then would one not be prompted to believe in the existence of a maker of said watch? Similarly, recognizing how absurdly well the organisms roaming earth are built, then would one not have to believe in the existence of an organism-maker?

But Darwin then proposed an argument that made presuming any supreme being obsolete while still explaining the astounding adaptations [38]: All organisms are evolving, and by natural selection, adaptation is achieved without the need for plan nor purpose. And this is how we understand today the origin of adaptation of all living beings.

Yet not just bodily aspects but all kinds of behaviors could be explained beautifully by natural selection; this constitutes the cornerstone of what is called behavioral ecology [106]. Behaviors that were thusly studied include parenting in birds, ritualized fights of deer, and life in colonies of ants and bees. Without a doubt, behaviors of Man, including social ones, are not excluded from the grip of the Darwinian logic.

> See him from Nature rising slow to Art!
> To copy instinct then was reason's part;
> Thus then to Man the voice of Nature spake–
> 'Go, from the creatures thy instructions take' [4]

The adaptations of the mind are not as fix as laws of nature. Not seldom they require Man to change his ways according to the circumstances. Thus learning is a crucial part of what really defines him. Some innovations are assuredly made during a human's lifetime but the lion's share of traits he learns are learned observing others. Although Man may not often copy what he sees in other creatures, he may still be quite adept at imitating other

[3] Ibid., 109–110.
[4] Ibid., 169–170.

people. Through imitation, augmentation, Man was capable of learning traits that in a lifetime no Man could have learned alone.

But if mankind is able to acquire many a copied trait, well does that not imply that the Darwinian laws apply no more?

Don't judge too fast, the way Man learns has been prepared and biased by his evolutionary past. To study how Man aggregates and transmits cultural traits is thus to study the Darwinian logic of these faculties. One great endeavor lying straight before us is to see if, similar to Newton's finding, one great mathematic framework will allow us to consolidate the field of human evolution with the field of human culture.

> Great Nature spoke; observant men obey'd;
> Cities were built, societies were made [5]

Apart from Man, there is no animal that forms complex societies whose building blocks are socially transmitted traits. And this may very well present the only property that makes Man so unique among all animals that populate this planet. To know what makes us human, therefore, will require us to understand and to explain the origin of human culture.

1.1.2 ... to the calculus of culture

Until not so long ago, cultural and biological explanations for many human behaviors were seen as separate. This culminated in the infamous nature vs nurture debate. Take a certain trait, such as the food taboos found in many societies – why do the French eat horsemeat but the British abhor the very notion? The one side would claim that the trait can be explained by biological adaptation that served to increase Darwinian fitness. The other side would claim that the trait can be explained solely by cultural specificities. And some would claim that the trait can be explained partly by biology and partly by culture.

What this debate eschews is the question why humans possess the faculty for culture in the first place. When we accept that evolutionary adaptations allowed us to develop culture, suddenly all the positions above become moot. It makes no sense to say that a trait is caused 56% by culture and 44% by nature. For many traits, which option to pick is not determined by human biology but the choice is culturally transmitted; the mode of cultural transmission is, however, not arbitrary but strongly shaped by evolution.

This way, humans have two inheritance systems, one vertical, involving the transmission of genes from parents to offspring; the other, more varied, involving the transmission of cultural traits. The pioneers establishing this dual inheritance theory are Cavalli-Sforza and Feldman [31] on the one hand and Boyd and Richerson [22] on the other hand. The former were especially involved in developing how concepts of evolution like inheritance, mutation, and drift of genes would translate to transmission of cultural traits. The latter established a wide range of possible mechanisms that bias cultural transmission and explained how many features could be understood through the lens of dual inheritance. Most of the work on this topic that came thereafter could be seen as a footnote to Boyd and Richerson's influential book. Overall, these authors showed that a rigorous formal framework is required to generate testable predictions, and how such a framework could look like.

A basic assumption of most models of social learning is that adaptation takes place on two timescales. First, there is adaptation of cultural traits

[5] Ibid., 199–200.

during an individual's lifespan. For example, this could consist of deciding whether one wants to align with party A or B. Next, there is adaptation of the learning strategy on an evolutionary timescale. The learning strategy could consist of doing what one's parents did, of aligning oneself with the majority, or of making up a mind of one's own. Although genetic evolution is generally slower than cultural evolution, not all models require the two timescales to be completely separated. This is why the field is often called gene-culture coevolution.

In simple gene-culture coevolution models, the content that is learned is determined by culture, while the mode of learning is determined by Darwinian selection. There is thus a split between *what* is learned and *how* it is learned. That does not mean, however, that the two can be studied in isolation. If one's learning strategy is to imitate cultural traits of others, one's choice becomes dependent on what others do, which in turn depends on their genetically inherited learning strategy. Therefore, the cultural and evolutionary dimension become linked and feedbacks enter the stage. These feedbacks in turn will produce a host of unexpected problems, as will be explored in this chapter.

1.2 Rogers' model

Our work will mainly focus on the different social learning strategies and their properties. But before we begin our venture into the complex world of social learning strategies, it is helpful to analyze a simple model of social learning that is still instructive. For that cause, we use a model initially proposed by Rogers [146] and modify it slightly to fit for our purposes.

1.2.1 Rogers' model with random copying

In the model proposed by Rogers, it is assumed that there are several behaviors, each of which may or may not result in a benefit, depending on the environment. The environment is assumed to be such that there is only one correct adaptive behavior, whereas all other behaviors are maladaptive. After each generation, with a probability $1 - q$ $(0 < q < 1)$, there is an environmental change, after which a new behavior becomes adaptive and the formerly adaptive behavior becomes maladaptive. Therefore, organisms need to constantly learn what is best.

If an individual adopts the correct behavior for the given environment, she receives a benefit of 1 and else she receives nothing. There are two phenotypes in the population, individual learners and social learners. Individual learners learn on their own which behavior is best. Social learners imitate randomly – they choose one random individual from the population and adopt this individual's behavior. Individual learners pay a cost c for learning $(0 < c < 1)$ and always acquire the currently best behavior, but we introduce the possibility of making an error, which occurs with probability ε. Imitators do not pay any costs and acquire the best behavior with probability p_{SL}, to be calculated later.

We assume that fitness is only determined by the probability that an individual adopts the currently best behavior times the benefit, minus the cost of learning. Hence, for individual learners, fitness w_{IL} is equal to

$$\begin{aligned} w_{IL} &= (1-\varepsilon)\cdot 1 + \varepsilon \cdot 0 - c \\ w_{IL} &= 1-\varepsilon - c \end{aligned}$$

For social learners, the fitness w_{SL} is simply:

$$\begin{aligned} w_{SL} &= p_{SL}\cdot 1 + (1-p_{SL}) \cdot 0 - 0 \\ w_{SL} &= p_{SL} \end{aligned}$$

The important question is now: What is the probability p_{SL} that a social learner adopts the correct choice? If the frequency of social learners is x and the frequency of individual learners is $1-x$, a social learner has a chance of $1-x$ to adopt the behavior of an individual learner. We assume that social learners copy choices from the previous generation. Therefore, there is also a chance of $1-q$ that the environment has changed in the meantime, rendering the copied behavior incorrect.

Furthermore, a social learner may observe another social learner, which happens with probability x. What is the probability that this imitated social learner chose correctly? Well, this is the same question as we have asked in the beginning, so we have a recurrent iteration, or chain, as shown in figure 1.1. If we assume that cultural evolution is fast whereas genetic evolution is slow, we can make the approximation that changes in x are negligible over the time periods. But we have to keep in mind that if a social learner imitates another social learner who imitated an individual learner, the information is not one but two generations old. Therefore, the probability that the environment changed in the meantime is not $1-q$ but $1-q^2$. Similar arguments can be made for three, four, five steps.

From the information given above, we can compute the probability p_{SL} of a social learner to choose correctly. All probabilities are assumed to be independent, so we can simply multiply the probabilities of the different

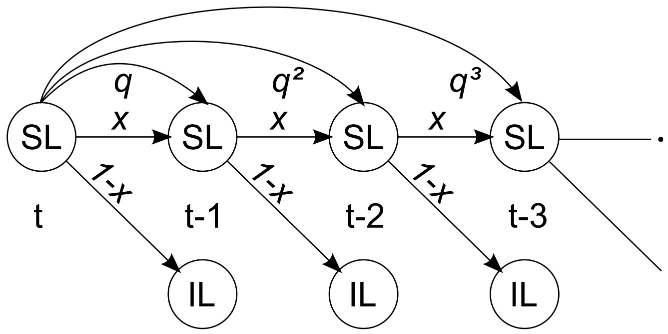

Figure 1.1: Chain showing the use of information by social learners (SL). They either imitate an individual learner (IL) or another social learner. The environment stays stable with probability q after each generation. A social learner might have imitated another social learner who in generation $t-1$ might have imitated another social learner etc., but at some point in time, the source of information must have been an individual learner, who is right with probability $1-\varepsilon$. Therefore, the chain is rooted in the choice of an individual learner but the farther in the past this root is, the lower the probability that the environment has not changes in the meantime. For one generation, the probability is q, for two generations q^2, etc.

events:

$$\begin{aligned}
p_{SL} &= q \cdot (1-x)(1-\varepsilon) + \\
&\quad q^2 \cdot x \cdot (1-x)(1-\varepsilon) + \\
&\quad q^3 \cdot x^2 \cdot (1-x)(1-\varepsilon) + \dots \\
&= (1-x)(1-\varepsilon) \sum_{\tau=1}^{\infty} q^{\tau} x^{\tau-1} \\
&= (1-x)(1-\varepsilon) \sum_{\tau=0}^{\infty} q \cdot q^{\tau} x^{\tau} \\
&= (1-x)(1-\varepsilon) \cdot q \cdot \sum_{\tau=0}^{\infty} (q \cdot x)^{\tau}
\end{aligned}$$

As we have $0 \leq q, x \leq 1$, it follows that $0 \leq q \cdot x \leq 1$. The equation can thus be simplified to:

$$\begin{aligned}
p_{SL} &= (1-x)(1-\varepsilon) \cdot q \cdot \frac{1}{1-q \cdot x} \\
&= \frac{(1-x)(1-\varepsilon) \cdot q}{1-q \cdot x} \\
&= w_{SL}
\end{aligned}$$

Having derived the required fitness functions, we can now plot the fitness of individual learners and of social learners as a function of the frequency

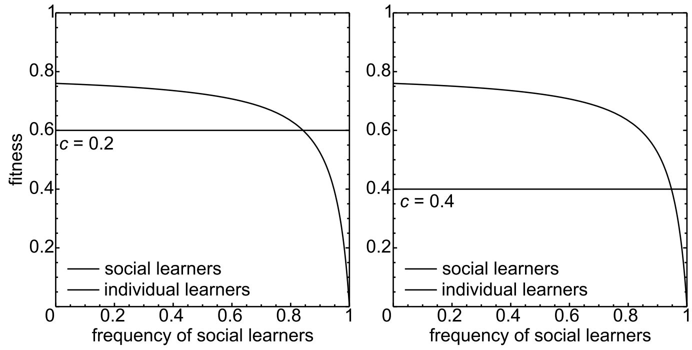

Figure 1.2: Fitness of individual learners (dashed line) and social learners (solid lines, left for $c = 0.2$, right for $c = 0.4$) as a function of the frequency of social learners. $\varepsilon = 0.2$, $q = 0.95$. Social learners copy randomly. Their fitness is monotonically decreasing with their frequency. The fitness of individual learners is constant with regard to the frequency of social learners. It decreases with c and ε.

x of social learners, as done in figure 1.2. The fitness of social learners (solid line) is negatively frequency dependent – the more social learners there are, the more they imitate one another, increasing the reliance on outdated information. On the other hand, the fitness of individual learners (dashed line) is independent of the frequency of social learners, as individual learners only rely on their own judgment.

The equilibrium frequency x^* of social learners is:

$$x^* = \frac{q \cdot (1 - \varepsilon) - (1 - \varepsilon - c)}{q \cdot c}$$

It can be shown that this equilibrium is always stable. The equilibrium frequency of social learners increases with c and ε. This is to be expected, since these two variables decrease fitness of individual learners. Insofar, the effect of costs and errors on the equilibrium frequency is trivial. Moreover, the equilibrium frequency of social learners increases with environmental stability q. The more stable the environment, the more reliable social information is even when old, so this result is not surprising either.

An interesting aspect of social learning is how it performs. Performance is defined here as the probability to choose the correct option. For social learners, it is simply p_{SL} and for individual learners, it is $1 - \varepsilon$. Costs of learning do not influence performance. If we only compare performance, we find that social learners never have a higher performance than individual learners. This should be expected, because social learners always lag behind individual learners, copying information that is at least one generation old, while individual learners always learn the correct choice for the given generation. In addition, social learners have no mechanism to correct the errors that individual learners make – a social learner will blindly copy any

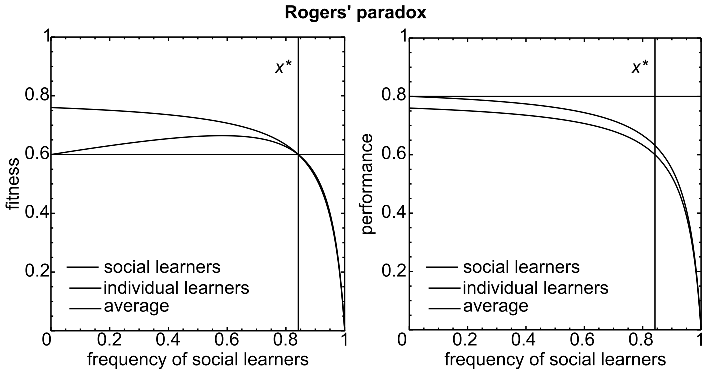

Figure 1.3: Rogers' paradox. Left: fitness; right: performance. Fitness/performance of individual learners (dashed line), social learners (solid line), and the population mean (dotted line) are shown as a function of the frequency x of social learners. $\varepsilon = 0.2$, $q = 0.95$. At the equilibrium frequency x^*, fitness of individual learners and social learners are equal, corresponding to the fitness of individual learners. The mean population fitness transiently exceeds the fitness of individual learners, but in equilibrium, it is always the same. The group is eventually not better off than it was before social learners invaded. This finding is called Rogers' paradox [23, 146]. Additionally, we find that in equilibrium, the population performs even worse on average than in absence of social learners. This finding adds to Rogers' paradox, as it suggests that in equilibrium, it is less likely that the better option be chosen than in absence of social learners.

error they observe. Since individual learners always outperform social learners, the latter could not invade if it were not for the additional cost that individual learners pay.

When social learners are very rare, their fitness is $q \cdot (1 - \varepsilon)$, which will be greater than the fitness of individual learners, as long as c is sufficiently large. When social learners are very very frequent, their fitness is reduced to 0, which is worse than the fitness of individual learners (except if ε and c are absurdly high). If the fitness of social learners is greater than that of individual learners when rare and lower when frequent, given that their fitness decreases monotonically with their frequency and fitness of individual learners is constant, there must be one frequency and one frequency only at which social learners and individual learners have equal fitness. This is obviously the equilibrium frequency x^*, as shown in the left panel of figure 1.3.

The fitness of individual learners being constant over x, it is the same at equilibrium as it was in absence of social learners. At equilibrium, the fitness of social learners is by necessity also the same as that of individual learners. Therefore, *in equilibrium, the average fitness of the population is the same as it had been before social learning was introduced.* This finding is called **Rogers' paradox**. Although it might seem trivial, its interpretation is actually our first important conclusion.

The “paradox” consists of the fact that although social learning, by virtue of saving the cost of learning individually, can indeed improve the population fitness, this improvement is not maintained. As long as social learners have a fitness advantage, they become more frequent and thereby deteriorate the accuracy of social information. At the stable equilibrium, the population fitness is exactly the same as in absence of social learners. Although natural selection would favor social learning to evolve, it would contribute nothing to the lot of the population as a whole.

The finding by Rogers is robust to some of its assumptions. Boyd and Richerson [23] showed that if spatial variation is added or if social learners are allowed to preferentially copy individual learners, Rogers’ conclusion still holds. They also showed possible solutions to Rogers’ paradox, which we will discuss in the third chapter.

When the adaptedness of culture is studied, the understanding is that culture is indeed very adaptive and allows humans to cope with situations they could not otherwise cope with [144]. This runs counter to Rogers’ findings. But Rogers’ findings have to be interpreted carefully. The conclusion is not that socially acquired traits (or “culture”) are useless and do not improve mankind’s lot; the conclusion is that under very general conditions, social learning would not improve mankind’s lot, so one or several of the initial assumptions must be wrong. Rogers deserves merit for pointing out that something in the study of the adaptedness of culture is amiss.

A further conclusion from Rogers’ work has not generally been recognized in the literature. He showed that in equilibrium, population fitness is the same as in absence of social learners. The disadvantage of relying on social information becomes so great that it equals the cost of learning individually. However, it is certainly true that the adaptedness of culture does not consist solely in saving the cost of learning individually. In our opinion, one should expect social learning to actually improve upon the performance of individual learning.

With simple random copying as in Rogers’ model, however, we actually find a decrease in performance. This is shown in the right panel of figure 1.3. There, we plotted performance of individual learners, social learners, and the mean population performance as a function of x. Performance, opposed to fitness, is calculated by leaving out the costs. Social learners always perform worse than individual learners, so that at equilibrium, the population as a whole performs worse than in absence of individual learners. After the invasion of social learners, we would thus find individuals to more often pick the wrong option than before. An important question we will address in this work is therefore whether there are social learning strategies that can actually improve performance in equilibrium.

sample	prob. to observe	prob. of 3 correct	prob. of 2 correct
0 SL, 3 IL	$(1-x)^3$	$q \cdot (1-\varepsilon)^3$	$3 \cdot q \cdot (1-\varepsilon)^2 \cdot \varepsilon$
1 SL, 2 IL	$3 \cdot x \cdot (1-x)^2$	$q \cdot p_{conf} \cdot (1-\varepsilon)^2$	$2 \cdot q \cdot p_{conf} \cdot (1-\varepsilon) \cdot \varepsilon +$ $q \cdot (1-p_{conf})(1-\varepsilon)^2$
2 SL, 1 IL	$3 \cdot x^2 \cdot (1-x)$	$q \cdot p_{conf}^2 \cdot (1-\varepsilon)$	$2 \cdot q \cdot p_{conf} \cdot (1-p_{conf})(1-\varepsilon)+$ $q \cdot p_{conf}^2 \cdot \varepsilon$
3 SL, 0 IL	x^3	$q \cdot p_{conf}^3$	$3 \cdot q \cdot p_{conf}^2 \cdot (1-p_{conf})$

Table 1.1: Conformists who sample three individuals. x is the frequency of conformists in the population, q is the probability that the environment stays constant, ε the error rate of individual learners, and p_{conf} the probability that a conformist chooses the correct option. Any number between 0 and 3 individuals in the sample of a conformist may be other conformists. If 2 or 3 sampled individuals choose correctly and the environment stays constant, a conformist also adopt the correct option; if not, a conformist chooses the wrong option.

1.2.2 Rogers' model with conformism

Social learners who engage in random copying can never outperform individual learners, no matter how inaccurate individual learners are. However, other forms of social learning may lead social learners to have a performance that exceeds the performance of individual learners. We extend Rogers' model by implementing another social learning strategy, conformism. We define a conformist as a social learner who samples three individuals and then adopts the choice that the majority of the individuals adopted.

For simplicity, a conformist is assumed to always observe 3 individuals. Among these 3, they may observe any number between 0 and 3 individual learners, while the rest are social learners. Let p_{conf} be the probability that a conformist makes a correct choice. We can then calculate p_{conf} using table 1.1.

For this calculation, we made the simplifying assumption that all incorrect options are lumped together. That way, if the majority of the sampled individuals choose the correct option, the correct option will be chosen, and if the majority of the sampled individuals choose any of the incorrect options, the incorrect option will be chosen. We furthermore ignore the possibility that a conformist chooses the wrong option but that the environment changes so that this option suddenly becomes correct.

By adding up the products of the third and fourth column of the table multiplied by weights in the second column and solving for p_{conf}, we arrive at three solutions; two are stable equilibria and one is an unstable equilibrium (see figure 1.4). All of the solutions are too cumbersome to write down here.

To show that conformists may perform better than individual learners, we can simplify the formula by assuming that conformists are very rare. For $x = 0$, the formula reduces to $p_{conf} = q \cdot (1-\varepsilon)^2(1 + 2 \cdot \varepsilon)$. For $\varepsilon = 0.2$

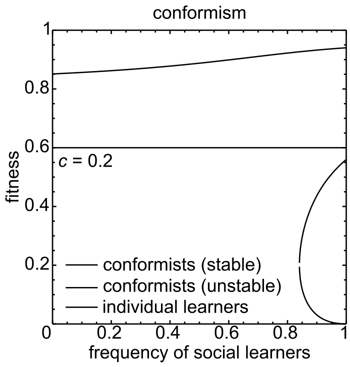

Figure 1.4: Fitness of individual learners (dashed line) and conformists (solid lines: stable equilibria; dotted line: unstable equilibrium) as a function of the frequency of conformists. $\varepsilon = 0.2$, $q = 0.95$. At lower frequencies, fitness of conformists increases with their frequency. When they are very frequent, depending on whether they adopt the good option or not, their fitness can either be very high or very low (two stable equilibria). Whether they end in the high or low equilibrium depends on which side of the unstable equilibrium they start on. The more conformists there are, the more likely it is that they start on the side of the unstable equilibrium that leads to the choiec of the inferior option.

and $q = 0.95$, we have $p_{conf} \approx 0.85$ and $p_{IL} = 0.8$, meaning that conformists outperform individual learners. In fact, as long as ε is sufficiently large and q sufficiently small, p_{conf} is always greater than p_{IL}, meaning that when conformists are rare and individual learners perform above chance level, conformists perform better than individual learners.

Paradoxically, the invasion threshold for conformists increases the *worse* individual learners perform. If individual learners have a high chance to pick the wrong option, mistakes will be amplified by conformists and lead to even more mistakes. For $q = 1$, this means that conformists cannot invade a population of individual learners if $\varepsilon > 1/2$. This, however, amounts to individual learners being wrong more than half of the times, which is unlikely. One would expect competition among individual learners to reduce ε, even if this meant that the population becomes prone to invasion by conformists.

Given that ε is sufficiently small, if conformists sample even more than just 3 individuals, their invasion fitness would be even greater. For example, if they sample 7 individuals and if $q = 0.95$ and $\varepsilon = 0.2$, the fitness of a rare conformist would be ≈ 0.92 instead of ≈ 0.85. This is because the more individuals are sampled, the less likely it becomes that the majority of them choose wrongly, given that each individual has a higher than 50% chance of being right.

Not coincidentally, this finding corresponds to Condorcet's Jury Theorem [36]. Condorcet started with the question how many members a jury should optimally have if it is about to make a binary verdict (like "guilty" or "innocent"). It is assumed that decisions are made by majority vote. If each juror has a higher than 50% chance to make a correct decision, the conclusion is that the more members the jury has, the more likely it is that the majority vote turns out right. Similarly, in our model, the jurors are the sampled individuals and the majority vote is realized by the conformist. The more individuals the conformist samples, the higher the probability that she actually chooses the right option. It is straightforward to take this theorem as support for democracy, and not coincidentally, the philosophe Condorcet was a major figure during the French Revolution [44].

Condorcet's theorem hinges on the assumption that each juror decides independently from the others. Similarly, we started with the assumption that conformists are rare, implying that they only sample individual learners, all of which indeed decide independently. What would happen if we dropped the assumption?

We will have to entertain the possibility that a conformist samples other conformists. As we saw, conformists are initially more likely to be correct than individual learners are. Therefore, if conformists sample other conformists, this should improve the performance of conformists even further. This is indeed the case. If instead of assuming that $x = 0$ we assume that $x = 0.1$, performance of conformists who sample 3 individuals increases by 0.01 to ≈ 0.86; for $x = 0.5$, it reaches ≈ 0.89. For conformist performance, we thus observe a positive frequency dependence, as shown in figure 1.4. The more conformists there are, the better they perform.

This does not mean, however, that conformists will become fixed in the population. As long as there is a sufficient number of individual learners, when there is an environmental change, conformists will also adopt. If there are too few individual learners, conformists become stuck with the now maladaptive option, though. This manifests as the second stable equilibrium in figure 1.4. As long as most individuals in the population choose the correct option, conformists will also choose it. If too many individuals choose an incorrect option, though, conformists will also choose the incorrect option (the dotted line separates the two stable equilibria).

Assuming environmental changes to be infrequent and conformists to be initially rare, we will thus see conformists outperforming individual learners and becoming more and more frequent. The more frequent they become, the better they perform. At some point, individual learners will become rare in the population. Sooner or later, however, there will be an environmental change. Afterwards, conformists are suddenly worse off and individual learners begin to rise in frequency. At one point, when there are enough individual learners, conformists will suddenly switch and choose the better option, again outperforming individual learners. Thus the cycle continues. A totally stable equilibrium frequency will never be attained; instead we observe constant ups and downs.

Even for the rather simple case of conformists with sample size 3, the model therefore generates complex outcomes. We cannot calculate an equilibrium frequency of social learners, as opposed to what we did when social learning consisted of random copying. This is why social learning theory often has to recur to simulations when strategies more complex than random copying are analyzed.

Although the frequency of conformists will never completely stabilize, there has to be a frequency around which it fluctuates. The average fitness of conformists after the occurrence of the bifurcation is evidently between the two stable equilibria and depends on the exact adoption dynamics. It is clear, though, that for $x = 1$, the fitness of conformists would be zero, as they would never choose the adaptive option after an environmental change occurs. Thus, somewhere between the x at which the bifurcation occurs and $x = 1$, the average fitness of conformists has to become lower than that of individual learners. This is the frequency around which conformists fluctuate in the population, and it is more or less constant depending on how fast evolution is compared to the social learning process.

Evidently, at this equilibrium, fitness of individual learners and conformists have to be equal. As individual learners still have the same fitness as they had before conformists invaded, the fitness at equilibrium must also take this value. Therefore, in equilibrium, the mean population fitness is the same as it was before the arrival of conformists. In other words, we have replicated Rogers' paradox. Conformism is not a social learning strategy that is able to solve Rogers' paradox.

Coming back to the question of learning costs, the same result holds as previously found for random copying. There is a more or less stable equilibrium of social learners and individual learners; when the cost of individual learning increases, the equilibrium frequency of individual learners decreases. The effect of costs is thus again trivial.

Overall, this exercise showed what we can learn from a simple social learning model. Social learning by random copying can never outperform individual learning, but conformism may do so. Still, conformism cannot resolve Rogers' paradox; in equilibrium, the mean population fitness is the same as in absence of social learning.

Moreover, it became obvious that learning costs have a trivial effect on the outcome. This is because costs only affect the fitness but not the performance of learners. However, performance is the more interesting aspect of social learning – in this work, we want to understand which social learning strategies lead to a better performance and not to cost savings. Therefore, we do not implement exogenous learning costs, in contrast to previous social learning models (e.g. [22, 23, 50, 61, 63, 76, 77, 100, 101, 103, 119–121, 143, 146, 174, 177]) but focus solely on performance instead.

We found that in the study of social learning, there are some similarities with early research on social choice theory. Condorcet studied the outcome of votes cast by a group of people [36]. In contrast to modern social choice theory [3], which is only concerned with whether the voting process correctly reflects the preferences revealed by the voters, Condorcet was additionally interested in whether the outcome was really the better outcome. Although in an election, it is debatable whether there are objectively better outcomes, in models of social learning, there are correct and incorrect outcomes. Therefore, Condorcet's studies are closer to our studies than is modern social choice theory.

In particular, we found that there is an analogy between the performance of conformists under the assumption that they are rare in the population, and Condorcet's Jury Theorem under the assumption that jurors make independent choices. Moreover, we showed that conformists could actually get stuck with choosing the wrong option if too many individuals in the population use conformism as a strategy. Is there an equivalent finding with regard to Condorcet's theorem when the assumption of independence is dropped?

We could assume e.g. that jurors have an incentive to be on the right side of the verdict. They may be afraid of being ridiculed if their personal decision turns out to be wrong at the end of the day. This would provide an incentive to imitate other jurors if possible. For instance, interesting effects may occur when jurors cast their vote sequentially. The first juror will obviously vote according to what she believes to be true, she has no better information. The second juror now knows the first juror's belief, as well as her own belief. Assuming that each juror's beliefs are equally accurate, if her own belief contradicts the first juror's belief, the second juror is confronted with a dilemma. Let us assume that in this case, a juror will always act according to her own belief. The second juror would thus decide the opposite of the first juror. If, however, the first juror confirms the second juror's belief, she will obviously cast the same vote as the first.

For the third juror it becomes tricky. She will observe either two contradicting votes or two conforming votes. In the first case, the two contradicting bits of information cancel out and the third juror is in the same position as the first juror was. If the two first jurors came to the same conclusion, however, the third juror knows that they held the same belief. If the third juror also holds this belief, she will happily concur. If not, though, what should she do? She could vote according to her own belief, but if her most important goal is to cast a vote that will turn out to be correct, is this her best option? Obviously, if all jurors are a priori equally likely to be correct, it is more likely that the third juror is misinformed than that the previous two jurors were both misinformed. It is thus rational for the third juror to ignore her own belief and to align with the other two.

The fourth juror will subsequently be in a similar position as the third juror. She knows that the first two jurors had the same belief, so that even if her own belief contradicts these jurors' believes, it is rational for her to concur. Interestingly, she cannot make any conclusions about the belief of the third juror, since the third juror will always align herself, regardless of her belief. The fourth juror is therefore not better informed than the third juror, even though she made more observations. This is evidently true for all subsequent jurors.

This finding has first been described by Bikhchandani, Hirshleifer, and Welch [16], a similar model was proposed by Banerjee in the same year [10]. The general phenomenon of people (or other organisms) aligning with others is called herding. The specific case studied here, when suddenly personal information is ignored and every subsequent observer aligns with the previous ones is called an informational cascade. It was shown by the authors that such a cascade will start as soon as one of the two propositions has a margin of two votes over the other. After a cascade has started, no further juror adds to the publicly available information, so that even the 100ths juror is not better informed than the first.

It has also been shown that in the limit, an informational cascade is bound to occur and that because of the alignment, the probability that the cascade ends with the correct verdict is only slightly higher than the probability of each individual juror to be correct [16, 17]. Compare this with the case of 7 independent jurors and an accuracy of 80% (i.e. $\varepsilon = 0.2$) and no environmental changes (i.e. $q = 1$). The probability of a correct decision would be $\approx 0.97\%$, the error rate thus being diminished to 1/6th of the initial value.

Coming back to Condorcet's theorem, this means that if jurors influence each other in the way just described, the verdict of a jury does not become more reliable the more jurors there are. Although the decision made by many jurors is slightly more likely to be true than the judgment of a single juror, the gain is minimal, especially when compared to what could be achieved if all jurors acted independently.

A second big conclusion, similar in spirit to Rogers' paradox, is, therefore, that if individuals act rationally to increase their own utility/fitness, they will choose to conform to the majority choice to a higher degree than is optimal (optimal in the sense of considering the overall result). The potential information gained by observing the behavior of others is thus almost obliterated. For the legitimacy of democratic processes, this has major implications: Polls, traditional and social media, opinion leaders, all create dependencies between the voters' opinions, which could undermine the 'wisdom of the crowd' [166]. Similarly, excessive social learning is suboptimal for the total fitness of the group.

1.2.3 What to expect from a social learning model

Rogers' findings has some striking similarities with those reached when studying a different type of social behavior, cooperation (see figure 1.5). In the tragedy of the commons [82], there is a large surplus (the common) that could be used to the benefit the whole group. However, selfish interest will lead to the over-exploitation of the common up to a point where the marginal benefit of exploiting the common even more equals that of not using the common. At the end of the day, there are thus no benefits to be had from having the common. Similarly, with social learning, there is a large benefit to be had – the aggregation of knowledge derived from observing others – but this benefit is exploited until the marginal benefit of using social learning equals that of using individual learning. Eventually, in both cases, the population is not better off in equilibrium than at the beginning, even though a better state clearly exists.

The problem of cooperation has long been identified, with the tragedy of the common being a prominent illustration, but with other problems like the prisoner's dilemma and the public goods game following suit. The great importance of Rogers' results [146] is that they make clear that a similar problem exists with the use of information in populations (societies). The importance of information has been ever increasing in the last centuries and decades and is nowadays arguably of equal importance as fostering and stabilizing cooperation. (For instance, besides spreading risk, aggregation of information is the main function of financial markets.)

Studying the problem of social learning provides a different challenge from studying the problem of cooperation. The latter is deeply linked to game theory, as the marginal impact of an individual's action on the payoff of another individual is considered to be large. That is, in most cases, the overall cost associated with a selfish behavior exceeds the personal benefit gained through this behavior – we are dealing with large externalities. Using the standard prisoner's dilemma payoffs[6] [6, 7], given that player A cooperates, player B's benefit from defecting is gaining 2 while player A loses 3; given that player A defects, player B's benefit from defecting instead of cooperating is gaining 1, while player A loses 4. In contrast, in social learning problems, the marginal impact that an individual has on another individual's payoff is typically relatively small. Although a player's best response ultimately depends on the strategies of other players, the tools provided by game theory are therefore rarely used to study social learning.

Besides cooperation, social learning is another big domain of the study of social behavior. In our opinion, it's paradigm needs to be developed further to make its study more successful. Although not the first to model social learning, Rogers was the first to present such a simple social learning model that it could be grasped immediately by a lay person while still providing valuable a insight (although the elaboration in the original paper was somewhat unclear and it was up to other scientists – notably Robert Boyd and Peter J. Richerson [23] – to clarify the importance of Rogers' results.)

Axelrod's tournament [6] sparked a huge explosion in the cooperation literature. Although the significance of the main results from this and the

[6]T (temptation) = 5, R (reward) = 3, P (punishment) = 1, S (sucker's payoff) = 0.

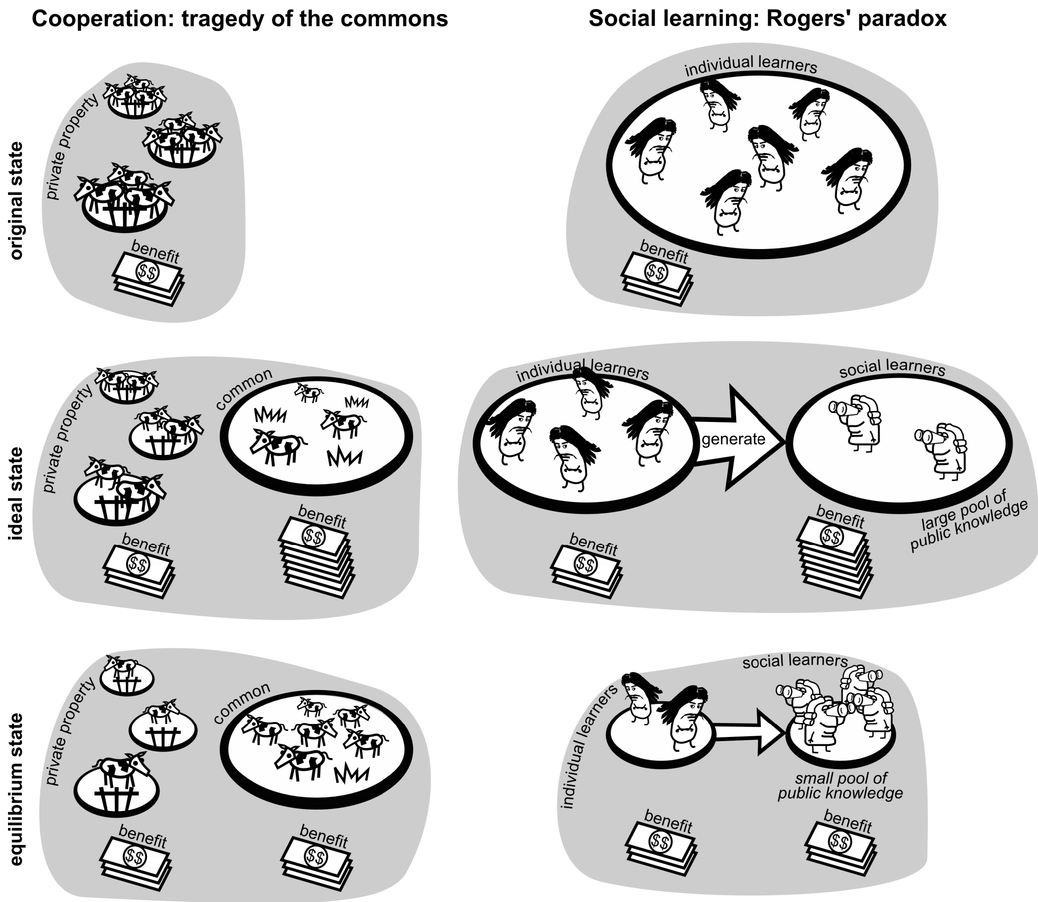

Figure 1.5: Similarities between the tragedy of the commons and Rogers' paradox. The tragedy of the commons is illustrated on the left hand side, Rogers' paradox on the right hand side. In a simplified form of the tragedy of the commons, in the hypothetical original state, herders have private pastures that are used to feed their cows. This generates a low benefit. Then a common is added that is free to use by everyone. In an ideal state, every herder puts a reasonable number of cows on the common and reaps the additional large benefit. However, self-interest will lead to the over-exploitation of the common, until in the equilibrium state, the common and private property generate the same, low benefit. In Rogers' paradox, in the hypothetical original state, there are only individual learners, who generate a low benefit. In the ideal state, some, but not too many, social learners exploit the public knowledge which is available from observing the behavior of others, maximizing the total population benefit. However, as social learners gain more than individual learners, they will become more common and deplete the pool of public knowledge. In the equilibrium state, they generate the same low benefit as individual learners.

following tournaments can be contested, it is still true that it provided a unified framework, in the form of accepting the iterated prisoner's dilemma game as the main paradigm for studying cooperation. Despite attempts to invigorate the field of social learning by copying Axelrod's premise [141], a true paradigm has not been established. However, doing so would be very helpful to advance the study of social learning.

Certainly, it can be argued that a paradigm will likely restrict researchers in an undue fashion. But the benefits from having one at this point would arguably exceed the disadvantages. Most notably, a paradigm allows the comparison of different results and also achieves to focus the critique of the field; without a paradigm, the weaknesses of the research will be easier to conceal. We will show in chapter 3 that many conclusions about social learning may have been premature, the reason being that only a small subset of possible social learning strategies were considered in the literature.

As mentioned, although Rogers was not the first to model social learning, his model was by the time the easiest model that still drove home an important point. Before this model, there were other models that, although adequate and useful for the respective purposes, were too specific to make a very broad impact beyond the field of social learning. Yet, while Rogers' model was almost perfect in achieving what the author meant to achieve, it is clear that it is too restrictive to breed the same "adaptive radiation" that was triggered by the adoption of the prisoner's dilemma or public goods game as the paradigm to study cooperation.

Other frameworks have been suggested for social learning, most notably by a team formed around Kevin Laland. Their game basically comes down to a multi-armed bandit with the option of learning individually, acquiring information about the choices and payoffs of other individuals, and exploiting the bandit [141]. This model encompasses more possibilities but is not intuitively as easy to grasp as Rogers' model. It may also be too specific, allowing savvy participants of the tournament to "game the system", thus providing little useful information about what to expect from social learning in the real world.[7]

In our opinion, a new framework is needed, although Rogers' model should serve as an exemplar. We propose that the framework should have the following features:

- Synchrony in choice, so that sequence effects can be ignored (in contrast to, say, herding models [10, 16]).
- Spatial patterns should be cast aside (at least at first), as they add unduly to the complexity.

[7] The problems of the original tournament have been recognized by the authors. By the time this text is written, a second tournament, which is supposed to amend the problems of the first tournament and to extend its scope, is being held.

- Dichotomy of choice, as many problems can be reduced to choosing the default state or the alternative (in contrast to more than two options or actions on a continuous space).
- Emphasis on performance of social learning; an important goal should be to study under which conditions social learning improves upon individual learning.

In our opinion, Rogers' model can be improved on four levels to come closer to what might be considered a good paradigm for the study of social learning.

First, we think that Rogers' model was too restricted in the way that social learning was allowed to take place. Social learning consisted of copying a random individual in the population. Instead, social learners should be allowed to sample more than one individual, they should be allowed to discriminate between individuals (for example, just copy the most successful individual), they should be allowed to facultatively switch their mode of learning (e.g. only copy when not sure which alternative is best), etc. This will allow for the needed flexibility in the study of social learning.

Next, in Rogers' model, it was assumed that individual learners always know instantly which choice is currently the correct one. Later models relaxed this assumption by introducing the possibility for errors. In our opinion, the goal should be to get closer to real learning processes. If not learning from others, humans and other animals learn by trial-and-error, so a good model should be able to at least accommodate strategies like reinforcement learning.

Following from the two previous modifications, Rogers' original model contained a variable that becomes unnecessary and should thus be omitted if not specifically needed: the cost of learning individually. The reason why Rogers required individual learners to pay an extra cost was that the only edge social learners had over individual learners was to save this very cost. Individual learners were perfect in their choice, so some form of handicap had to be added. By requiring individual learners to really learn instead of granting them omniscience, the possibility for making mistakes is included; by allowing for more sophisticated social learning strategies, social learning can have an advantage over individual learning apart from saving costs.

In the world of models, costs have a trivial effect on the results, namely that the higher the cost of a certain strategy is, compared to the cost of another strategy, the lower this strategy's equilibrium frequency. Empirically, the cost of individual learning might be more or less easy to measure, but the cost of social learning less so. What exactly is the additional cost of modifying the cognitive apparatus to allow social learning to happen? How much time does an individual need to sacrifice to follow others around and copy their choices? As learning costs only affect model results trivially and are hard to measure, parsimony would suggest to omit them as long as they are not specifically required for a given problem.

Finally, Rogers had to assume for his model that cultural learning happens at a much faster pace than genetical evolution. When determining how old social information is, he had to assume that the frequency of social learners was approximately constant. Though appropriate for the given question, the speed of cultural and genetical evolution should optimally be endogenous to the model, as it might easily be true that the most interesting phenomena in the evolution of social learning emerge in times when cultural and genetical evolution act on a comparable time scale.

Obviously, we tried to implement these four modifications to Rogers' model in the model that we will propose. We thus hope that our model strikes the right balance between generalizability and specificity, flexibility and restriction, allowing us to make significant progress.

Although the model arguably improves upon Rogers' model, especially in allowing for more flexible and realistic strategies, these improvements come at a cost. As is almost inevitable, higher flexibility leads to more parameters. Most notably, the fashion in which the environment changes over time will greatly expanded. In the original model, one single parameter determined the environment: the probability that it changes after each period. In our model, greater flexibility is allowed, such as that both options being good at the same time or bad at the same time.

Regarding the strategy space, it is clear that it expands substantially. Even if the sampling of a social learner from the population is random, even if observed individuals cannot be distinguished from each other, and even if sample size is limited to three individuals, there are already 20 possible observations for dichotomous choices (all three observed individuals chose option A and were successful, all three chose option A and two were successful, etc.). A strategy is only completely described by attaching to each possible observation a probability to choose one or the other option. With higher sample sizes or more choice options, the number of possible combinations quickly gets out of control. We will later discuss ways to restrict the complexity in a meaningful way.

How will the field of social learning develop? Again, the analogy to the cooperation literature may serve as a guide to what we could expect. Modeling approaches will give us a hint as to what we should expect to see in the real world. Only if we know what to look for, only if we have specific hypotheses, will the observations we make be interpretable. Often, we will find that observations refute our hypotheses, but then at least we have a hint at where we actually made mistakes. The study of cooperation has given rise to concrete policies aimed at improving cooperativeness in communities – just consider the work of the late Nobel prize laureate Elinor Ostrom [137]. Similarly useful conclusions could be drawn from studying social learning.

It is debated whether what we think we know about cooperation covers what we want to know to a reasonable degree, or whether cooperation is still not well understood. As an example, kin selection [79] and direct reciprocity

[170] are often considered to be keys to breed and sustain cooperation, yet these concepts were criticized in the recent years (see [80] and [134], respectively). Currently, punishment is hotly debated as a mechanism that supports cooperation [58] but it gives rise to second order free-riding and other problems [45]. Similarly, the study of social learning will be characterized by not knowing whether the gained knowledge is already very comprehensive or whether astounding discoveries are still lurking around the corner.

1.3 Introduction to our model

After this long prelude, we should now cut to the chase. The following section will expose the model we use throughout this work. We will derive some simple results and discuss them, before presenting more intricacies in the next chapters.

1.3.1 General outline

Individuals in our model face what seems to be a fairly simply task, namely repeatedly choosing between two options, A and B. These options can be thought of as e.g. hunting antelopes or wildebeests. Each option has a certain probability to lead to success, in which case fitness is increased. If two individuals choose the same option, they have the same probability to be successful, but each success is drawn independently. Therefore, even if both choose to hunt antelopes, one of the individuals may end up being successful and the other not.

The underlying success probabilities are unknown to the individuals. To allow them to learn, they have the opportunity to make several decisions per lifetime (a generation). Since they receive feedback from their previous choices – whether they succeeded or not – they can make better informed decisions over time. This way, more realistic forms of learning are possible in our model compared to previous models.

In order for the task to be challenging, the success probabilities of the options A and B, denoted p_A and p_B respectively, change over time. If option A was associated with many successes in the past, that does not necessarily mean that it is in the present. p_A and p_B behave independently of each other but always according to the same underlying algorithm.

We show an example of how p_A and p_B vary over time using our default parameter settings in figure 1.6. p_A is shown as the thick line and p_B as the thin line. After each period, p_A and p_B change by a small amount. But in contrast to a real random walk, they will tend to revert to the mean, which is 0.5 in this example. At the beginning, A is the better choice, as p_A exceeds p_B; later on this trend reverses. Between period 100 and 150, the difference between p_A and p_B is small, so choosing the worse option is not

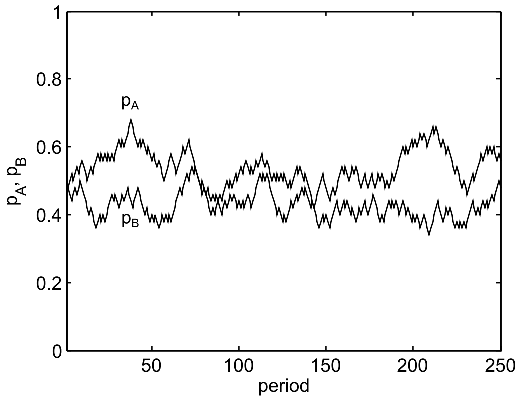

Figure 1.6: Example of how p_A (thick line) and p_B (thin line) vary over time. p_A is the probability that a choice of option A will result in a success, p_B that B results in a success. p_A and p_B fluctuate in small steps after each period but stick close to the mean of 0.5. At the beginning, A is the better choice but later on, B becomes the better choice.

backbreaking. Around period 200, though, B is significantly better than A, meaning that it is very costly to choose A in that situation.

There are several possibilities to tinker with the parameters of the environment. The next chapter will deal with this. In this chapter, however, we only use the default parameters. They are set so that each individual has 50 choices per lifetime. After each time period, the value of p_A (and p_B) changes by 2 percentage points. Whether p_A increases depends on the current value of p_A: the probability is simply $1 - p_A$ (same for p_B). Therefore, if p_A is greater than 0.5, it is more likely to decrease in the next period, and if it is less than 0.5, it is more likely to increase in the next period. This is why we wrote that the values of p_A and p_B tend to "revert to the mean".

In this chapter, we model only two types of learning strategies:

Individual Learners base their decision only on their own experience and not on the behavior of others. In particular, they will practice reinforcement-learning [97], i.e. they pick their assets according to a trial-and-error-method. Since p_A and p_B are fluctuating, the choice between A and B has to be updated constantly. Individual learners therefore have a memory that allows them to draw on earlier experiences.

Conformists do not rely on their own experience. Instead, they base their choice purely on the observation of other individuals. In each time period, after having observed a fixed number of individuals, a conformist will always pick the option that has been chosen by the majority of the observed individuals.

All individuals start with a base fitness of 10. After each time period, an individual's fitness is increased by 1 if the individual was successful or stays the same if not. After 50 periods, one generation ends. The frequency of the strategies will then evolve, with strategies that have a higher mean fitness being more likely to become more frequent in the population. This process of natural selection is accomplished with a standard method from population genetics, the Fisher-Wright process [62, 181], which preserves the total population size. Generations are non-overlapping, there are no mutations, and we assume non-sexual haploid inheritance.

The whole model is is shown schematically in figure 1.7. The exact details should not matter here. In general, each strategy is characterized by a decision mechanism. This mechanism has to be exactly defined – for each possible observation, there must be a rule that determines the reaction. From this mechanism, an aggregate behavior can be derived, which describes what option a strategy adopts; for social learners, adoption probabilities depend on what other individuals in the population chose. Moreover, for all strategies, the behavior may depend on the environment, i.e. the values of p_A and p_B. The interplay between the strategies' behavioral responses and the variable environment leads to cultural adaptation within a generation. If a strategy is especially apt at choosing the better option, it will likely have a higher fitness than others. This will lead to an increase of this strategy's frequency between generations, i.e. genetic adaptation. The frequencies of the strategies will in turn influence their faculty for cultural adaptation; genes and culture coevolve.

1.3.2 Learning strategies

In the following, we will consider in more detail the learning strategies. Social learning will take the form of conformism. Individual learning will take the form of reinforcement learning with exponential discounting. Bayesian learning would suggest itself as an alternative individual learning strategy. We will briefly discuss Bayesian learning and explain why it is not very adequate for the learning task.

1.3.2.1 Individual learning trough reinforcement

In our model, individual learning is modeled in a more realistic fashion than previously. This is possible because an individual faces more than just one decision per lifetime and can therefore rely on past experience to make informed decisions.

We chose to implement individual learning as reinforcement learning [97]. This is a rather broad term that includes several possible implementations, so we have to go into more details. The most simplistic implementation of reinforcement learning would be a strategy that sticks with the same option as last period when it was successful and switches options when unsuccessful – this strategy is called win-stay lose-shift [133, 145].

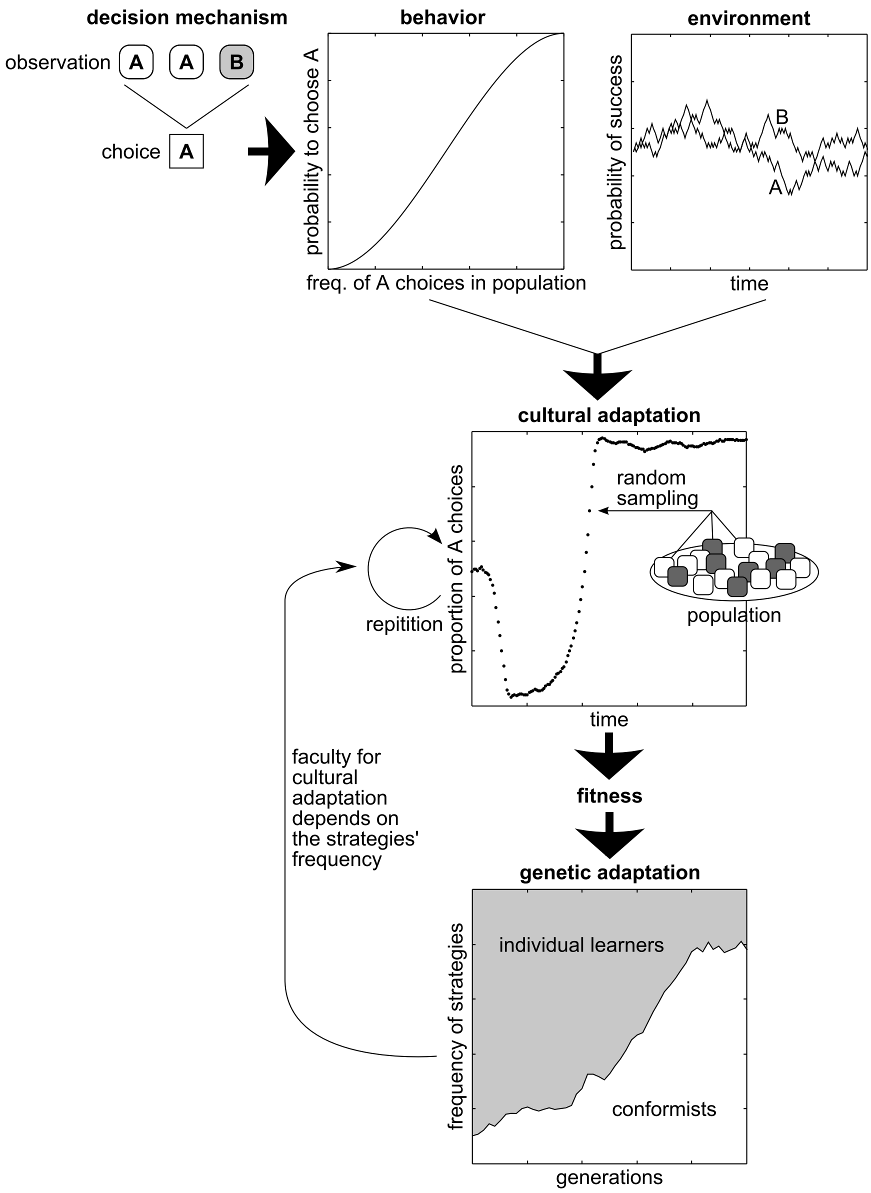

Figure 1.7: Overview of the model. Strategies are characterized by a **decision mechanism**. This mechanism results in a certain **behavior** that, for social learners, is dependent on how others in the population behave. Furthermore, the behavior of the strategies depends on the **environment**, which varies over time. Combining the behavioral responses of the strategies with the environment results in **cultural adaptation**. Social learners sample randomly from the whole population. Depending on the choices, strategies accumulate **fitness**. Strategies with higher fitness tend to increase in frequency, resulting in **genetic adaptation**. Changes in the strategies' frequency feeds back on cultural adaptation.

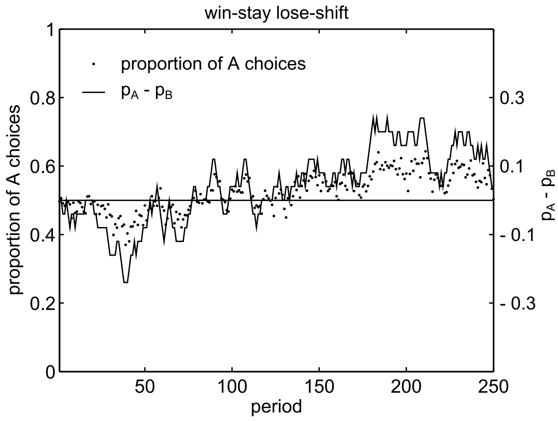

Figure 1.8: The behavior of win-stay lose-shift over time. The proportion of A choices made by this strategy (•) is measured on the left y-axis. The solid line represents the environment in the form of $p_A - p_B$, measured on the right y-axis. The horizontal line serves as a guide to the eye. A perfectly performing strategy would always choose A as long as $p_A - p_B > 0$, that is, when the $p_A - p_B$ line is below the horizontal line. It would never choose A in the opposite case of $p_A - p_B < 0$. Win-stay lose-shift tracks changes in the environment but is very "conservative" because it never deviates very far from 50% A choice. This behavior hampers the performance of this strategy.

An illustration of the behavior of win-stay lose-shift can be found in figure 1.8. Shown are the aggregate behavior of 1000 individuals who use this strategy and at the same time how the environment develops over time. Behavior is shown as the average proportion of A choices made be the individuals, measured on the left y-axis. The environment consists of p_A and p_B, but to simplify matters, we only plotted the difference $p_A - p_B$, which is measured on the right y-axis.

To understand why win-stay lose-shift is not a good performer, we first have to understand how a perfect strategy would behave. Obviously, the perfect strategy should always choose A when $p_A > p_B$ and B when $p_A < p_B$ (if $p_A = p_B$, the choice is irrelevant). When we look, e.g., at period 200, we find that p_A is larger than p_B but still only approximately 60% of the population choose A. Win-stay lose-shift is very conservative in that it will stick close to 50% most of the time, at the detriment of its performance.

Still, win-stay lose-shift is capable of tracking changes in the environment reasonably well, resulting in a performance significantly better than expected from chance. Performance is defined here as the degree to which a strategy approaches optimal behavior. If a strategy has a performance of 60%, that could mean, e.g., that in each period, 60% of the individuals using this strategy choose the better option. But it could also mean that during 60% of the periods, all individuals choose the better option and during 40% of the periods, none of them chooses the better option. In reality, we will always find a mixture of these extremes. The performance of random choice would be 50%, performance of win-stay lose-shift is 54.40% (0.06% standard error of the mean), which is significantly higher.

Essentially, win-stay lose-shift is a reinforcement learning strategy with a memory of only one period length. It cannot take into consideration what happened two, three, or more periods in the past. However, the environment tends to vary slowly. Say that A has led to success in the past 10 periods except in the very last period. Win-stay lose-shift would immediately switch to B after this turn of events, as it ignores everything except the last observation. However, such a string of successes makes it very likely that p_A is high and should not prompt a premature switch. This is where memory comes into play.

We model reinforcement learning with memory by assuming that individuals have a propensity, P_i for each option i, $i = \{\mathrm{A,B}\}$. The propensities are updated in each period by increasing them by an amount R_i if option i is reinforced, with R_i being equal to 1 if option i was chosen and yielded success or -1 if it was chosen and did not yield success; it is 0 if the option was not chosen in this period. Moreover, the propensity for an option depends on the discounted propensity of the last period, with the discount factor denoted as q. In sum, the propensity for option i in period t is given as:

$$P_i(t) = q \cdot P_i(t-1) + R_i(t-1)$$

The probability $\mathrm{Pr}_i(t)$ of choosing option i in period t is then given as:

$$\mathrm{Pr}_i(t) = \frac{\exp\left(P_i(t)\right)}{\sum_j \exp\left(P_j(t)\right)}$$

We only face two options, A and B. Therefore, we have:

$$\begin{aligned}
\mathrm{Pr}_{\mathrm{A}}(t) &= 1 - \mathrm{Pr}_{\mathrm{B}}(t) \\
&= \frac{\exp\left(P_{\mathrm{A}}(t)\right)}{\exp\left(P_{\mathrm{A}}(t)\right) + \exp\left(P_{\mathrm{B}}(t)\right)} \qquad (1.1) \\
&= \frac{1}{1 + \exp\left(P_{\mathrm{B}}(t) - P_{\mathrm{A}}(t)\right)} \\
&= \frac{1}{1 + \exp\left(-\Delta P(t)\right)} \qquad (1.2)
\end{aligned}$$

with

$$\Delta P(t) \equiv P_{\mathrm{A}}(t) - P_{\mathrm{B}}(t)$$

Moreover, for $\Delta P(t)$, we have:

$$\begin{aligned}
\Delta P(t) &= P_{\mathrm{A}}(t) - P_{\mathrm{B}}(t) \\
&= q \cdot P_{\mathrm{A}}(t-1) + R_{\mathrm{A}}(t-1) - q \cdot P_{\mathrm{B}}(t-1) - R_{\mathrm{B}}(t-1) \\
&= q \cdot (P_{\mathrm{A}}(t-1) - P_{\mathrm{B}}(t-1)) + R_{\mathrm{A}}(t-1) - R_{\mathrm{B}}(t-1) \\
&= q \cdot \Delta P(t-1) + R_{\mathrm{A}}(t-1) - R_{\mathrm{B}}(t-1) \qquad (1.3)
\end{aligned}$$

When choosing A in the last period led to a success, $R_{\mathrm{A}}(t-1) = 1$, while a failure with A results in $R_{\mathrm{A}}(t-1) = -1$; the same is true for B. On the other hand, the option that was not chosen is not reinforced. Therefore, $R_{\mathrm{A}}(t-1) - R_{\mathrm{B}}(t-1)$ equals 1 if A was chosen and yielded success or if B was chosen and did not yield success. In contrast, $R_{\mathrm{A}}(t-1) - R_{\mathrm{B}}(t-1)$ equals -1 if A was chosen and did not yield success or if B was chosen and did yield success. We can thus define $R(t-1)$ as:

$$R(t-1) \equiv \begin{cases} 1 & \text{A successful or B unsuccessful in } t-1 \\ -1 & \text{B successful or A unsuccessful in } t-1 \end{cases}$$

Inserting this in equation 1.3, we get:

$$\Delta P(t) = q \cdot \Delta P(t-1) + R(t-1) \qquad (1.4)$$

To calculate the probabilities, we thus insert equation 1.4 in 1.2:

$$\begin{aligned}
\mathrm{Pr}_{\mathrm{A}}(t) &= \frac{1}{1 + \exp\left(-\Delta P(t)\right)} \\
&= \frac{1}{1 + \exp\left(-q \cdot \Delta P(t-1) - R(t-1)\right)}
\end{aligned}$$

From these equations, it is clear that for the first period of the first generation, a probability to choose A or B cannot be defined. This is why for all strategies, the very first choice is defined to be random. For the first period of subsequent generations, offspring use as first choice the last choice of its parent. This is the only form of vertical transmission in our model and it is also applied to all other strategies. Individual learners do not inherit the propensities of their parents, though, instead starting fresh with propensities of 0.

Often, an additional modification to reinforcement learning is made by introducing a sensitivity factor λ (see e.g. [104]). This parameter alters the steepness of the probability of choosing an option as a function of the difference in propensities, so that:

$$\mathrm{Pr}_{\mathrm{A}}(t) = \frac{1}{1 + \exp\left(-\lambda \cdot \Delta P(t)\right)}$$

Low λ imply that the probabilities to choose A or B are rather insensitive to changes in propensity. The higher λ, the steeper the adjustment of the probability – even small differences in propensities would lead to large differences in the probabilities to choose A and B. In the extreme case of $\lambda = \infty$, this results in a simple threshold function that prescribes to choose A if propensity for A exceeds propensity for B and vice versa:

$$\mathrm{Pr}_{\mathrm{A}}(t) = \begin{cases} 1 & \Delta P(t) > 0 \\ 0.5 & \Delta P(t) = 0 \\ 0 & \Delta P(t) < 0 \end{cases}$$

We found that this simple threshold function leads to a performance that is almost the best that an individual learner can achieve. For this reason, and for reasons of simplicity and parsimony, we adopt this threshold function and do not further analyze λ.

The only free parameter that is left is thus q, the discount factor applied to previous propensities. For $q \leq 0.5$, propensities could never exceed 1 in absolute values in finite time, so that the reinforcement R would always exceed previous propensities. In other words, for $q < 0.5$, previous experience counts so little that it never affects the current decision. If previous experience counts that little, we effectively deal with win-stay lose-shift.

When $q > 0.5$, however, experience past the very last period may well affect the current choice, and the more so the higher q. We show an example of how individual learners with $q = 0.9$ behave in a given environment in figure 1.9. Individual learners with $q = 0.9$ too track the environment quite well but behave less "conservatively" than win-stay lose-shift (the right panel of the figure is a reproduction of fig. 1.8). For example, at around period 200, p_A is clearly greater than p_B, yet only approximately 60% of individuals using win-stay lose-shift pick A. In contrast, approximately 80% of individual learners with $q = 0.9$ pick A. On the flip-side, individual learners sometimes fall a little behind, e.g. during the changes between period 50 and 100. We will later analyze how performance of individual learners depends on q.

1.3.2.2 Bayesian individual learning

A natural alternative to individual learning based on exponential discounting reinforcement learning would be a Bayesian learning algorithm. A Bayesian learner has a prior belief about p_A and p_B and updates it after each realization of an outcome. For example, assume that in period t, the learner has a prior belief that p_A equals 0.6 of 0.7, $\Pr(p_A = 0.6) = 0.7$, and consequently chooses A. However, the learner is not successful with A. Then her posterior belief about A is:

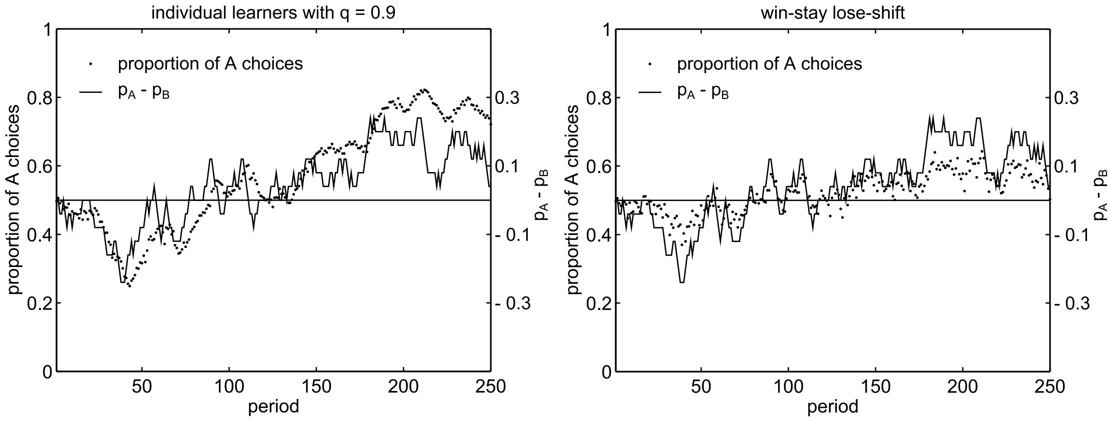

Figure 1.9: The behavior of individual learners with $q = 0.9$ (left panel) and of win-stay lose-shift (right panel) over time. The proportion of A choices made by the strategies (●) is measured on the left y-axis. The line represents the environment in the form of $p_A - p_B$, measured on the right y-axis. Individual learners with $q = 0.9$ track changes in the environment quite well and are less "conservative" than win-stay lose-shift but lag a little behind.

$$\Pr(p_A = 0.6|\text{A fails}) = \Pr(\text{A fails}|p_A = 0.6) \cdot \Pr(p_A = 0.6)/\Pr(\text{A fails})$$

The probability of A not succeeding when $p_A = 0.6$ is equal to 0.4. Assuming that $\Pr(\text{A fails}) = 0.5$, we have:

$$\begin{aligned}\Pr(p_A = 0.6|\text{A fails}) &= 0.4 \cdot 0.7/0.5 \\ &= 0.56\end{aligned}$$

Using this approach, it is possible to design a Bayesian individual learner.

The problem that occurs is that a Bayesian learner has to start with some kind of prior. We will be very generous: We supply the Bayesian learner the full probability distribution of p_A and p_B, as determined by simulating those values for a length of 10^7 periods (see figure 1.10). In reality, no learner would live long enough to be able to draw on such extensive data, so this should give a Bayesian learner a huge advantage. Even more, other learners in our model do not even know that there is a constant probability distribution of p_A and p_B, but as we said, we wanted to be generous towards Bayesian learners.

In each period, the Bayesian learner chooses an option, then this option is realized, resulting in either success or failure. After realization, the Bayesian learner updates her beliefs about the probability distribution of p_A if A was chosen or p_B if B was chosen. The belief about the option that was not chosen is not updated. In the next period, the Bayesian learner chooses the

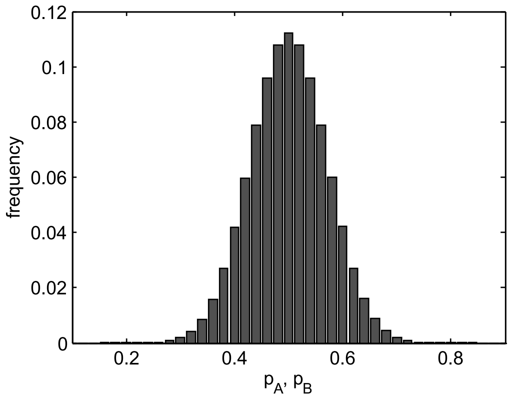

Figure 1.10: Probability distribution of p_A and p_B, the environment, for the default parameters.

option that, according to her beliefs, has the higher expected value and the cycle continues.

We made three further adjustments to Bayesian learners. As after each period, p_A and p_B change their values by one incremental step, they take distinct sets of values depending on whether the period is currently even or odd. For example, in odd periods, p_A can take the value 0.5 or 0.54 but never 0.52, whereas in even periods, the opposite is true. We allowed Bayesian learners to take this into account.

Another adjustment is to allow Bayesian learners to make inferences from their first period's choice, which is the same as their parent's last period's choice. If the parent chose A in the last period, it is more likely than not that $p_A > p_B$. We simulated the final priors of p_A and p_B conditional on whether the final choice was A or B and allowed Bayesian learners to use these biased priors as their initial priors, depending in whether they inherited A or B as their first choice.

Moreover, it is possible to give Bayesian learners the ability to predict the next step of the environment when they are given knowledge about how the environment changes from period to period. For example if they assign a probability to the event that currently $p_A(t) = 0.5$ and a probability to the event that currently $p_A(t) = 0.54$, they can use these probabilities to calculate the probability of $p_A(t+1) = 0.52$ when they know how the environment changes. This way, Bayesian learners not only have huge statistical knowledge about the environment, as well as the possibility to bias their priors according to their parent's last choice, but also a complete understanding of the process that generates the environment, a further advantage that no other strategy has.

The remaining conditions for Bayesian learners are the same as for individual learners. The very first choice is determined by chance. After each generation, offspring inherit their first choice from their parent. However, they do not inherit the beliefs of their parents but instead again start with the prior as determined initially. The situation is similar to the situation of reinforcement learners who also do not inherit the propensities of their parent.

1.3.2.3 Social learning through conformism

In this chapter, we do not want to present any breathtaking new insights derived from our social learning model but instead establish its validity. Therefore, before we begin to test new social learning strategies, we contend with studying a social learning strategy that is well explored already, conformism.

Conformism has always been an important part of the theory of social learning [21, 22, 86]. There is no single definition of conformism that is used by all researchers across disciplines, or even within disciplines. Boyd and Richerson write that "[Conformist] Individuals are assumed to be disproportionately likely to imitate the more common behavioral types among their cultural parents" [21], which is essentially how conformism (or "hyper-conformity", as some would call it [34]) is used in this work.

"Disproportionally" in the quotation above means that the probability of adopting the most frequent option is higher than this option's share in the population's choice. For example, say that option A is chosen by 60% of the population and option B by 40%. If social learning consisted of random choice, the probability to adopt A would also be 60%, not more or less than the option's frequency in the population. This would therefore not qualify as conformism. If a social learning strategy had an adoption probability of *more* than 60%, even if only of 61%, it could be conformism.[8] Similarly, since option B is chosen by only 40% of the population, it's adoption probability has to be *less* than 40% for a strategy to qualify as conformist.

For our purposes, we define conformism as a strategy that samples a fixed number of individuals and then chooses the option that the majority of the sampled individuals chose. For example, if three individuals are sampled and two or three of them chose A, the conformist will also choose A. We could implement a sensitivity parameter λ. This parameter would smooth out the step function-like mechanism we designed. When two of three individuals choose A, we could, e.g., say that the conformist also chooses A with 80% probability instead of 100%. But as for reinforcement learning, we found such a parameter to have little positive effect, and often even negative effects, so it is more parsimonious to drop it completely. This implementation of conformism thus corresponds to the example we discussed earlier in this chapter.

[8] A strategy with a less than 60% adoption probability would be called "anti-conformism" or a "maverick" [46].

1.3.3 Modeling details

We chose a population size of 10,000 individuals, which is thought to be the effective population size of early humans in the last 1 million years or so [83, 167]. Base fitness is set to 10, while the average fitness gain of choosing randomly would be 25 for a period length of 50. By choosing a low base fitness, random drift is less likely to determine the evolutionary outcome [136]. This is convenient assumption for the purpose of the simulation but most likely unrealistic. The validity of our result should not depend on this assumption, though.

For simplicity, reproduction is assumed to be asexual and generations to be non-overlapping. Generally, there is only vertical transmission of cultural traits in our model. However, offspring inherit their very first choice from their parent, since else, the very first choice would not be defined. Furthermore, the values of p_A and p_B in the first period of a new generation correspond to their last values of the previous generation. This way, not only is the choice but also the environment continuous between generations. Other variables, such as the propensities derived from experience, are not inherited.

Equilibrium frequencies can be estimated by two different methods. First, one can simply run the evolutionary simulations and check the frequencies after equilibrium has been reached. Since there is always stochasticity in the model (due to choice, the environment, and success having random elements), full convergence towards the equilibrium will never be observed, though. A second method is to fix the frequency of the strategies at different levels and estimate their performance precisely by averaging it over a large number of independent simulations. This allows us to derive performance as a function of frequency. Because fitness only depends on performance (there are no costs), the performance curves are sufficient to derive equilibrium frequencies.

All analytical results were derived with Wolfram *Mathematica* 8 and numerical simulations were run on Matlab 7 (The MathWorks, Inc.).

1.4 Preliminary findings

1.4.1 Individual learning

First we have a look at the performance of individual learners and how it depends on the only free parameter, the discount factor q. Furthermore, we differentiate between different generation lengths t_{max}. The results can be seen in figure 1.11. The more trials an individual has during her lifetime, the more experience she can rely upon to make her decisions. That is why performance is generally higher for higher t_{max}. This finding was expected.

Given that there are no sudden jumps in p_A and p_B, past experience should

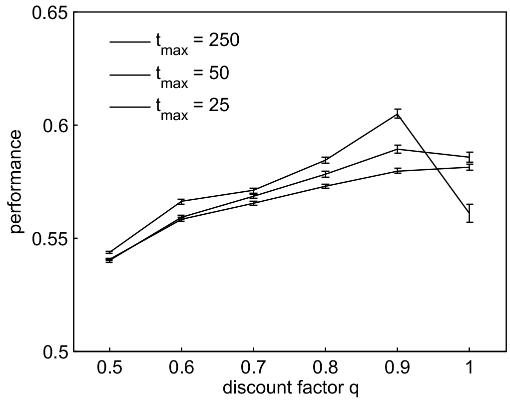

Figure 1.11: Performance of individual learners as a function of the discount factor q and for different lifetime lengths t_{max}. In general, performance is higher for longer t_{max} and higher q, except for $q = 1$. Error bars indicate standard errors of the mean. $q = 0.5$ corresponds to win-stay lose-shift. All simulations consisted of 250,000 periods and 1,000 individuals.

have a generous weight on the learners decision. If q is too low, experience is discounted too strongly and should lead to a lower performance. This is indeed what we observe. For the highest value of q, $q = 1$, we find, however, that performance is not higher than for $q = 0.9$. The reason is that for $q = 1$, experience even from 50 periods ago counts as much as the most recent experience. Yet the environment does vary over time, so recent experience should have more impact. We thus see a decline in performance for q values beyond 0.9.

If we have a low t_{max} of 25, we find performance not to differ significantly between $q = 0.9$ and $q = 1$. Here, even the oldest experience is never older than 24 periods and therefore, the disadvantage of not discounting it sufficiently does not manifest. The true optimal q is probably somewhere between 0.9 and 1, and we would actually see a decline in performance after this optimal q would have been reached. We wanted, however, to avoid overfitting the behavior of individual learners, which is way we did not test q to more decimals than the first.

For our default conditions ($t_{max} = 50$), we thus choose a discount factor q of 0.9. This leads to a performance of $58.96 \pm 0.16\%$ (standard error of the mean). Although this performance seems to be only little better than random choice (50%) or win-stay lose-shift (54.4%), it is already a vast improvement when one considers the difficulty of the task. In the default conditions, the mean difference between p_A and p_B is only 7.88 percentage points and 45.3% of the time, the difference is equal to or less than 4 percentage points. Considering the noisiness of the data, we have yet to see whether any social learning strategy is capable of besting individual learners.

When we regress the proportion of A choices of a population of individual

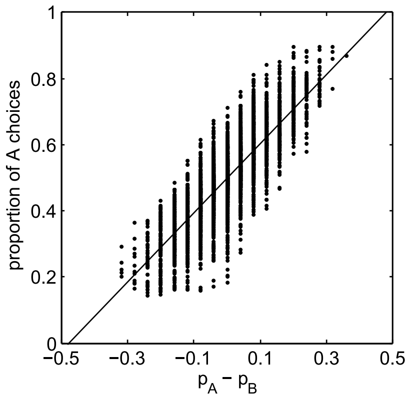

Figure 1.12: The proportion of A choices of individual learners ($q = 0.9$) as a function of $p_A - p_B$; simulation results (•) and linear fit (solid line). Individual learners "match" the probabilities in an almost 1 to 1 fashion. That is, when $p_A - p_B$ increases by 1 percentage point, the probability of an individual learner to choose A also increases by approximately 1 percentage point.

learner with $q = 0.9$ on the difference between p_A and p_B, we get an interesting finding (figure 1.12). The proportion of A choices made by the population corresponds to the probability of each individual learner to choose A. This probability seems to depend linearly on the difference between p_A and p_B. We fitted a linear model to the simulation data and found the slope to be 1.047 (1.041-1.054, 95% confidence bounds, $R^2 = 0.678$). This implies that if $p_A - p_B$ increases by 1 percentage point, we can expect a 1 percentage point increase in the probability of an individual learner to choose A.

This behavior corresponds to "probability matching" (with a power of 1 [14]). Probability matching is typically shown in experiments in the following way. The subject draws from two lotteries and after some exploration presumably reaches the estimate that lottery A yields a success, say, 70% of the time and lottery B 30% of the time (prizes are the same). What one should expect is the subject to continue drawing only from lottery A, as this would maximize her expected payoff. Instead, one finds that she draws from A roughly 70% of the time and from B 30% of the time [110, 118], hence the term "matching".

Although this deviation from optimal behavior seems irritating, it is consistent with optimal behavior when the fundamental probabilities are not fixed [72]. Our task corresponds to the probability matching task with variable environment and thus would suggest matching behavior.[9] This is indeed

[9] Note that in the numerical example above, the difference in success probabilities between A and B is actually 40% and would thus be matched with a 90% probability to choose A, not 70%, as indicated by the experiments. Our simulations reveal a slope of approximately 1, while the experiments suggest a slope of 0.5. If, however, we use $q = 0$ (win-stay lose-shift), we indeed get exact probability matching according to the exper-

what we find, supporting the idea that matching is caused by the need to explore in case of changing environments.

In the probability matching task, as in our model, optimal behavior consists of always choosing A when $p_A > p_B$ and always choosing B in the opposite case. The true values of p_A and p_B, at a given point in time, are, however, not known (except to the experimenter). Given the limited amount of knowledge that the subjects possess and taking into account the possibility of a varying environment, it may thus be optimal for an individual to choose according to a matching scheme. Subjects in the experiments may simply "import" their decision mechanism from the outside world, where exploration is often necessary, hence creating the apparent mismatch between optimal and actual behavior.

1.4.2 Bayesian learning

We discussed Bayesian learning as an alternative form of individual learning. For that, Bayesian learners could use the probability distributions of p_A and p_B as determined from 10^7 periods. These probability distributions are subsequently updated in a Bayesian fashion to estimate the values of p_A and p_B, taking experience into account. We found Bayesian learners and reinforcement learners to behave quite alike, but Bayesian learners actually performed worse. Therefore, we introduced several modifications to improve the performance of Bayesian learners, as discussed in the modeling section.

First, we allowed Bayesian learners to take into account whether they currently are in an even or odd period. This provided a very small but not sufficient improvement in performance. Second, we allowed Bayesian learners to use priors that are biased towards A or B, depending on whether they inherited A or B as their first choice. This led to a lower performance, probably because too much weight was given to outdated information about the environment. Starting with fresh, unbiased priors proved to be the better approach.

Third, we gave Bayesian learners full knowledge about the process that generates the environment. Using this knowledge, Bayesian learners can additionally predict the environmental state of the next period based upon their beliefs about the probabilities of the current period. An example of the resulting behavior is shown in figure 1.13. This change indeed improved the performance of Bayesian learners, which consequently reached 60.15% (±0.67%, S.E.M.). Therefore, Bayesian learners can be made to outperform reinforcement learners, albeit barely with a 1 percentage point advantage. A

iments, with a slope of 0.5. Also, analytically, the probability of win-stay lose-shift to choose A is:

$$\frac{1 - p_B}{1 - p_A + 1 - p_B}$$

This reduces to p_A if $p_B = 1 - p_A$, which is usually the case in the matching experiments.

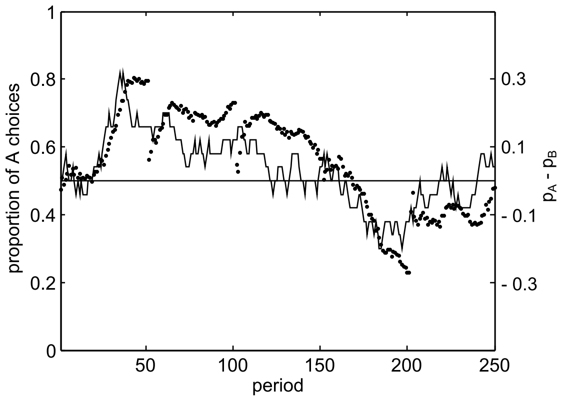

Figure 1.13: The behavior of Bayesian learners over time. The proportion of A choices made by Bayesian learners (•) is measured on the left y-axis. The solid line represents the environment in the form of $p_A - p_B$, measured on the right y-axis. Bayesian learners approximately match probabilities. Note that after every 50th period, when new generations start, behavior reverts back to the mean, reflecting that Bayesian learners start with fresh, unbiased priors.

linear regression of the matching behavior of Bayesia learners resulted in a slope of 1.083 ($1.075 - 1.091$, 95% confidence bounds, $R^2 = 0.580$), showing that Bayesian learners also match probabilities.

We varied the generation length t_{max} to see how it affects the performance of Bayesian learners. For $t_{max} = 25$, we find a performance of 59.42% (±0.55%), for $t_{max} = 100$ of 59.69% (±0.76%), and for $t_{max} = 250$ of 59.96% (±0.83%). These values barely differ. Generation length thus does not seem to have a strong impact on the performance of Bayesian learners.

In sum, even though we provided Bayesian learners with vast data about the model – information that is unavailable to any other strategy – and also with complete knowledge about the process that generates the environment, their performance can at best slightly exceed the performance of the much more simplistic reinforcement learners. The reason why Bayesian learners cannot do much better is that the environment is always changing. Bayesian learners chase after a target that moves faster than they are able to learn. It does not make much sense to calculate any supposed priors. Instead, it is better to discount old information, as reinforcement learners do.

Not only does Bayesian learning presuppose that the individuals have unrealistically deep insight into the environment, it is also very complex, requiring to remember a lot of information. For each distinct value of p_A and p_B, 2×51 possibilities for the default condition, the Bayesian learner has to know the unbiased prior in the first period, as well as the updated prior of the current period. After an outcome is realized, all priors have to be updated and then the expected values of the priors have to be computed and compared with each other. In contrast, reinforcement learning by exponential discounting only requires to remember the difference in propensities of A and B and the discount factor, and to multiply these two values. Reinforcement learning is thus far more parsimonious than Bayesian learning.

In sum, Bayesian learning presumes a lot of background information about the environment and on how it works, is far less parsimonious than reinforcement learning by exponential discounting, and requires a lot more computation, while still performing only minimally better. In our model, reinforcement learning by exponential discounting is thus the better candidate for an individual learning strategy and will henceforth be used instead of Bayesian learning.

1.4.3 Conformism

Conformists choose the most common option in their sample. An interesting consequence of this behavior is that it leads to uniformity of choice. If slightly more than half of a conformist population chooses A, eventually, all individuals will choose A (see top row of figure 1.14). Higher sample sizes make it easier to detect which option is actually the more frequent one and are thus associated with higher conformity, in the sense of a more aligned behavior of the conformists.

If transmission were unbiased (like for random copying), the frequency of A choices would not change (except by drift), and if transmission were anticonformist, it would tend to return to 50% a choice. A conformist population would, however, tend to exaggerate even small deviations. This exaggeration is even stronger for higher sample sizes. An illustration of this is found in the bottom left panel of figure 1.14, were the behavior of conformists of different sample sizes is shown. We simulated their behavior in the same environment when being paired with an equal number of individual learners, who are not shown. This example shows that instead of matching the environment, as individual learners would, conformists overmatch. This is validated by a linear regression (bottom right panel of figure 1.14), which results in a slope of 1.903 (1.853-1.953, 95% confidence bounds, $R^2 = 0.525$).

When a conformist has a higher sample size, she is more likely to actually choose the option that is most common in the population. As long as conformists are sufficiently rare, this should lead to a better performance and hence fitness. This can be seen in the left panel of figure 1.15. Conformists with higher sample size are, however, also more likely to collectively choose the incorrect option. Maladaptive behavior is possible earlier for higher sample sizes, and the incorrect option is more likely to be chosen once that becomes a possibility. Higher sample sizes thus have advantages and disadvantages.

Next we performed numerical evolutionary simulations starting with a population consisting predominantly of individual learners and a minority of conformists. We simulated conformists with sample size 3, 5, and 7.

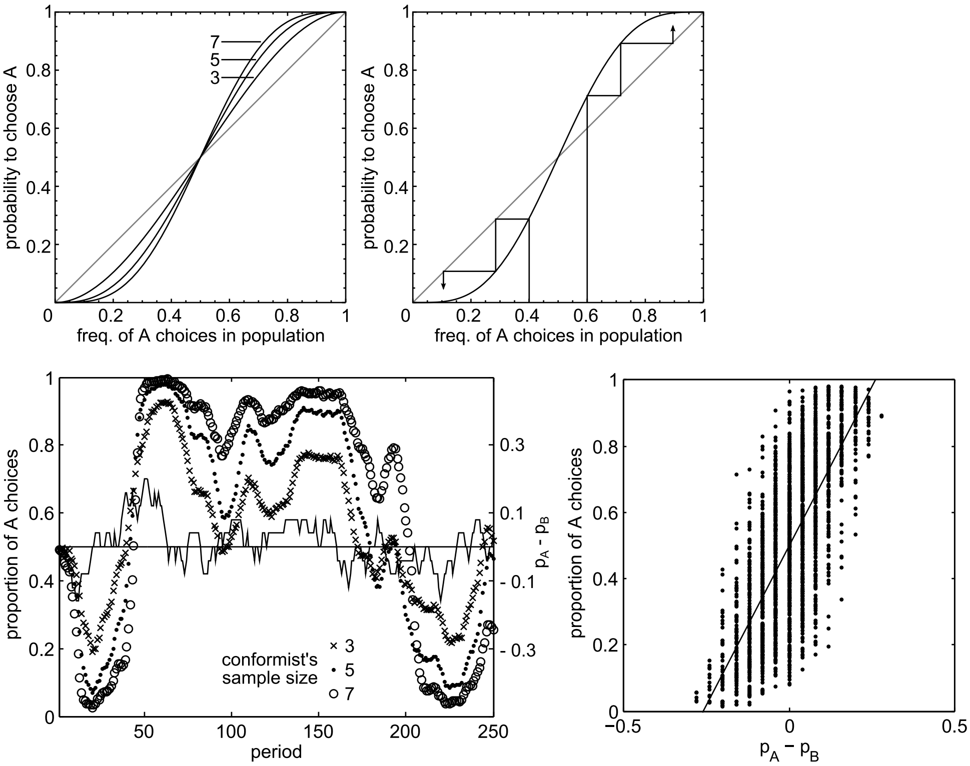

Figure 1.14: Top row: recurrence relation for conformists. Top left: comparison between different sample sizes (as indicated by the numbers). There are three equilibria, nobody chooses A, everybody chooses A ("conformist" behaviors, both stable), and 50% choose A (unstable). Higher sample sizes result in stronger conformity in the sense of faster convergence towards total conformity. Top right: The recurrence relation can be used to trace the convergence to conformity, as indicated by the arrows. Bottom left: Behavior of conformists over time. Conformists with sample size of either 3 (×), 5 (•), or 7 (◯) were paired with an equal number of individual learners (not shown). The proportion of A choices by conformists is shown on the left y-axis, the environment in the form of $p_A - p_B$ (solid line) on the right y-axis. Bottom right: Conformists (here with sample size 3) overmatch, since small changes in the differences between p_A and p_B may lead to large swings in the proportion of A choices.

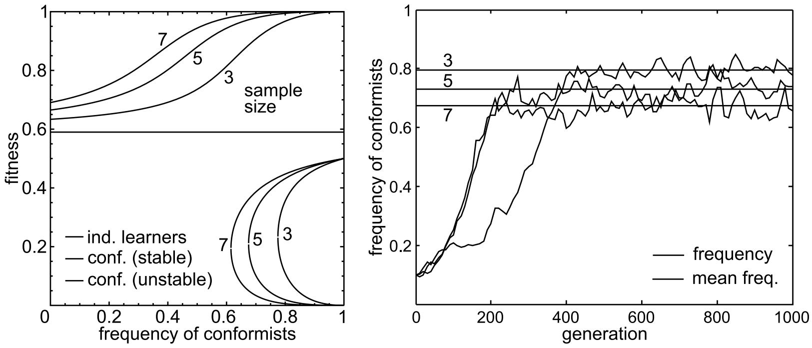

Figure 1.15: Fitness (left) and evolution (right) of conformists, depending on their sample size. Left: analytically derived fitness (stable: solid line, unstable: dashed line). Performance of individual learners (dotted line) was assumed to be 59%, as measured previously, and is independent of the frequency of conformists. Conformists with higher sample size are initially more likely to choose correctly than those with small sample size, and thus have a higher fitness at low frequencies. They are, however, also more likely to collectively choose the wrong option, as is evidenced by the earlier appearance of the bifurcation that marks the emergence of the maladaptive equilibrium. Right: Comparison of evolutionary simulations of the frequency of conformists (solid line) with different sample sizes as indicated when competing with individual learners (not shown). Conformists with lower sample size have higher equilibrium frequencies (as measured by the mean frequency during the last 500 generations, dash-dotted line) but take more time to reach it.

Our analytical findings suggest that 1) conformists with higher sample size invade faster because they initially have a higher fitness and 2) conformists with lower sample size have a higher equilibrium frequency because it takes more of them before maladaptive behavior occurs. In the right panel of figure 1.15, we show the simulation results. Although the invasion speed of conformists with sample size 5 and 7 is indistinguishable, both invade faster than conformists with sample size 3. Furthermore, the latter have the highest equilibrium frequency, calculated as the arithmetic mean of the last 500 generations, 79.6% to be precise. Sample size of 5 leads to an equilibrium frequency of 72.9% and 7 of 67.4%. All differences are significant (all $p < 10^{-11}$, Wilcoxon rank sum test). The numerical simulations therefore reflect our predictions from the analytical findings.

Instead of directly simulating the evolutionary path of conformists, we can use a different method to gauge their evolution. For this, we fix the frequency of the involved strategies and simulate their performance a large number of times. This will let us estimate precisely the performance of the strategies at different frequencies, and since we know that a strategy with a higher performance than its competitor also has higher fitness, we can infer the evolutionary path of the strategy from this. When performance is equal, we should expect that the corresponding frequency is an equilibrium. From

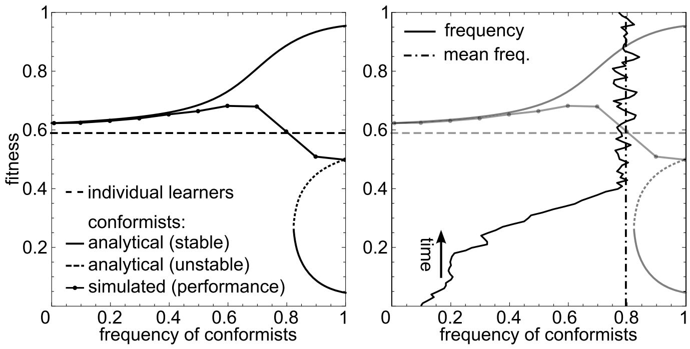

Figure 1.16: Comparison of different approaches to infer evolutionary outcome of conformists with sample size 3. Left panel: Comparison of performance measures with analytical results. The performance approach correctly predicts conformist fitness for lower frequencies. When the frequency of conformists become too high, maladaptive equilibria may sporadically occur, leading to the observed decline in performance. Right panel: Additional comparison with evolutionary simulation (left panel and evolutionary path superimposed). The performance approach correctly predicts the equilibrium frequency, finding equality in performance at the equilibrium frequency determined by evolutionary simulations. It also allows us to correctly infer the stability of the equilibrium. Standard error of the mean of the performance measures vary between 0.12% and 0.67% for conformists and are not shown.

the form of the curve – whether higher frequencies lead to better or worse performance – we can infer the stability of the equilibrium. The advantage of this method compared to direct evolutionary simulations is that since we measure performance repeatedly, we have more confidence for each data point. Moreover, we may find very high or low equilibria that might get lost due to stochasticity in the evolutionary simulations. (We will indeed encounter such cases in later chapters.)

To ensure that this approach leads to convergent results, we tested it with conformists of sample size 3. To do this, we used the error rate of individual learners and the frequency of environmental changes as derived from the simulation to fit the model. Fitness of conformists was predicted by the analytical model. The result is shown in figure 1.16. In the left panel, we see that the derived fitness almost perfectly predicts simulated fitness up to a frequency of 0.5. Afterwards, we observe some deviations, which are probably caused by conformists occasionally converging towards the worse option.

On the right panel of figure 1.16, we additionally superimposed the frequency of conformists as derived from the evolutionary simulation shown in figure 1.15. This simulation suggested an equilibrium frequency of 79.6%. The performance measure indicates that for a frequency of 80% conformists, performance should be about equalized (59.52 ± 0.51% for conformists vs

$58.96 \pm 0.16\%$ for individual learners, $\pm$S.E.M.); it thus correctly predicts the equilibrium frequency. Additionally, as conformists perform better when less frequent and worse when more frequent, the performance measure allows us to correctly infer that the equilibrium is stable. All in all, the results from the performance approach corroborate our earlier findings and thus validate the approach.

For sample sizes of 5 and 7, we simulated the performance of conformists at the equilibrium frequency suggested by the evolutionary simulations. We found performance to be $58.11 \pm 0.55\%$ and $59.20 \pm 0.55\%$, respectively, very close to the performance of individual learners ($58.96 \pm 0.16\%$). Therefore, the performance approach would yield the same equilibrium frequency as the evolutionary simulations for the given sample size, confirming this approach to be a good substitute for evolutionary simulations.

1.4.4 Principal findings

As we said, the main goal of this chapter was not to provide major new insights into social learning. Instead, our focus was to present some of the more interesting findings of the field and our own approach. Still there are some take-home-messages in this chapter that can be summarized as follows:

- Rogers' paradox is akin to the paradox of cooperation; in our opinion, uncovering how social learning works is as important for our modern world as uncovering the conditions and mechanisms of cooperation.
- The implications of Rogers' paradox are even more severe when focusing on performance – in equilibrium, performance is, by necessity, worse than before the arrival of social learners.
- Conformism, although capable of boosting performance transiently, is not a solution to Rogers' paradox – in equilibrium, the population's mean performance must be the same as that of individual learners.
- Individual learners could impede the invasion of conformists by performing *worse* but competition among individual learners will probably prevent that.
- Conformists can reach a higher equilibrium frequency if they sample less individuals but will take more time invading a population of individual learners.

Furthermore, we presented a social learning model that we believe to be quite useful as a social learning paradigm. As far as we could, we validated the results by comparing them with analytically derived predictions.

1.5 Discussion

1.5.1 The importance of Rogers' paradox

To completely understand the implications of Rogers' paradox, an analogy to another paradox uncovered by theory may help. Sexual reproduction is a common mode of reproduction in nature. For a long time, nobody wondered why organisms reproduce sexually. As John Maynard Smith noted [162], however, this reproductive mode cannot be justified by simple means. In a sexually reproducing species, an asexual mutant could, all else being equal, contribute twice as much genetic material to the next generation as its sexually reproducing conspecifics. Therefore, sexually reproducing species bear the cost of having only half the fitness of an otherwise identical but asexual mutant. Sexual reproduction leads to even further costs caused, e.g., by fights for mates and by sexually antagonistic evolution. This finding was called the two-fold cost of sex.

The conclusion from this theoretical finding is of course not that sexual reproduction is maladaptive. There have to be advantages, although it is not quite trivial to explain how these advantages can overcome the two-fold cost. The real importance from Maynard Smith's finding is that it made obvious that there is a lack in our understanding of sex, a lack that has to be investigated.

Another important theoretical finding of the 20th century was that cooperation is not easily explained. Non-cooperative game theory suggests [116, 130] that cooperation is not an equilibrium behavior (neither minimax nor Nash). Similarly, in nature, individual selection usually trumps group selection, so that competition should trump cooperation. Here too, the conclusion is not that that the scrutinized behavior, cooperation, is maladaptive, but rather that there is a lack in our understanding.

This perspective should be taken when considering Rogers' paradox. The straightforward conclusion is that social learning does not, in equilibrium, increase human adaptability. The real conclusion, however, is not that culture does indeed not contribute to adaptability, but that there is a lack in our understanding, a hole in the theory that needs to be filled.

The finding that the evolutionary origins of both sex and cooperation cannot be easily understood has sparked scientific endeavors so numerous that we could not even begin to cite all the resulting literature. Leaving aside the question whether these two riddles have been sufficiently solved, it is clear that the initial puzzles have attracted the much needed brain power to address them. On the other hand, the evolutionary origin of culture is studied by only a few scientists. The reason may be that it is the most recent of the three puzzles; it may be that it is less attractive a problem; it may be that the general impression is that it has already been solved (we later show why we think it has not); or it may be that the importance has not (yet) been sufficiently recognized. Whatever the reason, the evolutionary origin of culture is the least studied of the three phenomena, even though they are all considered "major transitions in evolution" [163].

1.5.2 Social learning and culture

Two of the most important researchers of the evolution of culture, Peter J. Richerson and Robert Boyd, define culture as follows [144]:

> Culture is information capable of affecting individuals' behavior that they acquire from other members of their species through teaching, imitation, and other forms of social transmission.

This very broad definition of culture encompasses the phenomena we study in this work. The cultural trait we study is whether option A or option B is adopted. Most people would probably not consider this to be culture. They would rather think of more elaborate cultural techniques, such as cave paintings, ornaments, or music. Also, technological progress would be considered as culture, such as ever more efficient arrow heads or kayaks. These cultural traits are seen as such because they show a level of sophistication that exceeds what one individual could find out by herself during her whole lifetime [144].

These cultural traits are what the literature calls "cumulative", i.e. they can be enhanced and modified over time. Without cumulative culture, the spread of humans to regions ranging from deserts to the Arctic could not be understood [26]. Mankind's greatest achievements are based on cumulative culture, whether they be Newton's *Principia*, Bach's fugues, or landing a robot on Mars.[10]

Yet these forms of cumulative culture are not what we are interested in here. Models of cumulative culture usually assume that cultural traits are accumulated or can be strictly enhanced (e.g. [23, 24, 48, 51, 111]). Our model instead just proposes two options that alternate in being the best option. This approach is clearly not appropriate to model cumulative culture.

Why did we decide not to model cumulative culture? From our point of view, cumulative culture is a different topic that raises different questions. It deals with different questions. Do humans occupy a specific "cognitive niche" that allows us to innovate especially well; or do we occupy a "cultural niche", with our ability to spread adaptive cultural traits rapidly [26]? Modeling cumulative cultural evolution has to somehow grapple these opposing theories. It has to make clear what the origin of innovations and their mode of diffusion is [147].

[10] Newton, despite not being known for his modesty, famously proclaimed that "If I have seen further it is by standing on the shoulders of giants."

Two aspects are in our opinion crucial in differentiating between models of social learning as ours and models of cumulative culture. First, by their very nature, innovations in cultural traits are rare. They may not even occur once per trait and generation, or at least they used to during much of human's history. As recently as until the beginning of the 19th century, technology growth was so low as to be barely measurable [35] and the very notion that mankind would make progress was considered outlandish.

The second difference is that innovation means improvement. Some traits are simply superior to others. They allow to generate a greater benefit or to save production costs, thus increasing productivity. The question is then not whether it is right or not to adopt the trait, but rather whether, when, and under what circumstances adoption occurs. The perhaps most famous study of innovation was the adoption of hybrid corn [148]; hybrid corn produced superior yields but was still slow to be accepted by farmers in Iowa, which sparked the question why innovations often take so long to be adopted.

Taking these features of cumulative culture into account would require quite a different model from ours. In our model, there is not one option that is strictly superior over time. Furthermore, we assume that both options are commonly known, meaning that we allow an individual to adopt an option even if she has never seen anyone else choosing that option.[11] Our model does not bother with the question how individuals come up with new options, which a comprehensive model of innovation eventually has to. As the options in our model vary in payoff, individuals have to constantly adopt during their lifetime. After acquiring a superior trait, it is not at all certain that it will still be superior ten periods later. Constant learning is required.

One could say that cumulative culture deals with traits whose adoption occurs rarely in life but has huge fitness impacts. Non-cumulative culture as we model it deals with adoption choices that occur frequently in life but that have, each by themselves, little fitness consequences.

Since our model neither accommodates hybrid corn nor space flights, could it not be argued that we actually model nothing of importance? We would, of course, object. Instead, we claim that although we model something different, it is still something important. These tiny life choices we describe may each have little fitness value but as they are so common, they could still add up to have a large impact.

Consider foraging, a daily task in most small scale societies. There is not only the choice between different forms of foraging, like hunting and gathering, but also between what items to search for, and how to prepare them for meal. These choices have direct and indirect rewards, but those can be noisy. One day the hunt will be successful, the other day it is not. Relying purely on one's own experience could be less informative than taking into account how others did.

[11] This defuses the criticism raised by Eriksson et al. [52].

Apart from food choices, many other common decisions could be influenced by social learning. What clothes one should wear, which direction to take, what manners to display, what words to use, what products to buy, whom to choose as interaction partner, whose writing style to imitate. In the modern world, such choices may have tremendous consequences for everyone on the planet. Just think of bubbles in financial markets, which are arguably caused by social learning (more on that in the last chapter).

As we argued, these many small decisions are equally interesting to study as the few large decisions. Should one expect different social learning strategies to be used depending on which of the two types of decisions are concerned? This question cannot be easily answered. On the one hand, there is no obvious reasons why the same strategies should work equally well. For example, when making many small decisions, conformism is often a good strategy in the face of uncertainty. If we deal with innovations, however, conformism will initially slow down progress because an innovation, per definition, is rare when it is first discovered, and thus unlikely to be adopted by a conformist.

On the other hand, it could be argued that the same social learning strategies apply to all domains. First, our brains could be constrained to use the same learning strategy for everything, not being able to readily switch. Second, the difference between small and large decisions could be obvious in theory but in the real world, options are not labeled according to how fast they vary or how often innovations are expected.

Ultimately, it is an empirical question when which social learning strategy is applied. Our theoretical work can be used as a guide to what to look for, and whether there are huge gaps in our understanding of the problem. Empirical research has to uncover which social learning strategies are actually used.

1.5.3 Choice of the model

We think that our approach to model the evolution of social learning is more appropriate to meet the goal of modeling non-cumulative culture than previous approaches. In previous approaches, it was often assumed that an environmental state is characterized by one choice being correct and the other(s) wrong. The image that is often invoked was that of a mushroom that could be poisonous or not [90, 100, 101]. The state would, however, reverse from time to time. But imagine that we had a social learner who copies the choice of others. If she sees others eating the mushroom, she will also give it a try. If we are in the environmental state where the mushroom is poisonous, clearly, she will become sick. Nevertheless, according to previous models, she will continue to eat the mushroom and pass this behavior on to following generations. Only when an individual learner appears who comes up with the idea of not eating the poisonous mushroom will social learners who copy this individual stop eating the mushroom. This whole description seems totally unrealistic. The problem is that if outcomes allow the learner to clearly distinguish between good and bad, as in the mushroom example, learning is trivial and there is nothing worth modeling.

One has to think of other situations where the outcomes are not of the all-or-nothing type. Perhaps the mushroom example could be modified so that the mushroom either has a slightly positive or slightly negative energy balance. Then the bad outcome is not so obviously bad. But this does not resolve the implausibility. The difference has to be detectable – if it were not, individual learners could not learn it. The social learner, after having copied a choice, should thus also be able to detect the difference and adopt the best behavior afterwards. There is no specific reason why the social learner should ignore her private information (except if one assumes that social learners totally lost their ability to learn individually). So we still cannot find a plausible application of this model assumption to the real world.

What would happen if the outcomes of the choice options were noisy? That is, although one option is at least temporarily better than the other, for each individual, the outcomes vary so much that the probability distributions of the two options overlap. Then, if a social learner chooses A and receives a worse outcome than her choice of B yielded, she can still reasonably choose A if her social information indicates that A is better (as e.g. in models of informational cascades [10, 16]). She thus ignores her personal information because it could be faulty. This, for us, is the most plausible scenario that would allow social learning strategies to evolve. As indicated earlier, examples that would fit this description include foraging, clothing styles, manners, consumption decisions, or the choice of interaction partners.

A more detailed example may help illustrate what situations our model captures. Imagine a teenager who wants to inquire whether wearing yellow shirts increases his attractiveness. If he wears a yellow shirt and encounters the target group, he receives feedback that allows him to infer his perceived attractiveness. This feedback, however, is noisy; another target group may respond differently, and there may not be sufficiently many opportunities to reduce the variance to a reasonable level, especially since fashions change. Therefore, this teenager may instead opt to wear shirts of the color that his peers most frequently wear or that a person he admires often wears. Such situations can be captured well by our model. One might object that since there are so many different shirt colors, a binary choice task is too simplistic. However, humans often search sequentially (e.g. [29]), so that our teenager might first test yellow against blue, then the (supposedly) better of the two against red, etc.

Another example that is probably closer to the life of a scientist might be whether publishing in journal X will lead to more citations. Imagine a scientist who has published an article in X with the result that few people read and cited this article. She concludes that it does not pay to submit to X. But maybe, this failure was due to other factors such as the particular topic not being fashionable at that time. If other people did have success with publishing in journal X, as suggested for instance by a high impact factor, the scientist may be best off ignoring her own experience and should continue submitting to X. (But perhaps even scientists can relate more to the former example.)

We also think that it is important that models of gene-culture coevolution allow for multiple decisions per lifetime. Having to make more than just one choice gives rise to intergenerational dynamics of decisions. It could thus happen that at the beginning, A is better than B, followed by periods of about equal payoffs from both choices, and ending with B being better than A. The choice of the individuals could or could not reflect this pattern, depending on how the individuals learn. Such dynamics cannot be studied if there is only one decision per generation.

Having more than one decision per lifetime also allows us to model the learning process itself in a more realistic fashion. Previously, it was assumed that individual learning consists of paying an exogenous costs to receive the information which of the choices is better, sometimes allowing for the possibility of a fixed error rate. Phenomena, such as that even individual learners do not immediately notice sudden changes, cannot be captured by simplistic models. However, if there are several trials and errors per lifetime, a more realistic learning mechanism can be implemented, e.g. learning strategies such as reinforcement learning [97]. As a byproduct, there is no more need for exogenous costs of individual learning; the cost of individual learning is the opportunity cost of not using social learning.

Dropping exogenous learning costs has the additional advantage of allowing us to drop two otherwise crucial parameters – the cost of learning individually and the cost of learning socially. First of all, the values of those costs are hard to measure. Moreover, our results do not rely anymore on the assumption that individual learning is more costly than social learning. Social learning requires a sophisticated cognitive apparatus and it requires time to follow others around to observe them. Individual learning, however, relies on a system that is already there and probably optimized by evolution. One can at least doubt the assumption that social learning is under all circumstances the cheaper way of learning. Our choice to exclude exogenous learning costs should, however, not be interpreted to mean that we believe there to be no costs of learning. We just believe it is best to leave this question open until more is known.

Interestingly, figure 1.15 reveals that as long as conformists are not too frequent, increases in their frequency actually improve their performance – there is positive frequency-dependent selection. This contrasts with the usual findings (figure 1.3) that social learners perform worse the more social learners there are. Consequently, we think that in our model's framework, it is inappropriate to think of individual learners as "information producers" and social learners as "information scroungers" or free-riders [69, 70, 101]. This terminology implies that social learners do not contribute to the information pool used by the population, which is contradicted by the transiently positive frequency-dependence.

Our model shows that for certain frequencies, social learners make more accurate decisions than individual learners. Their advantage stems from the performance boost and not from saving the cost of learning individually, which are non-existent in our model. Therefore, at least hypothetically, there can be an equilibrium involving social learning which leads to more accurate decision making. The adaptedness of culture would not simply stem from saving the costs of learning individually but also from increasing overall performance.

1.5.4 Outlook

In this chapter, we have established how we model social learning and why we made certain choices. Importantly, we also established what we do not try to model. Furthermore, we have presented some interesting findings from the social learning literature, most importantly Rogers' paradox.

A social learning model as ours permits an almost uncountable number of distinct social learning strategies. It would be impossible to treat them all. In the second chapter, we therefore focus our attention to a special class of social learning strategies, and characterize different types of strategies within that class. We will analyze the peculiarities of these strategies, as well as their potential.

Among those studied strategies, some will have the potential to solve Rogers' paradox. In the third chapter, we submit those strategies, in addition to other strategies proposed by the literature, to the test. We will find that strategies that have been proclaimed to solve Rogers' paradox do not actually solve it. Even worse, we find that in finite populations, there is the danger that social learning actually makes populations worse off. We conclude by showing the very narrow conditions under which some social learning strategies solve Rogers' paradox and discuss evidence for different forms of social learning.

There is no question that social learning profoundly shapes our everyday lifes, even in modern societies. The fourth chapter therefore addresses the question of how behavior in populations and societies is affected by the ubiquity of social learning. We will find that social learning alters behavior in a way that has striking similarities to anomalies observed in financial markets.

2 Payoff-biased social learning

2.1 Introduction

In the last chapter, we presented the basic outline of our model. Our learning strategies were, however, reduced to reinforcement learning by exponential discounting, Bayesian learning, and conformism. Although we believe we have found in reinforcement learning a satisfying candidate for the individual learning strategy, there are many more social learning strategies to be explored. This chapter will deal with a certain class of such strategies.

2.1.1 The problem

Individual learners, conformists, and other similar strategies do not observe the outcome of other individuals' choices, even though these observations may yield valuable information. An individual learner relies solely on her own experience. A conformist bases her decision on what other individuals do but does not care whether the outcome of these actions is good or bad.

An example should illustrate why outcomes are important. Suppose that a conformist observes seven individuals. Of those, four choose A and three choose B. A conformist (according to our definition) will then choose A. If the A choosers, however, all failed to reap the reward, whereas the B choosers were all successful with their choices, an argument could be made that it is wiser to opt for B instead of A. This is the rationale for introducing payoff-biases social learning (PBSL).

PBSL requires to observe the payoff that an option generates (hence the name) and to hopefully make good use of this additional bit of information. It requires to integrate more information than conformism, information which might or might not be available in real life situations. For now, we will not be concerned with the question of how realistic these assumptions are but will try to give an exhaustive picture of PBSL (as we define it).

The number of possible embodiments of PBSL is large, much larger than the possible embodiments of conformism. For the latter, given that there two options and no sequence effects, there are only four possible observations: {AAA}, {AAB}, {ABB}, and {BBB}. Each observation is associated with a probability to choose A. The way we model conformism, observation of {AAA} and {AAB} would lead to choosing A with probability 1, observation of {ABB} and {BBB} would lead to choosing A with probability 0. There are thus four possible parameters to choose in order to define conformists.

For symmetry reasons, however, the probability to choose A after observing {AAA} must be the same as the probability to choose B after observing {BBB}; the same holds for {AAB} and {ABB}. This reduces the number of free parameters to only two. Given that these parameters are probabilities, their values have to be between 0 and 1, and for the strategy to show a conformist bias, their values are even more restrained. To sum things up, conformists, who only observe what is chosen and who totally ignore the outcomes, can be characterized by a vector with two entries in real values.

Payoff-biased social learning broadens the number of possibilities considerably. Not only is the observed action important, but we have also consider the outcome. In our case, outcomes can either be positive (success) or negative (failure). Therefore, instead of two possible realizations per observation, A or B, we have four possibilities, A+success, A+failure, B+success, B+failure.

Restricting ourselves to a sample sizes of only three, there are already 20 possible observations. Of those, half can be eliminated for symmetry reasons, leaving us with 10 possible observations. Instead of being able to characterize a strategy by a vector with two entries, we thus have to resort to a vector of 10 entries, giving us much more possibilities. Assuming that every observation leads to choosing A with a probability of either 0, 0.5, or 1, we would already deal with $3^{10} = 59,049$ distinct strategies. If we were to allow for a higher sample size of, say, seven observations, the number of possible combinations would skyrocket to 60 possible observations (when symmetry holds) and $\approx 4 \cdot 10^{28}$ distinct strategies. Although many of the possible strategies would be nonsensical, there are still too many possibilities left to analyze all of them.[1]

Additionally to the fact that there are too many possible PBSL strategies to study them all, there is also the concern of plausibility. In the end, each learning strategy should be reasonable from a psychological point of view. A strategy that basically is defined as "If you observe A succeed twice and B once, choose A with probability 0.7; if you observe A succeed once and fail once, and B succeed once, choose A with probability 0.4; if..." is not a very realistic decision mechanism. If a strategy can be phrased as a heuristic, however, it makes the strategy more plausible. For instance, conformists, as we define them, work according to the heuristic "Choose the option that you observe most frequently".

This simplicity is arguably one of the reasons why conformism has often be the focus of social learning studies. We will try to find PBSL strategies that have a similar heuristic allure. Two classes in particular are considered in the following, the first based on scoring the different options, the second

[1] Unfortunately, to our knowledge, there is no easy way to deduce the performance of a strategy from its definition. We have to use simulations, which use up computation time; this is why we cannot just analyze all possibilities.

	A	$\bar{\text{A}}$	B	$\bar{\text{B}}$
conformists ($\equiv$ weights [1/1])	+1	+1	+1	+1
PBSL with weights $[1/-1]$	+1	−1	+1	−1
PBSL with weights $[4/-1]$	+4	−1	+4	−1
PBSL with weights $[1/-4]$	+1	−4	+1	−4
PBSL with weights [1/0]	+1	+0	+1	+0

Table 2.1: Scoring rules for different strategies. A indicates that the sampled individual chose A and succeeded, $\bar{\text{A}}$ that she chose A and did not succeed, B that she chose B and succeeded, and $\bar{\text{B}}$ that she chose B and did not succeed. A scoring-type PBSL is characterized by a pair of weights $[x/y]$, where x is added to the score of A (B) each time A (B) is observed being chosen successfully, and y is added to the score of A (B) each time A (B) is observed being chosen unsuccessfully. Conformism can be narrowed down to a special case of scoring-type PBSL with weights [1/1].

on averaging the payoffs of the options.

2.1.2 Scoring-type payoff-biased social learning

To understand how scoring works, we first have a look at a strategy that does not obviously rely on scoring. Conformism was defined as "choosing the option that is observed most frequently". But this strategy can be reformulated in a slightly different and less elegant way as "each time you observe A, add a point to the score of A; each time you observe B, add a point to the score of B; finally, choose the option with the higher score". This principle is shown in the first row of table 2.1.

Using the logic of adding scores for A and B and then choosing the option with the highest score, we introduce the class of scoring-type PBSL. Conformism is just a special case of that class. In general, it encompasses all strategies that obey the rule "each time you observe an individual who chooses A (B) and is successful, add the amount x to the score of A (B); each time you observe an individual who chooses A (B) and is unsuccessful, add the amount y to the score of A (B); finally, choose the option with the highest score". The numbers x and y could be any real number, but it is sufficient to restrict ourselves to (positive and negative) integers.

This principle is more easily understood by giving an example. Suppose that a scoring-type payoff-biased social learner puts the weight 1 on observed successes and the weight -1 on observed failures. This individual observes $\left\{\text{A}\bar{\text{A}}\text{B}\right\}$, i.e. one individual choosing A and being successful, one individual choosing A and being unsuccessful, and one individual choosing B and being successful. The total score S_A in favor of A is thus $S_A = +1 - 1 = 0$, and the total score in favor of B is $S_B = +1$. The higher scoring option, B in this case, is chosen.

Note that in this example, even though A was observed more often, B is eventually chosen. Therefore, this specific PBSL strategy would not make the same decision as a conformist. This would, however change, if we assumed another pair of weights. Say the strategy puts a weight of 2 on observed successes and a weight of 1 on observed failures (abbreviated as [2/1]). The observation $\left\{\mathrm{A}\bar{\mathrm{A}}\mathrm{B}\right\}$ would result in a score for A of $S_A = +2 + 1 = 3$ and a score for B of $S_B = +2$. A PBSL strategy with these weights would thus choose A, as would conformists.

All scoring-type PBSL strategies can be characterized by the pair of weights $[x/y]$. The principle of scoring the options and choosing the option with the highest score is reminiscent of the Borda count in social welfare theory [20].[2] Scoring-type PBSL does, of course, not encompass all possible PBSL strategies, but our first goal was to restrict the number of possibilities. This goal is therefore achieved. Our second goal was to present strategies that from a psychological point of view are somewhat plausible. As scoring-type PBSL can be characterized by just two parameters, we would argue that this goal is also achieved.

2.1.3 Averaging-type payoff-biased social learning

The class of PBSL strategies introduced above is of course not the only sensible class. McElreath et al. [120] proposed a variant that works according to the heuristic "Calculate the average payoff generated by A, calculate the average payoff generated by B, choose the option with the higher average payoff" (average meaning arithmetic mean in this case). They call this strategy simply "pay-off-biased social learning". In case of a draw, this strategy chooses the options according to their frequencies; e.g., if {AAB} is observed, both A and B would have the same average payoff, and the strategy would choose A with probability 2/3 and B otherwise.

Furthermore, they proposed a variant of this strategy, called "pay-off conformity". This variant decides exactly as the original, but in case of a draw chooses the more frequent option (i.e. {AAB} leads to choosing A with probability 1). Both variants often decide in the same manner, but "pay-off conformity" was shown to be more successful, so we focus on this strategy only. However, we found the name "pay-off conformity" to be imprecise,[3] hence we will call this strategy "PBSL McElreath" from now on.

To reverse-engineer the strategy proposed by McElreath and colleagues, we had to made the assumption that an unobserved option is considered to always produce a lower average payoff than an observed option. For example,

[2] The Borda count is a way of voting. Each voter ranks each candidate, then each candidate receives a score that is higher the higher she is ranked. After all scores have been added up, the candidate with the highest score wins.

[3] The name is imprecise because there are many PBSL strategies that display a conformist bias – remember that conformism is but a special case of scoring-type PBSL. We will elaborate on this later on.

observing A being chosen unsuccessfully three times and not observing B leads to the choice of A. In principle, one could also make the assumption that the unobserved option produces a payoff somewhere between total success and no success, which would then lead to the choice of B. However, this is not how McElreath et al. implemented their strategy, so we stuck to their choice.

As we mentioned earlier, all PBSL strategies could be characterized by specifying the probability to choose a certain option, given a certain observations of successes and failures. For a sample size of 3, there are thus 20 different observations. The choice made by scoring-type PBSL with different pairs of weights and by PBSL McElreath, given these observations, is shown in table 2.2. Additionally, rows in which a conformist would choose A are shown in gray. It can be seen that although the strategies behave similarly, they are all distinct, as no two strategies would choose the same in all situations.

2.2 Model description

In the previous chapter, we have given a hint at how our model works. Here we explain more details of the model.

2.2.1 The environment

The task that the strategies face will be to adapt to the environment, so it is important that we understand how it works. In previous works, the environment often only consisted of one option that is correct and will (almost) always yield a benefit on the one hand, and of one or several other options that are incorrect and do not yield the benefit on the other hand. In our model, the environment consists of two options A and B (if not stated otherwise) that are characterized by having certain probabilities, $p_A(t)$ and $p_B(t)$ respectively, to yield success in period t. These probabilities fluctuate over time and are independent of each other. This means that it is entirely possible that both A and B are good at the same time or bad at the same time; it is possible that they are very close and distinguishing between the better and worse option is unimportant, or that they take quite different values, so that distinguishing is very important. In this sense, our environment is more complex than in previous models but, arguably, also more realistic.

In the following, we characterize the parameters that influence the environment.

2.2.1.1 The length of a generation, t_{max}

During one generation, each individual has to make several decisions. This allows the individual to learn over time. Within each generation, there are

$A\bar{A}B\bar{B}$	$[1/-1]$	$[4/-1]$	$[1/-4]$	$[1/0]$	McElreath
3 0 0 0	1	1	1	1	1
2 1 0 0	1	1	0	1	1
1 2 0 0	0	1	0	1	1
0 3 0 0	0	0	0	0.5	1
2 0 1 0	1	1	1	1	1
1 1 1 0	0	0	0	0.5	0
0 2 1 0	0	0	0	0	0
1 0 2 0	0	0	0	0	0
0 1 2 0	0	0	0	0	0
0 0 3 0	0	0	0	0	0
2 0 0 1	1	1	1	1	1
1 1 0 1	1	1	1	1	1
0 2 0 1	0	0	0	0.5	1
1 0 1 1	1	1	1	0.5	1
0 1 1 1	0	0	0	0	0
0 0 2 1	0	0	1	0	0
1 0 0 2	1	1	1	1	1
0 1 0 2	1	1	1	0.5	0
0 0 1 2	1	0	1	0	0
0 0 0 3	1	1	1	0.5	0

Table 2.2: Probability to choose A of different payoff-biased social learning strategies with sample size 3. A indicates that the sampled individual chose A and succeeded, $\bar{A}$ that she chose A and did not succeed, B that she chose B and succeeded, and $\bar{B}$ that she chose B and did not succeed. Rows that correspond to samples with a majority of A choices are colored gray. The first column describes the observation, the second to fifth column the choices of scoring-type PBSL with weights as indicated, the last column the choice of averaging-type PBSL as described by McElreath et al.

exactly t_{max} decisions to be made, with $t_{max} = 50$ as our default setting. Between generations, there is evolution, meaning that the frequencies of the strategies may change, while they are fixed within a generation.

For the very first period, we set $p_A = p_B = 0.5$. Furthermore, the environment "remembers" the last value of p_A and p_B of the current generation and uses it as the initial value of p_A and p_B in the subsequent generation, creating a smooth transition.

In this section, we are only interested in performance and not in evolution. Each strategy is assumed to be fixed in the population and we analyze how they perform for a given environment. Also, PBSL strategies have no memory. Therefore, t_{max} does not play a role in this chapter but may well do so when strategies with memory are involved.

2.2.1.2 Mean success rate of the environment

p_A and p_B fluctuate around a mean, but the value of this mean may vary. There is no particular reason to assume that p_A and p_B should be 0.5 on average, though. Instead, if we imagine option A and B being two hunting grounds and the decision being which of the two to visit, one would expect the mean success probability to be below 0.5. If instead we dealt with gathering berries, the success probability could well be above 0.5. It is thus crucial to vary the mean success rate of p_A and p_B and to check whether payoff-biased social learning strategies can cope with this variation.

In our simulations, we will vary the mean p_A and p_B by adding the same value Δp to both of them. We illustrate the change in the joint distribution of p_A and p_B in figure 2.1. The shift in the means only results in a shift of the joint distribution, while the shape of the distribution is preserved. The reason why both p_A and p_B are shifted by the same amount is that if, say, p_A were on average greater than p_B, even if only by a very small amount, it would often be the best response to always choose A, discouraging any form of learning.

2.2.1.3 Reversion factor r

p_A and p_B vary over time but they will tend to revert to the mean, which is 0.5 by default. This reversion trend means that if p_A (p_B) is greater than 0.5, it is more likely than not that it will decrease in the next period; if p_A (p_B) is less than 0.5, it is more likely than not that it will increase in the next period. The reasoning for this is that if p_A and p_B were just random walks, there would be very long stretches of one option staying better than the other. Rarely would a switch occur. But if switches are too rare, strategies hardly need to adapt to the environment, which would strongly favor strategies that are slow in adapting.

The probability that a p_i increases ($i = \{\mathrm{A}, \mathrm{B}\}$) is calculated as $0.5 -$

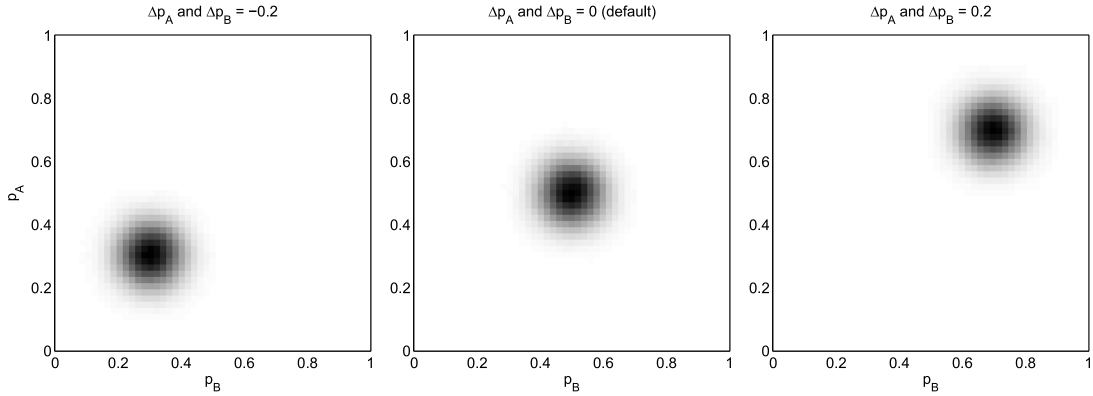

Figure 2.1: Influence of the mean value of p_A and p_B on the joint distribution of p_A and p_B. Darker shades indicate higher probability densities. The shape of the distribution is not affected by higher or lower Δp but the distribution is shifted to the bottom left (left panel, $\Delta p = -0.2$) or to the top right (right panel, $\Delta p = 0.2$). p_A and p_B, being probabilities, are never allowed to be less than 0 or greater than 1.

$r \times (p_i - 0.5)$. For example, if $r = 1$ and $p_i = 0.6$, the probability that p_i increases in the next period is 40%, so it will tend to decrease towards 0.5. If $p_i = 0.4$, the probability that p_i increases is 60% instead, so it will tend to increase towards 0.5. A stronger reversion factor of 3 would lead to a tendency of 80% to increase, instead of 60%, so reversion will be stronger; a weaker r of 0.1 would instead only lead to a tendency of 51% to increase, so reversion will be very weak. Overall, r thus has an important role for the probability distribution of p_A and p_B, as illustrated in figure 2.2.

An interesting measure of how the reversion factor influences the environment is the switch rate. This is the rate at which one option that was the best in the last period becomes the worse option in the following one or two periods. For $r = 0.01$, a switch occurs on average once every 51 periods, for $r = 1$ once every 25 periods, and for $r = 3$ once every 15 periods. Higher r are thus associated with more frequent switches, making it necessary to be a quicker learner.

2.2.1.4 Speed of environmental change, p_{incr}

The parameter p_{incr} indicates how probable it is that the environment changes at all from one period to the other. Change probabilities of p_A and p_B are independent of each other. For the default setting, the environment changes after each period, that is, $p_{incr} = 1$. We chose this high number because we modeled the environment so as to only change by small steps anyways, so this high speed still means that there are sufficiently long stretches where A is better than B or the other way round.

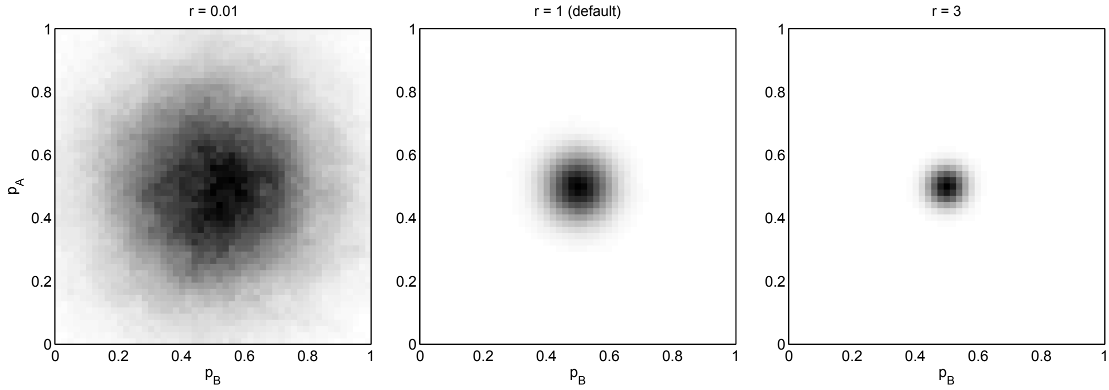

Figure 2.2: Influence of the reversion factor r on the joint distribution of p_A and p_B. Darker shades indicate higher probability densities. A low r of 0.01 (left panel) causes p_A and p_B to wander far astray from their mean of 0.5. The higher r, the tighter p_A and p_B will stick to their mean. For $r = 1$ (middle panel), p_A and p_B will very rarely if ever wander beyond ± 0.3 of their mean, for $r = 3$, they will rarely if ever wander beyond ± 0.16 of their mean.

2.2.1.5 Step size k_{incr}

Whenever p_A or p_B change from one period to the other, the absolute size of the increment is determined by the parameter k_{incr}. We have plotted the joint distribution of p_A and p_B for different values of k_{incr} in figure 2.3. The smaller k_{incr}, the finer the 'resolution' of the joint distribution. This is because smaller k_{incr} imply that the joint distribution of p_A and p_B can take more discrete value pairs.

The first influence of greater k_{incr} is that from one period to the other, changes in p_A and p_B are greater and therefore, strategies have to adopt more quickly. The second influence of the factor k_{incr} is not as easy to spot. We have to keep in mind that p_A and p_B tend to revert to the mean, so if they are far above the mean, it is more likely that they decrease and vice versa. If k_{incr} is large, it happens more quickly that p_A and p_B step far away from the mean and they will thus revert to it more frequently than for low k_{incr}. For example, for $k_{incr} = 0.005$, we can expect on average one switch from $p_A > p_B$ to $p_B > p_A$ every 50 periods, for $k_{incr} = 0.02$ one switch every 25 periods, and for $k_{incr} = 0.1$. one switch every 12 periods. The environment changes more frequently when k_{incr} is large, implying that strategies that are fast to adopt should fare better.

2.2.1.6 Three choice options

Until now, we have assumed that there are only two options to choose from. This assumption was also made in most previous works on this topic, or, with a similar effect, it was assumed that at each point in time, there is only

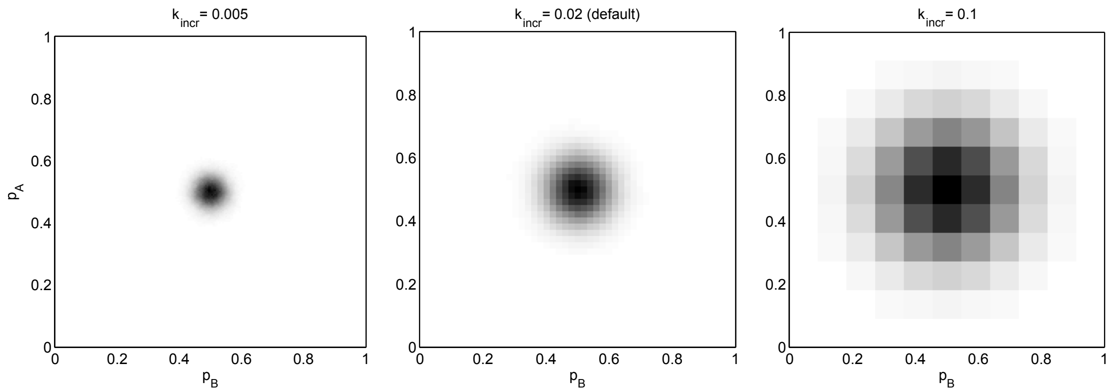

Figure 2.3: Influence of the step size parameter k_{incr} on the joint distribution of p_A and p_B. Darker shades indicate higher probability densities. The 'pixelation' seen in the figure, especially in the right panel, are not due to a malfunction of the printer but are there by construction. The higher the step size, the less discrete value pairs p_A and p_B can take.

one correct option that yields a benefit while all other options are incorrect. Relying on this simplification can be justified because it is not inconceivable that humans often compare just two options at the same time and only after discarding one option move on to the next. This way, only binary choices have to be made.

Still we felt it necessary to implement more than just one choice option to verify that our results are not dependent on this assumption. We reasoned that if stepping up from two to three options would not qualitatively affect the results, neither would increasing the number of options to four, five, or more. Thus we introduced a third choice option called C that behaves exactly as A and B do.

Strategies had to be slightly adapted to be able to cope with three choices. If two options are tied for the highest score/average payoff, each is chosen with probability 1/2; if three options are tied for the highest score/average payoff, each is chosen with probability 1/3. As there were no ambiguous cases, the adjustment of scoring-type and averaging-type PBSL could be done in a straightforward fashion.

A problem occurs when working with 3 choice options. The equations derived to calculate the choices of strategies with 3 options are by necessity longer and more complicated than when working with 2 options. As a consequence, computation time increases considerably when dealing with 3 options. Roughly, an otherwise identical simulation takes 15 times longer when one option is added. Therefore, when dealing with 3 options, we did not extensively check the parameter space, as this would have exceeded our available computational power.

2.2.1.7 Recapitulation

The process that generates the environment can be summed up as follows:

$$p_i(t+1) = \begin{cases} p_i(t); & 1 - p_{incr} \\ p_i(t) + k_{incr}; & p_{incr}\,(0.5 - r \cdot (p_i(t) - (0.5 + \Delta p))) \\ p_i(t) - k_{incr}; & p_{incr}\,(0.5 + r \cdot (p_i(t) - (0.5 + \Delta p))) \end{cases}$$

With probability $(1 - p_{incr})$, the environment remains static, with probability p_{incr}, it changes. If it changes, the increment of size k_{incr} will be added with probability $0.5 - r \cdot (p_A(t) - (0.5 + \Delta p))$ and else k_{incr} will be subtracted. Moreover, p_i is bound between 0 and 1. For the default values, the process reduces to:

$$p_i(t+1) = \begin{cases} p_i(t) + k_{incr}; & 1 - p_i(t) \\ p_i(t) - k_{incr}; & p_i(t) \end{cases}$$

The parameters that affect the environment are listed in the table below. We also list their default values and include a short explanation.

parameter	default value	description
t_{max}	50	Number of time periods per generation
Δp	0	Change in mean p_A and p_B
r	1	Strength of reversion to the mean
p_{incr}	1	Probability that environment changes per period
k_{incr}	0.02	Step size at which environment changes

2.2.2 Modeling details

In this chapter, we do not take evolution into account but instead only model populations consisting of one strategy. As the PBSL strategies we study do not have a memory past the last period, it is unnecessary to differentiate between generations. In general, we simulated the strategies for 100,000 periods. For scoring-type PBSL and a high sample size of 7, we had to restrict ourselves to 10,000 generations, as those simulations take so much longer.

We should justify briefly why we did not derive analytical results. In principle, we should be able to calculate equilibrium choice proportions of the strategies. However, by only focusing on equilibria, we would neglect the choice dynamics. For example, after the environment has switched, conformists will mostly choose the worse option. Even if they re-converge to the better option, this process will take some time. And in the meantime, they perform worse than they would do in equilibrium. In contrast, individual learners choose more conservatively (closer to a 50:50 split for A and B) but adopt much faster. If we assumed that the choices of all strategies reach equilibrium immediately, we would thus overestimate the performance of conformists. The approximation would only be valid if environmental change is very slow, which would severely restrict the applicable domain of the model.

But if we want to consider dynamics, it is hard to obtain analytical results and thus have to rely on simulations. Moreover, it is already quite hard to even determine the equilibrium choices of such a simples strategy as conformism with a sample of three because solutions contain roots of polynomials of the third degree. Degrees rise linearly with sample size, so trying to solve for sample sizes of, say, 7 is simply not feasible.

As we assume an infinite population size and as strategies are deterministic, the performance of the strategies is also deterministic. There is one source of stochasticity, though, which is the environmental change. Sometimes we will encounter many periods with one choice being better than the other, sometimes there will be several switches per generation. Some test runs with the same initial conditions but different random numbers led to the same equilibrium frequencies, so the stochastic elements do not seem to push the frequencies towards different equilibria.

Moreover, we built in a limit so that each option is chosen by at least 0.1% of the population at all times. Why this arbitrary choice? In principle, when the population starts with a certain mixture of A and B choices, these frequencies can never reach 0 or 1, so that there will always be some demonstrators that choose the rarer of the two options. Similarly to genetic variants following evolutionary replicator dynamics with infinite population size, choices will never become "extinct". This is important because if choices could die out, some strategies would never again choose this option, as it has to be demonstrated in their sample in order to have a chance to be picked. A computer program simulating numerical values will, however, cease at some point to be precise enough, so that very small numbers are rounded down to zero. So due to hardware constraints, there will always be an arbitrary lower bound. Therefore, it is better to control it than to let it be controlled by the hardware.

We have chosen to limit the minimum choice frequency for the options to 0.1%. We found that changing this limit by one order of magnitude in either direction does not greatly affect performance. Also, decreasing this limit even further, up until 0, rarely if ever affects the performance of scoring-type PBSL. However, it does affect the performance of averaging-type PBSL. For this type of strategy, due to rounding errors, it may indeed happen that all individuals at some point choose the same option and never recover from that choice. By introducing the limit, we prevented this from arbitrarily occurring.

To keep matters fair, we thus decided to use the choice limit and to apply it to all strategies. One may, however, interject that this limitations has very real implications for how strategies perform and that in absence of any evidence that such a limit really exists, one should not use this limitation for the simulations. On the other hand, the limit could be justified as being caused by errors or innovations.

2.2.3 Performance calculation

This chapter deals entirely with the question how the studied strategies perform in isolation and why that is so. For this task, it is not important how they would evolve. Instead, we assume that the whole population only plays one strategy and then we observe the outcome. Of course, the results from this chapter still allow us to make inferences regarding the evolution of social learning strategies – it is reasonable to assume that strategies that perform well will also have a greater chance to invade a population or to fend off other invaders, although this still will have to be tested. Given the vast number of possible strategies, an important role for this chapter is therefore to weed out implausible and redundant strategies, so that we can focus our efforts on studying the most promising strategies in an evolutionary context.

How exactly do we define performance? Performance is supposed to be a measure of how well a strategy adapts to the environment. We therefore defined performance as the fraction of a population that chooses the better of the two options, averaged over time (periods when both options are equally good are excluded). As an example, consider a population where at all times, 60% of the individuals choose the better option; the strategy would have a performance of 60%. Or consider another population where all individuals choose the better option 60% of the time and the worse option 40% of the time; the performance would also be 60%. The performance is thus a measure that is averaged over individuals and over time, so these two extreme cases result in the same performance. In practice, we will always find a mixture.

When a strategy would consist of flipping a coin and choosing according to this result, we would observe random choice. Since a priori, A and B are equally likely to be the better option, random choice will lead to a performance of 50% on average. A social learner will thus have to exceed a performance of 50% in order to be interesting at all. (We will encounter "paradoxical" strategies that indeed perform worse than 50%.)

Another way to think of performance is that it measures how well the population fits the "effective environment". Here "effective environment" means that when $p_A > p_B$, one should always choose A and else always B. Making a linear fit of the proportion of A choices to an effective environment, which takes the value 1 when $p_A > p_B$ and 0 when $p_A < p_B$, let the steepness of the curve be α. Performance is then equal to $0.5 + \alpha/2$, which for an optimal $\alpha = 1$ would give a performance of 1, for $\alpha = 0$ (no correlation, as for random choice) gives a performance of 0.5, and for $\alpha = -1$ (always the wrong choice) gives a performance of 0. This measure of performance is equivalent to the previous.

Another possibility to measure performance would be to calculate the expected payoff of a choice. Say that in period t, we have $p_A = 0.5$ and $p_B = 0.6$ and a strategy that chooses A with 90% probability and B with 10% probability. The expected value of this choice is thus $0.9 \cdot 0.5 + 0.1 \cdot 0.6 = 0.51$. Our previous measure would result in a performance of 0.1. The advantage of the second measure is that if, say, p_A were 0.1 instead of 0.5, this would be reflected in the performance measure, with performance dropping to 0.15. Our previous measure, on the other hand, would still result in the same performance of 0.1. The second approach using expected values thus factors in how much worse the worse option is compared to the better option.

A disadvantage of the second approach is that it is dependent on the mean values of p_A and p_B, which we will want to alter later on. If the means of p_A and p_B increase, so would performance, even if the strategies are not better at making decisions. However, we want performance measures to solely reflect the goodness of the decisions not the environment's properties, and more importantly, we want them to be comparable across conditions. Therefore, we chose the first approach. In several tests, we found that performance according to the first definition and performance according to the second definition are monotonically and positively correlated as long as means of p_A and p_B are held constant. This means that our results are not distorted by the choice of how to define performance.

2.2.4 Computation

Before we go on, we want to shortly discuss how the behavior of the PBSL strategies were computed. One might have noticed that the PBSL strategies were defined by a set of rules. These rules, for computational purposes, have to be translated into equations and then, numerical simulations have to be conducted. We proceeded in two steps to achieve the result.

First, we had to be able to translate the rules into equations. As Mathematica is a mathematical software specialized in symbolical computation, we used this software (version 8) as a starting point. We assumed that the population that the social learners draw their observations from is well mixed. As there are four possible observations per draw, A+success, A+failure, B+success, and B+failure, we could determine the probability of each observation, depending on the sample size, by using multinomial distributions. Given that the probability that A generates a success in period t is $p_A(t)$ and the frequency of A choices in the population is $x(t)$, we could express the observation A+success as $p_A(t) \cdot x(t)$, A+failure as $(1 - p_A(t)) \cdot x(t)$, B+success as $p_B(t) \cdot (1 - x(t))$, and B+failure as $(1 - p_B(t)) \cdot (1 - x(t))$.

Next we defined a rule that 1) for scoring-type PBSL strategies would assign scores to the options according to how often they are observed generating gains or losses, and 2) for averaging-type PBSL computes the average payoff generated by a given observation. In the next step, these results are translated into choices by assuming that the strategies simply always choose the option with the highest score/highest average payoff. Given the multinomial distribution as a function of $p_A(t)$, $p_B(t)$, and $x(t)$, we can translate the rules into a choice probability f for a given strategy to choose A in period $t + 1$.

This way, we could calculate the choice probability of any scoring-type PBSL, regardless of weights and sample size, and any PBSL McElreath, regardless of sample size (which is their only parameter). For example, for scoring-type PBSL with weights [1/0] and sample size 3, we have:

$$\begin{aligned} f_{[1/0]} &= \frac{1}{2}[1 - p_B^3\,(1-x) + 3p_A \cdot x - 3p_A^2 \cdot x^2 + \\ &\quad 3p_B^2\,(1-x)^2\,(p_A \cdot x - 1) - 3p_B\,(1-x)\left(1 - p_A^2 \cdot x^2\right)] \end{aligned}$$

The (t) argument was omitted for simpler reading. For PBSL McElreath with a sample size 3, we would have:

$$\begin{aligned} f_{\text{McElreath}} &= x \cdot [3p_A^2 \cdot p_B(1-x)x - 3p_A\left(1 - p_B^2\right)(1-x)^2 + \\ &\quad (3 - 3 \cdot p_B\,(1-x) - 2 \cdot x)\,x] \end{aligned}$$

The choice made for period $t + 1$ is always based on observations made in period t, so we deal with simple difference equations. Yet we also deal with polynomials of degree 3 or higher (the degree equals the sample size) and both $p_A(t)$ and $p_B(t)$ are not easily expressed in a static way, which means that we have to resort to numerical simulations to simulate how exactly the different strategies behave. As Matlab is the better software for numerical simulations, we transferred the equations from Mathematica to Matlab 7 and worked with this software from there on. Codes are available from the author by request.

2.3 Results

2.3.1 Scoring-type PBSL

2.3.1.1 Weights

Before starting to analyze scoring-type PBSL, we must ask what range of weights we should include. In principle, there is no limit; a payoff-biased social learner could weigh successes by a factor of 757 and failures by a factor of -312. However, many combinations of weights will be redundant. For instance, it makes no difference whether one chooses the pair [5/0] for the

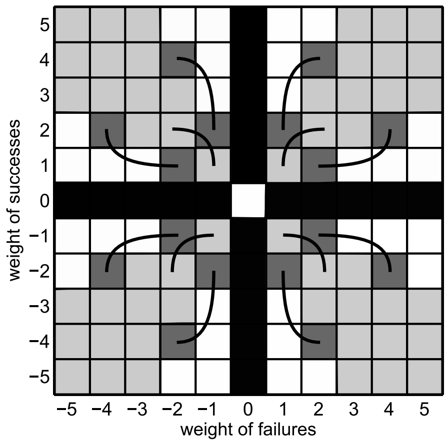

Figure 2.4: Redundancy of scoring-type payoff-biased social learning strategies. All pairs of weights of successes (y-axis) and weights of failures (x-axis) from -5 to 5 are shown. If two neighboring cells share the same color or if two cells are linked by a line, they result in the same decision rule. In this example, of the 121 weight pairs, only 25 are different strategies.

weights of successes and failures or [1/0]; the resulting choices, and hence performance, will be identical. This is illustrated in figure 2.4. All combinations of weights of successes and failures from -5 to 5 are shown, resulting in $(5+5+1)^2 = 121$ different pairs. However, some pairs are redundant. To showcase this, we colored neighboring pairs that are redundant with the same color and linked non-neighbors with a line if they are redundant. At the end of the day, there are only 25 non-redundant strategies, roughly a fifth of the initial number.

If we extend the range of weights of successes and failures to -20 to 20 (or, for that matter, any number above 4), we would still end up with only 25 different strategies. But for weights from -20 to 20, we would have to simulate $(20+20+1)^2 = 1681$ strategies, which would take a lot more time. It is therefore more efficient if we limit the scope of weights. Moreover, smaller pairs of numbers are arguably more plausible from a psychological viewpoint. In the following, we will settle for weight pairs from -5 to 5.[4]

One reason for the redundancy of strategies is that we started with a low sample size of 3. The higher the sample size, the more likely it becomes that a certain combination of observations may make a difference for the final choice. For example, weights of [3/3] and of [3/2] will deliver the exact same choice pattern for a sample size of 3 but to a different choice pattern for a sample size of 7 (6 differences in 120 possible distinct observations). For a

[4]For a sample size of 3, we could technically reduce the range to -4 to 4 and would still end up with all 25 strategies. But for sample size 7, which we will analyze later, that would only cover 49/105 possible non-redundant strategies, whereas the range of -5 to 5 covers 73/105 possible non-redundant strategies, a fair compromise.

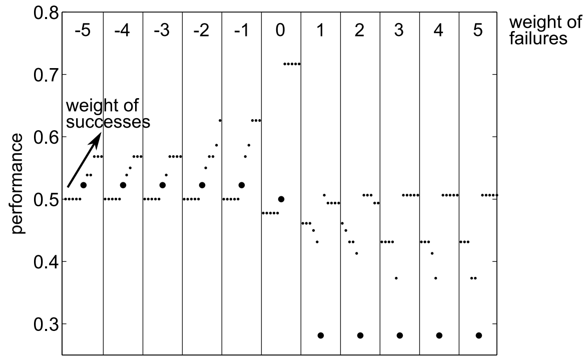

Figure 2.5: Performance of PBSL as a function of the weight of successes and failures. Each dot presents a simulated performance for a given combination of weights. From left to right, increasing weights of failures from -5 to 5 are shown in columns. Within these columns, weight of successes also increase from left to right, as indicated by the arrow. The large dot in each column presents the median, i.e. a weight of 0 for successes. In general, higher weights of successes are followed by better performance.

sample size of 7, there will already be 73 different strategies in the weight range of -5 to 5, up from 25 for a sample size of 3. A higher sample size will therefore lend itself to make more finely nuanced differences for certain observations, and will thus produce a higher number of different strategies. But for now, we will focus on sample sizes of 3 and leave higher sample sizes for later.

So for this section, with a sample size of 3, all strategies will be covered when using weights for successes and failures in the range of -5 to 5. Figure 2.5 shows how for a fixed weight of failures, increasing weights of successes influence the performance. This figure is separated in 11 columns, each presenting a fixed weight of failures, ranging from -5 to 5. Within the columns, the weight of successes increases from left to right from -5 to 5. This way, the trend in the performance as a function of the weight of successes becomes clear. Regardless of the weight of failures, higher weight of successes is associated with better performance. For example, when the weight of failures is fixed at -2, weight of successes of -5 lead to a performance of 50.0%, weight of successes of 0 to a performance of 52.2%, and weight of successes of 5 to a performance of 62.6%. In general, it thus pays to put more weight on successes.

Another pattern in the results is that positive weights of failures (right side of figure 2.5) are associated with lower performance. At best, PBSL strategies with positive weight on failures have a performance of 50.6%, hardly above chance level, and at worst of 28.1%, clearly below chance level. For negative weights of failures, performance is much better, 50.0% at least and 62.6% at best, so always at least as good as chance level performance.

The explanation for this finding is not very hard. A positive weight on failures means that the more negative outcomes are observed in relation to an option, the more likely it will become to pick that option, a behavior that is paradoxical and very unlikely to be found in nature. Positive weight of failures are thus unlikely to be part of a viable PBSL strategy.

The highest performance, perhaps surprisingly, is found for positive weight of successes and zero weight on failures. In that case, performance climbs up to 71.7%, 9.1 percentage points above the second best result. So a payoff-biased social learner who counts successes positively (the exact weight does not matter) and completely ignores failures has the highest performance. The surprising element here is that one type of information, observation of failed attempts, is completely neglected. The perhaps more intuitive guess that failures should count negatively towards the choice of the observed option is not true for the given parameters. We thus deal with a situation of "less-is-more" [73], where ignoring one type of information actually increases performance.

2.3.1.2 Mean success rate of environment

One of the most important features of the environment is the mean success rate. Preferentially, strategies should show a stable performance regardless of the mean value of p_A and p_B. To study this, we varied mean p_A and p_B from 0.25 to 0.75 in steps of 0.05 (mean p's of 0.5 are our default value). The results are illustrated in figure 2.6. In panels A) to C), performance results for all combinations of weights from -5 to 5 are shown; the better the performance, the darker the color. Clearly, strategies that place a positive weight on successes and no or negative weight on failures (top left areas of the grids) perform best. Strategies that place no or negative weight on successes and positive weight on failures (bottom right areas of the grids) perform worst. This finding is independent of mean p_A and p_B.

There are only two different types of strategies that perform best in some region of the parameter space. On the one hand, we have PBSL that considers successes positively and ignore failures (i.e. [1/0], [2/0], etc.), on the other hand, strategies with weights $[3/-1]$, $[4/-1]$, $[5/-1]$, and $[5/-2]$ (these weight pairs all produce the same choice pattern and are thus identical). Plotting the performance of these two types of strategies (panel D of figure 2.6) reveals that strategies with weights [1/0] and other identical strategies perform best when mean p's are between 0.25 and 0.6, whereas strategies with weight $[3/-1]$ and identical strategies perform best for mean p's of 0.65 to 0.75.

When just looking at the best performance, regardless of which pair of weights attains it, there is a clear trend. The higher the mean value of p_A and p_B, the better the performance of the best performing strategy. For means of 0.25, the best performance is 61.3%, for means of 0.5, it is 71.7%,

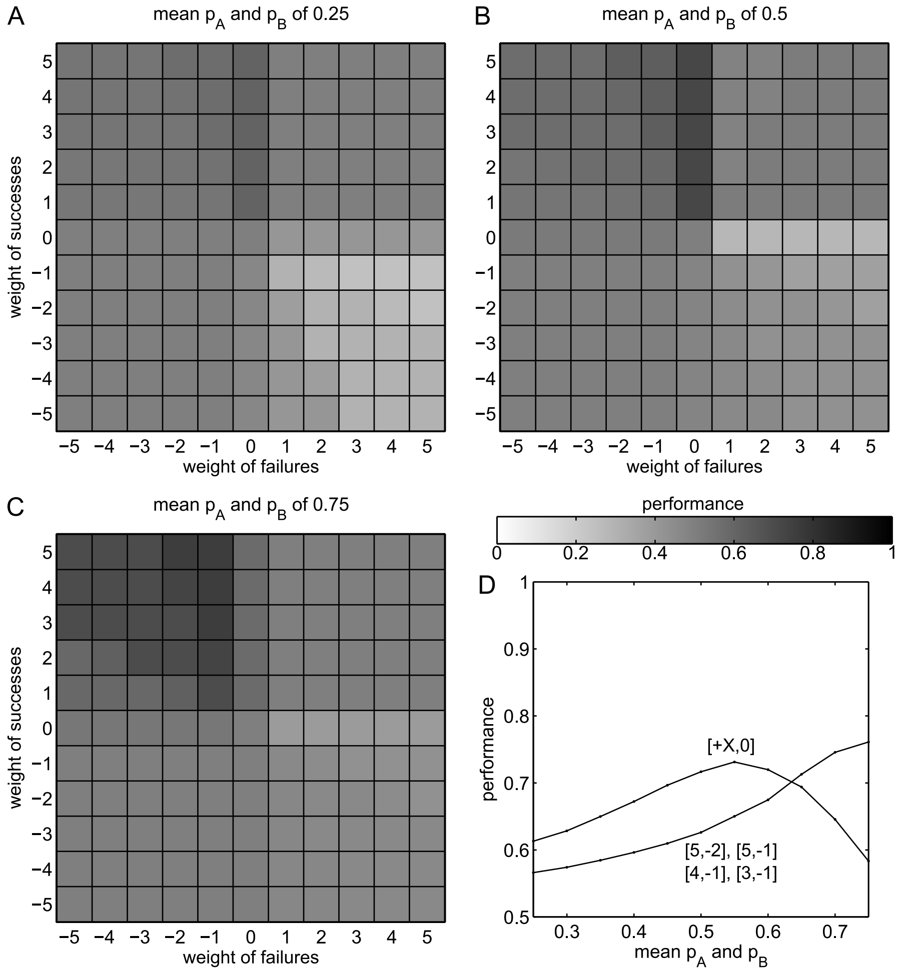

Figure 2.6: The performance of scoring-type PBSL according to their weight for different values of mean p. A) Mean p_A and p_B of 0.25. B) Mean p_A and p_B of 0.5. C) Mean p_A and p_B of 0.75. Performance is color coded, with darker shades indicating better performance. Generally, placing negative or zero weight on failures while placing positive weight on successes leads to the best performance. D) Best performance is found either for strategies that weigh successes positively and ignore failures (i.e. [+X/0]) or for strategies with weights $[5/-2]$, $[5/-1]$, $[4/-1]$, and $[3/-1]$ (all of which are redundant). The former perform best for low to medium values of mean p, the latter for high values of mean p. At no point in the parameter space do other combinations of weights result in the best performance.

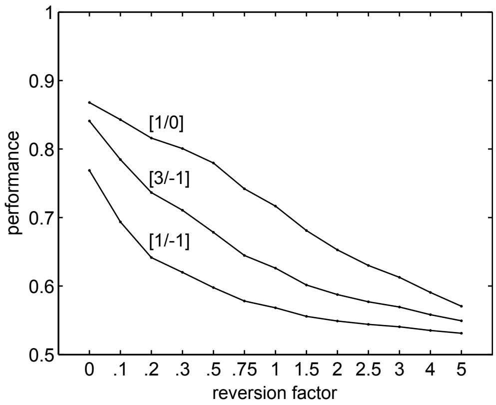

Figure 2.7: Influence of the reversion factor r on performance. Three exemplary weight pairs are shown. In general, performance drops with higher r, with all strategies being affected in the same way. Therefore, the rank of the strategies stays the same. PBSL of the type [+X/0], e.g. [1/0], perform best over the whole range of r.

and for means of 0.75, it is 76.1%. Scoring-type PBSL therefore does not perform very robustly over the parameter range.

2.3.1.3 Reversion factor

The reversion factor r could potentially have a large influence on performance. This factor regulates how strongly p_A and p_B revert to the mean of 0.5, with $r = 0$ meaning no reversion at all and higher r meaning stronger reversion. If reversion is strong, the success rate of A and B will rarely if ever lead far astray from the mean, oscillating irregularly around it instead. For this reason, the stretches of time in which one option remains superior to the other become shorter, switches become more frequent.

Our simulations showed that r has little influence on the overall result. Although higher r, implying frequent switches in best choice, lead to a decrease in performance, qualitatively, nothing changes. PBSL with score [1/0] and other identical strategies perform best over the whole range; PBSL with score weights of $[3/-1]$ and other identical strategies perform second best over the whole range, etc. (see figure 2.7). Thus, qualitatively, the results remain the same for all r.

As mentioned earlier, there are paradoxical strategies, e.g. strategies that place positive weight on failures and negative weight on successes. Those strategies still perform below chance level over all r, but actually become *better* for higher r. This is because the faster the environment changes, the harder it becomes to pick the worse option “on purpose”.

2.3.1.4 Speed of environmental changes

We varied the parameter p_{incr}, the speed of environmental changes, but found it to have almost no influence on performance. Strategies with weights [1/0] and equivalent performed better for lower p_{incr}, with performance increasing by 5.1 percentage points when p_{incr} drops from 1 to 0.1. Still, over this whole range, weights of [1/0] and equivalent performed better than any other weight pairs. PBSL with weights $[3/-1]$ and equivalent strategies were second best over the whole range, with performance varying between 61.82 to 62.61%. This variation showed no consistent pattern and could be due to stochasticity. Similarly, there is no trend in performance over p_{incr} in general. Overall, p_{incr} thus had little influence on performance.

2.3.1.5 Step size

For increasing step size k_{incr}, we observe that variance in performance between strategies becomes smaller. On the analyzed range, PBSL with weights [1/0] and equivalent performs best, except for $k_{incr} = 0.1$ (the highest studied value), were weights $[3/-1]$ and equivalent become the better choice, if only marginally so. There is no general trend of how performance correlates with k_{incr}. Overall, the performance of scoring-type PBSL is rather robust to k_{incr}.

2.3.1.6 Sample size

Until now, it was assumed that payoff-biased social learners could only observe three models. It is, however, possible that more observations are made when real beings make decisions based on social information. We therefore re-analyzed the performance data but instead of a sample size of 3, we took a sample size of 7. Note that for a sample size of 7, there are more quantitatively different strategies than before. For example, weights $[3/-1]$ and $[4/-1]$ result in the same behavior for sample size 3 but not for sample size 7.

A direct comparison between the two sample sizes is shown in figure 2.8. On the left hand side, for sample size 3, PBSL with weights [1/0] and equivalent show the highest performance. PBSL with weights $[3/-1]$, $[4/-1]$, $[5/-1]$, and $[5/-2]$ performed second best. On the right hand side, performance results for a sample size 7 are shown. Here, weights of $[4/-1]$ lead to the best performance. Second comes $[3/-1]$, third $[5/-2]$, and fourth $[5/-1]$ (due to the increase in sample size, these weight pairs are not redundant anymore). Strategies with weights [1/0] and equivalent only come in 10th, with other strategies that place positive weight on successes and negative but less weight (in absolute numbers) on failures ranking in-between.

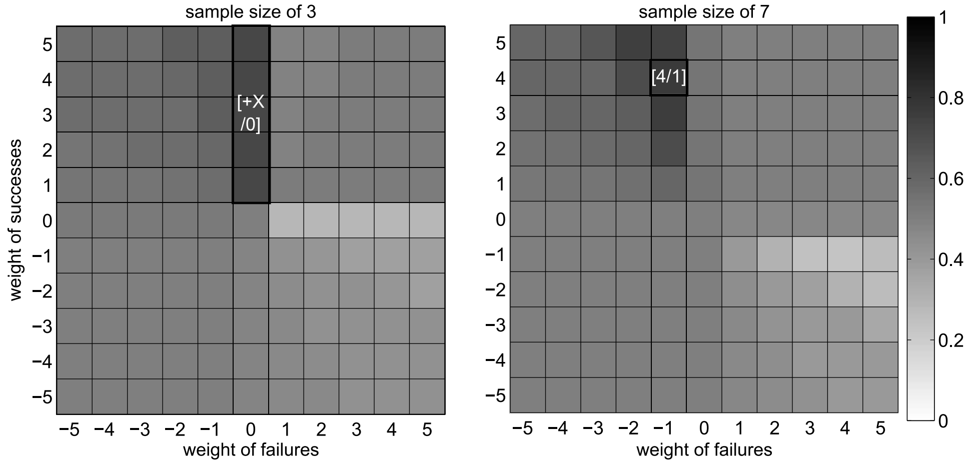

Figure 2.8: Performance of scoring-type PBSL according to their weight for sample size 3 (left panel) and 7 (right panel). Performance is color coded, with darker shades indicating better performance. For a sample size of 3, PBSL with weights of [1/0] and equivalent perform best. For a sample size of 7, PBSL with weight [4/ − 1] performs best.

With regard to the absolute performance, there are some interesting findings. On the one hand, strategies with weights [4/ − 1] and similar perform better with higher sample size. For example, weights [4/ − 1] lead to a performance of 62.6% for sample size 3 and to a performance of 76.8% for sample size 7, a 14.2 percentage point increase; for weights [3/ − 1], the difference is 13.7 percentage points. On the other hand, strategies with weights [1/0] perform worse, with a sample size of 3 leading to a performance of 71.7% and a sample size of 7 leading to a performance of only 53.7%, an 18 percentage point decrease. We should thus expect evolution towards higher sample sizes for the former but not for the latter. For those strategies, the principle "less is more" not only relates to the fact that they ignore information about failures completely, but also to the fact that they should not sample too many individuals.

Previously, we saw that some of the environmental parameters had surprisingly little influence on the performance of scoring-type PBSL, especially when it comes down to ranking the weight pairs. The only environmental parameter that had a large influence was the mean value of p_A and p_B. Therefore, we looked again at how Δp influences the result, this time for sample sizes of 7.

For sample size 3, we found that weights of [1/0] and equivalent were optimal, except for mean p_A and p_B of 0.65 or higher, where strategies with weights of [3/ − 1], and other identical strategies, performed best. For sample size 7, the picture is more nuanced. [1/0] is now only best for very low mean p_A and p_B of 0.35 or less. Afterwards, strategies that count successes

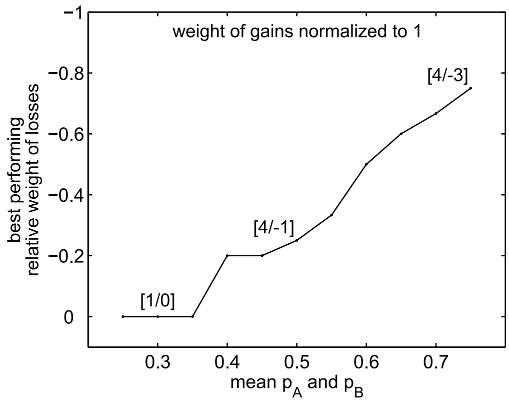

Figure 2.9: The best performing relative weight of failures as a function of mean p_A and p_B. We normalized the weight of successes to 1 and plotted the weight of failures that, for the given Δp, results in the highest performance. The higher the mean of p_A and p_B is, that is, the higher Δp, the more weight in absolute terms has to be put on failures. Some weight pairs are indicated for illustration.

positively and failures negatively are the best performers. However, the best weight ratio of successes to failures varies. To better illustrate this, we plotted the weight of failures, given that the weight of successes was normalized to 1 (figure 2.9). For example, all strategies with weights [+X/0] corresponds to [1/0], $[4/-1]$ corresponds to $[1/-0.25]$, etc. As can be seen, depending on mean p_A and p_B, various combinations of weights for successes and failures can now lead to the best performance. For low mean p_A and p_B, it is best to completely ignore failures. For a mean of 0.5, failures should count one quarter the weight of successes, corresponding to $[4/-1]$. For a mean of 0.75, failures should count three quarters the weight of successes, corresponding to $[4/-3]$. The overall trend is simple: The higher the mean of p_A and p_B, the more weight should be put on failures.

Interestingly, although the optimal combination of weights for successes and failures varies, the performance of the best combination is quite stable for the tested parameter range. This was not true when sample size was restricted to 3. With a sample size of 7, however, scoring-type PBSL attains a performance of at least 73.6% and at best of 78.5%. There is no correlation between performance and mean p_A and p_B. In summary, the performance of scoring-type PBSL with sample size 7 is quite robust on the parameter range, if optimal adjustments of weights are allowed for.

2.3.1.7 Three choice options

We tested whether the performance of scoring-type PBSL is robust with regard to the number of choice options. For that, we increased the number of options from two to three and re-assessed the performance. We found

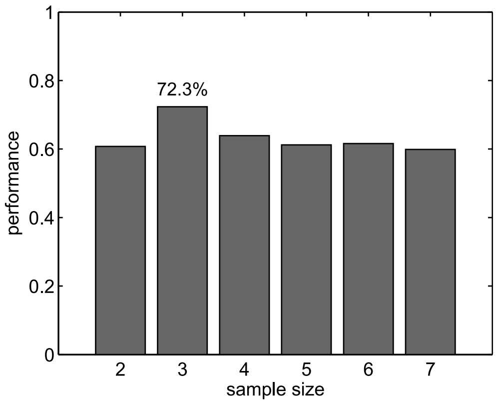

Figure 2.10: Performance of PBSL McElreath under default conditions for sample sizes ranging from 2 to 7. The best performance, 72.3%, is reached for a sample size of 3.

that there is essentially no difference between the results. Of the 121 pairs of weights tested, 50 performed better and 70 worse (the weight pair [0/0] generated the draw). The mean difference being only 1.3 percentage points and the deviations not being systematic, we conclude that the number of choice options is unimportant for performance of scoring-type PBSL.

2.3.2 Averaging-type PBSL

2.3.2.1 Sample size

A specific form of averaging-type PBSL was first proposed McElreath et al. [120], which is why we call it PBSL McElreath for short. This strategy works differently than scoring-type PBSL. It takes the average generated by options A and B as observed in the sample and then chooses the option with the higher performance. A greater sample size might seem favorable to this approach, as a small sample could be more subject to randomly assigning the higher average payoff to the worse option. In the original work, only sample size 3 was tested, so we were left wondering whether higher sample sizes increase performance. To check this, we simulated PBSL McElreath for sample sizes ranging from 2 to 7. The resulting performance using our default parameters is shown in figure 2.10. In contrast to the expectation, higher sample sizes do not lead to higher performance. Instead, the best performance by a fair margin is reached for a sample size of 3, the sample sizes studied in the original work. Here, as was the case for scoring-type PBSL with weights [1/0], the principle seems to be: less is more.

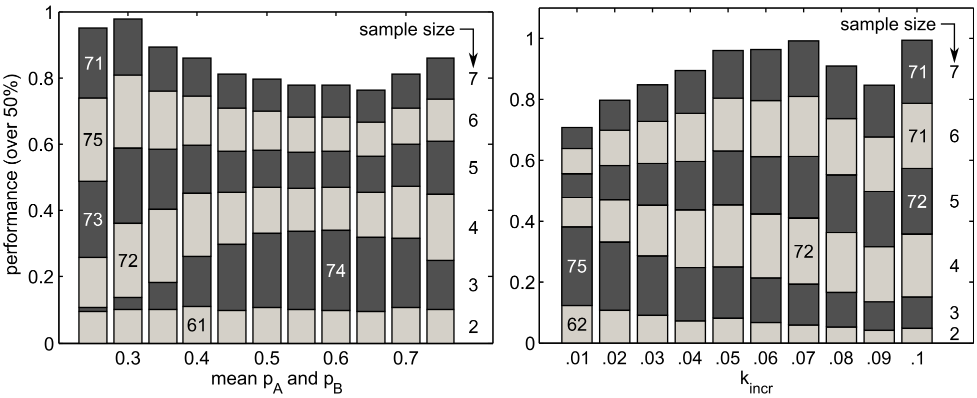

Figure 2.11: Mean performance of PBSL McElreath as a function of the mean success rate of the environment (left panel) or as a function of the step size k_{incr} (right panel). Shown is the excess performance over chance level (50%) depending on sample size, each stacked on top of one another. Numbers in the bars indicate the best performance achieved for each sample size, bars with thick borders indicate the best performance achieved for each mean parameter value. Left: A sample size of 2 leads to a rather stable but low performance ($\approx 60 - 61\%$) over mean p. For a sample size of 3, very good performance ($\approx 74 - 75.5\%$) is attained for medium to high mean p (0.45 to 0.65). The reverse trend in performance is found for higher sample sizes; for instance, the best overall performance (77.38%) is found for sample size 6 and mean p of 0.25. Right: Sample size 3 is favored for low k_{incr}, while sample sizes of 4 and higher are favored for higher k_{incr}.

2.3.2.2 Mean success rate of the environment

It is crucial to study how strategies perform depending on the mean success rate provided by the environment. The left panel of figure 2.11 shows the performance of PBSL McElreath as a function of sample size and mean p_A and p_B. To emphasize the differences, we only show performance above the chance level of 50%. For each mean p, we stack the performance for all sample sizes on top of each other. For low or high mean p, higher sample sizes provide the best performance. For example, for a mean p of 0.25, sample sizes of 6 yield the best results; and for a mean p of 0.75, a sample size of 4 yields the best performance. A sample size of 2 generally leads to low but stable performance. A sample size of 3 leads to good performance in the p range of 0.45 to 0.65 but for low mean p also leads to the lowest observed performance, hardly exceeding chance level.

2.3.2.3 Reversion factor

We simulated the influence of the the reversion factor r on performance. For PBSL McElreath, we found that performance decreases with increasing r, regardless of the sample size. For all r except for $r = 0$, sample size of 3 delivered the best performance; for $r = 0$, performance peaked at a sample size of 4. The advantage over a sample size of 3 was, however, only 2.4 percentage points for $r = 0$ and was on average 5.7 percentage points lower for all other r. It is thus safe to state that a sample size of 3 results in the highest performance and that this is robust with regard to the parameter r.

2.3.2.4 Speed of environmental change

We varied p_{incr} from 0.1 to 1 in steps of 0.1 and found no large differences to the previous results. Sample size of 3 remains the best choice by far over the whole parameter range. Generally, when changes become slower (lower p_{incr}), performance is stable or increases. The only important change is that for environments that change very slowly ($p_{incr} \leq 0.5$), PBSL McElreath with sample size of 2 see a performance boost and becomes the second best choice. PBSL McElreath with sample size 3 is the only strategy to outperform scoring-type PBSL that ignores failures (i.e. with weights [1/0] and equivalent); slightly for high p_{incr} and more so for low p_{incr}.

2.3.2.5 Step size

We varied the step size k_{incr}, which reflects the absolute change in p_A and p_B each time they change. The results are shown in the right panel of figure 2.11. A sample size of 3 leads to the best performance as long as k_{incr} is small ($k_{incr} \leq 0.03$). For higher values, $0.04 \leq k_{incr} \leq 0.08$, a sample size of 4 is best. For even higher values, $0.09 \leq k_{incr} \leq 0.1$, a sample size of 5 is best. It thus seems that the higher k_{incr}, the higher the optimal sample size.

2.3.2.6 Three choice options

We tested the performance of PBSL McElreath for three instead of two choice options. The results are shown in figure 2.12. Overall, PBSL McElreath, regardless of its sample size, improves its performance when there are three options. Strategies with sample size 4 profit most, showing an increase of 6.8 percentage points. However, sample size 3 still remains the best choice, delivering the best performance for both two and three choice options.

Of the environmental parameters, we found Δp, which determines the mean value of p_A, p_B, and p_C, to be the most important. Therefore, we re-simulated performance for different values of Δp, this time with three choice options. The results were almost identical to those found for two choice options. Interesting to note, over the whole parameter range of Δp, performance improved for each studied sample size when there were three instead of two choice options; with only 3.3 percentage points on average, these improvements were rather small, though, and they never exceeded 9 percentage points.

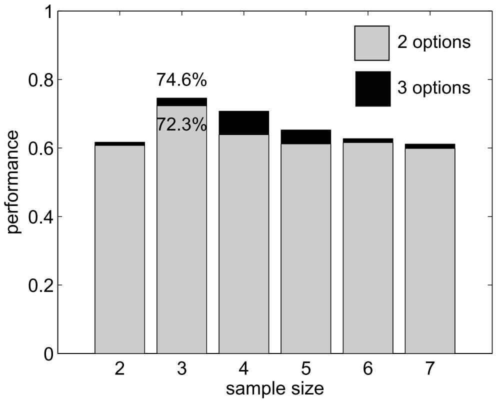

Figure 2.12: Performance of PBSL McElreath for sample sizes 2 to 7 and for 2 choice options (gray bars) and 3 choice options (black bars). For all sample sizes, 3 choice options result in better performance, although the differences are small. A sample size of 3 is still the best choice.

2.3.2.7 Payoff-conformism trade-off

PBSL McElreath compares the mean payoff of two options and chooses the option with the higher payoff. In case of a tie, it chooses the more frequent of the sampled option. This strategy thus only resorts to conformism as a tie breaking rule. We implemented a modification of this strategy which actually used a trade-off between payoff and conformism. This means that the conformist element enters the stage not only when the mean payoffs are identical but also when the mean payoffs are sufficiently close.

We do not want to delve too much into the details but rather give an example to illustrate the new modification. Previously, if a learner using PBSL McElreath observed three successes with A, one failure with A, and one success with B, she would choose B, as the mean payoff of B is greater. With the new payoff-conformism trade-off, this individual would instead choose A, because A is observed so much more often in the sample than B. After tinkering with the magnitude of the trade-off, we found one that gave reasonable results. The chosen magnitude in favor of conformism is, however, quite small, so that it only affects the results if the sample size of PBSL McElreath exceeds 4.

With this new payoff-conformism trade-off introduced, we managed to find improved performance. Under the default conditions, the best that PBSL McElreath could produce was a performance of 72.3%, which was achieved by choosing a sample size of 3. After modifying the strategy, the best performance was 78.4%, which was achieved for a sample size of 6. This is a remarkable improvement which under default conditions is only matched by scoring-type PBSL with weights $[4/-1]$ and a sample size of 7 (78.5%).

We looked again at how the mean value of p_A and p_B would affect the

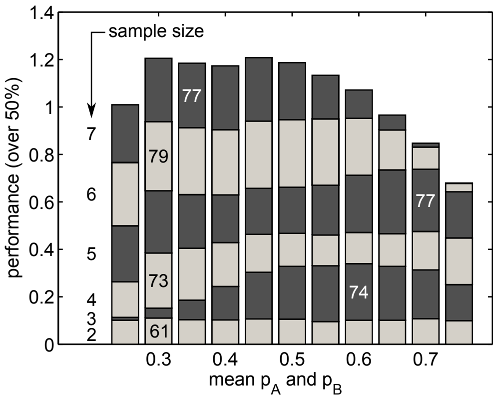

Figure 2.13: Mean performance of PBSL with a payoff-conformism trade-off as a function of the mean success rate of the environment and sample size. Shown is the excess performance over chance level (50%), with sample sizes stacked on top of one another. Numbers in the bars indicate the best performance for each sample size, bars with thick borders indicate the best performance for each mean p. For low to moderately high mean p_A and p_B, a sample size of 6 yields the best result, for high mean p_A and p_B, a sample size of 5 is better, and for very high mean p_A and p_B, a sample size of 4 is best. Sample size of 3, which was dominant without the trade-off, is never the best.

performance. The result is shown in figure 2.13. Previously, we found that PBSL McElreath with a sample size of 3 produced the best performance, except for low or very high mean p_A and p_B. After including the payoff-conformism trade-off, a sample size of 3 never leads to the best performance. Instead, a sample size of 6 is the best choice for mean values of p_A and p_B between 0.25 and 0.55. For higher values, smaller sample sizes are better. In summary, introducing the trade-off increases performance considerably and shifts the best sample size upwards.

We furthermore analyzed the modified social learning strategy under the condition that there are three choice options. The results were very similar to those found for two choice options. Increasing the number of options does therefore not affect performance in an important way.

2.3.3 Behavioral results

Until now, we have focused our analysis on the performance of the strategies as a function of different parameters. To get a better grasp on why strategies sometimes perform well and sometimes not, we will have a closer look at their actual behavior.

2.3.3.1 Conformist bias

We discussed how we use the term “conformism” qualitatively earlier in this work. For our current purposes, we need a quantitative definition of conformity, though. Therefore, we chose the approach to define the degree of “conformist bias” as how closely a strategy sticks to a true conformist. A first impression of how to approach this can be gained by again looking at table 2.2. Gray rows correspond to samples in which the majority of the sampled individuals chose A. Pure conformists would choose A exactly for these samples and B otherwise. Of the PBSL strategies, those with weights $[1/-1]$ would act in line with conformists in 50% of the possible samples; those with weights $[4/-1]$ in 60%, those with weights $[1/-4]$ in 40%, those with weights $[1/0]$ in 75%, and PBSL McElreath in 80% of the possible samples.

A social learner who imitates the choice of a random individual in the population preserves the choice frequencies. In contrast, conformist bias leads to polarization – if one option is more frequent than the other, a conformist is more likely to adopt this option than expected from random choice, as is shown in panel A of figure 2.14. For example, if A is chosen by 2/3 of the population, a conformist with sample size 3 (solid line) will chose A with a probability of almost 3/4. The more individuals are samples, the more conformist bias a conformist will show. For example, with a sample size of 7 (dashed line), the same conformist would have a probability of almost 5/6 to choose A instead of 3/4. If the whole population were sampled, a conformist would always choose the more frequent option (dotted line).

When we look at PBSL and whether they display a conformist bias, we have to keep in mind that PBSL strategies act differently when payoffs differ. For example scoring-type PBSL with weights $[4/-1]$ (panel B of figure 2.14) do not show a conformist bias when p_A and p_B are rather low (dashed line) but do so when they are rather high (solid line). PBSL McElreath (panel C) shows a conformist bias when p_A and p_B are low (solid line) but less so when they have intermediate values (dashed line). It would thus be interesting to quantitatively measure the degree of conformist bias of different PBSL strategies as a function of p_A and p_B.

We define the conformist bias as the tendency to increase the proportion of A choices if A is also chosen by the majority of the population and to increase the proportion of B choices if B is chosen by the majority (see panel D of figure 2.14). Put mathematically, if $f(x)$ is the proportion of A choices made by a PBSL strategy as a function of the frequency x of A choices made by the population, then the conformist bias is defined as

$$\text{conformist bias} = \int_0^{1/2} (x - f(x))\,\mathrm{d}x + \int_{1/2}^{1} (f(x) - x)\,\mathrm{d}x$$

Using this definition, conformist bias is bound between -0.75 and 0.25 (note: negative values do not imply an anti-conformist bias). Conformists, whose behavior is independent of p_A and p_B, have a conformist bias of 0.0625 for a sample size of 3 and of ≈ 0.113 for a sample size of 7.

The conformist bias of scoring-type PBSL as a function of p_A and p_B is shown in figure 2.15. The thick line presents the isocline at which conformist bias is 0, thin lines are isoclines with 1/8th difference between each. We colored areas with positive values of conformist bias in gray. PBSL with weights [1/0] (and equivalent) and sample size 3 (top left) have a rather high conformist bias for most of p_A and p_B. But conformist bias becomes much more severe for sample size 7, with positive values for the better part of the area. PBSL with weights $[4/-1]$ have a rather low conformist bias compared to PBSL with weights [1/0] but conformist bias also increases with higher sample size.

We applied the same analysis to PBSL McElreath and PBSL with payoff-conformism trade-off (figure 2.16). Both strategies show a high degree of conformist bias only for high or low mean p_A and p_B. For a sample size of 3, both strategies are identical and thus show the same behavior. For sample size 6 they differ. Most notably, PBSL McElreath become substantially less conformist, while PBSL with payoff-conformism trade-off is still almost as conformist as before. The fact that we see the latter strategy being more conformist is not surprising, as we designed the strategy in a way that conformism has more impact on decision making.

Combining these findings with our performance measures, there is no simple correlation between the degree of conformism and performance. We may suspect, however, that some amount of conformist bias is good, but too much conformity is detrimental for performance. For instance, for sample size 3, PBSL with weights [1/0] is more conformist than PBSL with weights $[4/-1]$ and our earlier analysis also showed that it performs better. If sample size is increased to 7, however, PBSL with weights $[4/-1]$ becomes moderately conformist and also performs better, whereas PBSL with weights [1/0] becomes very conformist and performs worse. Is it therefore possible that there is an optimal degree of conformist bias?

To get more data, we needed to make an assumption. To calculate the conformist bias as a single number, we took the joint distribution of p_A and p_B using the default parameter values. We calculated the conformist bias of all scoring-type PBSL strategies that have positive weight on successes,[5] the conformist bias of PBSL McElreath, and the conformist bias of PBSL with payoff-conformism trade-off. Next we plotted their performance as a function of their average conformist bias, shown in figure 2.17. Confirming

[5]We excluded negative weight on successes, since these strategies are often paradoxical in the sense that they are more likely to pick the worse option than to pick the better option. As they perform so horribly, these strategies are uninteresting in general.

our suspicion, we observe that performance first increases with increasing conformist bias, only to drop after a certain point. The optimum is reached for a conformist bias close to zero but slightly negative (remember that a conformist bias of 0 still implies a moderately high degree of conformity). Therefore, there is indeed an optimal degree of conformity.

To check for robustness of the findings, we also plotted performance as a function of conformity under the assumption that the mean p_A and p_B either were 0.25 or 0.75. The results did not differ substantially, we still observe the peak in performance for a conformist bias close to 0. The general shape of a curve fitting the data is somewhat odd – first performance increases modestly with conformist bias, then it suddenly drops after optimal conformity is reached. An interpretation of this could be that after this point is reached, maladaptive equilibria may establish, as we observed when studying conformism in the previous chapter.

For PBSL McElreath and PBSL with payoff-conformism trade-off, there is more variance left to explain than for scoring-type PBSL. This implies that more is going on than just performance altering as a function of conformist bias. Alternatively, our measure of conformist bias may simply be imperfect and that a better measure would reduce unexplained variance.

To sum up, a strategy does not need to be a conformist to display a conformist bias. This is why we refrained from adopting the nomenclature used by McElreath et al. [120], who called their strategy "pay-off conformity" – it would be confusing, as many PBSL strategies show conformity, even if they are not specifically designed to be conformist. Payoff-biased social learners often behave in a conformist manner, which is especially true for those PBSL strategies that also show a high performance. Therefore, conformism is an ingredient of a successful social learning strategy. It is, however, possible to be too conformist, which is detrimental to performance. Finding the right level seems to be key to success.

2.3.3.2 Frequency-information vs payoff-information

In the previous section we exclusively dealt with how social learners react to frequency-information, i.e. information about how often an option is chosen in the population. A prominent feature of payoff-biased social learning is, however, that it takes payoffs into account. We can therefore characterize a PBSL strategy according to how it reacts to frequency-information vs payoff-information.

In general, there is always a trade-off between incorporating frequency-information and incorporating payoff-information. For example, assume that a strategy samples three individuals, two of whom chose A but were unsuccessful and one of whom chose B and was successful. Frequency-information gives a 2:1 advantage to A; payoff-information gives an advantage to B. A strategy can either rely more on the former type of information, leading to

A as choice, or on the latter type of information, leading to B as choice. It is impossible that both types of information have at the same time a high impact on choice. Strategies can thus be characterized by how much they rely on one or the other type of information.

First we considered scoring-type PBSL strategies. To gauge how they react to the two kinds of information, we plotted their frequency of A choices as a function of 1) the frequency of A choices in the population and 2) as a function of the values of p_A and p_B, assuming that $p_B = 1 - p_A$. The results are shown in figure 2.18. First we have an example of a strategy that only reacts to frequency-information, namely pure conformism. Conformism (which corresponds to PBSL with weights [1/1]) does not distinguish between successes and failures and thus completely disregards payoff-information. In contrast, PBSL with weights $[1/-1]$, in this example, only reacts to payoff-information and completely disregards frequency-information. In effect, under the assumption that $p_B = 1 - p_A$, this strategy is the complete mirror image of conformism. Scoring-type PBSL thus encapsulates both extremes, sole reliance on frequency-information and sole reliance on payoff-information.

Of course, relying but on one source of information is not necessarily wise, and indeed we see that strategies that rely on both are more successful. We already saw that PBSL with weights [1/0] performed quite well. Figure 2.18 shows that this strategy is about equally sensitive to frequency- as to payoff-information. PBSL with weights $[4/-1]$ and with sample size 7 was also quite successful; our results reveal that the gradient is steeper in the direction of payoff changes. This strategy thus also relies on both kinds of information but reacts more strongly to payoff-information.

Next we turned to PBSL McElreath and PBSL with payoff-conformism trade-off (figure 2.19). For a sample size of 3, when both strategies are identical, we see a high gradient in the direction of frequency-information, meaning that they rely more on this type of information. If we increase the sample size to 6, PBSL McElreath changes so that the gradient is much higher in the direction of payoff-information. This is especially valid when $0.3 < p_A, p_B < 0.7$, which is true 99% of the time for the default parameters. For PBSL with payoff-conformism trade-off, we do not see such a dramatic change when sample size is increased to 6. Instead, this strategy relies almost as much on frequency-information as it does on payoff-information. This reflects the fact that this strategy was designed so as to give frequency-information more weight.

2.3.4 Principal findings

This subsection recapitulates the most important results presented on the last pages. First, we find that among scoring-type PBSL, many pairs of weights of successes and weights of failures are redundant, allowing us to focus on just a subsample of the possibilities. Second, among the environmental parameters, we found that only the mean value of p_A and p_B has a strong influence on the result, and the reversion factor r has a moderately strong influence. Even though most parameters often change the outcome quantitatively, they leave the rank of the strategies and thus the qualitative findings untouched.

Of the whole range of tested strategies, we found three strategies that consistently had a decent showing: scoring-type PBSL with weights [1/0], with weights $[4/-1]$ (and redundant ones), and PBSL McElreath with sample size 3. Although none of the strategies performed well under all circumstances, they reliable outperformed other strategies in most conditions. A comparison of the performance of these strategies under different conditions is shown in figure 2.20. Shown is the performance for default values (standard) and for varied environmental parameters (reversion factor r and mean p_A and p_B). Furthermore, we show how an increase of the sample size from 3 to 7, as well as an increase of the number of choice options from 2 to 3, affect the result.

All three strategies consistently outperform strategies of the same type but none of them consistently outperforms the other. While the environmental parameters and the number of choice options are exogenous and therefore not amenable to natural selection, the sample size arguably is. We may thus predict that PBSL with weights $[4/-1]$ will evolve to have a higher sample size, as this boosts performance, whereas PBSL with weights [1/0] and PBSL McElreath fare better with a sample size of 3.

Limiting the scope of the analysis to the default conditions, scoring-type PBSL could achieve a performance of 71.7% if they only counted successes and ignored failures. PBSL McElreath achieved a performance of 72.3%. The best overall performance, 78.5%, was achieved by PBSL with weights $[4/-1]$ and sample size 7. A similar performance, 78.4%, was achieved by PBSL McElreath with sample size 6 after we modified the strategy so that instead of using conformism solely as a tie-breaking rule, it really traded off payoff-information and conformism. This modification allowed the strategy to consistently increase performance at higher sample sizes.

2.4 Discussion

Empirical findings supporting or refuting the existence of certain types of social learning strategies are mainly discussed in the next chapter, after having introduced some other interesting strategies. In this chapter, we instead discuss why payoff-biased learning generates the behavior and performance we observed, and what this tells us about social learning in general.

2.4.1 Robustness of the strategies

The environment is characterized by several parameters that affect, for example, the magnitude of changes, how often the environment varies, and how far these variations tend to stray away from the mean. All these parameters did, however, not affect the strategies' performance qualitatively. Sometimes, they would improve or decrease overall performance, but the rank of the strategies was unaffected. Similarly, it made no substantial difference whether two or three choice options were considered. In summary, these results speak in favor of payoff-biased social learning because this class of strategies is quite capable of coping with many different environments.

There is one parameter, though, that has a large impact on the outcome. This parameter is Δp, which modifies the mean value of p_A and p_B. Finding this parameter to be so crucial is especially interesting because we know of no other study that analyzes social learning and takes this – or an equivalent – parameter into account. The reason is that in most other models, the probability that the better option and that the worse option(s) lead to a success is fixed. We would argue, however, that in many real life examples, this is not true. The example we often use is the choice between two foraging grounds. If, for instance, we deal with hunting large game, the success probability of the options could indeed be very small, while still one option is better than the other. If we deal with gathering fruits, the success probability of the options could easily be very large. One of the most common examples in the social learning literature, the adoption by farmers of hybrid corn [147, 148], would also fit the description that both the current technology and the new technology can simultaneously have high (or low) success probabilities. In our opinion, whenever it is possible, this factor should be taken into account.

2.4.2 Scoring-type payoff-biased social learning

In this chapter, we introduced a new form of social learning, which we called scoring-type payoff-biased social learning. In the following, we want to explain some of the phenomena we observed relating to this class of strategies.

2.4.2.1 Positive responsiveness to frequency-information

Scoring-type PBSL may at first be an unintuitive form of learning. We can improve our understanding of it by comparing it with statistical tests. There are some differences, though, so let us use an example to illuminate the similarities and differences. Imagine a standard test of whether a medical treatment will cure a disease. Typically, the patients are divided randomly into a test group and a control group, and at the end of the run, it is determined how many patients were cured. This information is best presented in a contingency table such as the following:

	cured	not cured
treatment group	61	39
control group	58	42

Evidently, in this case, the difference in treatment success between the treatment and the control group are not significant. Now what would happen if the group sizes could not be controlled? For example, if a company specialized in polls wants to know whether being Republican or Democrat has an influence on the endorsement of capital punishment, it may proceed by questioning random people on the street or by phone. However, it cannot usually control the number of Republicans or Democrats it encounters. A contingency table could then look like the following (the data are all made up):

	pro capital punishment	contra capital punishment
Republican	205	97
Democrat	104	45

Here, again, there is no significant difference between the two groups; in both groups, capital punishment is endorsed by two thirds of the subjects. A good statistical test, being robust to the relative sizes of the groups, would thus show that partisanship has no significant effect on endorsement of capital punishment.

Now we come back to the task of social learning. Imagine that both options A and B have the same probability of leading to success but that A is chosen more frequently than B, as in the following table:

	success	failure
A	35	32
B	17	16

A statistical test would show that treatment (choosing A or B) has no significant effect on outcome (success or failure). For example, Fisher's exact test would lead to a highly insignificant p-value, $p > 0.99$. A social learner, if she were as "fair" as a statistical test, should consequently choose both options with equal probability. However, a look at how the different social learning strategies would choose, given that they sample from this population, reveals quite a different behavior, as presented in table 2.3.

Some strategies behave consistently with a "fair" statistical test while others do not. For instance, PBSL with weights $[1/-1]$ would choose A with a probability close to 50%, reflecting the relative success rate of A. In contrast, conformists, PBSL with weights $[4/-1]$, PBSL with weights $[1/0]$, and PBSL McElreath would choose A with probabilities ranging from 61% to 75%. This, as we saw earlier, is because they have a high conformist bias. For example, PBSL with weights $[1/0]$, which ignores unsuccessful choices, would

strategy	prob. of choosing A
conformists	74.5%
PBSL, weight $[1/-1]$	51.5%
PBSL, weight $[1/-4]$	41.1%
PBSL, weight $[4/-1]$	60.9%
PBSL, weight $[1/0]$	66.5%
PBSL McElreath	66.3%

Table 2.3: Probability to choose A by different strategies (all PBSL with sample size 3). It is assumed that the sample population of 100 individuals consists of 35 who chose A successfully, 32 who chose A unsuccessfully, 17 who chose B successfully, and 16 who chose B unsuccessfully.

only look at the first two rows of the contingency table, completely ignoring the third row that contains the information regarding failures. Since option A is twice as prominent as option B among the successes, players who use this strategy would end up choosing A with probability 2/3. This strategy, as most of the strategies we studied, does not reflect the success rates in an unbiased fashion.

These strategies display a conformist bias, but as we have seen, they have proven to be very successful. Why then is it desirable for social learners to respond positively to frequency-information, whereas a statistical test, which is also supposed to tease out which of the two options/treatments is better, should not respond positively to frequency-information? The reason is that in ordinary statistical testing, the sample sizes are created exogenously and are uncorrelated with the dependent variable. In the medical treatment example, for instance, test groups are created randomly; in the capital punishment example, we also assumed that partisanship is predetermined. In the domain of social learning, though, the proportion of A choosers is endogenous. It is, in general, positively correlated with the success rate of A, so using frequency-information to bias the decision is rational.

A similar logic would apply if we assumed that in the example of capital punishment, partisanship was not predetermined. If we assume that Republicans are in general in favor of capital punishment whereas Democrats are not, and if random sampling revealed a high proportion of Republicans in the population, we could be confident that a high proportion of the population is in favor of capital punishment. In the case of social learning, if choice is somehow positively correlated with success, it would be unreasonable to disregard the frequency-information. And as making good choices is favored by natural selection, we should expect this positive correlation to generally exist.

Relating to previous theoretical work, Eriksson et al. [52] found that introducing more than two choice options is detrimental for conformity. In their model, an individual first observes a fixed number of individuals, then with a certain probability finds out which of the observed variants is the best and adopts it; if she does not find out, she instead adopts the most common variant (conformism) or copies at random. If the authors then introduce an infinite number of choice options that have to be invented in a step-wise manner (innovation), conformists do poorly.

In our model, increasing the number of choice options (from two to three) did not affect the performance of scoring-type PBSL in any significant way, even though this strategy often behaves in a conformist way. PBSL McElreath, a strategy that at times is very conformist, even benefited consistently from increasing the number of options to three. In general, conformity is an obstacle to innovation, as new variants are per definition rare and therefore unlikely to be adopted through conformism. If the payoff derived from a choice is taken into account, though, a new innovation might still be adopted, thus overcoming the conformist bias. This might partly explain the discrepancy between our findings and the findings of Eriksson et al.

2.4.2.2 Putting more weight on successes

PBSL that overweigh successes compared to failures display a conformist bias that is found to be adaptive. But why does overweighting successes produce conformity? To better understand this, consider a sample of 10 individuals of which 6 choose A and 4 choose B. Assume further that for both options, half of the individuals are successful, meaning that the social learner observes $\left\{\mathrm{AAA\bar{A}\bar{A}\bar{A}BB\bar{B}\bar{B}}\right\}$. If successes and failures were weighted equally, they would cancel out and the given sample would lead to a draw. If, however, successes count more than failures, the more frequent option will prevail, which is A in this example. Therefore, putting a higher weight on successes than on failures induces payoff-biased social learners to have a conformity bias. Conversely, if a social learner were to weigh failures more than successes, they display an anti-conformist bias (see PBSL with weights $[1/-4]$ in table 2.3).

There are other reasons why weighting successes more than failures is beneficial. For example, say that both A and B are very unlikely to lead to success but that B is still better than A. Furthermore, assume that more individuals choose the better option B. As B is more likely to lead to failure than to success, it is also more likely that an observer gathers more negative evidence with regard to B than with regard to A. She will thus be inclined to choose the worse option A.[6] However, if positive evidence is weighted more heavily than negative evidence, the described misjudgment will be unlikely to occur. Overweighting of successes thus allows to pick the better of two bad options.

[6]This effect is akin to the Simpson's Paradox from statistics.

For the opposite problem, when A and B are very likely to produce success, placing higher weight on successes than on failures might be disadvantageous. Say that B is better but the majority chooses A. Then, as successes weigh more than failures, a payoff-biased social learner will probably gather more evidence in favor of A and will thus not switch to the better option B. This is why we observe that the higher the success rates of A and B, the less overweighting of successes should be expected (see figure 2.9). However, we presumed that the worse option is chosen more often than the better option, which rarely occurs anyway. Therefore, this opposite case is less problematic, so that the positive effects of overweighting successes should more than compensate for the negative effects.

2.4.2.3 Influence of the mean value of p_A and p_B

We found that environmental parameters have surprisingly little influence on the performance of strategies, especially on how the strategies rank. One parameter, however, the mean value of p_A and p_B, was found to severely influence the outcome. A good social learning strategy should, if possible, perform well regardless of this mean value. However, payoff-biased social learners often do not perform equally well for high and low mean p_A and p_B (see figures 2.6, 2.9, 2.11, and 2.13 for reference). This is perhaps not surprising, given that payoff-biased social learners pay special attention to the payoff, which is greatly affected by the mean value of p_A and p_B – the other parameters only affect the variation of the payoffs but not their mean. Can we explain why payoff-biased social learners react the way they do to Δp?

With regard to scoring-type PBSL, we found that the higher the mean value of p_A and p_B, the higher should be the weight placed on negative outcomes, and vice versa. A short reflection shows this to be sensible – observing the rarer event is simply more informative. A numerical example may help to illustrate this. Assume that either p_A or p_B equals 0.2 and the other 0.3. So both options are bad but one is still better by 50%. Moreover, assume that a social learner observes that an individual chose A and was unsuccessful. We can now calculate the posterior probability that B is the better choice, that is, $\Pr\left(p_A = 0.2 \& p_B = 0.3 | \bar{\mathrm{A}}\right)$. Using Bayes' theorem, we arrive at:

$$\Pr\left(p_A = 0.2 \& p_B = 0.3 | \bar{\mathrm{A}}\right) = \frac{\Pr\left(\bar{\mathrm{A}} | p_A = 0.2\right) \cdot \Pr\left(p_A = 0.2\right)}{\Pr\left(\bar{\mathrm{A}} | p_A = 0.2\right) \cdot \Pr\left(p_A = 0.2\right) + \Pr\left(\bar{\mathrm{A}} | p_A = 0.3\right) \cdot \Pr\left(p_A = 0.3\right)}$$

with "$\bar{\mathrm{A}}$" designating the evidence that A led to failure. Assuming that our prior belief was that p_A and p_B are equally likely to be 0.2 or 0.3, we arrive at:

$$\begin{aligned}
\Pr\left(p_A = 0.2 \& p_B = 0.3 | \bar{\text{A}}\right) &= 0.5 \cdot \frac{0.8}{0.8 \cdot 0.5 + 0.7 \cdot 0.5} \\
&= 0.5 \cdot \frac{16}{15} \\
&= 0.5333\ldots
\end{aligned}$$

So the observation of a failure with A increases our posterior belief that B is indeed the better option by a factor of 16/15. What would happen, though, if we had instead observed an individual who chose B and was successful? The probabilities now change to:

$$\begin{aligned}
\Pr\left(p_A = 0.2 \& p_B = 0.3 | \text{B}\right) &= \\
\frac{\Pr\left(\text{B}|p_A = 0.2\right) \cdot \Pr\left(p_A = 0.2\right)}{\Pr\left(\text{B}|p_A = 0.2\right) \cdot \Pr\left(p_A = 0.2\right) + \Pr\left(\text{B}|p_A = 0.3\right) \cdot \Pr\left(p_A = 0.3\right)} &= \\
0.5 \cdot \frac{0.3}{0.3 \cdot 0.5 + 0.2 \cdot 0.5} &= \\
0.5 \cdot \frac{6}{5} &= \\
0.6 &
\end{aligned}$$

with "B" designating the evidence that B led to success. Observing a success, the rarer of the two events, thus increases our posterior belief that B is better by a factor of 6/5. This corresponds to 18/15 and is greater than 16/15, the update in probabilities we found before. That is, we learn more by observing the rarer event (the success) than by observing the more frequent event (the failure). Expressed differently, when failures are more frequent, the self-information of observing a success is greater than the self-information of observing a failure; entropy is reduced by a higher degree.

Turning back to our model, this conclusion explains why social learners should preferentially overvalue the event that is rarer in the given environment. Therefore, as shown in figure 2.9, it is optimal to put a relatively higher weight on failures when mean p_A and p_B rise.

2.4.3 Adverse effects of higher sample sizes

In several instances, we found that increasing the sample size does not improve performance but instead deteriorates it. This finding may be surprising at first, since taking into account more information should not lead to a worse result. There are, however, theoretical reasons why sometimes ignorance is a bliss. For example, knowing too much may prevent the usage of the recognition heuristic. American students were better at estimating which of two German cities has the larger population than which of two

American cities has; similarly German students were better with American cities than with German cities [73, 74].[7]

Other reasons have been found for why knowing less can be advantageous. For example, sampling only a few times can lead the payoff difference between the studied variables to appear larger than it actually is [93]. Correlations between variables may appear larger and thus more distinct if the sample size is smaller [102]. These findings make sense under the assumption that an individual needs to cross a certain threshold before taking action; sampling less may result in the threshold being crossed more easily. The findings do not explain why the individuals do not simply lower the required threshold instead of sampling less. This makes it dubious whether less sampling is really adaptive in these cases and not rather a constraint.

This section deals with explaining why in the context of social learning, we should not always expect that as much information as possible is acquired, even if information acquisition is costless.

2.4.3.1 PBSL with weights [1/0]

First, we found that PBSL that only values successes (i.e. with weight pairs [1/0], [2/0], etc.) perform worse when they sample too many individuals. Given standard conditions and for a sample size of three, this strategy attained a performance of 71.7%; for a sample size of 7, performance dropped to only 60.8%. Other scoring-type PBSL strategies, such as those with weights $[4/-1]$, did not suffer this decline.

To better understand why we see these opposite effects of higher sample size, we had a closer look at the behavior of the different strategies. As an example, we assumed that $p_A = 0.55$ and $p_B = 0.45$. As A is slightly better than B, one would expect that in the best case, a strategy would converge towards only choosing A. PBSL with weights $[4/-1]$ and sample size 3 indeed converges towards predominantly choosing A, but only with a probability of 65%; approximately one third of the population would incorrectly choose B (see left panel of figure 2.21). With a sample size of 7, we see instead that 94% of the population will eventually choose the better option A. For PBSL with weights $[4/-1]$ and similar, it becomes easier to pick the better option when they observe a greater sample.

We applied the same conditions to PBSL with weights [1/0]. For a sample size of 3, 85% would converge towards choosing the better of the two options (right panel of figure 2.21), explaining why they perform so well for this sample size. For a sample size of 7, it is possible that 99.8% of the population converges towards the better option, which is even better. There is, however,

[7]The recognition heuristic works this way: If one of the compared cities is known by name but not the other, one should guess that the known city is larger, as this is typically true. In the best case, one knows half of the cities; knowing too many cities would prevent the subjects from using the heuristic.

also a strong conformist bias. If 66% or more of the population choose the worse option, the conformist bias would lure more individuals to choose the worse option and in the end, 98.8% of the population would do so. PBSL with weights [1/0] and sample size 7 are therefore often very decidedly on the right side but can also very decidedly end up on the wrong side. The conformist bias is too strong for a sample size of 7 and therefore, this high sample size is an impediment for high performance.

2.4.3.2 PBSL McElreath

We found that for PBSL McElreath, too, a sample size of 3 results in higher performance than higher sample sizes in most conditions. Here we explain why this is the case. PBSL McElreath works by averaging the observed payoffs of the choice options and choosing the option with the highest average payoff. This appears to be a reasonable heuristic but this heuristic may also fail in some conditions. Assume for example that the social learning strategy observes 100 individuals that chose A and 1 individual that chose B. Furthermore, assume that the one B choice led to a success and that of the 100 A choices, 99 were successful and 1 unsuccessful. It would now be very sound to deduce from this observation that A must have a very high chance of success, while we know almost nothing about B. Knowing nothing about B, we should conclude that the very successful option A is the better choice and thus confidently choose it over B. Yet, if we take the average payoff of A (assuming a success gives 1 and a failure 0), it will be 0.99, while the average payoff of B is 1 and therefore higher. PBSL McElreath would choose B, the option that is almost certainly worse.

A similar argument as above can be made for very unsuccessful options. If A is observed 100 times, of which 99 were unsuccessful and one successful, and if B is observed once and was unsuccessful, then B is in almost all situations the better option. Still PBSL McElreath would calculate the average payoffs, 0.01 in case of A and 0 in case of B, and consequently choose A. This example is of less importance, though, since it is a priori less likely that the worse option is chosen overwhelmingly often in the population.

These examples showcase why the strict comparison between average payoffs can sometimes go awry, especially if the observed sample is large. Even in less extreme examples, if A is successful 5 times and unsuccessful once, while B is successful once, A is arguably the better choice but PBSL McElreath would choose B. Indeed, given this setup, if we assume that the better option has a success probability of 60% and the worse option of 40%, then A is the better choice in 10 out of 13 cases. The success probability of the better option would have to exceed 98% in order for B being more likely to be the better choice.[8]

[8]This is because if the better option has a very very high success probability, it becomes unlikely to observe even one failure in a sample of 6.

The approach of simply comparing means is flawed for high sample sizes. When many individuals choose the same option, because of the described effect, it becomes *less likely* that an additional individual will follow their lead. PBSL McElreath will display anti-conformist behavior when many individuals are sampled (see left panel of figure 2.22). This effect occurs even though McElreath et al. [120] included a bias towards choosing the more frequent option. This bias is, however, only applied in fringe cases – only if there is a tie in average payoffs – and thus cannot correct for these suboptimal decisions. Now it is also clear why observing less leads to better results for PBSL McElreath. If the sample consists of only three or four observations, the described error cannot occur.[9]

Baldini showed more formally that anti-conformism is expected for higher sample sizes when averaging-type PBSL is used [8]. He therefore called into question the usefulness of averaging-type PBSL as a social learning strategy in general. However, as we showed, the problem can be circumvented by using small sample sizes. In effect, performance was even quite high despite small sample sizes. There is no reason to believe that evolution would not settle for smaller samples, as sampling is not exogenously fixed but presumably subject to natural selection.

We argued that if one observes 5 successes with A and one failure with A, while observing one success with B, it is in most circumstances better to choose A. To implement this change, we modified PBSL McElreath so that if the average payoffs of the observed options are close but not necessarily equal, the more frequent of the two options will be chosen. In theory, this payoff-conformism trade-off should cure the described weakness of PBSL McElreath with high sample sizes. We found this change to remedy the anti-conformist behavior observed for higher sample sizes (see right panel of figure 2.22), resulting in an improvement in performance. For example, under default conditions and for a sample size of 6, performance increased from 61.6% (without trade-off) to 78.4% (with trade-off).

In summary, we thus find that if an option is chosen very frequently in the population, this can actually lead to that option becoming less likely to be chosen by PBSL McElreath when they sample many individuals. Therefore, PBSL McElreath performs better when only sampling 3 individuals. The adverse effect can, however, be amended by introducing a true payoff-conformism trade-off (instead of a tie-breaking rule). As a consequence, we find the performance at higher sample sizes to dramatically improve, exceeding performance at sample size 3.

[9] If one observes A being successful twice and unsuccessful once, and B being successful once, then choosing B is not obviously wrong.

2.4.4 Rationality of payoff-biased social learning

Economists are very interested in the question whether certain behavior can be reconciled with a rational choice model. For example, they might want to know whether the preferences that subjects reveal are transitive. In a slightly changed fashion, we can apply a rationality analysis to the behavior to social learners. For instance, borrowing from social choice theory, there is an interesting theorem by Peyton Young [182] that has consequences for the strategies we study. (Social choice theory typically deals with the question of how the preferences of different voters can be reconciled to choose a single winner, or a list of winners, from a set of candidates.)

Before we present Young's results, some definitions are required. First, he used two axioms, symmetry and consistency. Symmetry means that each voter choice is treated equally. Consistency means that if f is a social choice function, w and w' are disjoint subsets of voters, and $f(w) \cap f(w') \neq \emptyset$, then $f(w) \cap f(w') = f(w + w')$. In other words, if one group of voters w, according to their preferences, would choose X and a second group of voters w' would also choose X, the two groups combined, $w + w'$, should choose X as well. Young further defined a scoring rule as following: For each alternative, a score is assigned by a voter and the alternative(s) with the highest score is (are) part of the social choice set (the list of winners). A social choice function is a function that assigns to each voting profile with a finite amount of voters a social choice set.

Young showed that if the axioms of symmetry and consistency are followed, any social choice function is a scoring rule [182]. The contrapositive then also holds: Every social choice function that is not a scoring function violates at least one of the two axioms.

An analogy can be made between social choice theory and the study of social learning. The voter profiles of the former correspond to the sampled observation and the social welfare function corresponds to a social learning function that translates observations into a choice. Although the proof by Young would have to be adjusted to our case, it is quite likely that the same theorem would hold. The implication is then that for consistency and symmetry to hold, we have to require that the social learners obey a scoring function. Conformists and all other scoring-type PBSL strategies we studied obey to a scoring function, meaning they conform to the proposed axioms.

Not violating the symmetry axiom is a trivial task. The definition of the strategies does not allow them to differentiate between individuals in their sample. If three individuals are sampled, all are treated equally. It is possible to change this, one could e.g. define a strategy that gives more weight to the first sampled individual. But the strategies we used necessarily have to comply to symmetry, so that we can focus on whether the consistency axiom holds.

If the sample of a scoring-type PBSL is broken into two disjoint samples

w and w', and if the choice function f chooses option X for both samples, it is equivalent to say that the score S_X of X is higher than the score S_Y of Y (were X and Y, interchangeably, correspond to A or B):

$$(f(w) = X) \quad \Leftrightarrow \quad (S_X(w) > S_Y(w)) \tag{2.1}$$
$$(f(w') = X) \quad \Leftrightarrow \quad (S_X(w') > S_Y(w')) \tag{2.2}$$

Furthermore, as scores are simply summed up, we have:

$$S_X(w + w') \quad = \quad S_X(w) + S_X(w') \tag{2.3}$$
$$S_Y(w + w') \quad = \quad S_Y(w) + S_Y(w') \tag{2.4}$$

The choice derived from the joint sample $w + w'$ is X if and only if:

$$(f(w + w') = X) \Leftrightarrow (S_X(w + w') > S_Y(w + w'))$$

which, using 2.3 and 2.4, is equivalent to:

$$(f(w + w') = X) \Leftrightarrow (S_X(w) + S_X(w') > S_Y(w) + S_Y(w'))$$

Using all the above, we can conclude that:

$$\begin{aligned} ((f(w) = X) \wedge (f(w') = X)) \quad &\Leftrightarrow \quad ((S_X(w) > S_Y(w)) \wedge (S_X(w') > S_Y(w'))) \\ &\Rightarrow \quad (S_X(w) + S_X(w') > S_Y(w) + S_Y(w')) \\ &\Leftrightarrow \quad (f(w + w') = X) \end{aligned}$$

Or, in short:

$$((f(w) = X) \wedge (f(w') = X)) \Rightarrow (f(w + w') = X)$$

In words, this means that for scoring-type PBSL, if a subsample would result in choosing X and the other subsample would result in choosing X, the joint sample would also result in choosing X. This is in agreement with the consistency axiom, meaning that scoring-type PBSL obeys both symmetry and consistency and thus, in one sense of the word, behaves rationally.

Does the same hold true for PBSL McElreath [120]? This strategy acts according to the following rule: Calculate the average payoff generated by the two options in the observed sample, then choose the option with the higher average payoff; in case of a draw, choose the more frequent option. This rule seems to be reasonable but is it also rational?

As we said, the symmetry axiom cannot be violated. To check whether consistency is violated by PBSL McElreath, we have to stretch a little. Let f be the choice function that translates the observations of a social learner using PBSL McElreath into a personal decision. Moreover, let $\{A\}$ be an observation of an individual who chose A and was successful and $\left\{\bar{\mathrm{A}}\right\}$ be an observation of an individual who chose A and was unsuccessful; the same holds for $\{\mathrm{B}\}$ and $\left\{\bar{\mathrm{B}}\right\}$. According to the definition of McElreath et al. [120], $f\left(\left\{\bar{\mathrm{A}}\right\}\right) = \mathrm{A}$. Furthermore, owing to symmetry, $f(\{\mathrm{AB}\}) = \{\mathrm{AB}\}$, meaning that observing one success with A and one success with B may either lead to A or B choice. For the two disjoint samples, only one option overlaps, namely A. Therefore, one should expect A to be chosen from the joint sample – $f\left(\left\{\bar{\mathrm{A}}\right\} + \{\mathrm{AB}\}\right) = \mathrm{A}$. However, according to McElreath et al. ([120], see also table 2.2), $f\left(\left\{\bar{\mathrm{A}}\right\} + \{\mathrm{AB}\}\right) = \mathrm{B}$. This is a violation of consistency. With regards to the system we study, PBSL McElreath does thus not behave according to one of the two axioms. In this narrow sense, PBSL McElreath acts irrationally.

We have proven that scoring-type PBSL but not PBSL McElreath obeys the two axioms; we have not proven though that a necessary condition for a strategy to behave rationally requires that strategy to be a scoring-type PBSL. Such a statement would require a more lengthy proof in the spirit of Young [182], which we will not provide here.

PBSL McElreath may not obey the axiom of consistency, but does this make them "irrational". To show that acting according to this strategy would seem unreasonable on the surface, consider the following example: Say that a social learner encounters either i) $\{\mathrm{AB}\}$ or ii)$\left\{\bar{\mathrm{A}}\bar{\mathrm{B}}\right\}$. In either case, symmetry requires that these observations must leave the social learner indifferent between choosing A and choosing B. Now, imagine that later the social learner encounters $\left\{\bar{\mathrm{A}}\right\}$. If she was in situation i), her total sample would be $\left\{\mathrm{A}\bar{\mathrm{A}}\mathrm{B}\right\}$, making her choose B [120]. If she was in situation ii), her total sample would be $\left\{\bar{\mathrm{A}}\bar{\mathrm{A}}\bar{\mathrm{B}}\right\}$, making her choose A [120] (see also table 2.2). So starting from the same assumption, indifference between A and B, the same additional information, $\left\{\bar{\mathrm{A}}\right\}$, will in the first situation tilt the choice towards A and in the second situation towards B. According to one's standpoint, this may or may not appear to be irrational behavior. Now it is of course possible that empirically, humans use the same evidence in one context to justify X and in another context to justify not-X but it is hard to deny that people acting according to McElreath et al.'s PBSL strategy would appear to be somewhat unreasonable.

There are findings in the psychological literature showing that subjects can respond in opposing fashions after being exposed to the same evidence. A famous study by Lord et al. [115] began by separating subjects into those who favor and those who disapprove of capital punishment. Then the subjects were exposed to two studies, one supporting the hypothesis that capital punishment has a deterrent effect, one refuting it. Interestingly, exposure to these studies had a polarizing effect. Those in favor of capital punishment were even more in favor after the exposure, whereas those against capital punishment opposed it even more afterwards. Empirically, the same evidence can be used to rationalize X, as well as not-X.

Analogously, the observation of $\left\{\bar{\mathrm{A}}\right\}$ carries ambiguous information. On the one hand, if another individual chose A, this individual presumably had a good reason to; on the other hand, if A did not lead to success, maybe B is really the better option. This is similar to the ambiguous evidence provided by Lord et al. However, a major difference to our example is that in their study, subjects were already biased towards one of the two opinions beforehand, making their case an example of motivated reasoning [107]. For social learners, the observation of $\left\{\bar{\mathrm{A}}\right\}$ cannot be used to support a pre-established opinion, as there is no reason to favor A or B after observing just$\{\mathrm{AB}\}$ or $\left\{\bar{\mathrm{A}}\bar{\mathrm{B}}\right\}$. Motivated reasoning cannot explain the differential outcomes.

Another case of inconsistency is the violation of the sure thing principle observed by Shafir and Tversky [153]. In one example, a game was proposed were a coin flip that shows heads results in the subject gaining \$200 and tails results in him or her losing \$100. Most subjects would prefer to only play the game once instead of twice or more. However, after they have won once or have lost once, most subjects were willing to take a second gamble. This is inconsistent because if both outcomes make the subject prefer to play a second game, they should also prefer a second game before they know the outcome. Different observations lead to the same outcome, even though different outcomes should be expected.

In our case, the same observation leads to different outcomes, even though the same outcome should be expected. The cases are therefore similar but reversed. This indicates that we should not dismiss the possibility that real subjects behave in such an inconsistent manner when using social learning, but we are not aware of any finding that would directly support or refute our example. Still, since we know that PBSL McElreath shows very high performance, it is interesting to contemplate that a behavior that appears to be irrational from an axiomatic point of view can prove to be favored by natural selection.

2.4.5 Undervaluation of the advantages of conformist bias

One assumption we made in this chapter was that we have homogeneous populations, meaning each strategy only interacts with copies of itself. This shortcoming may be especially detrimental for strategies with high conformist bias. For example, pure conformists (corresponding to PBSL with weights $[+\mathrm{X}/+\mathrm{X}]$) only performed at chance level in this chapter. The reason is that they will completely converge towards only choosing A or only choosing B. Afterwards, they will never switch again, regardless of how the environment behaves, since they simply imitate each other's choices. Given that each option is better than the other 50% of the time, their performance is thus 50%. Similarly, other strategies with a high conformist bias could be at a disadvantage when using the approach of this chapter.

In this light, it is not surprising that conformists and similar strategies were found to perform poorly compared to other strategies. If there were, however, other strategies to be imitated in the population, the picture could change dramatically. Although conformists, who only use frequency-information, could never overtake a whole population, they could well attain a high equilibrium frequency. To address this possibility, we will use evolutionary simulations containing mixed populations in the next chapter.

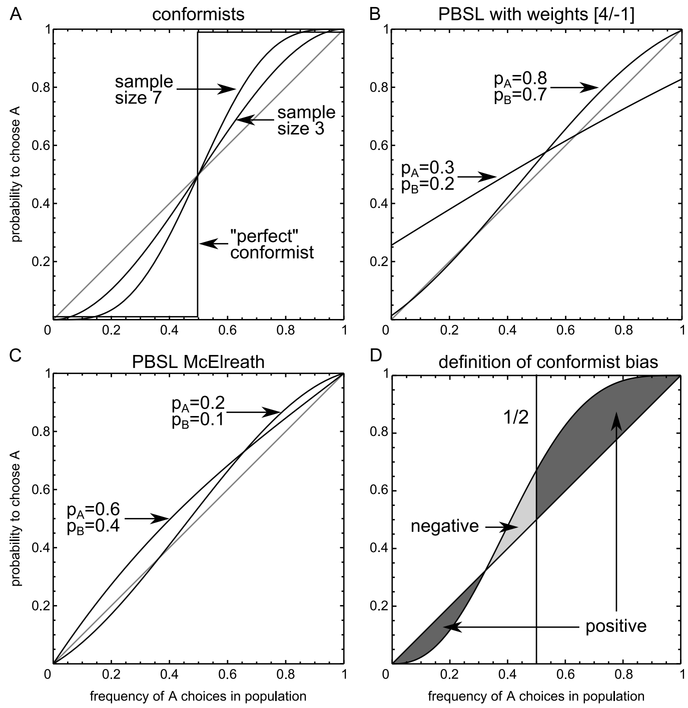

Figure 2.14: Conformist bias of different strategies. For all panels, the probability to choose A is shown as a function of the frequency of A choices in the population. A) Pure conformists with sample size 3 (solid line), sample size 7 (dashed line), and "perfect" conformists (dotted line). Random choice would result in preserving the A choice frequency in the population, as shown as the dashed gray line. B) PBSL with weights $[4/-1]$ and $p_A = 0.8$, $p_B = 0.7$ (solid line) or $p_A = 0.3$, $p_B = 0.2$ (dashed line). C) PBSL McElreath with $p_A = 0.2$, $p_B = 0.1$ (solid line) or $p_A = 0.6$, $p_B = 0.4$ (dashed line). D) Definition of conformist bias. Conformist bias is greater the more a strategy tends to decrease its probability to choose A when A is chosen by the minority of the population and the more it tends to increase its probability to choose A when A is chosen by the majority of the population. The strategy shown in this example generally tends to act according to conformity (dark gray area) but will sometimes increase its probability to choose A even if A is the minority option (light gray area).

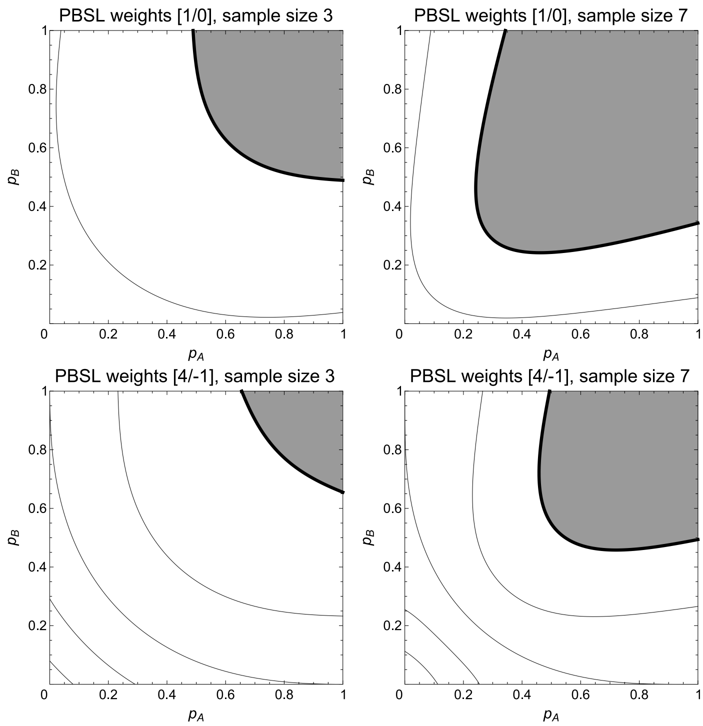

Figure 2.15: Conformist bias of scoring-type PBSL as a function of p_A and p_B. Isoclines are shown for differences of 1/8th, the thick isocline is the zero isocline. Gray areas correspond to positive values of conformist bias, white areas to negative values of conformist bias. Top left: PBSL with weights [1/0] and sample size 3; top right: PBSL with weights [1/0] and sample size 7; bottom left: PBSL with weights $[4/-1]$ and sample size 3, bottom right: PBSL with weights $[4/-1]$ and sample size 7. PBSL with weights [1/0] is more conformist. Conformist bias increases with higher mean p_A and p_B and higher sample size.

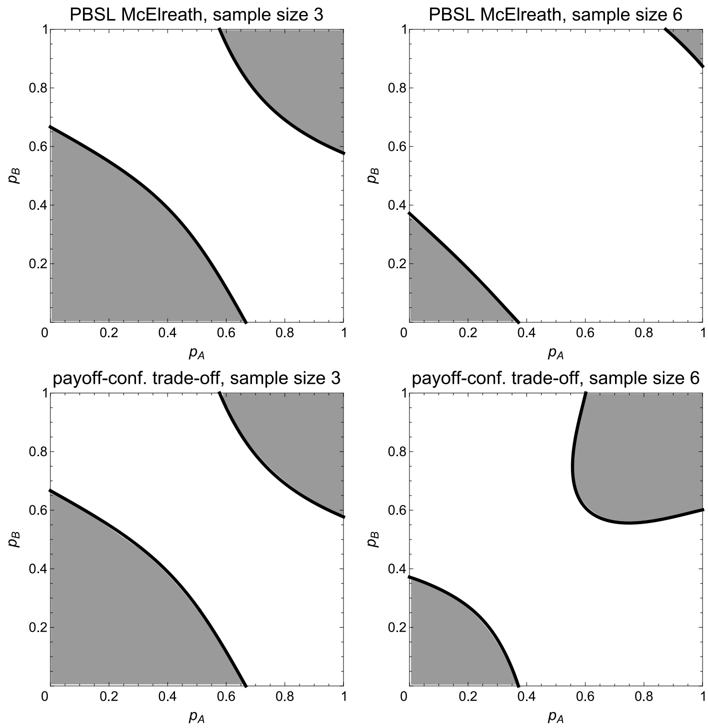

Figure 2.16: Conformiist bias of averaging-type PBSL as a function of p_A and p_B. Isoclines are shown for differences of 1/8th, the thick isocline is the zero isocline. Gray areas correspond to positive values of conformist bias, white areas to negative values of conformist bias. Top left: PBSL McElreath with sample size 3; top right: PBSL McElreath with sample size 6; bottom left: PBSL with payoff-conformism trade-off and sample size 3; bottom right: PBSL with payoff-conformism trade-off and sample size 6. Both strategies are more conformist when mean p_A and p_B are very high or very low. For higher sample size, PBSL McElreath becomes substantially less conformist but PBSL with payoff-conformism trade-off become only slightly less conformist.

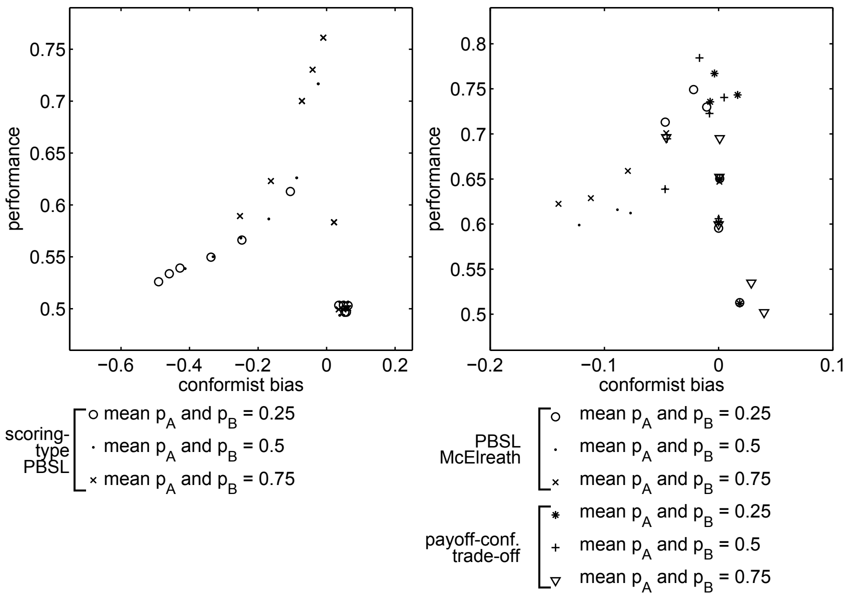

Figure 2.17: Performance of different PBSL strategies as a function of conformist bias. Conformist bias was calculated using the joint distribution of p_A and p_B, with their means set to 0.25, 0.5, and 0.75. Left panel: Scoring-type PBSL with positive weight on successes; right panel: averaging-type PBSL. Across conditions and regardless of the specific strategy, performance first increases as a function of conformist bias. After a certain point, for a conformist bias slightly below 0, performance suddenly drops. There thus appears to be an optimal degree of conformist bias.

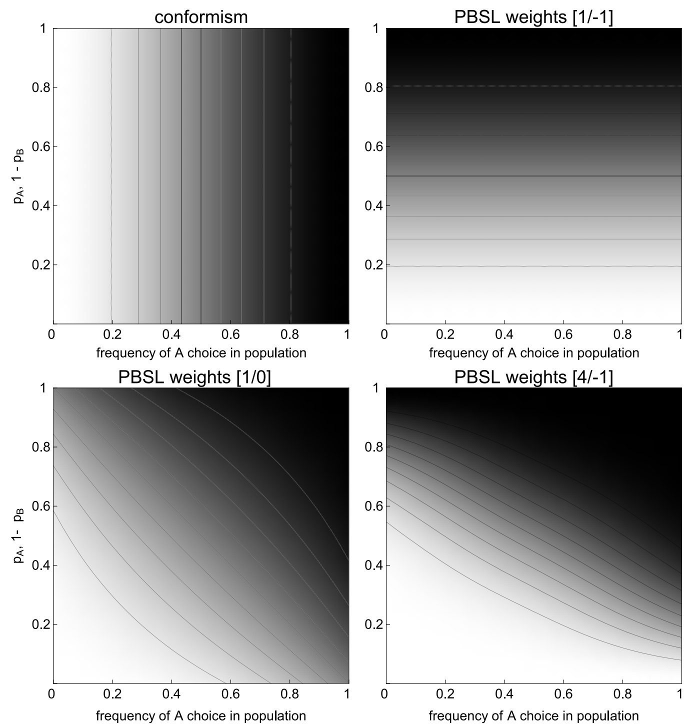

Figure 2.18: Contour plots of the probability of a PBSL strategy to choose A as a function of frequency-information (x-axis) and payoff-information (y-axis; it is assumed that $p_B = 1 - p_A$). Isoclines are at 0.1 distance, darker shades indicate higher probabilities to choose A. Top left: Conformism with sample size 3. This strategy only reacts to frequency-information. Top right: PBSL with weights $[1/-1]$ and sample size 3. This strategy only reacts to payoff-information. Bottom left: PBSL with weights [1/0] and sample size 3. This strategy reacts to both kinds of information about equally strongly. Bottom right: PBSL with weight $[4/-1]$ and sample size 7. This strategy reacts more strongly to payoff-information.

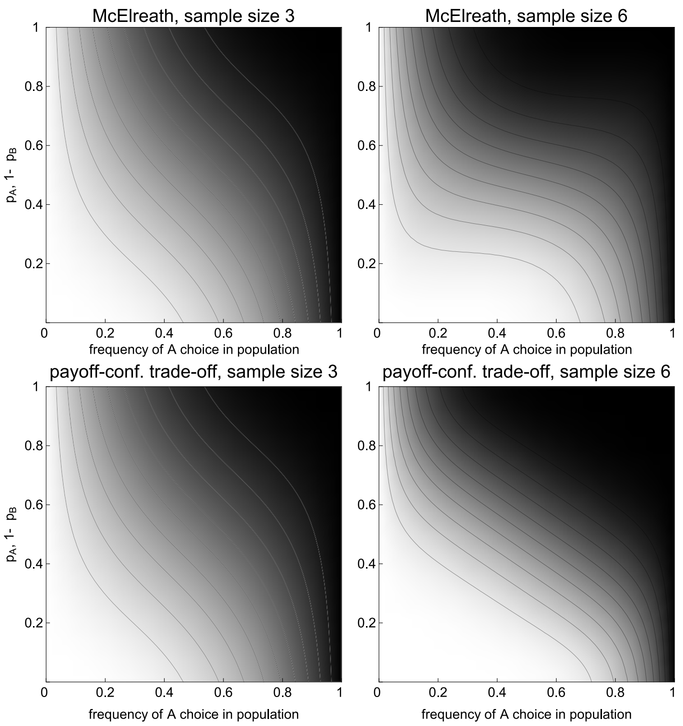

Figure 2.19: Contour plots of the probability of a PBSL strategy to choose A as a function of frequency-information (x-axis) and payoff-information (y-axis; it is assumed that $p_B = 1 - p_A$). Isoclines are at 0.1 distance, darker shades indicate higher probabilities to choose A. Top left: PBSL McElreath with sample size 3. This strategy reacts strongly to frequency-information. Top right: PBSL McElreath with sample size 6. Unless p_A and p_B are very low or high, this strategy reacts very strongly to payoff-information. Bottom left: PBSL with payoff-conformism trade-off and sample size 3. This strategy reacts strongly to frequency-information. Bottom right: PBSL with payoff-conformism trade-off and sample size 6. This strategy relies slightly more strongly but not too much on payoff-information.

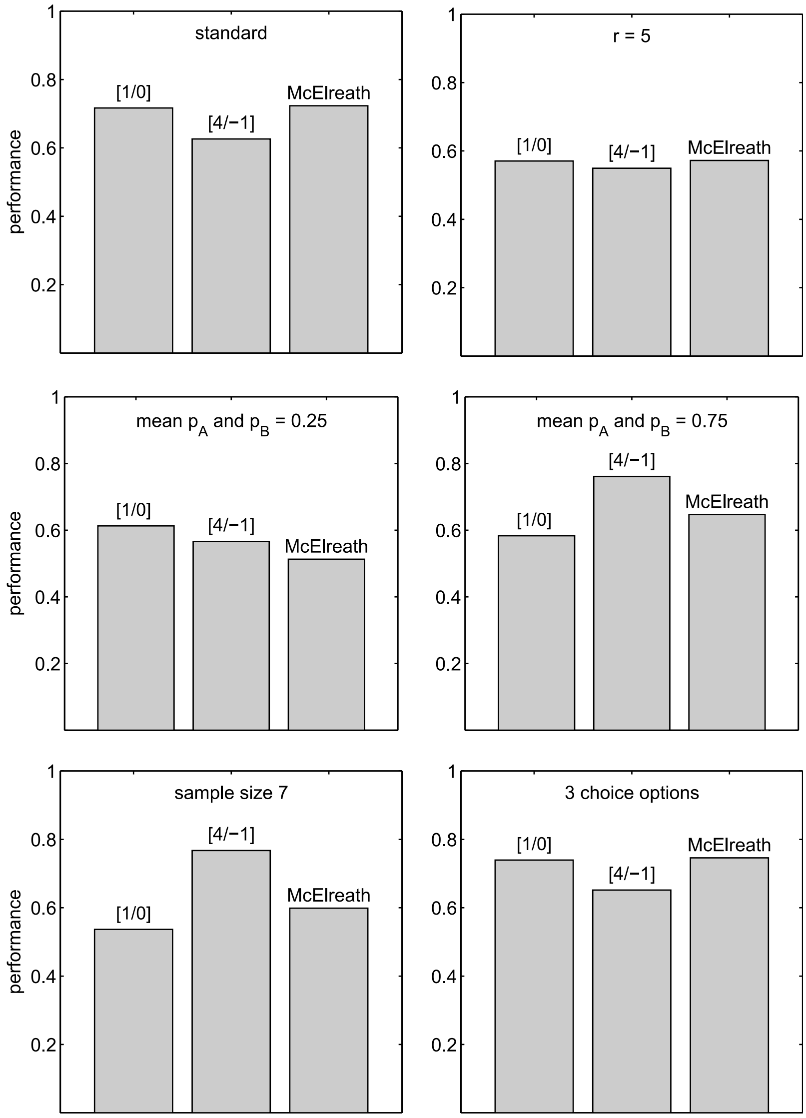

Figure 2.20: Comparison of the three strategies that have shown the most consistently high performance: scoring-type PBSL with weights of [1/0], scoring-type PBSL with weights $[4/-1]$, and PBSL McElreath (all with sample size 3 except if indicated otherwise). Conditions are as indicated, with all else held equal.

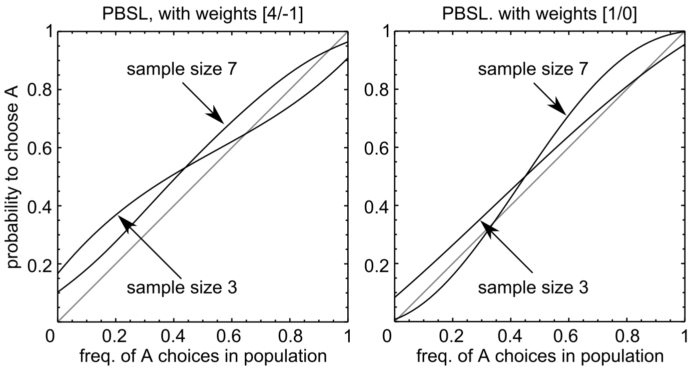

Figure 2.21: Probability to choose A as a function of the frequency of A choices in the population. It is assumed that $p_A = 0.55$ and $p_B = 0.45$. Left panel: PBSL with weights $[4/-1]$ and sample size 3 (dashed line) or sample size 7 (solid line). Right panel: PBSL with with weights [1/0] and sample size 3 (dashed line) or sample size 7 (solid line). The bissectrix (dashed gray line) is shown for reference. PBSL with weights $[4/-1]$ and sample size 3 converges towards mostly choosing the better option, though a large minority will choose the worse option. With sample 7, however, the strategy will converge almost completely towards choosing the better option. PBSL with weights [1/0] and sample size 3 shows reasonable behavior and converges overwhelmingly towards the better option. With sample size 3, however, the strategy shows strongly conformist behavior and may get stuck choosing the worse option.

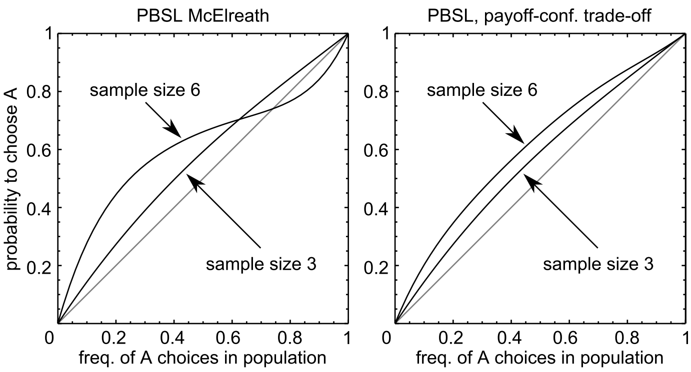

Figure 2.22: Probability to choose A as a function of the frequency of A choices in the population. It is assumed that $p_A = 0.6$ and $p_B = 0.4$. Left panel: PBSL McElreath with sample size 3 (dashed line) and sample size 6 (solid line). Right panel: PBSL with payoff-conformism trade-off with sample size 3 (dashed line) and sample size 6 (solid line). The bissectrix (dashed gray line) is shown for reference. PBSL McElreath with sample size 3 show reasonable behavior that suggests complete convergence towards the better option. With sample size 6, anti-conformist behavior occurs, meaning incomplete convergence towards the better option. With payoff-conformism trade-off, the anti-conformist behavior disappears for a sample size of 6; the strategy instead converges completely towards the better option, even faster than it would with sample size 3.

3 Rogers' paradox and informational breakdown

3.1 Introduction

3.1.1 Rogers' paradox

Social learning has been invoked to explain the adaptedness of human culture [22, 144]. For example, assume that there are two prey animals, say antelopes and wildebeests, that yield different expected values when hunting them. If person 1 is an individual learner, she will try hunting both antelopes and wildebeests and will eventually learn which prey is more profitable, albeit at a cost. Assume that antelope is the better choice. Person 2, a social learner, observes person 1's behavior and infers that antelopes are more profitable to prey on. If person 1 had opted for wildebeests, person 2 would have inferred that wildebeests are more profitable. So person 2 gained information about the prey animals without ever trying for herself.

If antelopes were always more profitable to hunt, one would expect natural selection to "teach" humans to better hunt antelopes. But if the environment is variable (in space or time), profitability may change. For example, environmental conditions might have favored wildebeest growth in recent years, increasing their abundance and thus making their hunt more profitable. Or excessive hunting might have depleted the pool of available antelopes, making them harder to find. The environment thus being variable, natural selection could be too slow to adapt to the relentless changes. One would thus be better off if one could learn, or if one could observe others in order to learn from their mistakes and save the costs of trying. Gaining the benefit while saving the cost should lead to a net profit.

So it seems that in varying environments, the existence of social learning makes it possible to avoid the costs of individual learning, improving the lot of the population as a whole. However, Rogers showed [146] that this reasoning is fallacious. Remember Rogers' model as we presented it in the first chapter. The fitness of individual learners was constant for all frequencies x of social learners. Social learners, however, did well when rare but worse than individual learners when too frequent. There was thus a certain frequency x^* at which social learners and individual learners did equally well, the equilibrium frequency of social learners. Yet, the fitness of individual learners being constant, the fitness of social learners at $x = x^*$ is exactly the same as the fitness of a population consisting only of individual learners. The population is not better off after the arrival of social learning. Others have shown that Rogers' results are quite robust [23].

In equilibrium, the mean fitness of the population is therefore not higher than initially, in absence of social learners. The curious conclusion that Rogers and others drew from his findings is that overall, simple forms of social learning do not lead to a higher adaptedness of the population. Yet it is plainly obvious that socially learned traits are beneficial for humans, for instance allowing us to survive in almost all terrestrial habitats on earth [26, 144]. The conclusion thus has to be that there is a catch, a possibility for social learning to improve adaptedness after all.

3.1.2 Proposed solutions to Rogers' paradox

There are several possible ways out of Rogers' paradox. Intuitively, they are easy to understand. In Rogers' original model, the mean fitness of the population transiently increases during the invasion of social learners. In equilibrium, social learning is, however, used so often that it cancels out the benefit of its use. It is clear that if social learning were used less often, the average fitness must be greater than that of a population of pure individual learners. To stop social learning from becoming more common, one could introduce a strategy that uses a fixed mixture of social and individual learning, involving less social learning on average than in Rogers' equilibrium.

Strategies that mix social and individual learning have been proposed in the past. Boyd and Richerson [23] and Kameda and Nakanishi [101] proposed strategies that are based on this idea. The proposed strategies rely on the information derived from learning individually only if this information is very clear with regard to which option is better. If the own experience leaves one indecisive, one imitates a random individual instead.

Enquist and coworkers [50] considered a "critical social learning" strategy that learns socially first but then learns individually if social learning proved unsatisfying. We will study a similar strategy. The reverse strategy, "conditional social learning" [50, 143], was also proposed; it consists of using individual learning first but switching to social learning if individual learning proves unsatisfying.

We are interested in whether the presented strategies solve Rogers' paradox in the context of our model. In our model, there are no individual learning costs, so random choice is not a particularly successful social learning strategies. We therefore changed the presented models so that when they employ social learning, they actually use conformism, a social learning strategy that we know is capable of invading a population of individual learners even in absence of learning costs. This alteration should, however, keep the spirit of the strategies alive and if it affects the performance, it does so in a positive way. We thus make the strategies even better than before, giving them a better shot at proving that they can resolve Rogers' paradox.

3.1.3 Informational breakdown

Rogers' paradox shows that the existence of social learning by itself does not help explain why culture is adaptive. Social learning would evolve, since as long as there are but few social learners, it is individually beneficial to learn socially. But in equilibrium, everybody is back to square one. So at least, one could say, social learning does not hurt the population. However, one may ask whether there are a social learning strategies that would make the population worse off.

Earlier, we explained that although the mean population fitness in equilibrium is the same as the fitness of individual learners in Rogers' model, this was not true for performance. Performance, measured as the probability of choosing the better option, was actually worse in equilibrium than before the invasion social learners. In contrast, one would expect social learning to produce better performance, and not to only save costs. We could show in the previous chapter that there were indeed social learning strategies that have the potential of vastly improving performance. This chapter will test whether this finding remains true in an evolutionary setting.

A further possibility for the population to be worse off through social learning would be if social learners became fixed. Assuming that these social learners cannot learn about the environment without the help of other strategies, they would only perform at chance level. Such a population consisting only of social learners would thus be worse off than one consisting solely of individual learners. We call this phenomenon informational breakdown because it entails that the decisions of all individuals in the population become self-referential, no information is drawn from the environment.

Some works find informational breakdown. Rendell et al. [143] found in their model that when there is spatial structure, social learning can become fixed. The reason for the extinction of individual learners is that at the borders where they interact with social learners, social learners are better off and thus proliferate, even though social learners that are surrounded only by social learners are worse off. For social learning to sweep through the population, the authors needed to assume specific spatial structures that are fixed (no mixing) and the cost of learning individually to be very high (greater than 30% of the potential benefit). Moreover, when further learning strategies were introduced, vanishing of individual learning was prevented. Therefore, their model shows that informational breakdown only occurs in very specific conditions.

Whitehead and Richerson [177] showed that social learning can become fixed when the environment varies according to red noise. In that context, red noise means that there can be long periods of environmental stability, resulting in gradual loss of costly individual learning. In their simulations, social learners became fixed in 84 of 100 runs with high individual learning costs. But sometimes, the environment would suddenly change, rendering the behavior of social learners maladaptive and thus leading to informational breakdown. However, again, the authors needed to assume high exogenous costs of learning individually. Introducing a richer set of social learning strategies prevented informational breakdown in most simulation runs. Therefore, informational breakdown again was but a very unlikely occurrence.

Informational breakdown could thus be found, but only under very restrictive conditions. These findings may lead to the conclusion that it should not be expected that a population becomes detached from the environment. However, there might be circumstances under which the equilibrium frequency of social learners can become very high. In finite populations, random drift could then lead to the fixation of those social learners, who suddenly find themselves only imitating other social learners. At this point, informational breakdown would follow.

How can such a high equilibrium frequency of social learners be attained? In Rogers' model, by increasing the cost of learning individually to almost match the benefit, the equilibrium frequency of social learners can be pushed close to 1. One could object, however, that such high costs of learning individually are implausible, since they would make the evolution of individual learning very unlikely in the first place (in contrast to just guessing). But we will show that certain social learning strategies are capable of attaining very high equilibrium frequencies without requiring learning costs at all. More so, our findings hold under very general conditions. Informational breakdown is thus not just a fringe case that one should never expect to occur but a real possibility.

So not only has it been shown that social learning is not beneficial, it is even possible that it can be detrimental for a population. This finding places an even higher burden on the hypothesis that social learning increases adaptedness of populations. An additional goal of this chapter will thus be to find out when informational breakdown occurs and when not.

3.1.4 Outline of this chapter

This chapter has three principal goals. First we are interested in whether the solutions to Rogers' paradox proposed in the literature do indeed solve Rogers' paradox. We will argue that they cannot. The reasons are 1) that in equilibrium, the strategies are not even present in the population when competing with other social learning strategies, and 2) that even if those other strategies were excluded, simple social learning strategies such as conformism severely hamper the proposed strategies' performance in equilibrium. The previous works did not test for these possibilities, which is why the limitations did not become manifest.

Second, we turn to strategies that in the context of our model may indeed solve Rogers' paradox. We show that payoff-biased social learning (PBSL), a type of social learning that we detailed in the previous chapter, offers some strategies that indeed lead to considerably better performance. Furthermore, these strategies should be expected to dominate the population in equilibrium, meaning that they are more plausible candidates for solving Rogers' paradox. They are, however, also cognitively more demanding, which is why we scrutinize the literature for evidence of payoff-biased social learning.

Third, we have to analyze the possibility of informational breakdown. We will show that under very general conditions, social learning strategies will become fixed in the population that are not, of themselves, capable of learning from the environment. This fixation is possible because the equilibrium frequency of those social learners is so high that random drift will fix them in a finite population. If that happens, the population's behavior will become completely detached from reality and performance will drop to chance level. We will discuss the implications of this finding.

3.2 Model description

3.2.1 Agent-based modeling

In the previous chapter, we followed a more or less classical modeling approach in that we calculated the aggregate behavior of different strategies, given how they behaved in the last period. Then we placed these strategies in a generated environment and simulated their behavior. In this chapter, however, we will choose a different approach to model the same task. Instead of modeling strategies in an aggregate manner, we will employ what is called agent-based modeling.

Agent-based modeling treats each individual of the population separately. The strategy of an individual is formulated as a behavioral rule. For example, instead of defining that a conformist adopts A with probability $(3 - 2 \cdot x) \cdot x^2$, with x being the proportion of A choices in the population, we define a conformist as an individual who obeys the rule, "observe three individuals and choose A if two or more of them chose A". At the end of the day, the two approaches should lead to the same result, but there are advantages and disadvantages of using either of the modeling approaches:

- Advantage of using aggregate modeling: Whole groups of individuals are modeled, instead of simulating each individual separately. This process takes less computational time – simulations of aggregate models are roughly two orders of magnitude faster than those of agent-based models with a population size of 10,000. Aggregate modeling thereby allows us to test different parameter values more systematically, which was our main concern in the previous chapter. Also, being closer to previous models, the obtained results can be compared more easily.

- Advantage of using agent-based modeling: It is very hard to include an individual's history into aggregate models. Learning rules that are path-dependent can therefore not be considered [19]. Using agent-based models allows us to implement these path-dependent strategies. The continuity of aggregate models may also be problematic. For instance, the fraction of individuals who choose A will never become 1 but only converge arbitrarily close to 1. In agent-based models, however, it can well happen that all individuals choose A in the same period. As human effective population sizes were, historically, rather small ($\approx 10,000$ [83, 167]), agent-based models may actually capture reality more accurately.

3.2.2 Individual learning, conformism, and mixed forms

By use of agent-based models, we can keep track of the individual histories of each agent in the population. This allows us to incorporate several strategies that could not be modeled before. Additionally, we will include strategies we have already encountered earlier.

Reinforcement learning: In the first chapter, we described that we model individual learners as using reinforcement learning with exponential discounting. Agent-based modeling allows for each individual's path to matter, so that they can make choices based on their private experience. Although very simplistic forms of individual learning like win-stay lose-shift (reinforcement learning without memory) can easily be modeled in the aggregate, this is much harder for reinforcement learning with memory.

Conformism: Conformism is one of the most studied social learning strategies and there are good reasons why one should expect conformism to be used (see e.g. [21, 22, 86]). We found in the first chapter that conformists can invade a population of individual learners and attain a high equilibrium frequency, especially when they sample few (3) individuals. In our model, conformism is defined as sampling a group of individuals and always choosing the option that is most common among these individuals.

Of course, we also want to test how "critical social learning" [50] and "conditional social learning" [50, 143] perform. One criticism that has been leveled against these strategies (Robert Boyd, personal communication) is addressed here. The criticism is that the way these strategies were conceptualized, they inherently have an edge over just individual learning or social learning. A critical social learner was defined as to first acquire a trait through social learning, and if this trait is unsatisfying, to learn individually. This implies that the critical social learner has the ability to assess the acquired trait and relearn it if the trait does not yield a good payoff. By definition, this makes the critical social learner better than a pure social learner, who has to stick with the first acquired choice. Simple individual and social learners are not granted the possibility to re-evaluate, even though it would be a possibility. An individual learner, e.g., could first learn the trait individually and if she does not succeed in acquiring the good trait try to learn individually again.

If the possibility to re-evaluate were granted to other strategies, it would not be true anymore that critical social learners are always superior to social learners. However, the authors do not allow for this possibility. The reason that critical social learning is superior in their model is therefore not that it combines first social than individual learning, but that it learns twice, while social learners learn only once.

In our model, we prevent this unequal battle. This is possible because our model consists of making several decisions during an individual's lifetime. Therefore, it is possible to incorporate a critical social learning strategy that learns socially in period t and then, if unsatisfied, learns individually in period $t + 1$, instead of learning twice in period t. This way, we avoid the unfair advantage of some strategies being allowed to learn twice each round while others could only learn once.

Furthermore, we changed the mode of social learning of critical and conditional social learners. In the original works, they used random copying when learning socially. In the context of our model, we know, however, that random copying is not a good learning strategy. The reason is that random copying results in lagging behind individual learners while the choice frequency in the population is preserved. Since there are no learning costs to be saved, there is no advantage of copying randomly. For this reason, we substituted conformism for random copying. We know that conformism, if not used too frequently, is capable of leading to very high performance, so mixing it with individual learning should result in reasonable learning strategies.

Since we changed how critical and conditional social learning work compared to the original papers, we chose to rename them. Furthermore, "critical" and "conditional" do not transport the message what exactly these strategies do; in fact, the names could be reversed and would fit equally well. We thus prefer names that describe the behavioral rule of the strategies more precisely. In our model, we call an individual who uses individual learning by default but switches to conformism if unsatisfied "Opportunistic Conformist" (OC). In contrast, an individual who uses conformism by default but switches to individual learning if unsatisfied is called "Opportunistic Individual Learner" (OIL).

Opportunistic Conformism (OC): This strategy consists of individual learning by default. If in period t, the chosen option yields a success, the strategy will continue to use individual learning. If in period t, the chosen option does not yield a success, the strategy uses conformism in period $t + 1$. Note that the propensities for A and B, on which a strategy relies for individual learning, are updated even if conformism is used. This strategy is similar to "critical social learning" [50].

Opportunistic Individual Learning (OIL): This strategy consists of using conformism by default. If in period t, the chosen option yields a success, the strategy will continue to conform. If in period t, the chosen option does not yield a success, the strategy uses individual learning in period $t + 1$. Note that the propensities are updated even if conformism is used. This strategy is similar to "conditional social learning" [143].

In addition to these two strategies, we also want to incorporate another strategy proposed in earlier works [23, 101]. This strategy consists of using individual learning only when it is certain about which option is superior and to use social learning otherwise. In our model, individual learners have propensities to choose A and B. If these propensities are close, both options are equally reinforced, meaning that the individual is uncertain about which option is the better one. If one option has a much higher propensity than the other, this option has proven more successful in the individual's past. Therefore, using the difference in propensities as indicator of certainty is most natural in the context of our model. Again, we choose conformism to play the part of social learning. Because there is no consensus on how to call such a strategy, we use a name that best describes the behavioral rule, "In Doubt, Conform" (IDC).

In Doubt, Conform (IDC): This strategy consists of individual learning by default. If, however, the individual is in doubt about which option is superior, she will rely on conformism instead. Uncertainty is measured as the difference between the propensities of the two options. If the difference is less than 2, an individual is considered to be uncertain. Note that the propensities are updated even if conformism is used. This strategy corresponds to strategies studied in earlier works [23, 101].

3.2.3 Imitate The Wealthiest

Since we use agent-based models, we can keep track of individual characteristics of the members of the population. For instance, we know at each point in time how often an individual has been successful in the past. An argument can be made that a social learner should preferentially imitate those who are especially successful. We therefore propose a strategy that, among the sampled population, imitates the individual who has had the highest aggregate number of successes so far (in case of a tie, a random individual among the tied individuals is imitated). As the aggregate number of successes could be interpreted as the "wealth" of an individual, we call this strategy "Imitate The Wealthiest" (ITW). In our model, the number of successes adds to the fitness of the individual, so that in effect, the fittest individual is imitated. In the Pleistocene, this could be e.g. the individual with the most children, whereas in modern times, it could be the individual with the biggest real estate.

Earlier works have studied strategies called "imitate the best" [2, 96, 152, 173] but this strategy corresponds more closely to payoff-biased social learning as we defined it, since only the last period's payoff and not the total payoff is used as criterion. ITW is instead more akin to prestige-biased social learning [87], but note that our model does not include idiosyncratic differences except for learning strategies and personal histories of decisions and outcomes. In our model, prestige can therefore not depend on an individuals intelligence, hunting skills, etc.

Imitate The Wealthiest (ITW): ITW consists of adopting in each period the last period's choice of the individual who, among the sampled individuals, had the highest aggregate number of successes so far. This individual is the "wealthiest", hence the name of the strategy. It can be modified so that one does not imitate if one has a higher fitness than all sampled individuals (making it more similar to previously studied strategies [151, 152]). In preliminary tests, this modification did not alter the results, though, so we did not implement it. Moreover, these tests showed that ITW fares better the more individuals are sampled.

3.2.4 Payoff-biased social learning

In the previous chapter, we studied in detail two classes of payoff-biased social learning (PBSL), scoring-type PBSL and averaging-type PBSL. In short, the former consists of having two scores, one for A and one for B, which are updated according to how successful the options were in the observed sample, then choosing the option with the higher score. Conformism is a special case of scoring-type PBSL. The second class, averaging-type PBSL, consists of averaging the payoff of A and of B, and choosing the option with the highest average payoff. Mixed in is some degree of conformism.

Of all the strategies we tested, four promised to be exceptionally good. Of the scoring-type PBSL strategies, we found those who weighted successes four times as much as failures and counted the latter as negative (short: weights $[4/-1]$) to perform quite well, especially for higher sample sizes. We therefore include this strategy in our tests. Furthermore, scoring-type PBSL that completely ignores failures (short: weights $[1/0]$ or, more generally $[+X/0]$) were also quite successful, especially for a sample size of 3.

Of the averaging-type PBSL strategies, PBSL McElreath [120] with sample size 3 performed on a high level. This strategy uses conformism as a tie-breaker – if both options have the same average payoff, the more common option is chosen. In addition, we found the averaging-type PBSL with payoff-conformism trade-off and sample size 6 to be very successful. For this strategy, conformism can sometimes override the choice suggested by the payoff average, hence the trade-off. These two strategies were also included in our tests. For more information on them, refer to the previous chapter.

3.2.5 General points

When social learning occurs, one has to decide how many individuals are sampled at a time by a social learner. Often, we will find that the more individuals are sampled, the better informed the decision of the social learner will be. Therefore, we have to put a limit on this number, a limit which is more or less arbitrary. As previously, we choose "the magical number 7" [128] for this purpose. However, we also found that some strategies are better off with lower sample sizes. For instance, conformists acquired a higher equilibrium frequency for sample size 3; scoring-type PBSL with weights [1/0], as well as PBSL McElreath and PBSL with payoff-conformism trade-off had better performance for lower sample sizes too (3, 3, and 6, respectively). In these cases, we used the optimal sample sizes instead of 7. As Opportunistic Conformists, Opportunistic Individual Learners, and In Doubt, Conform all make use of conformism, their sample sizes are also 3. That means that the only strategies with a sample size of 7 are scoring-type PBSL with weights $[4/-1]$ and ITW.

We included three strategies that rely sometimes on individual and sometimes on social learning, namely Opportunistic Conformism, Opportunistic Individual Learning, and In Doubt, Conform. One may ask why these mixing strategies always combine individual learning with conformism and not with ITW or PBSL. The reason is that we found that conformists can transiently lead to a very strong improvement of mean population performance. In equilibrium, of course, Rogers' paradox manifests and conformists perform as good or bad as individual learners. But as they show a high potential, it is reasonable to make them part of the mixing strategies. Moreover, conformism is the simplest form of social learning that we study in this chapter and thus, a priori, most likely to be found in nature. Strategies that mix other types of social learning could be studied in later works.

Strategies can be categorized along different dimensions. The most important category is whether a strategy is *autonomous* or *dependent*. An autonomous strategy can perform above chance level even in the absence of other strategies in the population because it has a direct way to learn about the environment. This can be tested in the following way: Assuming that all members of the population were to choose A, is there a real chance that they will eventually move to B when B becomes the better choice? If there is, the strategy counts as autonomous but otherwise, it is dependent on other strategies. For example, individual learners cannot become stuck with one option; if this option results in too many negative outcomes, they will decide to choose the other option.

In contrast, imagine that there were only conformists in the population and that they all choose the same option. As there are no dissenting role models that could convince the conformists to switch, conformists would never abandon their choice. Individual learners are thus autonomous and conformists dependent; all other strategies can be characterized according to this characteristic as well. Of course, there are strategies that are autonomous but do not incorporate environmental information, such as choosing randomly. But such strategies are not studied here and are thus of no concern.

3.2.6 Evolution

We model populations with a size of 10,000, which is thought to correspond to the effective population size of humans during the last million years or so [83, 167]. Generations are non-overlapping, genetics haploid, and reproduction non-sexual (there is only one locus). Updating the strategies' frequencies is modeled by the Fisher-Wright process. That means, after each generation, parents contribute a number of offspring equal to their fitness to the offspring pool. Of this pool, 10,000 individuals are drawn at random to form the next generation. This way, population size is held constant and individuals with higher fitness on average contribute more offspring to the next generation.

p_A and p_B vary continuously over time. This means, after a generation has ended, the last value of p_A and p_B in that generation constitutes the first value of the subsequent generation. In the very first period, p_A and p_B equal 0.5. Analogously, offspring inherit their very first choice from their parent; this is the only form of vertical cultural transmission in our model. In the very first period, individuals start with a random choice. Offspring genetically inherit the strategy from their parent, mutations do not occur.

3.3 Results

3.3.1 Behavior of the strategies

3.3.1.1 Opportunistic conformists and opportunistic individual learners

We introduced three strategies in this chapter that make use of a mix of individual learning and conformism. An opportunistic conformist (OC) uses individual learning by default and switches to conformism when her choice did not yield success; an opportunistic individual learner (OIL) does the reverse. As both rely on some mixing of individual learning and conformism, it should be expected that their behavior is also a mixture of both learning strategies. In figure 3.1, we see on the left panel the behavior of individual

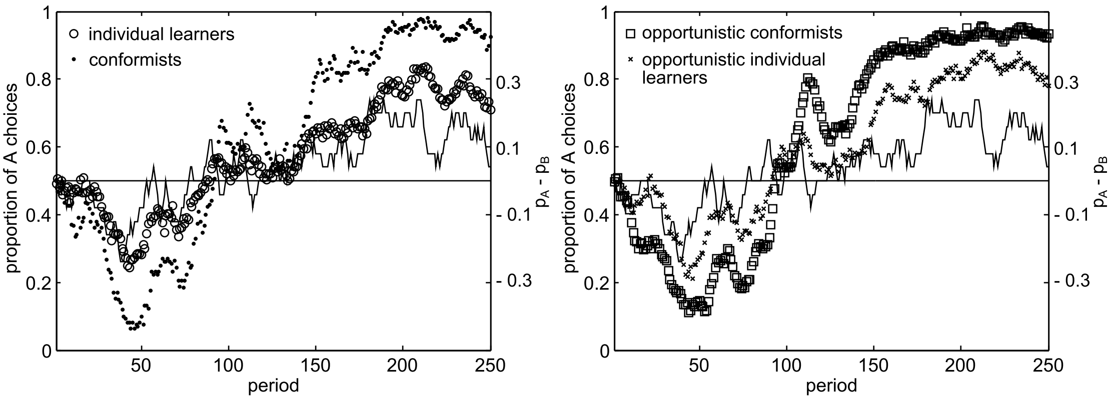

Figure 3.1: Left: Proportion of A choices made by the learning strategies, measured on the left y-axis, and the environment in the form of $p_A - p_B$ (solid line), measured on the right y-axis, over time. Left panel: Individual learners (◯) and conformists (•) in a population consisting of 50% of both. Individual learners match, conformists overmatch probabilities. Right panel: OC (□) and OIL (×), simulated separately but superimposed for comparison. Their behavior is a mixture of individual learning and conformism. OC overmatches more strongly than OIL.

learners (◯) and conformists (•) in a population consisting of 50% of both strategies. As we already know, individual learners are more conservative, whereas conformists strongly overmatch. Remember that overmatching generally is good, as it allows more individuals to choose the better option, but often comes at the cost of lagging behind environmental changes, which is of course bad for performance.

On the right panel, we see the behavior of OC (□) and OIL (×'s), both simulated separately and using the same environment as in the left panel. As expected their behaviors are a mixture of the behavior of individual learners and conformists. OC overmatches more strongly than OIL.

3.3.1.2 In Doubt, Conform

The third strategy that relies on mixing individual learning and conformism uses the former when it is certain about which option is the better one and the latter otherwise. We called it In Doubt, Conform (IDC). As has been argued Boyd and Richerson [23], the more certain an individual needs to be about her choice (the higher the difference in propensities), the more often she will be correct when she actually uses individual learning.

We tested whether this holds in our model by varying the threshold difference in propensities. Pure individual learning produced a performance of $\approx 59.0\%$; if the difference in propensities has to be 1 or more, individual learning results in a performance of $\approx 61.5\%$, if it has to be 2 or more, performance is $\approx 64.3\%$, and if it has to be 3 or more, performance is $\approx 68.8\%$

(likewise, performance is especially low when the difference in propensities is low). So performance is indeed higher for higher thresholds. At the same time, the higher the requirements for certainty are, the less likely an individual is to use individual learning in the first place. For a difference in propensities of 1, individual learning is used $\approx 64\%$ of the periods, for a difference of $2 \approx 38\%$ of the periods, and for a difference of $3 \approx 20\%$ of the periods. So there is a trade-off here – the more certain an individual wants to be before relying on her own experience, the more likely it is that she actually chooses the better option, but the less often she will actually rely on her experience.

From this, it becomes clear that the required difference in propensities should neither be too small nor too large. If it is too small, IDC would use individual learning too often, even if uncertain, thus lowering performance; if it is too large, it would use individual learning too rarely and rely too much on conformism, causing the known troubles. In figure 3.2, we show the behavior of IDC for a given environment and for different levels of required certainty. If propensities have only to differ by 1 (•), the strategy mostly relies on individual learning, as can be seen by the strong tendency to closely match the environment. If propensities have to differ by 2 (◯), there is stronger overmatching, as conformism becomes more ubiquitous. If propensities have to differ by 3 (□), there is strong uniformity in behavior, which is expected if conformism is the main form of learning. The more conformism is used, the stronger a strategy overmatches (which is good) but the slower it reacts to environmental changes (which is bad).

In preliminary tests, we found that a threshold of approximately 2 results in the best performance of IDC if competing in a homogeneous population. Therefore, we use this threshold for our analysis in this chapter. The threshold generates a behavior that is very close to the behavior of opportunistic conformists (compare figures 3.1 and 3.2). This is because by construction, both strategies become more likely to rely on conformism after a failure.

3.3.1.3 Imitate The Wealthiest

We introduced a new strategy in this chapter that we called Imitate The Wealthiest (ITW). This strategy samples 7 individuals and imitates the choice of the individual with the highest aggregate amount of successes so far. This strategy is in some way very different from the other social learning strategies, especially conformists and other scoring-type payoff-biased social learners.

To understand this, we have to distinguish between strategies that use a *compensatory* decision rule and strategies that use a *noncompensatory* decision rule (these terms are borrowed from [67]). "Compensatory" means that all the gathered bits of information have approximately the same weight in shaping the final decision, whereas "noncompensatory" means that some

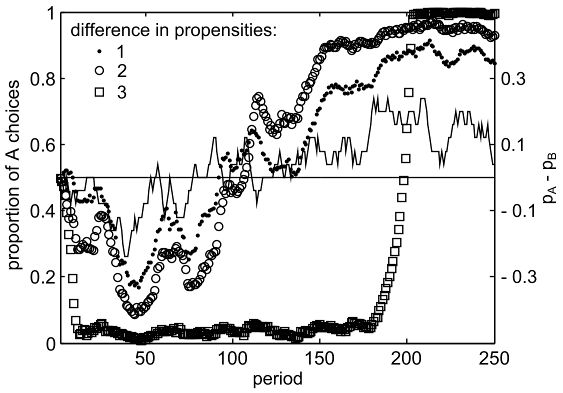

Figure 3.2: Proportion of A choices made by In Doubt, Conform (IDC, measured on the left y-axis) over time as a function of the required absolute difference in propensities. The strategies were simulated separately in a homogeneous population, the results are overlaid for comparison. The environment (solid line) is shown as $p_A - p_B$, measured on the right y-axis. For a required difference of 1 (•), the strategy uses individual learning even if fairly uncertain about which option is better. For a required difference of 3 (□), the strategy mostly uses conformism and individual learning only when very certain. A required difference of 2 (◯) presents a compromise between the two more extreme strategies. It can be observed that a low required difference leads to more "conservative" behavior (probability matching) associated with individual learning, whereas a high required difference leads to more extreme and uniform behavior associated with conformism.

part of the gathered information counts much more than the rest. For example, a conformist counts all the sampled individuals who choose A and compares them with all the individuals who choose B; all observations have the same weight in this calculation. Conformism is thus a compensatory decision rule. ITW is different. If the wealthiest among the sampled individuals chooses B, it does not matter what all the other sampled individuals do, even if they all choose A; although they are superior in numbers, they cannot "compensate" for the one dissenter. The difference between compensatory and noncompensatory social learning strategies is crucial, as we will see.

One implication of using a compensatory decision rule is that small differences in the sample are very unlikely to cause large differences in behavior. Therefore, it is also very unlikely that the presence of some individuals with a different strategy can affect the aggregate behavior of a strategy with a compensatory decision rule. This is shown in figure 3.3 using the example of conformists and ITW. On the left panel, we show the behavior of conformists ($\times$) when either competing with 2% individual learners or when the whole population uses conformism ($\square$). As is readily seen, regardless of the presence of individual learners, conformists behave the same. Compare this with the behavior of ITW shown in the right panel. For ITW, it makes all the difference whether 2% individual learners are present ($\bullet$) or not ($\bigcirc$); in the first case, they react adaptively to the environment, in the second case they do not. Since for ITW, even the voice of a single sampled individual can completely reverse a decision, small minorities in the population have the potential to shift the behavior of the whole population.

3.3.1.4 Payoff-biased social learning

Last chapter, we encountered two types of payoff-biased social learning strategies. Those PBSL strategies were either of the scoring-type or the averaging-type. Among all the possibilities, we encountered some strategies that performed especially well. As those strategies were already analyzed in detail in the last chapter, we will not provide too much further analysis here.

A point of interest we have not discussed yet is whether the strategies are compensatory or not. Scoring-type PBSL consists of adding up the scores in favor or against options A and B, and then choosing the option with the higher score. If two individuals using scoring-type PBSL observe the same sample safe for one individual who behaves differently, it is thus very likely that the scores are similar and that the decisions are the same. Scoring-type PBSL is therefore a compensatory decision rule. (As we just saw, conformism, which is a special case of scoring-type PBSL, is also compensatory). This is illustrated in figure 3.4, where the behavior of PBSL with weights $[4/-1]$ is shown once in the presence ($\times$) and once in the absence ($\square$) of 2%

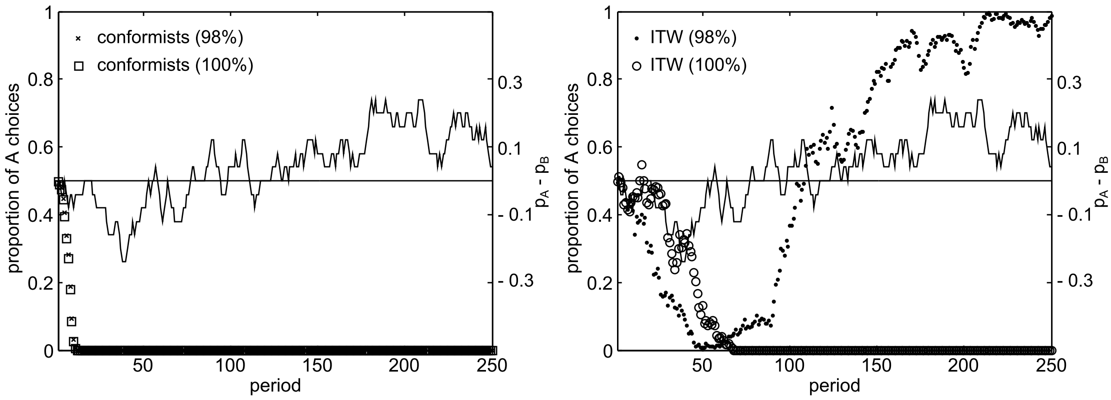

Figure 3.3: Left: Proportion of A choices (measured on the left y-axis) made by conformists if they make up 98% of the population (×) and the rest are individual learners (not shown), or if they make up 100% of the population (□). Right: Proportion of A choices made by Imitate The Wealthiest (ITW) if they make up 98% of the population (•) and the rest are individual learners (not shown), or if they make up 100% of the population (◯). The environment (solid line) is shown as $p_A - p_B$, measured on the right y-axis. Conformism is compensatory, resulting in the presence of 2% individual learners having as good as no influence on aggregate behavior. ITW, in contrast, is noncompensatory, leading to a huge difference in aggregate behavior in the presence of only 2% individual learners.

individual learners. For almost all periods, the presence of individual learners makes hardly any difference, although it must be noted that between periods 50 and 100, it does make a difference. Scoring-type PBSL is thus very compensatory but not completely insensitive to small changes.

Is averaging-type PBSL also compensatory? One simple example may elucidate this question. If an individual using such a strategy observes 6 individuals who chose A, 3 of which were successful, and one individual who chose B, it makes a difference whether this last individual was successful or not. If she was successful, B has a higher average score and is the choice that follows, and if she was unsuccessful, A has the higher average score and is the choice that follows. For scoring-type PBSL, whether with weights [1/0] or [4/ − 1], the lone B chooser would not make a difference. In contrast to scoring-type PBSL, averaging-type PBSL is therefore noncompensatory.

In the right panel of figure 3.4, we illustrate this. There, the behavior of PBSL with payoff-conformism trade-off is shown once in presence (•) and once in absence (◯) of 2% individual learners. In absence, the social learning strategy will reach 100% A choices shortly after period 50 and never move away thereafter. The presence of 2% individual learners convinces them, however, to abandon A and to predominantly choose B some periods later. A few individual learners can therefore make a huge difference, showing that averaging-type PBSL is noncompensatory.

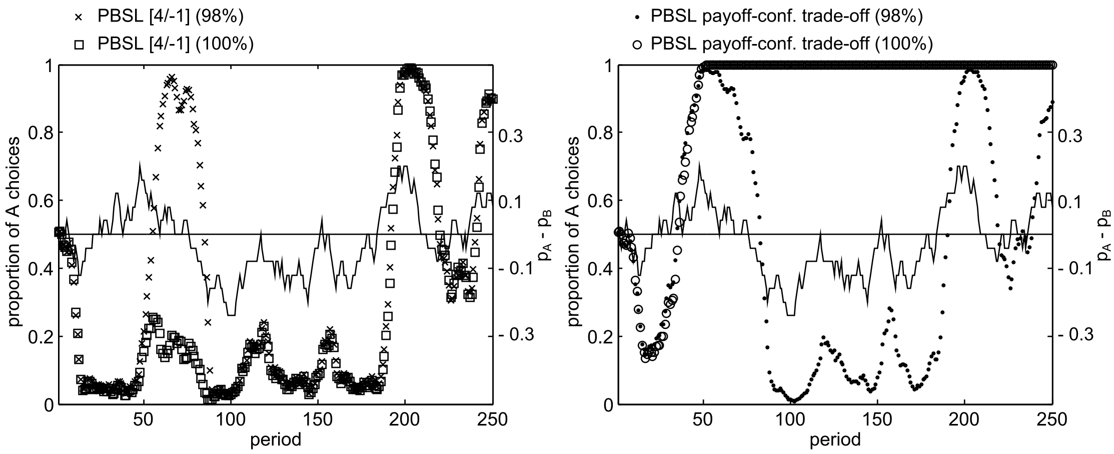

Figure 3.4: Left: Proportion of A choices (measured on the left y-axis) made by PBSL with weights $[4/-1]$ if they make up 98% of the population ($\times$) and the rest are individual learners (not shown), or if they make up 100% of the population ($\square$). Right: Proportion of A choices made by PBSL with payoff-conformism trade-off if they make up 98% of the population ($\bullet$) and the rest are individual learners (not shown), or if they make up 100% of the population ($\bigcirc$). The environment (solid line) is shown as $p_A - p_B$, measured on the right y-axis. PBSL with weights $[4/-1]$ is compensatory, resulting in the presence of 2% individual learners having almost no influence on aggregate behavior, except in between periods 50 and 100. PBSL with payoff-conformism trade-off, in contrast, is noncompensatory, resulting in a huge difference in aggregate behavior in the presence of only 2% individual learners.

3.3.1.5 Dependent strategies

A sufficient condition for a strategy to be dependent is if that strategy would never switch away from choosing A (B) if the whole population were to choose A (B). In such a case, the population will be stuck with this choice forever and there will not be any information flow from the environment towards the learners – there is informational breakdown. In other words, they require the presence of other strategies that do not get stuck with one choice in order to meaningfully react to the environment in the long run. Both options have to be present to form a “substrate” for their being chosen.

Among the strategies we test in this chapter, four are dependent strategies. These strategies are:

- conformism
- Imitate The Wealthiest
- PBSL McElreath
- PBSL with payoff-conformism trade-off.

Each time that an evolutionary simulation results in the extinction of all autonomous strategies, only leaving back some combination of the dependent strategies above, we have encountered an informational breakdown.

Notably, not all social learning strategies are dependent. This is obviously true for those who mix in some amount of individual learning, OC, OIL, and IDC. But it is also true for scoring-type PBSL with weights [1/0] and [4/ − 1]. A social learning strategy that completely ignores its personal information and instead only bases the choice on the observation of others can therefore still be autonomous.

3.3.2 Rogers' paradox: Evolutionary stability

As we said, there are several proposed solutions to Rogers' paradox [23, 50, 101, 143]. These solutions consisted of developing strategies that perform a mix of individual and social learning. We implemented these strategies in our model. In the following sections, we will show that these strategies are not persuasive solutions to Rogers' paradox. Our argument will use two lines of attack. First, we will show that these strategies are outcompeted by other strategies, especially by PBSL strategies and ITW. Second, even if PBSL and ITW were disallowed, the proposed solution strategies would be prone to invasion by pure conformists, in which case the attained equilibrium produces a performance that is hardly an improvement compared to the performance of individual learners. These findings cast a doubt on the claim that these strategies are convincing solutions to Rogers' paradox.

3.3.2.1 Default conditions

Our first line of attack consists of showing that the proposed solutions to Rogers' paradox, Opportunistic Individual Learning (OIL), Opportunistic Conformism (OC), and In Doubt, Conform (IDC) are prone to invasion by other strategies. To show this, we performed evolutionary simulations containing all 10 strategies, each starting with 1000 individuals, over 10,000 generations. Since we deal with a stochastic process in finite populations, outcomes will vary between simulations. For each condition, we therefore ran 10 simulations.

The first result shows the outcome using the default parameters. The frequencies of the all strategies that did not become extinct at least once in the 10 simulations is shown in figure 3.5. The thin lines present the results from all ten simulations, the bold line the average. The first thing to note is that individual learning, as well as the three proposed solutions to Rogers' paradox, namely OIL, OC, and IDC, always become extinct. PBSL with weights [1/0] also always goes to extinction. The strategies that remain at least once are conformism, ITW, PBSL with weights [4/ − 1], PBSL McElreath, and PBSL with payoff-conformism trade-off.

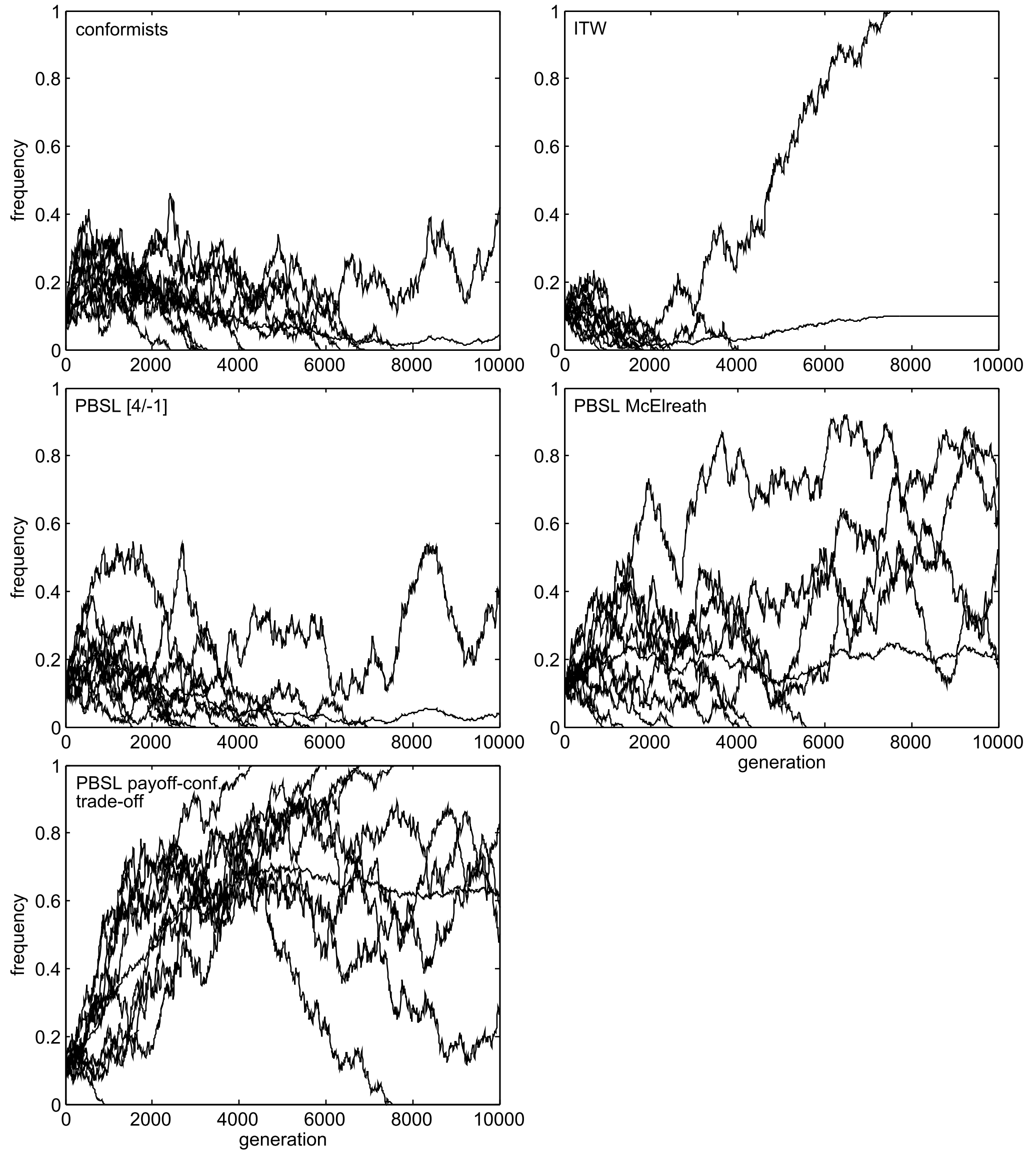

Figure 3.5: Evolutionary simulation of all 10 strategies. Shown are only the strategies that at least once did not become extinct in the 10 simulations. The thin lines are the distinct frequencies derived from the 10 simulations, the bold line the average frequency of those simulations. The only strategies that did not always become extinct are conformism, ITW, PBSL with weights $[4/-1]$, PBSL McElreath, and PBSL with payoff-conformism trade-off. Among those, the first three were only present in 1 of the 10 simulations, suggesting that only the latter two can be expected to be robustly present. Note that OIL, OC, and IDC, the proposed solutions to Rogers' paradox, always become extinct.

Taking a closer look at the results reveals that conformism, ITW, and PBSL with weights $[4/-1]$ were only present in one trial each at the end of simulation. PBSL McElreath and PBSL with payoff-conformism trade-off were, in contrast, present several times, with the latter attaining the highest frequencies most of the time. We repeated the simulations with a population size of 25,000 instead of 10,000 to check for robustness (see Appendix). It turns out that in this replication, only these two PBSL strategies, as well as conformists, achieved to not become extinct. Furthermore, PBSL with payoff-conformism trade-off proved to always be the dominant strategy.

3.3.2.2 Robustness

We showed that OIL, OC, and IDC always become extinct when competing with the other strategies. It is, however, possible that the default conditions are hostile to OIL, OC, and IDC, and that if we changed the parameters, outcomes would be more favorable towards them. Therefore, we repeated the evolutionary simulations with different reversion factors, different mean success rates of the environment, and increasing the environmental stability. These parameters had proven to have the most influence on performance (see the previous chapter for more details). Interestingly, none of the changes resulted in OIL, OC, or IDC to be able to compete with the other strategies (see Appendix). We can thus state with confidence that the observation that OIL, OC, and IDC always become extinct when competing with the other strategies is robust to changes in the environmental parameters.

3.3.2.3 Excluding PBSL

Another objection to our findings could be that some of the studied strategies are implausible from an empirical point of view. Especially payoff-biased social learning is, compared to the other strategies, cognitively more demanding. It requires some form of sophisticated integration of all observed outcomes, either in the form of adding them up (scoring-type) or averaging them (averaging-type). Therefore, we excluded some of the PBSL strategies and reran the simulations.

First, we excluded scoring-type PBSL. The results were very close to those of the default condition (see Appendix). Next, we excluded averaging-type PBSL, which has proven to be the most dominant type. Still, neither OIL, nor OC, nor IDC stood a chance against the other strategies (see Appendix). Excluding one of the two classes of PBSL strategies could thus not help our three candidate strategies.

Next, one could argue that PBSL is implausible altogether. Perhaps, it is impossible to observe the payoffs of others in the real world. Therefore, we ran the simulations again while excluding all four PBSL strategies. The result is shown in figure 3.6. Individual learners, OIL, and OC always became extinct before the end of the simulations. Conformism persisted in 2 of the 10 simulations, but only at low frequencies. IDC persisted in 9 out of 10 simulations, averaging at about 20% frequency. The most dominant strategy was ITW, which persisted in all simulations and consistently attained high frequencies.

Since IDC persisted, does this mean we have a potential solution to Rogers' paradox? To answer this, we need to have to look at performance. We show the mean population performance derived from the simulations on the bottom right hand side of figure 3.6. As performance is noisy, we applied a running average filter of 1000 generations length. First, we indeed see that the mean performance exceeds the performance of pure individual learning (dashed line). Over time, however, the frequency of the strategies evolve, entailing a decline of the mean performance. Finally, mean performance dips below the performance of individual learners, meaning that the population is actually worse off than before social learning was introduced. Therefore, Rogers' paradox is not solved at all.

In sum, even after excluding PBSL completely, only one of the three proposed solutions to Rogers' paradox, IDC, is capable of resisting total extinction. But even then, it remains the minority in the population, while ITW makes up the majority. Consequently, the mean population performance drops below the performance of individual learners. This is not a very convincing outcome for the proposed solution strategies.

3.3.3 Rogers' paradox: performance

Although showing some promise, neither Opportunistic Individual Learning, nor Opportunistic Conformism, nor In Doubt, Conform proved capable of competing with the other strategies in an evolutionary battle. In fact, they almost always became extinct. Now one could make the argument that all those other strategies that are oppressing OIL, OC, and IDC are not realistic and should thus be excluded. Here we will show that even then, the candidate strategies do not present convincing solutions to Rogers' paradox.

To prove this, we will proceed as follows. If it is plausible that either OIL, OC, or IDC are existing learning strategies, pure conformism must also be a possibility. This is because conformism is part of OIL, OC, and IDC, so if strategies exist that rely on conformism as part of their decision mechanism, there should also be a strategy that relies only on conformism. Consequently, one has to ask whether pure conformists could invade a population consisting of OIL, OC, or IDC. And if conformists do invade, what is the mean population performance when equilibrium is reached; specifically, is it still above the performance of individual learners? If it is not, Rogers' paradox is not solved.

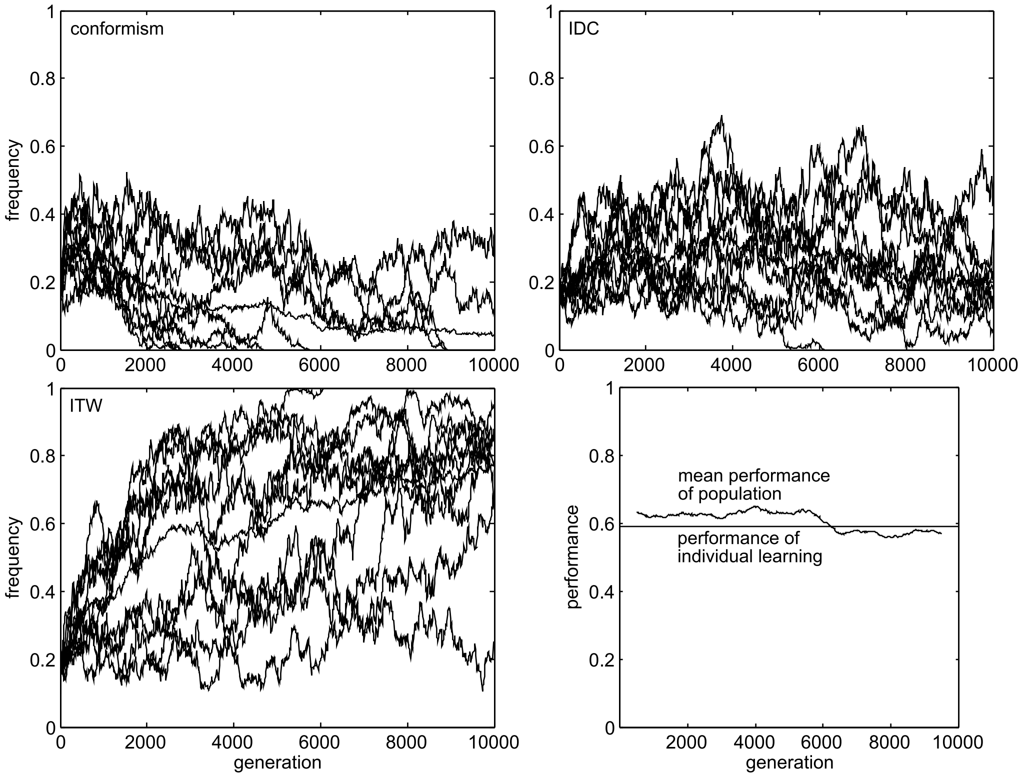

Figure 3.6: Evolutionary simulation of the 6 non-PBSL strategies. Shown are only the strategies that did not become extinct at least once in the 10 simulations. The thin lines are the individual frequencies derived from the 10 simulations, the bold line the average frequency of those simulations. The only strategies that did not always become extinct are conformism, IDC, and ITW. Among those, ITW was the most dominant, while IDC and conformism persisted at low frequencies on average. Bottom right: Mean performance derived from the same simulations. Performance was smoothed by a running average filter to reduce noise. Although the mean population performance (solid line) is initially higher than performance of individual learners (dashed line), it declines over time to a lower level. The population is thus worse off, meaning that Rogers' paradox is not solved.

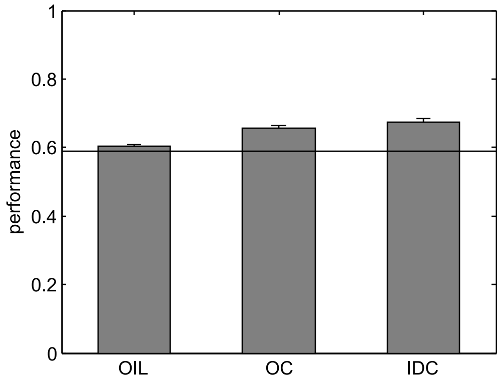

Figure 3.7: Performance of OIL, OC, and IDC (+S.E.M.). All three strategies beat the performance of individual learning, as shown by the dashed line. IDC displays the best performance.

3.3.3.1 Performance in homogeneous populations

First, we wanted to check a necessary (but not sufficient) condition for OIL, OC, and IDC to resolve Rogers' paradox, whether they can improve overall performance by themselves. We found that they do, as shown in figure 3.7. Pure individual learning results in an average performance of about $59.05 \pm 0.12\%$ (S.E.M.). Opportunistic Individual Learning results in a mean performance of $60.31 \pm 0.50\%$, Opportunistic Conformism in a mean performance of $65.61 \pm 0.75\%$, and In Doubt, Conform in a mean performance of $67.32 \pm 0.86\%$. Thus OIL, OC, and IDC do indeed improve upon the performance of individual learners, the difference ranging from 1 to 8 percentage points. Moreover, these strategies do not require individual learners in the population, meaning that even if they could completely replace individual learners without jeopardizing performance.

3.3.3.2 Performance when competing with conformists

The candidate strategies use a mixture of individual learning and conformism that involves sufficient individual learning to maintain good performance. However, we found that they are prone to invasion by pure conformists, as shown in figure 3.8. In the top left, the performance of individual learners competing with conformists, as obtained by simulations, are shown. The performance of individual learners being independent of the frequency of conformists, the classical paradox as formulated by Rogers obtains: In equilibrium, the population has the same average performance as a population consisting solely of individual learners (see also the first chapter of this work).

For OIL (top right), OC (bottom left), and IDC (bottom right), the results differ. As long as conformists are rare in the population, they outcompete these strategies and will thus invade. As conformists become more common,

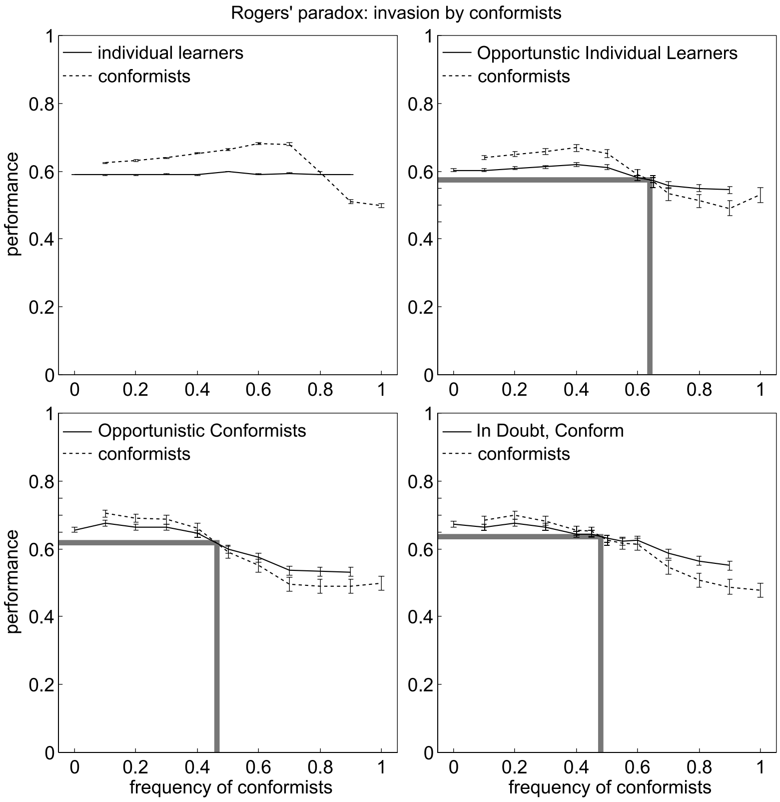

Figure 3.8: Rogers' paradox and the proposed solution strategies. The performance of conformism (dashed line) and the performance of individual learning (top left), OIL (top right), OC (bottom left), and IDC (bottom right, solid lines) are compared as a function of the frequency of conformists. The gray lines indicates the approximate equilibrium frequency and the corresponding performance. Error bars indicate standard errors of the mean. When individual learners compete with conformists, as the performance of individual learners is frequency-independent, the classical paradox described by Rogers [146] obtains. OIL, OC, and IDC have a higher performance than individual learners in homogeneous populations. When they compete with conformists, however, their performance drops as conformists become more frequent. In equilibrium, they perform worse than in absence of conformists. The equilibrium performance of OC and IDC are slightly above the performance of individual learners, which could be seen as a (weak) solution Rogers' paradox.

at a certain point, the performance of OIL, OC, and IDC starts to sink. The performance of conformists starts to sink even faster, though, so that at one point, performance curves of the two strategies intersect, which is where the equilibrium is expected.

In contrast to individual learners, the three candidate strategies have a lower performance at equilibrium than initially. For OIL, the estimated performance at equilibrium is roughly 58%. This means that even though this strategy, in a homogeneous population, outperforms individual learners, in equilibrium with conformists, it does not. It can therefore not solve Rogers' paradox, which would at least require it to do better at equilibrium than individual learners. For OC, the performance at equilibrium is roughly 62%, for IDC, it is roughly 64%. Those are 3 and 5 percentage point advantages over individual learning, respectively. One could thus argue that OC and IDC could indeed be solutions to Rogers' paradox. However, the initial advantages of 6 to 8 percentage points have vastly diminished. And it is questionable whether an increase of 5 percentage points total is really sufficient to explain the adaptedness of human culture.

In summary, we analyzed three candidate strategies, OIL, OC, and IDC, that were proposed in the literate as solutions to Rogers' paradox [23, 50, 101, 141]. In homogeneous populations, they would indeed lead to a better overall performance. From evolutionary perspective, however, we showed that there are better strategies that we would expect to drive these three strategies to extinction. But even if we assume that those other strategies did not exist, conformists would invade and deteriorate their performance. In equilibrium with conformists, OC would preserve a 3 percentage points advantage, IDC a 5 percentage points advantage over individual learners. This means that among the proposed strategies, IDC shows the most promise but it is questionable whether the meager performance increase shown by IDC is a satisfying solution to Rogers' paradox.

3.3.3.3 Payoff-biased social learners

If neither OIL, OC, nor IDC present convincing solutions to Rogers' paradox, maybe payoff-biased social learning do. We introduced this class of strategies in the previous chapter. Of the studied strategies, 4 proved especially promising, namely PBSL with weights [1/0], PBSL with weights [4/ − 1], PBSL as proposed by McElreath et al. [120], and PBSL with payoff-conformism trade-off. Therefore, we included these four strategies from the previous chapter in the present analysis.

First we have to check whether these strategies outperform individual learners. This is indeed the case, as shown in figure 3.9. PBSL with weights [1/0] have a performance of $68.98 \pm 0.65\%$ (S.E.M.), PBSL with weights [4/ − 1] have a performance of $74.16 \pm 0.83\%$. (Because the agent-based model deals with finite population sizes, the performance derived from it

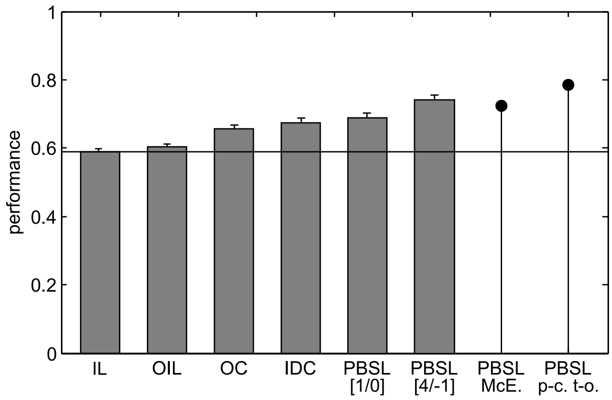

Figure 3.9: Performance of individual learning, OIL, OC, IDC, PBSL with weights $[4/-1]$, PBSL with weights [1/0], PBSL McElreath, and PBSL with payoff-conformism trade-off. As the last two strategies are not autonomous, their performance is shown as derived from the aggregate model (as discussed in the previous chapter), in contrast to the other performance values, which were derived using the agent-based model. All shown strategies do better than individual learners.

is not exactly the same as performance derived from the aggregate model.) These performance results are well above what individual learners achieve.

PBSL McElreath and PBSL with payoff-conformism trade-off are dependent strategies, meaning that in absence of autonomous strategies, they will eventually all choose the same option. This constitutes an informational breakdown and will, in the long run, reduce performance to chance level. Therefore, we cannot test them in a homogeneous population. Their performance was thus derived from the aggregate model in the previous chapter, where the choice limit prevented informational breakdown. For PBSL McElreath, we observe a performance of $72.34\pm0.23\%$, for PBSL with payoff-conformism trade-off, we observe a performance of $78.43 \pm 0.17\%$. As these numbers indicate, the four tested PBSL strategies outperform individual learners by a fair margin. Even more, they also outperform OIL, OC, and IDC. Therefore, the PBSL strategies are potentially better candidate strategies for solving Rogers' paradox.

As always, performance has to be considered in equilibrium. Therefore, we tracked performance over time when all strategies competed with one another. The evolutionary outcome was shown in figure 3.5. When we look at the mean performance derived from these simulations (left panel of figure 3.10), we see that at the start, when all strategies are equally represented, we indeed find mean performance to be quite high ($\approx$ 70%). Over time, though, the frequencies of the strategies adjust, which is accompanied by a decline in the mean performance. At the end, mean performance is well below the performance of individual learning. This means that even with PBSL, Rogers' paradox is not solved.

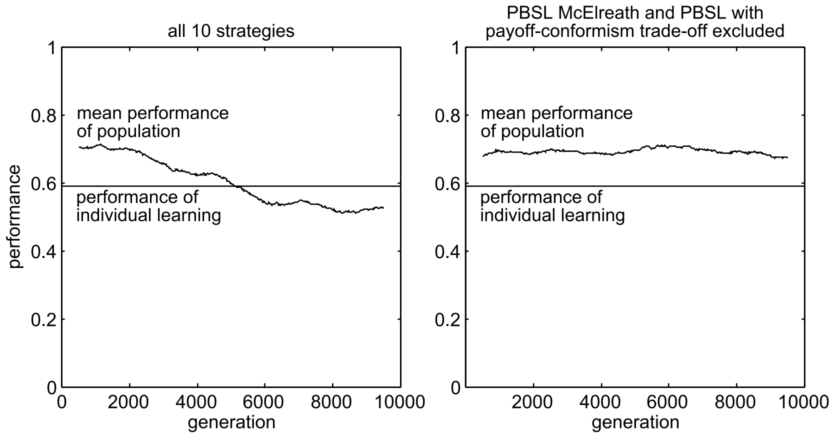

Figure 3.10: Mean performance (solid line) of the population as derived from evolutionary simulations over time. A running average filter was applied to the mean performance to smooth out noise. Performance of individual learning (dashed line) is shown for reference. Left: When all 10 strategies were tested, mean performance is initially well above the performance of individual learning. Over time, however, the composition of the strategies changes, resulting in a loss of performance. At the end, mean performance falls substantially below that of individual learning. Right: When PBSL McElreath and PBSL with payoff-conformism trade-off are excluded, mean performance remains higher than that of individual learning for the whole time.

We tested several adjustments of the parameter values to check whether this result is always obtained. The tested alterations could not bring about a qualitative change. There was, however, one possibility to maintain a high performance. When PBSL McElreath and PBSL with payoff-conformism trade-off were excluded, we find that PBSL with weights $[4/-1]$ becomes the dominant strategy while conformism is maintained at low frequency (see Appendix). This setup leads to a consistently high performance, as shown in the right panel of figure 3.10. In fact, average performance was $69.19 \pm 0.39\%$, 10 percentage points higher than performance of individual learning. There was no downward trend in the mean performance (linear model fit: slope of -0.02 (-0.156 to 0.116, 95% confidence bounds), $R^2 = 8.4^{-5}$, $p = 0.77$). This means that under the condition of excluding the two mentioned strategies, a solution to Rogers' paradox is possible.

3.3.4 Informational breakdown

In addition to studying Rogers' paradox, we were interested in the possibility of informational breakdowns. An informational breakdown occurs when none of the strategies that are left in the population are influenced in their behavior by the environment. This means that there is no flow of information from the environment towards the individuals. Consequently, the

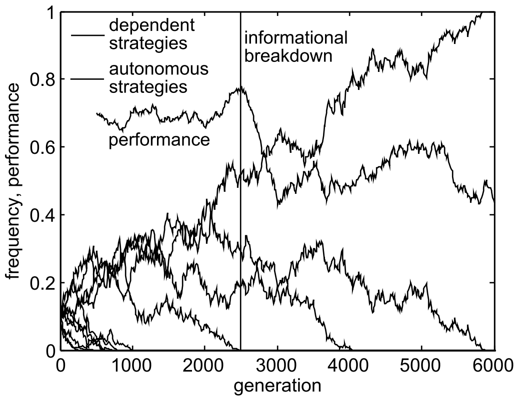

Figure 3.11: Evolutionary simulation with informational breakdown. The frequencies of autonomous (thin lines) and dependent strategies (thick lines) over time is shown. At period 2490, all autonomous strategies have become extinct, resulting in informational breakdown. Performance (dashed line, smoothed by a running average filter) is high before the informational breakdown but declines immediately afterwards.

population behaves in a way that is completely detached from reality.

The simplest way that an informational breakdown could occur is if all the present strategies choose the same option and if their decision mechanisms do not allow them to switch away from this option. For example, if every individual were a conformist, sooner or later all of them would choose the same option, say A. As soon as this state is reached, nobody will ever switch to B, since the learning mechanism requires that in order to switch to B, B must be present in the conformist's sample, which is impossible in this state. Similarly, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off are prone to reaching a state of informational breakdown.

An example of informational breakdown is shown in figure 3.11. In generation 2490, the last autonomous strategy becomes extinct, leaving only dependent strategies behind. Consequently, the mean population performance drops from 70.3% to 52.3%, a significant decrease ($p < 10^{-8}$, two-tailed Wilcoxon rank sum test). Therefore, if the population is ever composed of only the four dependent strategies or a subset of them, the population will sooner or later encounter informational breakdown, leading to chance level performance.

The fact that dependent strategies have to rely on autonomous strategies to learn about the environment makes our search for informational breakdowns quite easy. For example, we found that in the default conditions, in 9 out of 10 simulations, the population finally reached a state were all remaining strategies are dependent (see figure 3.5). Only in one of the simulations could PBSL with weights $[4/-1]$ persist, which prevents informational breakdown. Informational breakdown is thus expected in the default conditions.

We further checked our results for robustness by changing the parameter values (see Appendix). In general, informational breakdown was the predominant outcome. There were some exceptions, though. When we increased the reversion parameter, PBSL with weights [4/ − 1] persisted in the population 8 out of 10 times, mostly at high frequencies. A higher reversion parameter basically implies that the environment is more volatile, which requires a strategy to adapt very rapidly. We could thus expect informational breakdown to be less likely in rapidly changing environments.

Moreover, we found that if we increased the mean success rate of the environment to 0.75, PBSL with weights [1/0] persisted in 4 out of 10 simulations. Environments in which successes are very common could thus also possibly prevent informational breakdown.

Finally, we found that when PBSL McElreath and PBSL with payoff-conformism trade-off were excluded, informational breakdown was prevented. PBSL with weights [4/ − 1] then became the most dominant strategy, persisting in all 10 simulations with a mean frequency of almost 70%. Of course, it is not very surprising that informational breakdown could be prevented when two of the four dependent strategies were excluded from consideration.

3.3.5 Principal findings

In this chapter, we were interested in investigating whether social learners are capable of improving human adaptedness to the environment compared to a population of only individual learners. Rogers [146] found that when social learning consists of random copying, in equilibrium, there is no improvement – this is Rogers' paradox. Furthermore, we showed that one has to consider the possibility that in equilibrium, the population is actually worse off than before. This is possible when in finite populations, all autonomous strategies become extinct, so that the remaining social learners lose touch with reality, creating a state that we called informational breakdown.

In total, ten strategies were studied. Individual learning and conformism are standard strategies. Opportunistic Individual Learning, Opportunistic Conformism, and In Doubt, Conform are derived from strategies that were proposed in the literature as solutions to Rogers' paradox. We find that by themselves, these strategies indeed outperform individual learners. They are, however, prone to invasion by pure conformists, which depreciates the performance to a point where there is only a small improvement. In Doubt, Conform did best and, in equilibrium with conformists, achieved a performance improvement of 5 percentage points, which is neither bad nor remarkable.

There is a bigger problem, however. We introduced another strategy based on a simple heuristic, Imitate The Wealthiest. It turns out that this strategy outperforms the other strategies in most circumstances without providing enhanced performance in equilibrium. Even worse, when we introduced four payoff-biased social learning strategies, all of the three proposed solutions became extinct in equilibrium. They are thus not very likely candidates for solving Rogers' paradox.

Are there other strategies that might resolve Rogers' paradox? We found that under a wide range of parameter values, payoff-biased social learning with payoff-conformism trade-off outperformed all the other strategies. However, this strategy was not able to sustain a high performance in finite populations. In contrast, it led to a decline of performance to chance level owing to informational breakdown.

There were only few scenarios that allowed to escape Rogers' paradox and avoid informational breakdown. When we excluded averaging-type payoff-biased social learning, we found another strategy to thrive, scoring-type PBSL with weight 4 on successes and -1 on failures. In equilibrium, performance was 10 percentage points higher than performance of individual learners, which is a sizable improvement. Moreover, we found that Rogers' paradox and informational breakdown were prevented when the environment varies very rapidly or when successes or failures were very frequent.

3.4 Discussion

3.4.1 Solutions to Rogers' paradox

Rogers showed that in a simple and general model of the evolution of social learning, in equilibrium, the individuals are not better off than before social learning was introduced [23, 146]. Thus one important question in the study of gene-culture coevolution is whether social learning can nevertheless improve the lot of a population.

Solutions to Rogers' paradox often invoked strategies that relied on mixing both individual and social learning [23, 50, 101, 143]. Using the more realistic social learning model we introduced in the previous chapters, we implemented three incarnations of such strategies. Opportunistic Individual Learners (OIL) use conformism if successful in the last period and individual learning if unsuccessful; Opportunistic Conformists (OC) use individual learning if successful in the last period and conformism if unsuccessful; In Doubt, Conform (IDC) consists of using individual learning if the agent is sure which option is better and conformism if in doubt. Simulations of these learners show that they perform better than individual learners; OIL performs 1 percentage point better, OC 7 percentage points better, and IDC 8 percentage points better than individual learners in monomorphic populations.

In Rogers' classical formulation of the paradox [146], the performance of individual learners is frequency-independent, whereas the performance of social learners drops as their frequency increases. Hence, in equilibrium, the polymorphic population attains the same average performance as a monomorphic population of individual learners. The proposed solution strategies seem to resolve Rogers' paradox because they have a higher performance than individual learners without being dependent on them. However, one has to ask whether pure conformists can invade populations consisting of any of the strategies. Our simulations show that indeed they can. Interestingly, the more frequent conformists become, the worse the solution strategies perform. Therefore, in equilibrium, the polymorphic population performs worse than in absence of conformists.

How can the polymorphic equilibria be characterized? For OIL, the equilibrium frequency of conformists would be slightly above 60% and the average performance 57-58%. For OC, the equilibrium frequency of conformists would be slightly below 50% and the average performance 61-62%. For IDC, the equilibrium frequency of conformists would also be slightly below 50% and the average performance 63-64%. So at least for OC and IDC, we see a small improvement over the 59% average performance of individual learners. However, the question remains whether such a small improvement is really sufficient to conclude that this explains the adaptedness of human culture. We would be more conservative and state that the proposed solution strategies have not (yet) convincingly proven to be solutions to Rogers' paradox.

Why do our results differ from the earlier findings? One reason may be that our model is not exactly the same as earlier models. For instance, we have several periods per generation, allowing more realistic learning algorithms to exist. This should be considered as an advantage, though, and not as a weakness of our approach. Furthermore, as discussed earlier, there was a problem with the implementation of "critical social learners" ([50], corresponding to OIL) and "conditional social learners" ([143], corresponding to OC); they were allowed to learn twice, whereas individual learners and pure social learners were not, explaining the latter two strategies' inferiority. We changed this so that each strategy has an equal learning opportunity and predictably find that OC and OIL are not strictly better than pure social learning. Our model presents a fair ground for all strategies and, therefore, our results should be more believable.

IDC is analogous to strategies studied by Boyd and colleagues [22, 23, 86, 139]. Their implementation is, however, completely different. In essence, they suggest that all individuals rely to some extent on both individual learning and imitation. They receive a private signal from the environment about which option is better and, when the signal is not sufficiently clear, choose to rely on social information instead of relying on this cue. In our model, there is no exogenously generated cue. Instead, individuals have to learn by trial-and-error or by learning socially which option is putatively better. Past observations could be interpreted as a noisy cue to which of the two options is better. The cue in our model is thus generated endogenously and influenced by the behavior of the strategies; its reliability depends on the environmental parameters.

Furthermore, a difference to previous models is that we omitted costs of learning and focused on what performance is reached in equilibrium instead. This could make a difference, although it is unlikely. Finally, and most importantly, we included more learning strategies than the previous models did and checked what state is reached when all the strategies' frequencies settle in equilibrium. Our analysis is thus more comprehensive than previous ones, and it shows how important it is to include a wide range of social learning strategies. As it is impossible to cover all social learning strategies, it is possible that other undiscovered social learning strategies could invalidate our present findings. We would even expect this to be the case. Further studies are thus needed.

3.4.2 Solutions to informational breakdown

Several works have shown [63, 143, 177] that under certain circumstances, individual learners can be pushed out of the population. Although there are some social learning strategies – we call them autonomous – that are able to learn from the environment without the help of individual learners, these strategies may become extinct as well. As the other social learning strategies – we call them dependent – require autonomous strategies to learn about the environment, the extinction of the latter leads to total detachment of the behavior from the environment. We called this phenomenon informational breakdown and showed that it robustly occurs in our model. Informational breakdown reduces performance to chance level and is thus even worse than the equilibrium state described by Rogers.

One possible solution to prevent informational breakdown is to invoke strategies that rely on a mixture of both individual and social learning [143]. However, our simulations revealed that these strategies generally become extinct when competing with the other strategies. Two scoring-type PBSL strategies we studied here are also autonomous and could, in principle, prevent informational breakdown. But these strategies subsided to averaging-type PBSL (PBSL McElreath [120] and PBSL with payoff-conformism trade-off). This second type of social learning is dependent and therefore it is not a solution to the threat of informational breakdown.

Similarly to the conclusion drawn from Rogers' paradox, the conclusion drawn from our findings should not be that human populations are generally in a state of informational breakdown. Instead, Rogers' paradox and informational breakdown provide challenges to too simplistic notions of how social learning works. If a proposed social learning strategy can neither solve Rogers' paradox, nor prevent informational breakdown, there must be something amiss. Or, alternatively, one would have to show why our assumptions are inappropriate.

Our findings could be criticized on some grounds. We chose a population size of 10,000 for our model, since this was suggested to correspond to the effective population size in much of our recent history [83, 167]. Because the population size is finite, very small equilibrium frequencies cannot be realized. Therefore, it could be that when we allow for higher population sizes, our results would change. This is certainly a possibility. Testing it is not easy, since computational constraints set an upper limit on what we can do. We did, however, run 10 simulations in the default condition but with a population size of 25,000 (see Appendix). In this setting, informational breakdown still occurred in all of the simulations. The next chapter will provide more evidence for the robustness of our findings. It is thus unlikely that too small population sizes explain our findings.

Moreover, one could argue that informational breakdown is prevented when there is always a small possibility (e.g. through error) that an individual does choose the minority option. Then there would always be some dissenters who could be copied when their inferior choice becomes superior over time. However, evolution would act to reduce the error probability, since it is not individually optimal to make errors; those individuals who are most accurate would be selected. It does not matter that groups with less accurate individuals are better off in the long run, since individual level selection usually trumps group level selection. There would have to be an external constraint on accuracy to prevent informational breakdown.

Similarly, one could object that informational breakdown would be less frequent if offspring did not inherit their first period's choice from their parents. When they would start fresh with random choice, each option would be represented in the population, making informational breakdown less likely (although it may still occur within the 50 periods that a generation lasts). We would, however, expect such vertical transmission to evolve, since parents are more likely than not to make the right choice. Offspring would face a selective pressure to inherit their parent's choice and not to start by guessing. Furthermore, discrete generations are an artificial assumption made to simplify the model. In reality, there is never a completely new generation, horizontal, vertical, and oblique transmission will always act in concert. This makes the re-boot a purely hypothetical possibility.

3.4.3 Evidence for the use of social learning strategies

In the empirical literature, it is often important to distinguish between different forms of how social learning is actually performed. For our model, this question is mostly irrelevant. Take the example of choosing between hunting two different prey species A and B. An individual could learn socially by imitating others (i.e. faithful copying of means and ends), by emulating others (trying to achieve the same goal), or local enhancement (going where the others go and then act autonomously). The exact form of learning is not instrumental in achieving the final goal, choosing A or B based upon information derived from others. Consequently, we are also not interested in the role of speech for social learning. For our purposes, social learning is defined purely on a functional basis.

We also want to clarify that we will mostly ignore a large part of the literature that studies cumulative cultural learning. Among others, these works study how social transmission shapes the cultural traits themselves, and they are well summarized [124]. As we made clear earlier, there is a distinction between cumulative cultural change and non-cumulative cultural change, the former of which is not our topic. Therefore, there is no point in studying this literature here, despite it being of high general interest.

Furthermore, we are not focused on social learning in animals. For this literature, readers should have a look at the 2004 review by Kevin Laland [108]. Andrew Whiten and colleagues compared different forms of imitative behavior in human children and chimpanzees [178]. These works and references therein should be sufficient to satisfy the curiosity of those interested in social learning in animals.

3.4.3.1 Conformism

Conformism in the sense used in this work requires that an option is more likely to be adopted than by random chance if it is chosen by the majority of the population, and less likely to be adopted if it is chosen by the minority of the population. In an extreme case of conformism, the subject overrides her personal information when others unanimously choose a different option. This was famously tested by Asch in his conformity experiments [4, 5]. In these experiments, subjects had to tell which of three shown lines was the longest. Differences were so extreme that making a wrong call almost never occurred. However, most of the subjects in the experiment were really confederates. In some rounds, the confederates who announced their judgment before the test subject all chose one of the false options. Asch found that in one third of the cases, the test subjects conformed to the wrong judgment. Two thirds of the subjects conformed at least once.

Despite their general interest, these findings are of little value for our research. Test subjects had to make their judgments publicly. Therefore, there was a normative pressure to conform; if one contradicts the previous answers, one would publicly proclaim the previous subjects to be wrong. In fact, when the real test subject was allowed to write down the answer instead of making a public announcement, conformity vanished. This should be expected, as, by design, there was no uncertainty about which answer was correct. However, as we have argued in the first chapter, social learning is most likely to be applied when the goal is to reduce uncertainty about what option is correct.

Another classic study performed by Sherif involved reduction of uncertainty. The test consisted of showing the subjects a light dot and of having them estimate the distance the dot moves [154]. In reality, the dot did not move at all, the perceived movement was due to the autokinetic effect. After one session, subjects received a feedback of the estimates of other group members. In a subsequent session, judgments were more likely to fall within the "norm" established by the group. Questionnaires showed that the subjects' decisions were influenced by concern about how they are perceived by others, meaning that the main objective again was how the subjects would be seen by the others. Therefore, it is important to differentiate between normative and informational social influences [42]. In contrast to Asch and Sherif, we are only interested in the latter.

Later experiments on social learning paid more attention to prevent normative influence from altering the subjects' behavior. Efferson et al. [47] tested individual and social learning involving two options with subjects from the Sama Biological Reserve in Bolivia, who mainly engage in subsistence herding. Subjects could sample from two options with different payoffs and in some treatments were additionally provided with payoff- and frequency-information from other subjects. The best model fit of the data was achieved with individual learning models, payoff- and frequency-information did not seem to have an impact. The authors state that improvements in performance mainly took place within the first 6 or 7 of the 50 periods, which is not too surprising, as the quality of the options did not change over time. Within these few periods, there was some evidence for conformity but there were too few data points to find significant effects.

The same authors later tested the propensity for conformism in students from Zurich [46]. Again, there were two options that, within a session, had a fixed mean payoff. Some subjects only had individual feedback at their disposal, others only frequency-information. Of the latter, 12 self-identified as non-conformists and 28 self-identified as conformists, which the authors found to be validated quite well by the data. The conformists indeed behaved in a conformist fashion, displaying the typical sigmoidal adoption curve. Since these subjects could not respond but to the frequency-information, this experiment cannot tell us anything about the natural occurrence of conformism, though. It only tells us that even when conformism is adaptive and the only sensible learning strategy, one will still find a substantial proportion of subjects who reject conformism. One reason for this could be that the students did not want to appear to be conformists in the eyes of the experimenters.

Goeree and Yariv also tested conformity in a laboratory setting [71]. The basic task was to guess which of two colors was more common by drawing with replacement from an urn. However, the first subject was forced to guess and the two subsequent subjects were also not allowed to draw but could observe the choice of the previous subject(s). Clearly, these three judgments were not informative about the true outcome. Still subsequent subjects, who could choose between drawing from the urn or observing the choice of subjects who did not draw (i.e. uninformative actions) often chose to do the latter. 34% of subjects did in fact choose to observe the uninformative actions; this number even increased to 50% when stakes were raised. Of those subjects, more than 80% followed the majority choice. The later a subject was in sequence, the more likely it was that the subject chose to observe the social information. The authors claim that this puzzling behavior could not be explained by confusion, inequity aversion, or the desire to balance the choices. Instead, the authors concluded that subjects have an intrinsic taste for behaving in a conformist fashion, which might be adaptive in other situations. However, considering how much money was lost due to observing an uninformative signal, one has to consider the possibility that many subjects were confused on a level not detected by the authors.

Corazzini and Greiner conducted a follow-up experiment [37]. Their experiment was very similar to that of Goeree and Yariv but subjects did not have the possibility to draw from the urn. Instead, they had to rely purely on the uninformative social signal. The authors found that there was no sign of conformity. Quite in contrast, subjects were more likely to choose the option less chosen. Even if one option was inherently better, some subjects chose the inferior option. This could be interpreted as non-conformism but is probably rather due to probability matching. Overall, these experiments tell us little about conformism because we deal with models that assume that the social signal does provide information.

Claidière and colleagues studied conformity in a field experiment [34] by hosting a competition in a Scottish zoo involving real prizes. Subjects were asked to submit either a drawing or a text to the competition. The researches found them more likely to adopt an option if it is was displayed more frequently, but the relationship was mostly linear, which contradicts the use of conformity. There was, however, a content bias; even if one of the options was never shown, almost 40% of subjects chose that option. In addition, it is possible that some subjects had the impression that choosing the minority option would enhance their chance of winning. These and other possible confounds make the experiment hard to interpret.

Faria et al. [56] observed people waiting at a street light in Leeds, UK, under real world conditions. They found that when lights were red, pedestrians were 150% to 250% more likely to cross the street if their neighbor crossed first. Males were more likely to initiate imitation. Sometimes, pedestrians who followed others aborted their crossing attempt, which the authors interpret as occurrence of a maladaptive informational cascade.

Haun et al. [84] tested conformism in human 2-year old children, chimpanzees, and orangutans. There were three options to choose from. In the first study, three demonstrators of the same species chose option A and one demonstrator option B; each choice was rewarded. It was found that 72% of the chimpanzees' choices, 56.3% of the children's choices, and 36.1% of the orangutans' choices conformed to the majority. The authors concluded that this shows that chimpanzees and children show “majority-biased transmission”. One has to be careful, though, to not call this conformism. Remember that conformism requires that the probability to choose the majority option has to be greater than its representation in the sample. In this study, 75% of the demonstrators chose A. None of the species responded with choosing A with a higher probability than 75%. The observed behavior is therefore more akin to linear adoption (chimpanzees), total random choice regardless of observation (orangutans), or a mixture (humans).

In another treatment, Haun and colleagues had one demonstrator chose option A three times and one demonstrator option B once. All attempts led to a reward. Chimpanzees and orangutans were not significantly more likely to choose A than B but children were. In fact, the behavior of children and orangutans hardly differed between study 1 and study 2. In our model, due to synchrony of choice and lack of recognizability, it is not possible to observe the same individual three times, so we cannot make predictions whether this should lead to conformism or not. This would certainly be an interesting question for future studies.

3.4.3.2 When to imitate

Kameda and Nakanishi [100] tested social learning with undergrad students from psychology classes of the Hokkaido University, Japan. There were two options, one of which was correct and the other false, but this switched occasionally. After each round, subjects received information about the choice of three other subjects of the pool; additionally, they could purchase individual information, which was accurate 67% of the time. The cost of individual learning was either high or low. There was no immediate feedback about success after each round. Instead, after 5 rounds, a subject’s cumulative payoff, as well as the other group members’ cumulative payoffs, were displayed.

The experiments revealed that for low costs of individual learning, some individuals always and some never learn individually; most individuals learn individually some of the times. For high costs, many individuals never learned individually and few individuals did so more than half of the rounds. This finding is consistent with the basic prediction about the impact of learning costs, namely that higher costs of individual learning lower the reliance on individual learning.

Participants were more likely than not to adopt the majority choice when it opposed their own choice in the previous round, and more so when they did not seek individual information. However, the authors did not show how the probability of choosing an option varies, in general, as a function of the number of demonstrators of that option. Firm conclusions about the degree of conformism can thus not be drawn.

Additionally, one could criticize that in the high cost condition, the cost of learning equaled half the total reward of being right. Considering the rather low accuracy of individual learning, and depending on the frequency of switches in the environment (which was not indicated), one would not even expect subjects to ever learn individually. Instead, it could be possible that never or rarely learning individually and relying on guesses otherwise would have been the best response in absence of social learning.

Overall, most subjects engaged in both individual and social learning, although heterogeneity in behavior increased over time. The authors started their paper with a model of social learning in mind very similar to In Doubt, Conform. But in the experiment, the accuracy of the individual learning signal was held constant. Feedback was only given on occasion. Therefore, it was hard to test whether such a strategy was really employed on an individual level. This question could perhaps be addressed with the gathered data – do subjects engage in more social learning when they performed worse than the other subjects in the pool? – but the data were unfortunately not analyzed in this fashion.

McElreath et al. tested undergrad students from psychology classes of UC Davis, California [122]. In their task, there were two options whose payoffs were subject to variance, so that it was impossible to tell the better option with absolute certainty. Variation in the mean payoffs of the options, as well as changes in the payoff variance, were changed according to condition. In the first experiment, individual learning was tested and a model of exponential discounting handily beat two Bayesian learning models in fitting the data. This confirms our choice for the individual learning strategy.

In the second experiment, subjects could reveal the choice of one group member who played in the same environment. 20 of the 55 subjects never or almost never acquired social information. Moreover, social information use declined over time. Fitting three models, pure individual learning, random copying, and using social information as confirmation, showed that there was no clear winner. The authors conclude, however, that confirmation provides a better fit than random copying, but that overall, social learning was used very rarely.

In the third experiment, subjects were allowed to reveal the choice of all group members (size 4 to 7). Subjects were more likely to acquire social information in this experiment but its usage still declined over time. Random copying achieved the best fit, except when the environment fluctuated and variance was low, in which case conformity fitted best. Unfortunately, neither did the authors show the probability to adopt an option as a function of its frequency in the sample, nor did they analyze whether social information usage depended on previous outcomes. Firm conclusions with regard to social learning usage cannot be drawn.

In a study with 11th grade college students from Cambridge, UK, Mesoudi studied social learning of designing virtual arrow-heads [123]. Subjects mostly had to rely on individual learning but sometimes were allowed to learn socially. If they did, they could reveal the cumulative payoff of other players and then the choice of one of these players. This allowed them to use a strategy such as Imitate The Wealthiest. Interestingly, Mesoudi found that subjects who had low cumulative payoffs shifted their choice more strongly in the direction of the demonstrator than subjects who did better. This is supportive of a strategy that relies more on social learning when not very successful or uncertain about the better option, i.e. of OC or IDC. However, since subjects were forced to use individual or social learning in specific rounds, flexibility in switching was not possible; a strategy like OIL could not have emerged. We thus learn very little about which social learning strategies subjects would engage in were they free to choose.

More support in favor of IDC comes from Morgan et al. [129] with mostly students of the University of St Andrews, Scotland. They found more confident subjects to be less likely to use social information; on a Likert scale, subjects who self-rated most confident in a task were approximately 4 times less likely to use social information than those who self-rated least confident (confidence ratings reflected actual performance). But even the least confident subjects used social information only less than half of the times. Harder tasks led to more reliance on social information. When there was more consensus among the demonstrators, which should correlate with their being correct, the subjects' probability to rely on social information also increased. These findings support the idea that the reliability of a cue is factored in when deciding between individual and social learning.

Williamson and colleagues studied imitation by 3-year old children [180]. Two treatment groups differed in whether the task was easy or hard to solve. Next, a demonstrator showed an effective way to complete the task. The researchers found that children who faced the difficult task but not those who faced the easy task were more likely to engage in the same behavior as the demonstrator. This is consistent with opportunistic social learning.

In the described experiments, children always had prior experience with the task. In a later study, Williamson and Meltzoff introduced a variation of this experiment [179]. Here children either encountered an easy or a difficult task. Groups were further separated into a condition were subjects could either make the experience themselves or saw a demonstrator making the experience. After this phase, there was another demonstration of a novel behavior that led to an easy resolution of the problem. Children who faced the difficult task, even when they had no personal experience with it, were more likely to later imitate the demonstrated behavior that led to an easy solution. This is also consistent with opportunistic social learning.

Moreover, children who could only observe in the first phase, regardless of whether the task was difficult or not, were more likely to imitate the behavior from the second phase than those who had personal experience. This demonstrates a bias in favor of own experience over observed experience. Such a bias is consistent with a meta-analysis [176] that showed that personal experience is weighted more heavily than observations, approximately in a

2:1 fashion

3.4.3.3 Whom to imitate

ITW consists of imitating the individual with the highest aggregate number of successes so far, which could be interpreted as "wealth". Therefore, evidence in favor of ITW would require to show that individuals who somehow display that they were particularly successful in the past are imitated preferentially.

Henrich and Gil-White summarize evidence, mostly from anthropological literature, that skillful individuals usually are especially prestigious and are more likely to be imitated [87]. Chudek et al. [33] tested for such "prestige-bias" in children and found that children are more likely to imitate a model who was previously given more attention. This preference was domain-specific in the sense that prestige was not translated from a food-related task to an artifact-related task and vice versa.

Psychological studies by Hill and colleagues [95] further supported the existence of ITW-like strategies. The researchers presented participants with fictitious persons who could either elicit envy or not. Envy was triggered by signs of wealth and by attractiveness. It was shown that the participants spent more time examining the wealthy and attractive persons, and were later able to remember more details about them. Men and women were found to be equally envious of wealth but women were more envious of looks. These results suggest that special attention is paid to, and more details are remembered about, successful individuals. This in turn would allow to more easily imitate this person, but whether there was an increase of imitative behavior of successful individuals was not tested. Still the shift in attention is compatible with a strategy that relies on imitating the wealthiest.

Mesoudi and O'Brien tested social learning with students of the University of Missouri-Columbia, USA [126]. The task was to build a virtual arrowhead that could be changed along several dimensions. The multiple local optima in design were fixed. In the first treatment, subjects were shown the cumulative scores of six models and could copy one of the models. In total, 182 of 225 choices (81%) resulted in copying the most successful individual. In this treatment, subjects did, however, not have a choice to consult other information like individual experience; they were effectively forced to use this social learning strategy. Hence these results cannot be counted strongly in support of ITW.

In another treatment, subjects could learn individually and additionally could reveal the cumulative score of other subjects in specific rounds. After having revealed the scores, subjects could look at the choice of those other subjects. Overall, 82 of 87 decisions (94%) were preceded by looking at another player's choice. Of those 82 decisions to peek, 68 were directed at the player with the highest cumulative score (83%). Therefore, if subjects did not have to rely on ITW but could use individual learning instead, they mostly chose to acquire social information. Other social learning strategies such as conformism or PBSL were, however, not possible. Interestingly, subjects looked only at an average of 1.5 other choices during 5 rounds, although they could have accessed 1 choice per period. It thus seems that they used social information sparingly. These results were broadly replicated in a follow-up study [123].

Forcing participants to use individual or social learning in particular round made it difficult to tease out what learning strategy was actually preferred by the subjects. To correct this, Mesoudi performed another experiment that used the same setup, designing virtual arrow-heads, but allowed to freely choose the learning strategy [124]. Subjects were undergrad students from Queen Mary University, London, UK. The available learning strategies were individual learning, conformity, random copying, averaging of the choices (not payoffs!) of other group members ("blending transmission" [22], which is possible if the options are not binary, as in our model, but continuous), and copying of the choice of the most successful player (similar to ITW).[1]

The results showed that when subjects engaged in social learning, copying the most successful was most prevalent (84%), with the remaining social learning strategies being approximately equally likely. This result is easy to understand. Random copying is never a good choice when there are better social learning strategies available. Averaging was an inferior strategy by design, since there were multiple optima; when subjects cluster in the different local optima, averaging results in ending in a payoff valley. Conformism was also inferior to "ITW", since the learning process included very little noise, so that the best score was also associated with the best choice most of the time – there was no reason not to rely on this signal.

Most surprisingly perhaps was the finding that social learning was rarely used. Instead, subjects used individual learning in 77.5% of the rounds. This is surprising because higher frequency of social learning was associated with higher payoffs. The problem with this study was, however, the incentive structure. Each participant received 5 pounds as a show-up fee; additionally, if at the end of each of the three sessions a subject performed better than a control group, she earned 1 additional pound. This means that she could earn at best 3 more pounds, or 60% more. Now one has to consider that social learning in this experiment consisted of pressing one button and then the computer would automatically determine the arrow-head design. Individual learning, however, allowed to tweak each of the variables and then look at how this visually affected the virtual arrow-head. Therefore, individual

[1] Mesoudi calls a strategy that relies on this latter type of information "payoff-biased social learning", since it is biased by the cumulative payoffs. In our definition, payoff-biased social learning is instead targeted at copying choices that generated high payoffs specifically in the last period. We call the social learning strategy that looks at cumulative payoffs Imitate The Wealthiest instead.

learning was intrinsically more rewarding and may thus have crowded out social learning. This is a concern that potentially exists with all similar lab experiments.

3.4.3.4 Payoff-bias

Our simulations showed that payoff-biased social learning strategies were superior to other forms of social learning. Over the whole tested parameter range, individual learning, conformism, and the strategies mixing both almost always became extinct. Thus PBSL should always be the first choice if possible. Two obstacles could be in the way of using payoff-information of others to inform one's own choice. First, the payoff-information might simply not be available, depending on the situation. For example, when two restaurants are presented, one chock full of people and the other serving only few customers, one might argue that only frequency-information is available. However, if one could somehow see that all the customers of the full restaurant turn away from their food in disgust, whereas the customers of the other restaurant visibly enjoy their meals, then the valuable payoff-information *is* present. The availability of payoff-information is thus dependent on the specific situation.

Although payoff-information might be available, it could also be possible that human cognition is not able to make use of it. Although this might seem unlikely, there are precedents for this inability. For example, the food choice of Norway rats can be reinforced, and aversion be overridden, if a rat has contact to a model rat that ate the tested food item. Surprisingly, this reinforcement occurs even when the model rat was injected a nauseating substance, presumably indicating that the food item was toxic [66]. Rats thus seem to be incapable of taking the result (becoming sick) of a certain action (eating a food item) into account.

To extend this conclusion to humans would be premature, though. Quite in contrast, tests have been conducted showing that payoff-information is taken into account even by three- to four-year old children. In one test [18], two models (hand puppets) were presented to children. One model called four objects whose names were familiar to the test subject by the correct name, whereas the other model called four other familiar objects by a wrong name. Next, two objects with names unknown to the child were presented. The good performer called one of the objects a fantasy name X, the bad performer called the other object this same name X. The experimenter then asked the child to give her X. Children significantly more often handed over the object that was called X by the good performer. In a second test, the experimenter instead asked the child to give object Y, with Y being an unknown fantasy word. Children more often handed over the object not called X by the good performer. This test shows that even very young children are capable of distinguishing whether a certain action (naming an object) was carried out successfully or not and that they use this information to inform their own decisions. It thus seems unlikely that making use of payoff-information, if present, should be too hard a cognitive task for adults to perform.

Williamson et al. also tested social learning in 3-year old children [180]. As reported earlier, they found that when children had difficulty accomplishing a task, they were more ready to adopt the behavior of a demonstrator. This is consistent with OC. In another experiment, the treatment groups were separated in two: one in which the demonstrator successfully completed the task and one in which the demonstrator failed to accomplish the task. Subjects performed the demonstrated behavior more frequently after observing it but were significantly more likely to imitate in the former, successful case, showing that they take success/payoff into account. This also confirms that already at a young age, children show the potential for PBSL. The authors also show that the children do remember the demonstrated act even if it was unsuccessful, meaning that their choice to not use this act was deliberate and not because the act was absent from their repertoire.

Some experiments have shown that subjects are more likely to adopt an option that has been successful in the last period. Offerman and Sonnemans showed that when undergrad students from the University of Amsterdam, NL, had to make investment decisions that depended on estimating probabilities, subjects took the estimations of other players into account [135]. Unsuccessful players shifted their estimations more strongly, suggesting opportunistic social learning or social learning in case of uncertainty about the better option (OC or IDC). Furthermore, shifts were stronger in the direction of successful players, suggesting a payoff-bias. Similarly, Apesteguia et al. showed that in an oligopoly game, subjects were more likely to adopt an option if the difference between the payoff associated with this option and their own payoff was particularly high [2]. All these results are compatible with social learning strategies we studied and in accordance with previous theoretical findings [151, 152].

Further evidence for the use of PBSL comes from computer experiments with students. In the experimental part of their paper, McElreath et al. [120] present an experimental paradigm for testing social learning in the lab. Subjects were "farmers" and had to repeatedly choose between two "crops". After each period, their yield of the corresponding crop was revealed to the subjects. Furthermore, subjects could inspect choices and yields of other participants. They knew beforehand that one crop would have a higher mean yield but that variance was so high that the yields of both crops could not be distinguished with certainty.

(By the way, the experiment did not include exogenous costs of learning, and payoffs were designed to leave subjects uncertain, thus making social information a factor in reducing uncertainty instead of a means to save the cost of individual learning. In their theoretical analysis, in contrast, they assume that there is a cost of learning and that individual learners know with certainty which option is better. Thus, curiously, their experiment more closely resembles our model than their own model.)

McElreath et al.'s analysis of the participants' choice behavior revealed that the model that best explains the data consists of relying on payoff-information but falling back on frequency-information if payoffs of both crops are similar. They thus conclude that a payoff-biased social learning strategy with a payoff-conformism trade-off is most likely to be used.[2] However, as we saw, there are many ways to implement payoff-biased social learning strategies that also react positively to frequency-information. Their evidence can therefore not be used to distinguish between the different PBSL strategies.

As we noted, one general mechanism to implement a conformist PBSL strategy is to pay more attention to successes than to gains. Interestingly, McElreath et al. observed that in each of the 15 rounds, subjects were a little bit more likely to inspect the yields than to inspect the choice of other participants. This indicates that subjects first determine who the more successful participants are and then inspect those participants' choices, a behavior that would be consistent with paying more attention to successes. Unfortunately, McElreath et al. did not further analyze the behavior – neither the order of information acquisition, nor specific biases towards successes or failures were tested –, so that this possibility remains but speculation.

We found, to our knowledge for the first time, that the mean success rate of the environment is a crucial parameter in determining which social learning strategy prevails. This makes testing for specific social learning strategies in the real world more difficult than it might seem. For example, one prediction that follows from our simulations is that successes should be valued more heavily than failures; there should be an optimist bias. If one were to find, however, that both are valued equally, this would not conclusively refute our findings, because we saw that the more frequent successes are, the more attention should be paid to failures and vice versa. The optimist bias could thus still be present but balanced out by the mean success rate.

If we see that people in general pay more attention to the rarer event, this finding would be trivial, as it results from basic information theory. It would only be interesting if one could show that an optimist bias exists when success and failure are equally likely. The critical control would thus have to be the mean success rate as determined by the environment. Of course, this mean success rate could be manipulated in laboratory experiments, but if priors about the optimal relative weight are "imported" into the lab by drawing on real world experience, this manipulation would not be revealing.

Keeping these restrictions in mind, is there experimental evidence in favor

[2]Note again that the specific strategy analyzed in the theoretical part of their paper, which we call "PBSL McElreath", does not trade off payoff- and frequency-information, but instead uses frequency-information as a tie breaker. The strategy we call "PBSL with payoff-conformism trade-off" does, however, trade off both types of information.

or against the findings that relative weights should be adjusted for rarity and present an optimist bias? The first thing that will probably come to mind are the famous experiments by Kahneman and Tversky about loss aversion [98, 99]. They showed that relative to a subject's current status, losses produce greater disutility than gains of the same magnitude produce utility. For example, if after a fair coin flip, heads would result in losing $100 and tails in gaining $100, most people would decline to participate in such a gamble. A risk premium would have to be paid to make subjects participate. What Kahneman and Tversky found is that typically, gains have to be twice as large as losses to compensate subjects, so only if tails resulted in $200 or more would a typical subject agree to take the gamble.

These results seem to speak against our notion of an optimist bias but a straightforward application to our model is not possible. In Kahneman and Tversky's experiments, subjects responded to *future prospects* that would affect *their own* wealth; in our example, subjects respond to *past experience* that affected the wealth of *other individuals*. There are differences in both the time of realization (*future* vs *past*) and the persons from which experience is drawn (*self* and *others*).

There are more differences that might explain why Kahneman and Tversky's findings seem to contradict our predictions. In the context of individual learning, it has been shown repeatedly that it makes a difference whether subjects read a descriptive stimulus that describes what will happen to these subjects in terms of probability, or whether they instead draw on their own experience before making a decision [91, 92, 172]. For example, when subjects draw cards that yield payoffs from different decks, they tend to underestimate rare events, a finding that could not be attributed to sampling biases. Kahneman and Tversky found the opposite, though, namely that small probabilities are overweighted [99, 171]. In their experiments, subjects did, however, learn probabilities not from experience but from description. Thus the mode of learning completely changes the behavior of human subjects.

The other difference between loss aversion and the optimist bias we conjecture is the distinction between what happens to self and what one observes happen to others. While an equal number of successes and failures may push a person away from repeating a behavior because she is loss averse, observing this same proportion of success and failures in other persons may reinforce the behavior because of the optimist bias.

Bandura performed experiments that shed some light on how payoffs to others influence subjects. In one experiment [9], children that were on average 4-year old were exposed to a movie of a person mistreating a clown doll. Afterwards, the actor was i) rewarded, ii) punished, or iii) no consequences followed (control condition). Next, the children entered a room with the same clown doll and the number of aggressive acts directed towards this doll were measured. Bandura found that children who observed the actor being rewarded did not engage in more aggressive acts than children in the control group, whereas children who observed the punishment significantly reduced their rate of aggression. This experiment can be interpreted as *negative* outcomes (punishment) affecting children more than positive outcomes (rewards). This contradicts our hypothesis that positive outcomes should have a higher impact.

Several caveats exist for this interpretation. First, as we noted, the prior probabilities have to be taken into account. For instance, at the age that the children were tested, it is very likely that they know that aggressive behavior results in punishment. If the actor is not punished, even in absence of a reward, the expectation of a punishment would be contradicted. No punishment could hence be interpreted as a reward. If this were the case, one should not be surprised that the reward and the control condition resulted in similar outcomes.

A second caveat is that strictly speaking, our model only makes predictions if more than one person is observed. So an optimist bias could be expected if, after a child observed two actors being punished and one actor being rewarded for the aggressive behavior, the child would still engage in more aggressive behavior compared to the control condition (the appropriate control would be children who did not observe any actors at all). Again, though, the results would be contingent on unobservable priors. Bandura's experiments thus unfortunately tell us very little about our question, which is simply owing to his experiments being designed with completely different research questions in mind. An experiment that is supposed to test our predictions would have to be designed carefully and have good controls.

3.4.3.5 Social information use on the internet

Laboratory experiments on behavior have some limitations, many of which are summarized by Levitt and List [112]. These authors conclude that field experiments should play a bigger role in experimental economics. Despite the difficulties in executing more naturalistic experiments, we concur with them that many phenomena found in the laboratory may not transfer to the real world.

One possible route to take in order to study social learning in a more "natural" fashion is to look at the internet. Two advantages compared to lab experiments come immediately to mind here, namely that subjects are not supervised and behave according to their "natural" preferences and that the sample size is potentially enormous. In the following, we will show some examples of social information usage on the internet and explain how these examples relate to the studied social learning strategies.

The world wide web lends itself to the use of social information, as connections between different sites and users are its bread and butter. Data about the user behavior is routinely tracked to extract valuable information. Moreover, factors such as teaching or norms have less value on the internet, making it more likely that social information is foremost used for informational and not normative reasons, as is assumed in our model.

An important part of the time on the internet is spent on news sites. These sites routinely provide more articles for the readers than they could possible read; therefore, filters that extract the most interesting articles are often used. Which articles are more or less interesting could be determined by the editors of the journals or by some algorithm that evaluates the content of the article. This would amount to individually considering the relevance of an article as a guide for readers. However, this is not the only route news sites usually take. Additionally, they often harness the power of social information to decide which articles are most relevant for users.

For example, The New York Times, Le Figaro, and Spiegel Online, three popular news sites in their respective countries, provide information about the relevance of news articles based on the behavior of other users (see figure 3.12).[3] All three sites have a section that lists the most read articles. This type of information is akin to frequency-information. It simply reflects which articles have gathered the highest number of views in a certain stretch of time, probably with some discounting. Reading the articles that were most read previously corresponds to a conformist strategy, which adopts the option that has been most frequently adopted by others.

Moreover, two of the three listed sites show information about which articles have been recommended most often by other users. A recommendation can be understood as a positive outcome the recommending person associates with the article. If a reader recommends an article, she thinks that it is well written or relevant to the targeted person. Therefore, recommendations are similar to successful choices in our model. An article that is often recommended is like an option that generated many successes. From the studied strategies, scoring-type PBSL with weights [1/0] is a strategy that chooses the action that generated the most successes in the sample; failures are simply ignored. Similarly, a reader who reads an article because it has been recommended frequently relies on the positive outcome associated with the article but ignores the negative outcomes (for instance, how often the article was *read and not recommended*).

We had a look at the top 10 news sites in the USA according to Nielsen.[4] Sites were ranked according to the number of unique viewers from the US. We were interested in whether these sites provided social information on their main page to help the readers browse the presumably most relevant articles. From the list, two sites had to be excluded because they were listed twice (once as member of a group, once as single sites). From the Tribune

[3]Sources: *www.nytimes.com*, *www.lefigaro.fr*, *www.spiegel.de*. Information retrieved in June 2012.
[4]*blog.nielsen.com/nielsenwire/?p=32201*. Retrieved May 2012.

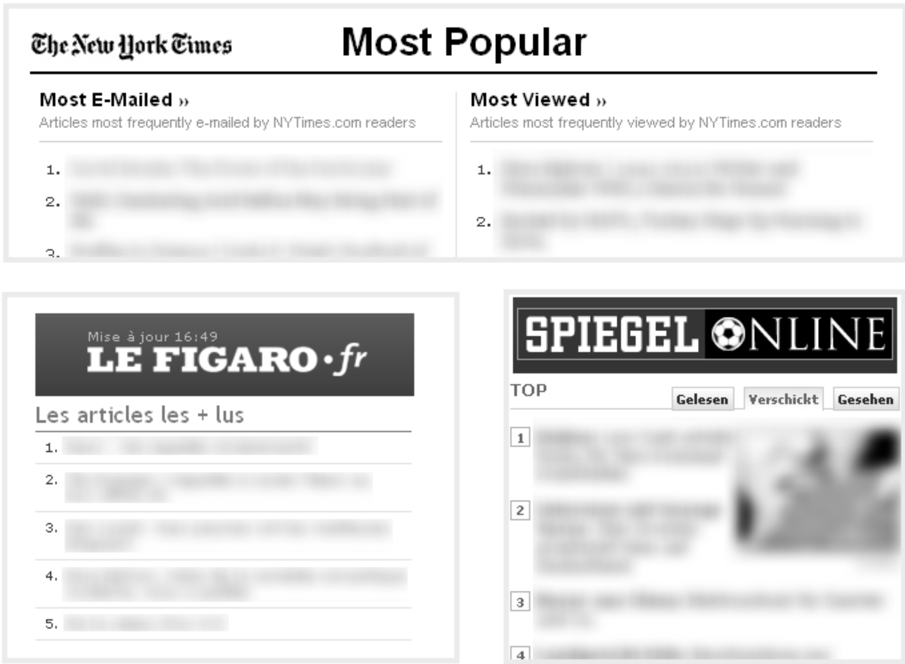

Figure 3.12: Popular news sites from the US (The New York Times), France (Le Figaro), and Germany (Spiegel Online). These sites provide the user means to filter the articles according to how relevant other users deemed the articles. The New York Times and Spiegel Online show which articles have been recommended most frequently ("most e-mailed" and "verschickt"). All three sites show which articles have been most read ("most viewed", "les articles les + lus", "gelesen").

Site	Social information
news.yahoo.com	yes but type unclear ("trending")
edition.cnn.com	yes, frequency
www.huffingtonpost.com	yes, frequency
www.nbcnews.com	no
www.nytimes.com	yes, frequency and payoff
www.chicagotribune.com	no
www.foxnews.com	yes but type unclear ("trending")
www.usatoday.com	yes, frequency

Table 3.1: The most popular news sites among US citizens as of May 2012 according to Nielsen. Six out of the eight sites provide some kind of social information. Frequency-information is the most frequent one.

Company group, we chose, before knowing the outcome, the Chicago Tribune website.

The results from the visited news site are listed in table 3.1. We found six out of eight sites to directly give access to social information on the main page. Frequency-information was made available in four cases, payoff-information in one case. Two sites provided frequency-information but were unclear about the exact type, designating it as "trending". We suspect that this also corresponds to frequency-information but one cannot exclude that payoff-information plays a role here.

It should be noted that even though we searched for explicit provision of social information, that does not mean that when this information is not provided, it is not used. Quite in contrast, most news sites determine the placement of head lines according to algorithms that take popularity and other types of social information into account. Therefore, what one sees on a news site is almost certainly influenced by what other visitors clicked on in the past, and not just by the judgment of the website's editors.

Next, we were interested in how other popular websites make use of social information. Again, we looked at the Nielsen data that listed the websites that had the highest number of unique US visitors as of May 2012.[5] In addition, we consulted Alexa, a web analysis company, which provides a daily update about the most popular sites of its users, who are not exclusively of US origin.[6] The results are shown in table 3.2. In the following, we will show that almost all of these sites make use of social information to improve their content.

Some of the sites do not need further discussion. *www.wikipedia.org* is an online encyclopedia that allows everyone to edit articles or write new ones, and is frequently praised as the prime example of "swarm intelligence" or

[5] *blog.nielsen.com/nielsenwire/?p=32201*
[6] *www.alexa.com/topsites*. Retrieved 8. February 2013.

Rank	Nielsen	Alexa
1	*www.google.com*	*www.google.com*
2	*www.facebook.com*	*www.facebook.com*
3	*www.yahoo.com*	*www.youtube.com*
4	*www.youtube.com*	*www.yahoo.com*
5	*www.hotmail.com*	*www.baidu.com*
6	*www.microsoft.com*	*www.wikipedia.org*
7	*www.aol.com*	*www.live.com*
8	*www.wikipedia.org*	*www.qq.com*
9	*www.amazon.com*	*www.amazon.com*
10	*www.ask.com*	*www.twitter.com*

Table 3.2: Top 10 websites with the highest number of unique viewers originating from the US (Nielsen) or among users of the Alexa web analysis tool. Regarding *www.hotmail.com*, Nielsen apparently pooled hotmail, MSN, WindowsLive, and Bing; we used hotmail as stand in.

"wisdom of the crowd" [166]. The Microsoft services (including Live) and the Yahoo! services also need not be further described. Providing an e-mail service (hotmail, yahoo mail) does not require social information, and we will treat search engines (bing, yahoo search) later. *www.aol.com* is, in essence, also a news portal, but since it provides additional services, it was apparently not listed in Nielsen's top news sites. As most other news sites, AOL provides direct information on its main site about the most popular articles and videos. *www.qq.com* is a Chinese news hub associated with a very popular instant messaging service. Due to the language barrier, we cannot properly evaluate this site.

Other sites are more interesting to discuss. Amazon is an online retailer that has a large inventory of products to sell. It is in this site's interest to provide the potential buyers useful information about which products could be especially relevant to them. This is achieved in several ways (see figure 3.13).[7] When the user types in a product type, a list is displayed which is ordered according to relevance, a criterion that is determined in an unknown way. Second, when the user chooses a product type from a list, it is ordered according to popularity, which allows her to assess frequency-information and thus the use a conformist strategy (this is basically a best-seller list). Last, each product has a five star rating (beginning at 1, with 5 being best), which is shown very prominently. This rating is simply the rounded arithmetic mean of all individual ratings of the product and thus corresponds to the average payoff of the product. A strategy akin to PBSL McElreath, which chooses the option with the highest average payoff, can thus be used based on this information.

[7]Information about *www.amazon.com*. Retrieved February 2013.

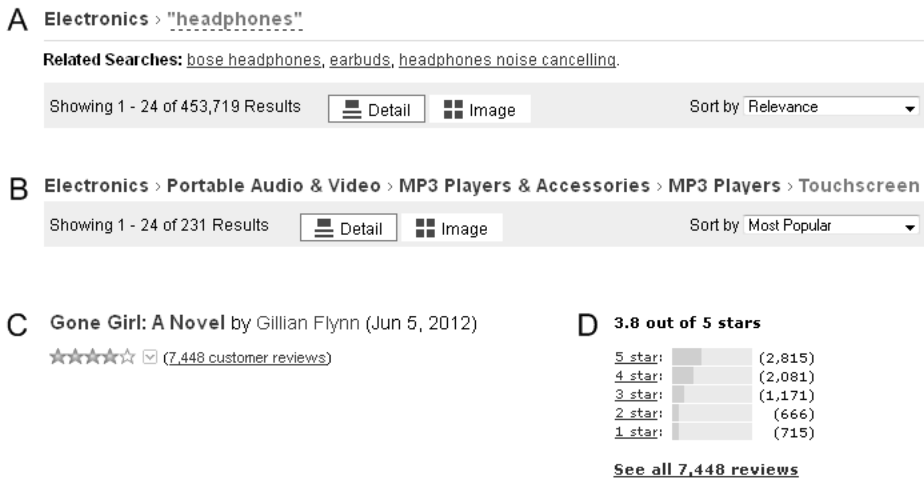

Figure 3.13: Social information displayed by *www.amazon.com*. A: When typing in a specific keyword, results are by default sorted by relevance using an unknown algorithm. B: When browsing by category, results are by default sorted by popularity. C: Each product that has been rated shows the average rating on a scale of 1 to 5 stars. Moreover, the total number of reviews is shown. D: The user may inquire more detailed information about the star ratings, in the form of a histogram (left) and the total numbers (right) of each rating.

The user can access the total number of ratings (frequency-information), as well as how often each rating has been awarded. Therefore, even more complex social learning strategies based on scoring could be employed. Amazon also allows users to write a custom review of the products, which is more akin to teaching or giving advice and thus not very interesting to us. Moreover, amazon uses an algorithm to determine similarities between users, which allows it to suggest products that similar users were interested in in the past. These idiosyncrasies are also not captured by our model.

What can all these findings from amazon teach us? Obviously, amazon is interested in providing information that boosts its sales. Therefore, the information has both to be relevant and it should be considered by the user as authoritative (whether consciously or not). It is unlikely that a company that intends to maximize revenue would provide useless information. Since amazon displays both frequency and payoff-information, the users probably make use of both types. Only amazon knows the weights that are attached to these information sources, so we are left guessing whether its customers use strategies that are similar to our social learning strategies. Popularity and average rating being the most salient forms of information suggests, though, that conformism, PBSL McElreath, or a mixture (PBSL with payoff-conformism trade-off) are mostly used. Although amazon also provides information that would allow the use of scoring-type PBSL, this information is less salient, suggesting that scoring-type PBSL is less relevant.

In an experiment with Taiwanese students, Chen tried to evaluate how different social cues affect decision making in a setup that mimics amazon's bookstore [32]. Subjects were confronted with a hypothetical decision to buy books using an interface that is similar to amazon's. Chen found out that in a binary choice, a book with a higher star rating that is otherwise similar to another book is considered more likely to be purchased. However, the quantitative difference in star ratings (2 vs 4 stars difference) did not significantly affect the likelihood to buy. This indicates that subjects use payoff-bias to inform their choice, but that the influence behaves in a step-like fashion, corresponding to how we modeled social learning and contradicting other models were choice scales with difference in payoff [151, 152].

Furthermore, Chen found out that higher sales volumes also resulted in a higher likelihood of subjects indicating a willingness to purchase a book. This finding indicates that frequency-information alters consumer decisions, which is more or less common knowledge. The researcher also tested how recommendations affect the subjects' choices. It was found that an expert recommendation improves potential sales compared to no recommendation, but that the recommendation from other consumers improves potential sales even more. With care, this could be interpreted as weighting the positive outcome of many similar individuals – payoff-biased social learning – more heavily than the positive outcome of a single, albeit knowledgeable individual – imitating the wealthiest.

Youtube is the largest video hosting platform on the internet.[8] Similarly to amazon customers, youtube users have access to much more content than they could possibly consume. Videos on the front page of youtube are chosen according to some relevance algorithm. This algorithm probably compares past behavior of the user and, through comparison with other similar users, tries to predict what the current user might be interested in. Thus social information is used here but cannot be assessed directly by the user.

Upon entering a video query, a list of possible results is displayed. Here too, relevance is determined by some internal algorithm, which takes into account several signals. The user has the possibility, though, to change the filter to rank the results, among others, according to popularity and average rating.

The most salient forms of social information shown to the user is the number of times a video has been seen (see figure 3.14A). It is thus very likely that frequency-information is the most relevant for youtube's users. Moreover, a rating is accessible, but only after a video has been selected (see figure 3.14B). This rating is more coarse than amazon's, only allowing for positive (thumbs up) or negative (thumbs down) ratings. The total numbers of positive and negative ratings are shown, as well as a bar that reflects the proportion of positive to negative ratings. The total numbers could be used for a scoring-type PBSL strategy, whereas the bar could be used for an averaging-type PBSL strategy. As always, we do not know how the users actually use these types of information, but the mere fact that they are present suggests that they are relevant for the user. In contrast to amazon, youtube makes the frequency-information more salient than the payoff-information, which is also less detailed on youtube.

Is there a certain reason why amazon and youtube differ in how they place the weights? First of all, amazon is probably reluctant to fully disclose the sales volume of each individual product, so that frequency-information has to be ordinal (ranked) and not cardinal (total number of sales). This reduces the information contained in the frequency-information and makes it less informative. On the other hand, the rating system of youtube is more coarse, maybe because a fine-grained rating of a video is harder than of a typical amazon product, making payoff-information less informative. This could explain the difference in information salience.

Alternatively, the relevance of payoff-information could be deteriorated when ratings are too similar. Personal experience suggests that rated youtube videos have almost always an overwhelmingly positive rating, so that this rating cannot be used to distinguish most video clips. However, most rated amazon products also have overwhelmingly positive ratings, meaning that the same argument could apply. But since amazon's ratings are more fine-grained, differentiation is still more readily possible.

[8]Information about *www.youtube.com* as of February 2013.

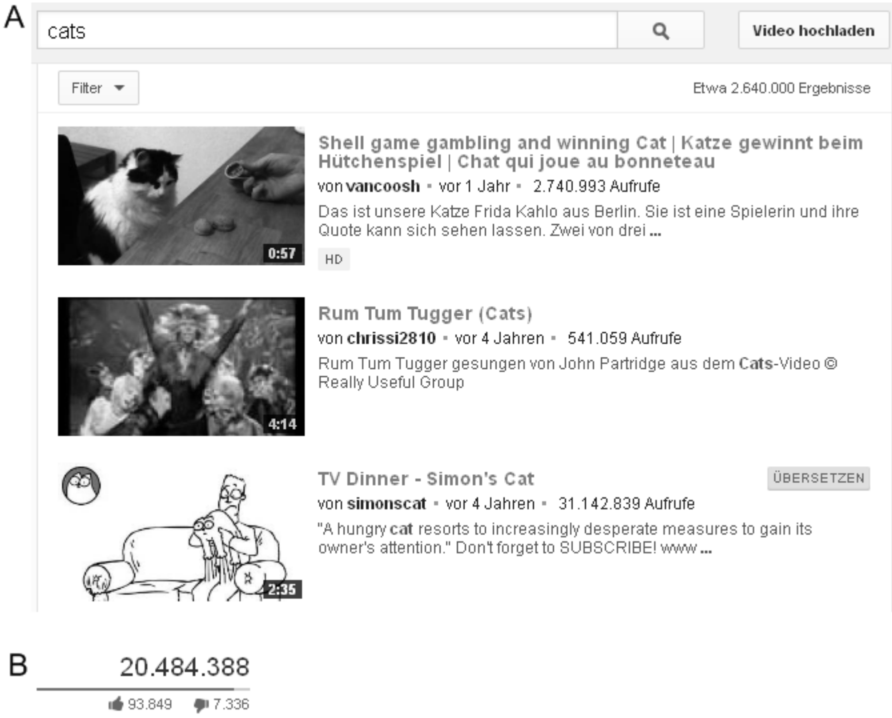

Figure 3.14: *www.youtube.com* provides two direct types of social information. A: The most salient type is frequency-information, which is shown after a query has been entered ("Aufrufe", meaning the number of views). B: Additionally, information about the rating is shown, but only after a video has been chosen by the user.

Facebook is a social network and therefore massively makes use of social information. For example, the users are ranked according to how relevant facebook deems them to be. This ranking is similar to the PageRank system designed by google, which we will explain later. Here we want to focus on a different aspect of social information use, though. Facebook allows its users to "Like" certain web pages, comments, etc. (see figure 3.15A). A "Like" is essentially a display of approval, a positive rating. The total number of Likes is then displayed. There is no frequency-information, as e.g. the number of Likes per visitor of a page. Neither is negative information shown. This very sparse type of information can only be used to perform a scoring-type PBSL strategy with weights [1/0]. Remember that this strategy only takes into account positive outcomes, ignoring negative outcomes. We saw that this strategy performs reasonably well under some conditions.

Why does facebook restrict the amount of given information to merely positive outcomes? One reason is certainly to not offend its users. When a user posts a comment that receives a lot of negative ratings, the user may be upset and quit using facebook. Facebook is not anonymous, so users would

Figure 3.15: A: The facebook “Like” button allows users to indicate that they like a certain web page, comment, picture, etc. The total number of “Likes”, here 929, is usually shown. B: For each user, twitter shows the number of tweets, the number of people being followed by that user (“folgt”), and, most importantly, the number of followers of that user. Information as of February 2013.

probably take negative ratings personally. In contrast, users of youtube and amazon typically do not know each other and therefore, there is less risk of being offended by a negative rating.

There could be another reason why facebook restricts the displayed social information to only positive outcomes. The highest share of Likes falls on comments, photos, web pages, etc. that have very few comments. This is linked to the very personal nature of facebook – there are simply not that many persons who are interested in a picture of your dog in the first place (personal observation). As we found out earlier, with small sample sizes, PBSL with weights [1/0] actually performs quite well, better than, say, PBSL with weights $[4/-1]$. Although PBSL McElreath also performs quite well with small sample sizes, giving an average of three ratings would look ridiculous. The sparse use of social learning on facebook may thus serve a purpose.

www.twitter.com is a micro-blogging service. Its foremost use is to write and read “tweets”, messages of no more than 140 characters. To sort through the mass of possible people to follow, twitter indicates for each user the number of followers she already has (see figure 3.15B). This allows the reader to quickly find the most popular twitter users, thus engaging in a conformist strategy.

In the lists of the most popular websites, we have several search engines, namely *www.ask.com*, *www.bing.com*, *www.yahoo.com*, *www.baidu.com*, and *www.google.com*. According to both lists, google is the most popular site (although, according to Nielsen, US citizens spend four times as much time on facebook as on google). We will focus on google search, which is by far the most frequently used search engine and also revolutionized web search in general.

Before google entered the market, search engines evaluated the relevance of a website – and thus its position after a search query – by trying to assess its content [113]. For example, if a user searched the word “dog”, a page that contains a lot of instances of this word or whose title contains the word would have a good shot at a top spot. This system was unsatisfying, though, since meaning was, and is, notably difficult to extract from a site. Also, the system could easily be gamed; just having the word "dog" appear a hundred times on a site, even if it did not deal with dogs, made it rise in rank.

The google search engine took a different approach to rank websites [113, 138, 166]. Put simply, a website received a higher score (the "PageRank") when many other websites linked to it. This is similar to the frequency-information in our model. The advantage of this system was that google search did not need to evaluate itself how good the content of a website was. If other people found the website relevant, they would be more likely to link to this site. Thus google made use of the individual experience of other people to improve its own results. This is social learning at its finest.

In reality, the PageRank is a little bit more complex than just described. Not all links are treated equally. Instead, links originating from websites that were themselves often linked to have a higher weight [113, 138, 166]. For example, many websites link to the homepage of the New York Times. Therefore, if the New York Times links to another website, that site's PageRank increases more than if just some minor website linked to it. Although the rule is recursive, google founders Page and Brin developed an algorithm that allowed to calculate a steady state, even if billions of links are included [138]. The PageRank improved the search results so considerably that despite existing competition, google quickly became the most popular search engine [113]. Other search engines nowadays use similar algorithms as google.

In our social learning model, there could, theoretically, be a social learning strategy that works similarly to the PageRank. Until now, all sampled individuals are treated the same. However, one could design a strategy that ranks individuals according to how often they have been copied, and then imitates the individual who has been copied most. In the next step, it could place different weights according to how often the copiers have been copied, and so on and on. The problem is, however, that it is very unlikely that this type of second order social information is available in most real world situations, not even mentioning higher orders. Not only is this information hard to come by, after some recursions, it is likely that the signal is drowned in noise, making it uninformative. The structure of the web allows algorithms like the PageRank to work; in the real world, things are unfortunately not that easy.

As in social learning models, the internet runs the risk of relying too much on imitation. Are websites often linked to because they are relevant or because they figure prominently in google search? Are popular products on amazon also the best products, or do users just buy products because other users did before them? The same dynamics that we describe in our models are possibly governing interactions on the web. It is not hard to imagine that online systems are prone to the same inefficiencies caused by Rogers' paradox as are social learning models.

3.4.3.6 Reinterpretation of experimental results in the light of social learning

Many experiments and observations indicate that social information is an important source for decision making. However, most experimental designs test subjects under the assumption that social learning plays no role. For many experiments, this will be true, but some experiments may be affected by social learning without the researchers being aware of it. This may be especially true for experimental economics, were several players interact and feedback about the other players' choices and payoffs is often provided. (Reviews of some of the most important findings from experimental economics can be found in the literature [28, 57, 65].)

In the Public Goods Game (PGG), subjects typically play anonymously with the same group, most of the times consisting of four players total, for several rounds. At the beginning of each round, they receive an endowment of, say, 20 monetary units (m.u.), of which they can contribute any amount to the public good. The total contribution is multiplied by, say, 1.6, and then redistributed equally among the players. Therefore, the group can maximize the reward by contributing everything but it is individually rational to contribute nothing. Over time, after players adjust, typically they notice this and lower their contributions to the lowest possible level.

Furthermore, subjects in these experiments are often allowed to punish other players; this punishment is costly for the punisher but even more costly for the punished player. As punishing is costly, rational choice would dictate to never punish, so no player should be afraid of being punished and the end result should be that nobody punishes and nobody contributes. The actual outcomes are different, though. Some players do punish, targeting especially those who contribute little, which results in contributions rising. In contrast to what should be expected, the possibility to punish increases the level of cooperation in the experiments.

Several explanations have been forwarded to explain this deviation from expectations. Subjects could act rationally but their utility could depend on the utilities of the other group members, as e.g. proposed by the inequity aversion model [60]. Alternatively, people may be motivated by reciprocity [53]. Although these kind of theories do take into account how other players behave, they do not account for social learning. They focus on the outcomes of actions, what payoffs they generate or what choice sets they produce, but they do not allow for the possibility that players instead prefer to imitate the actions of others.

We want to illustrate how social learning may affect outcomes in the PGG. In the experiment just described, subjects in the non-punishment condition are informed about the total contribution after each period. Say that player 1 contributes 20 m.u. to the public good and player 2 contributes nothing; player 3 and 4 contribute 10 m.u. each. After this decision, the total contribution is 40 m.u., resulting in the public good growing to 64 m.u., which amounts to 16 m.u. being awarded to each player. Player 1 now has 16 m.u. total, player 2 has 36 m.u. total, player 3 and 4 have 26 m.u.

Imagine that the players use some form of opportunistic social learning. Player 1 has made a loss compared to the group and is therefore likely to adopt another choice. For example, the average contribution of the other players was 20/3, so player 1 may contribute 7 in the next period. Player 2, however, has made a gain and will not change her strategy. Similarly, player 3 and 4 received exactly the average payoff and thus do not change their behavior.

In the next period, player 1 contributes 7 m.u., player 2 nothing, the other players stay at 10 m.u. Total contributions are 27 m.u., which amounts to each player receiving 11 m.u., so that player 1 has made 24 m.u., player 2 has made 31 m.u., and players 3 and 4 have made 21 m.u. each. Now those latter two players fared worse and could therefore imitate the other players' choices, which again would lower total contributions. This behavior would finally mirror the downwards trend in contributions for the non-punishment part of the experiment.

In the punishment part, players are informed about each single contribution before they can make their choice to punish [59]. They also are informed about how much they were punished but do not see the payoffs after punishment. Fehr and Gächter observed that those who contributed less to the public good than the average were most likely to be punished. Say that we see the same initial contributions as before, but now player 2 gets punished for defecting. Therefore, she receives a lower payoff and, instead of being imitated, may thus herself imitate the choice of other players. As the players cannot directly observe how costly punishing affects payoffs, they may not notice the opportunity of second-order free-riding (contributing to avoid punishment but not punishing others). This dynamic could thus lead to the observed upwards-trend in the punishment condition.

Of course, the explanation based on social learning is totally ad hoc. Although it could not explain some of the observations (for instance why there is punishment of free-riders in the first place, or why there are differences between cultural groups), it could still plausibly replicate these and other findings. Furthermore, social learning makes some unique predictions. For example, the way that feedback is given should influence the outcome. If, e.g., players would be informed about the decisions and cumulative payoffs of others, they should be more likely to imitate the most successful player in these types of experiments.

One could design an experiment where imitating the most successful is neither consistent with selfish payoff-maximization, nor with fairness concerns or reciprocity. This would provide valuable insights about the true reasons for the players' behavior. Another prediction derived from social learning is that in more complex tasks, subjects should rely more on copying. A social learning model should be developed that can accommodate the experiment so as to make precise predictions. Our model, for instance, only works for binary payoffs and would have to be modified to make sensible predictions.

The PGG is not the only game that could be affected by social learning. It should be interesting to develop a social learning model that is independent from specific game structure, makes predictions about the applied strategies, and then to test those predictions against the existing predictions, or to re-analyze existing data from the new point of view. Since we know that people learn socially when they are allowed to and when it makes sense, we find it unlikely that social learning has no influence on economic experiments.

3.4.3.7 Conclusion

There are several experiments that may help to elucidate which forms of social learning are used by humans. We found evidence for conformism, for imitation when uncertain or unsuccessful previously, imitation of the most successful, and payoff-bias. In short, we found evidence for almost all of the studied social learning strategies. This could mean that there is a lot of heterogeneity, either between individuals or between tasks (or both). The most consistent evidence in line with the studied social learning strategies is in favor of opportunistic social learning, i.e. social learning in case of uncertainty about the best decision.

Many experiments were tested with a specific social learning strategy in mind. Other social learning strategies could not be performed owing to restrictions in the experimental setup, or the authors just did not test their data for the possibility of these strategies. Especially payoff-bias has received too little attention. Our findings should help researchers to be more open-minded about different forms of social learning. An experiment specifically designed to test our hypotheses should be open towards which information the subjects use and then see what develops naturally. The experiment that came closest to this ideal was performed by McElreath and colleagues [120].

In general, the laboratory experiments showed high variability in subject behavior, as well as under-usage of social information most of the time (over-usage was rare). Often, random copying was observed when conformity could have been possible. One reason for these surprising findings may be that the experiments were not sufficiently incentivized. Also, most test subjects were "W.E.I.R.D."[9] students. Perhaps, students do not want to make the impression to follow others instead of thinking of their own, especially if the experimenters are their own professors. This may explain why they rely less on social learning than theory predicts.

For all these reasons, more naturalistic settings should be preferred. It is known that under naturalistic settings, people are very sensitive to social

[9] They come from "Western, educated, industrialized, rich, and democratic" societies, which are argued to be the least representative of humanities past [88, 89].

cues; the presence of images of eyes increases honesty [13] and charitable giving [140]. Field experiments such as the one about imitation in street-crossing behavior [56] are a good start. These experiments are of course harder to control but have higher external validity. We think that the indirect evidence gleaned from social information usage on the internet is also an indicator for its widespread use. It would be even more desirable to get direct access to databases such as those from amazon.com to study when which social information affects whom. Those data are hard to come by, though, because of privacy concerns and because companies are reluctant to share data with their competitors. We nevertheless believe that this avenue of research is promising.

Appendix

Evolutionary simulations

To check our results for robustness, we repeated the evolutionary simulations while changing some of the parameters.

Population size of 25,000

Usually, we use a population size of 10,000. However, there is a decent amount of stochasticity in the results. Therefore, it is cautious to repeat the simulations with greater population size, as this reduces the amount of stochasticity [136]. We simulated the default conditions again, this time with 25,000 individuals. The result is shown in figure 3.16. Only three strategies are shown to be able to resist extinction at least some of the time. Conformists became extinct in only 2 of the 10 simulations but their average frequency was below 20% most of the time. PBSL McElreath could persist in half of the simulations but if they did, their frequency also was only 20% on average. PBSL with payoff-conformism trade-off was the most successful, never becoming extinct and attaining frequencies of 70-80%.

When comparing these results with the simulations with a population size of 10,000, the outcome is mostly the same. With 10,000 individuals, ITW and PBSL with weights $[4/-1]$ managed to not become extinct both exactly once, but apart from this, the results are consistent. Most notably, we observe in both conditions PBSL with payoff-conformism trade-off to be the predominant strategy.

Reversion factor

The environment in our model is heavily influenced by the reversion factor. The higher it is, the more the environment tends to revert to the mean. This entails that the stretches of periods in which one option remains consistently better than the other are shorter for higher reversion factors. To see how this factor influences the outcome, we changed it from the default value of 1 to 3. The results are shown in figure 3.17. In contrast to the default condition, PBSL with weights $[4/-1]$ is the dominant strategy, with conformism and PBSL with payoff-conformism trade-off being able to persist at low frequencies on average. Higher reversion factors seem to favor PBSL with weights $[4/-1]$.

To check this in more detail, we increased the reversion factor even more to 5. The result can be seen in figure 3.18. Again, PBSL with weights $[4/-1]$ are most dominant, while PBSL with payoff-conformism trade-off is able to persist at a low frequency on average. This time, conformists always become extinct.

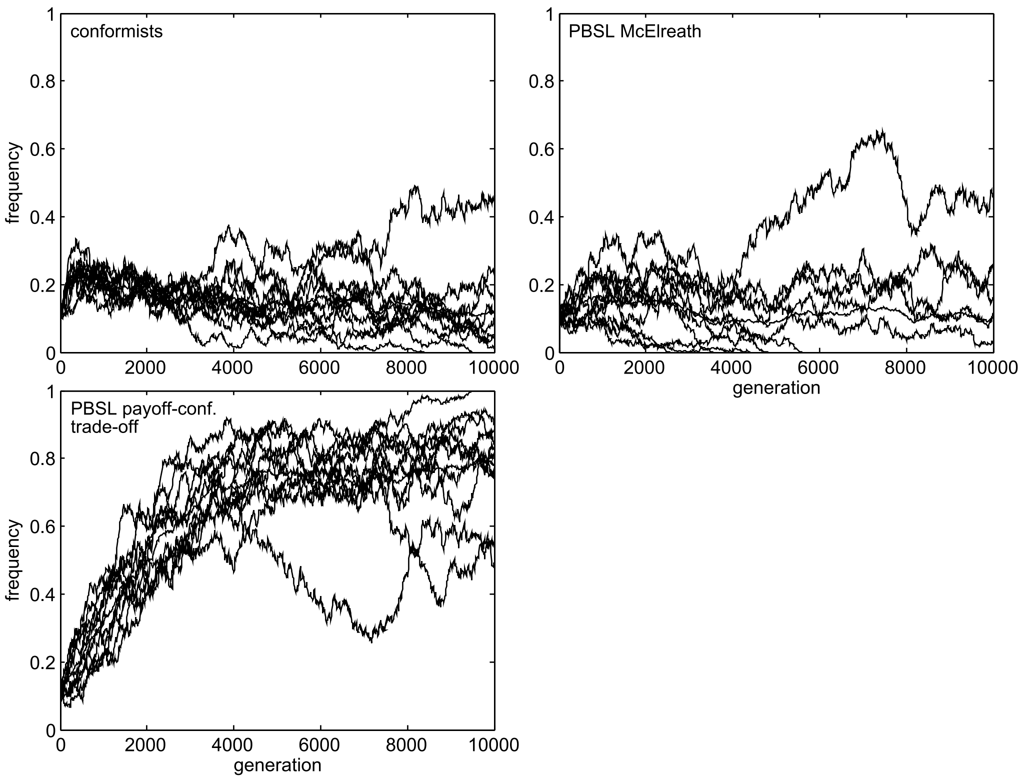

Figure 3.16: Evolutionary simulations with the default conditions but with a population size of 25,000. As always, thin lines show the individual frequencies of the ten simulations and the bold line the average frequency. The only strategies that do not always become extinct are conformism, PBSL McElreath, and PBSL with payoff-conformism trade-off. The latter is always dominant.

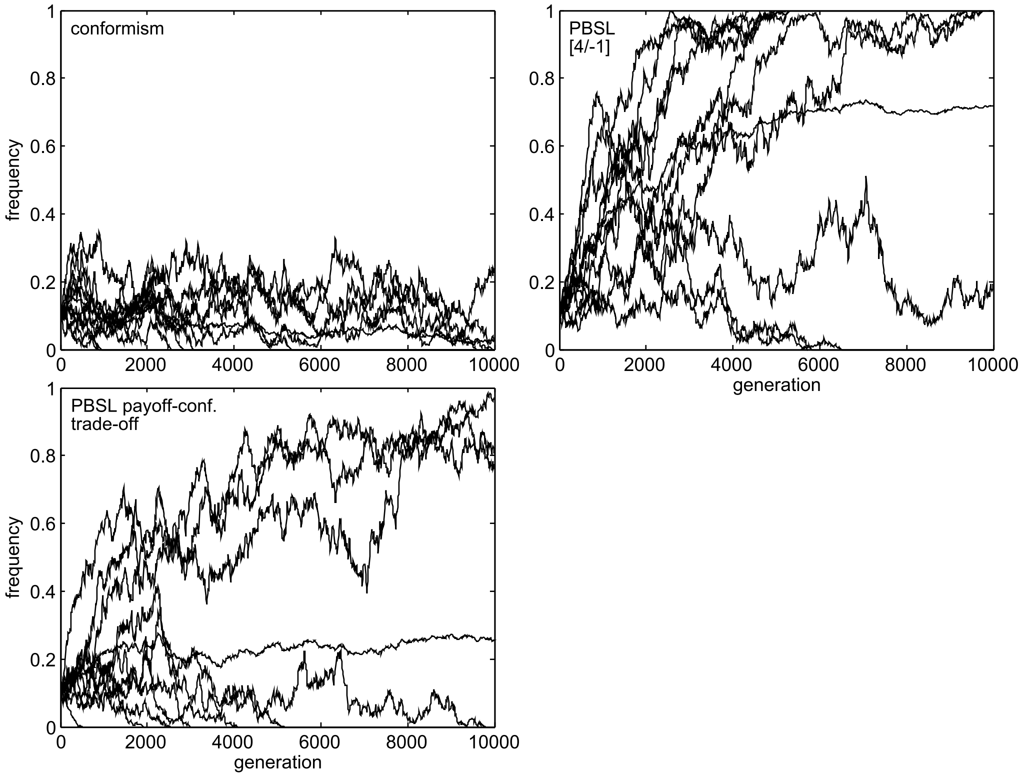

Figure 3.17: Evolutionary simulations with the default conditions but with a reversion factor of 3. The only strategies that do not always become extinct are conformism, PBSL with weights $[4/-1]$, and PBSL with payoff-conformism trade-off. The second strategy is the most dominant.

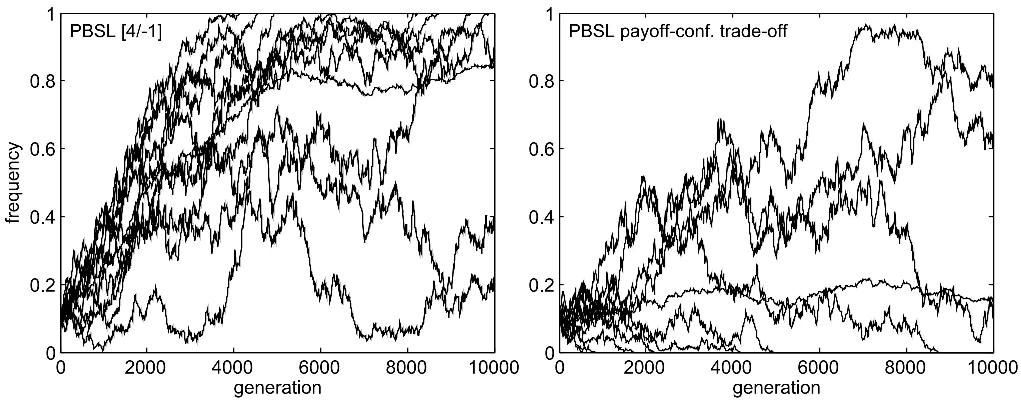

Figure 3.18: Evolutionary simulations with the default conditions but with a reversion factor of 5. The only strategies that do not always become extinct are PBSL with weights $[4/-1]$ and PBSL with payoff-conformism trade-off. The former is the most dominant.

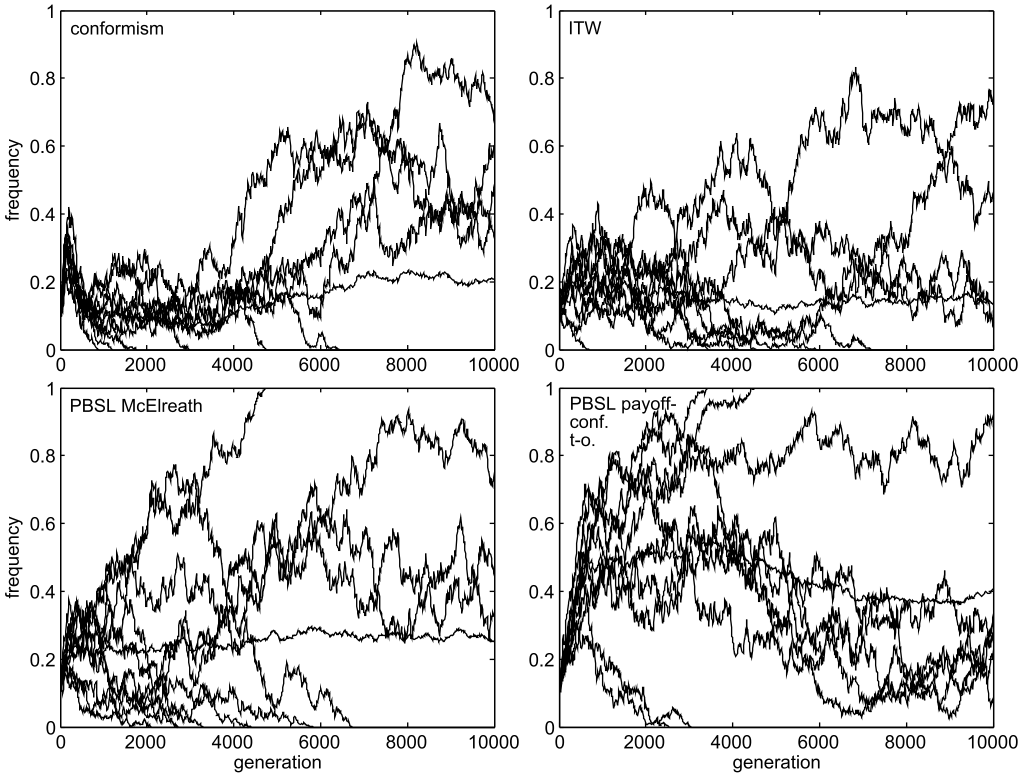

Figure 3.19: Evolutionary simulations with the default conditions but with a mean success rate of 0.25. The only strategies that do not always become extinct are conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off.

PBSL with weights $[4/-1]$ is an autonomous strategy. Therefore, with higher reversion factor, it is possible to reliably circumvent informational breakdown.

Mean success rate of the environment

In the last chapter, we saw that the mean success rate of the environment was the parameter that had the largest impact on the performance of PBSL strategies. To see whether the mean success rate has a substantial impact in the evolutionary simulations, we repeated the simulations with different success rates. First, we decreased the mean success rate from 0.5 to 0.25. The outcome is shown in figure 3.19. Strategies that survive at least in some of the 10 simulations are conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off. The only qualitative difference to the default condition is that ITW manages to not become extinct in 4 of the 10 simulations, instead of 1 of 10 simulations.

Next we performed 10 simulations with a mean success rate of 0.75 (figure 3.20). Strategies that survive at least some of the time are ITW, PBSL with

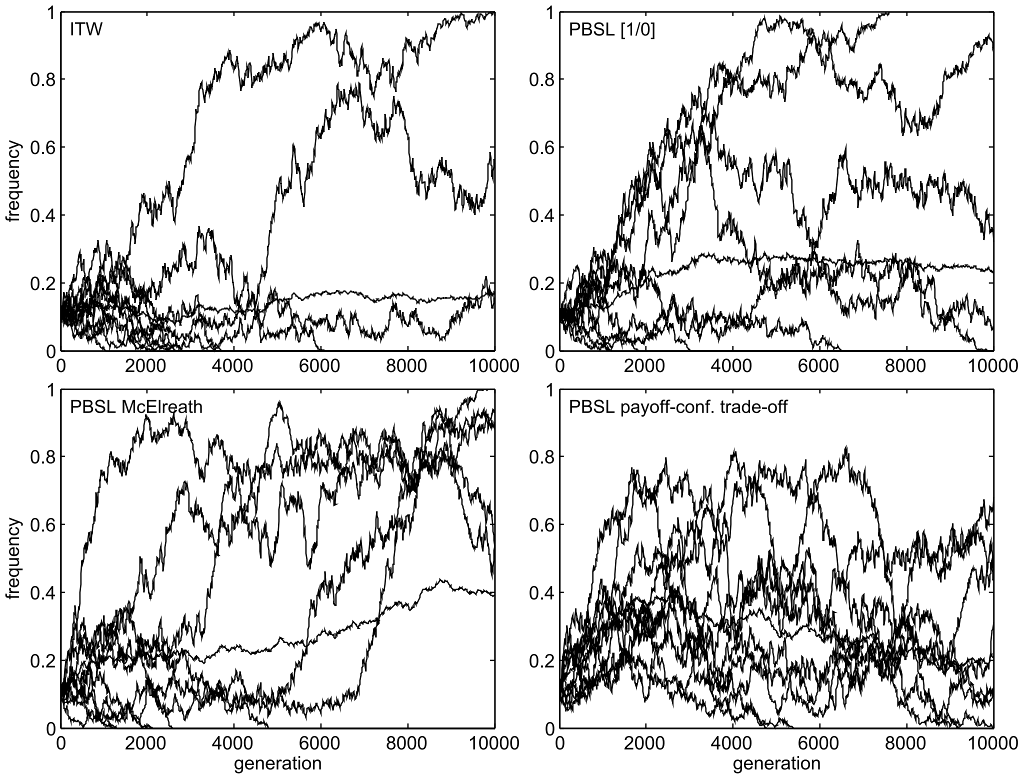

Figure 3.20: Evolutionary simulations with the default conditions but with a mean success rate of 0.75. The only strategies that do not always become extinct are ITW, PBSL with weights [1/0], PBSL McElreath, and PBSL with payoff-conformism trade-off. The latter strategy is the most dominant.

weights [1/0], PBSL McElreath, and PBSL with payoff-conformism trade-off. The only qualitative difference is the occasional presence of PBSL with weights [1/0]; this strategy persisted in 4 of the 10 simulations, compared to none in the default condition. Note that PBSL with weights [1/0] is autonomous, meaning that when it is present, there is no informational breakdown.

Lower probability of an environmental change

In the default condition, the environment changes after each period, albeit only at small steps. We changed this so that each environmental component, p_A and p_B, only change with a probability of 25% after each period (each independently). The result is shown in figure 3.21. The strategies that did not always become extinct were conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off. These are the exact same strategies as in the default condition. The probability of environmental change has thus no large impact on the outcome.

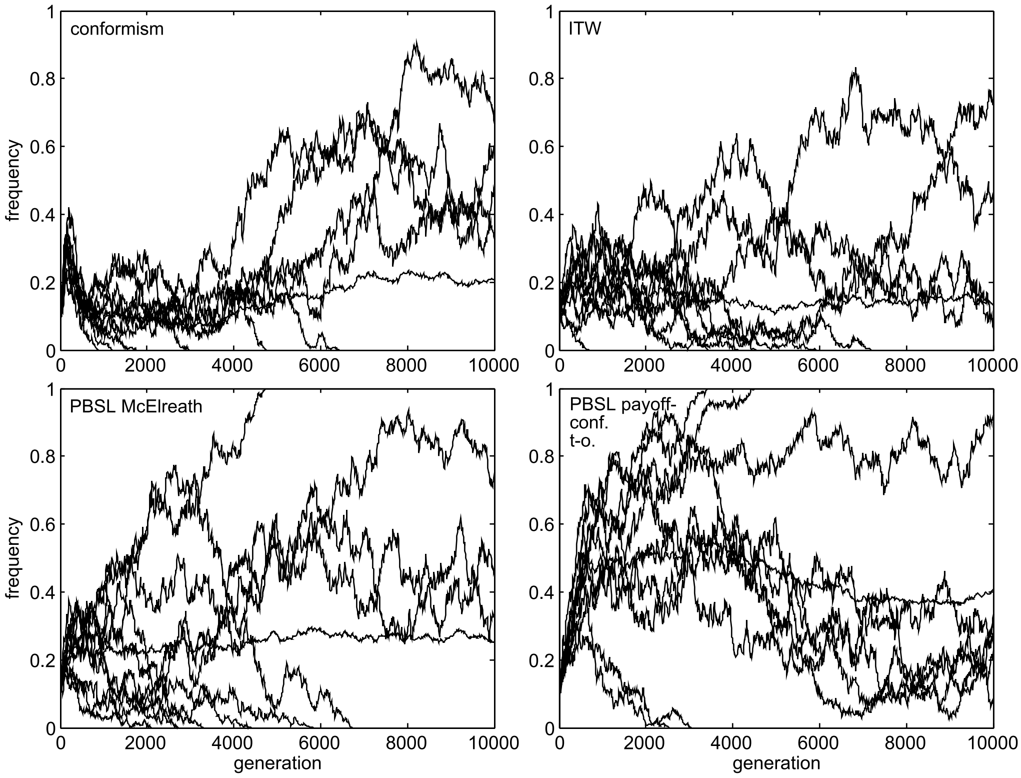

Figure 3.21: Evolutionary simulations with the default conditions but with a probability of only 0.25 for the environment to change at all after each period. The only strategies that do not always become extinct are conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off. The results are very close to those found using the default condition.

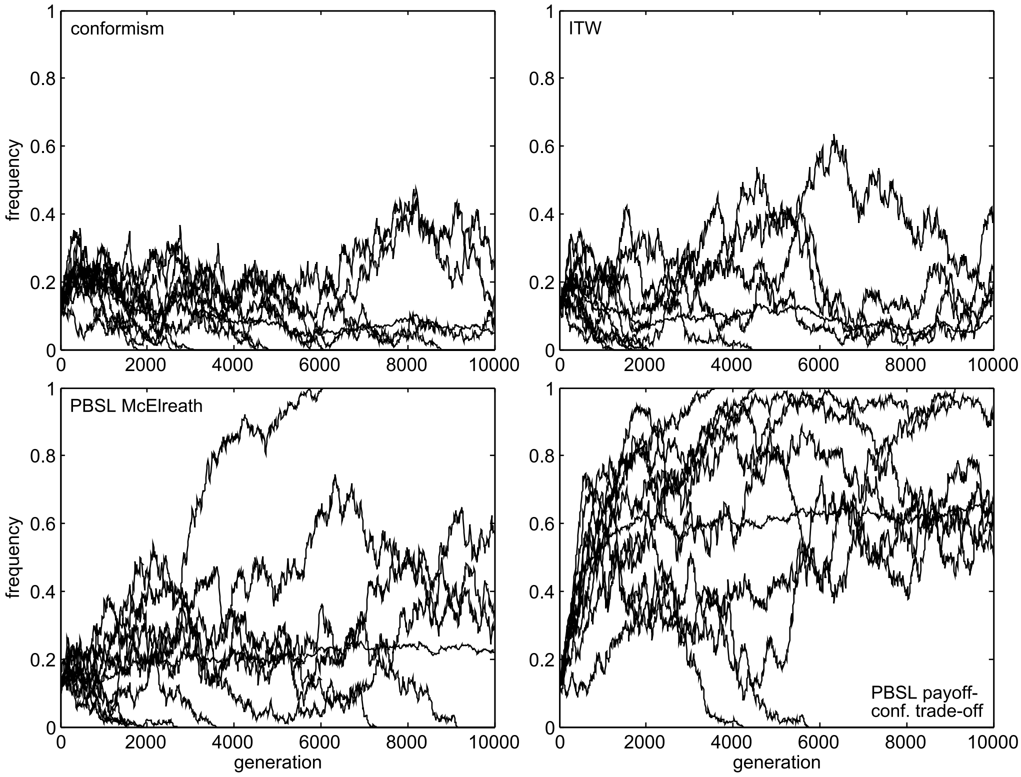

Figure 3.22: Evolutionary simulations with the default conditions but without scoring-type PBSL strategies. The only strategies that do not always become extinct are conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off. The results are very similar to those found using the default condition.

There are other parameters that affect the environment, but our analyses in the last chapter suggested that they do not have any meaningful impact on the results. Therefore, we did not test them here.

Simulations without scoring-type PBSL

We wanted to test how the outcomes are affected if we exclude the scoring-type PBSL strategies from the contest, namely PBSL with weights $[4/-1]$ and PBSL with weights $[1/0]$. The result is shown in figure 3.22. The only strategies that did not always become extinct are conformism, ITW, PBSL McElreath, and PBSL with payoff-conformism trade-off, with the latter being most dominant. This closely mirrors the results found for the default condition. These results were to be expected, as scoring-type PBSL strategies did not play a huge role in the default conditions anyway, so that there absence should not affect the results.

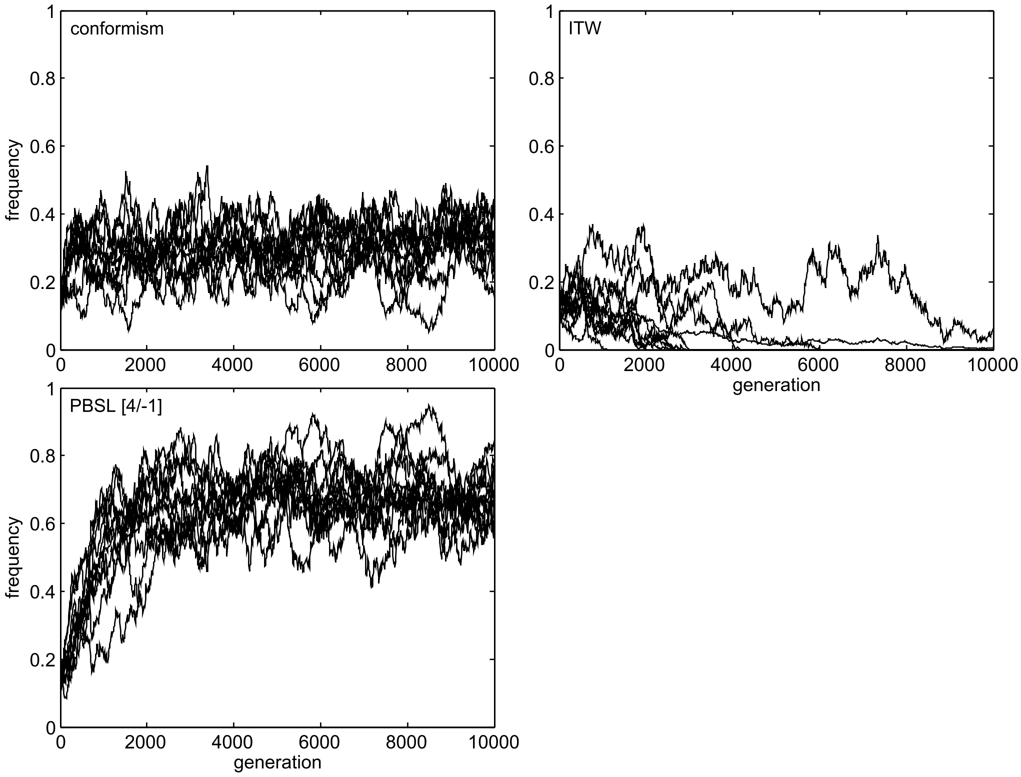

Figure 3.23: Evolutionary simulations with the default conditions but without averaging-type PBSL strategies. The only strategies that do not always become extinct are conformism, ITW, PBSL with weights $[4/-1]$. The latter is the most dominant.

Simulations without averaging-type PBSL

Furthermore, we tested the outcome if averaging-type PBSL in the form of PBSL McElreath and PBSL with payoff-conformism trade-off were excluded. The evolutionary outcome can be seen in figure 3.23. The strategies that did not always become extinct are conformism, ITW, and PBSL with weights $[4/-1]$. Among those, ITW managed to survive but in one simulation. PBSL with weights $[4/-1]$ and conformism, in contrast, seem to reach a stable equilibrium, with the former fluctuating around 70% of the population and the latter around 30% frequency.

Conclusion

We tested our results for robustness by changing the population size, the most influential environmental parameters, and by excluding some of the PBSL strategies. We were especially interested in whether in any of the conditions, one or more of the proposed solution strategies to Rogers' paradox

(OIL, OC, or IDC) were capable of establishing themselves in the population. We found, however, that these strategies always became extinct. Even pure conformism, which is supposed to be inferior to the solution strategies, managed to stabilize in some the conditions. Our line of argument to attack the proposed solutions, namely that they are outcompeted by other strategies, is thus robustly supported.

4 Consequences of social learning for society

Disclaimer: Many parts of this chapter are based on an article written in close collaboration with Ole Jann from the University of Copenhagen Department of Economics and Prof. Peter Hammerstein. The article is currently in submission (April 2013).

4.1 Introduction

In the last chapter, we addressed Rogers' paradox and how it could be resolved. Rogers' paradox mainly deals with whether culture is an evolutionary product that is adaptive for humans. Therefore, it could be argued that although the study of social learning is interesting from an academic standpoint, it has no bearing on our modern societies. Here we show why this is not the case, and why the study of social learning could be more important than it has ever been.

4.1.1 Social learning is ubiquitous

There is general consensus that our behavior is largely shaped by traits that were acquired through social learning (see e.g. [26, 144]). A simple introspection should make clear that almost every routine that we have is at least partly influenced by incorporating information derived from others. For example, the way we dress, the things we eat and how we prepare them, the manners we have, the skills we learn at work – all these traits would be hugely different if we had to learn them just by ourselves.

Among the things we learn socially, a large fraction is probably not learned through teaching [109]. Hewlett et al. [94] observed that among Aka and Bofi hunter-gatherers, although teaching was present, imitation was used more frequently. However, it is difficult to quantify how many traits are really learned through teaching and how many by unguided observation. It is a central finding from psychologists, though, that many decisions we make are not based on conscious reasoning, even if we believe otherwise [12, 132]. Therefore, we should assume that our decisions are heavily biased by social information even when we are not aware of it.

It is thus a safe assumption that human behavior is broadly influenced by social information that we obtained without direct guidance or teaching.

Of course, among those influences, many will be of the normative kind [42], which possibly follow a pattern that is different from the social influences we have described in the previous chapters. Still, not all behavior is normative, especially in modern Western societies. This is because these societies are generally more progressive and liberal, which de-emphasizes the weight of many norms [78], and more market-oriented, which emphasizes the importance of efficiency. For better or worse, an investment banker is more interested in revenues than in pleasing the public.

Since so many of our behaviors are influenced by social information, understanding its impact on behavior is of great importance. For example, it has been argued that the pattern of how innovations are adopted in modern societies is best explained by assuming that people learn from and influence one another [147]. Although it is safe to assume that culturally acquired traits are adaptive more often than not – if they were not, why would one have the ability to acquire them – they certainly produce maladaptations that would not occur without social learning [144]. For example, as Jared Diamond describes [43], the inhabitants of the Eastern Island completely deforested their island; a lot of the wood was required for practices such as cremation of the dead or production and transport of the elaborate Moai statues. Those practices have no direct fitness enhancing effect but contributed to the total deforestation of the Eastern Island, which in turn was a major reason for the collapse of the Eastern Island society.

4.1.2 Social learning may take different forms

Even in our modern societies, the actions we take are influenced by social learning. This leads to behavioral patterns that could not be explained by pure individual learning. Due to the findings of our last chapters, we also have a good idea about which social learning strategies are especially likely to play an important role. We will pick two strategies in particular, Imitate The Wealthiest (ITW) and payoff-biased social learning with payoff-conformity trade-off ("PC-PBSL" for short), that were particularly successful. PBSL was shown to dominate under almost all conditions, so it is the natural candidate for closer scrutiny. Yet it is cognitively and informationally very demanding, as it requires observing the short-term outcome of each sampled individual, calculating the average payoff of the options, and weighing this average against the relative frequency of the options. Therefore, we also included ITW in our analysis, as this is a rather simple heuristic.

Is it reasonable to assume that nowadays, imitation of the wealthiest is a practiced form of social learning? Wealthy individuals are likely to have made good decisions during their lifetime and therefore to be especially prestigious. There is good reason to believe that prestigious individuals should be imitated preferentially [87]. Furthermore, wealth is an especially important variable in social learning in modern societies, which have greater wealth inequalities than small scale societies (although the Industrial Revolution has attenuated some of the inequalities [35]).

If we look, e.g., at consumption behavior, our choices are arguably influenced by the consumption behavior of others, especially those who are wealthy. We may imitate the neighbor with the big house, the colleague with the expensive car, or the famous TV actor who displays his lush life style. Consumption behavior is, however, more dependent on total lifetime income (i.e. wealth) than on short term flows of income (i.e. payoff), as people tend to smooth out their income over their lifetime by accounting for expected future incomes [64]. Wealth is thus more readily reflected in consumption behavior, which in turn serves as an indicator of wealth. We therefore believe that in today's world, imitating the wealthiest is especially likely to be used as a social learning strategy.

In addition, we use payoff-biased social learning with payoff-conformity trade-off (PC-PBSL) in this chapter as a candidate social learning strategy. Remember that PC-PBSL consists of first calculating the average payoff generated by the two options as observed in the sample. These two averages are compared and the option with the higher average is chosen. When the averages are very close, however, the option that was more common in the sample is chosen. If, e.g., A is observed four times and was successful three of the times, while B was observed once and was successful, the average payoff of A is less than the payoff of B but PC-PBSL would still choose the more frequent A. The magnitude of the trade-off, as well as the optimal sample size, were chosen by us so as to maximize performance of PC-PBSL in the default condition.

In the last chapter, we showed some evidence that indicated that averaging-type PBSL is used as a social learning strategy but strong evidence is still lacking. Our main reason to use this strategy is rather that it performed best in all our simulations and across most conditions. It is robustly the best social learning strategy we encountered so far and should thus be studied this chapter.

PC-PBSL is on the other end of the spectrum of complexity when compared to ITW. The description of the decision mechanism should make it clear that this strategy is cognitively more taxing than ITW. While ITW requires only to copy the sampled individual with the highest wealth, PC-PBSL requires to observe the payoff, which is harder to observe, it requires to calculate averages, to determine the most frequent option, and to make a trade-off between payoff- and frequency-information. An additional reason why we chose these two strategies is thus that the one is more successful in evolutionary simulations, while the other is cognitively more simple. Hopefully, other social learning strategies are somewhere within this spectrum.

ITW and PC-PBSL are dependent strategies, as we showed in the previous chapter. Therefore, we have to assume that there is also an autonomous strategy in the population. For this role, we chose the trusty individual learners, the anti-thesis to the social learning strategies. Our procedure will thus be to start with a population consisting to some degree of individual learners and to some degree of either of the two social learning strategies. Then we simulate the behavior of the strategies and perform several analyses of their aggregate behaviors.

4.1.3 Social learning affects population-wide behavior

The presence of social learning should obviously have implications for the aggregate behavior of the strategies. Social learning, especially when having a conformist bias, should for instance increase variation in cultural traits between populations. Assume that one such cultural trait is whether an individual is cooperative or defective in a cooperation game, e.g. the public goods game. More variation through social learning means that there will be more groups consisting of mostly cooperators or of mostly defectors, despite migration between the groups. Therefore, the forces of group selection could act more strongly on such groups than would be the case in absence of cultural transmission. Group selection, which favors cooperation, could thus be stronger than individual level selection, which favors defection. Therefore, social learning could be a possible mechanism that allows cooperation to stabilize. This has been studied earlier works [21, 22, 25].

Social learning could also lead to a delay of the population's ability to quickly respond to environmental changes [146]. In Rogers' model, social learners have the advantage of saving the cost of learning individually but bear the cost of lagging behind the environment. For example, if a social learner imitates another social learner who imitated an individual learner, the information is already two generations old. This deficit is accentuated the more social learners there are in the population. Social learners will thus often find themselves in a position of using a cultural trait that is not adaptive anymore.

Individual learners will generally stick closely to the environment. In contrast, social learners show more variance in behavior and react to environmental changes with a delay. Combine these two patterns and one will often find the behavior of social learners to have little to no relationship with the environment. It will seem that, at least transiently, social learners live in their own world, which is completely detached from reality. Of course, this cannot last forever, since then individual learners would outperform social learners and become more frequent. There is some danger, however, that this transient detachment is grave enough to provoke a collapse of the population [177].

In this chapter, we thus want to study in detail how pervasively social learning affects these three aspects of the population-wide behavior. That is, what are the consequences social learning has for:

1. the variance in adoption of cultural traits,

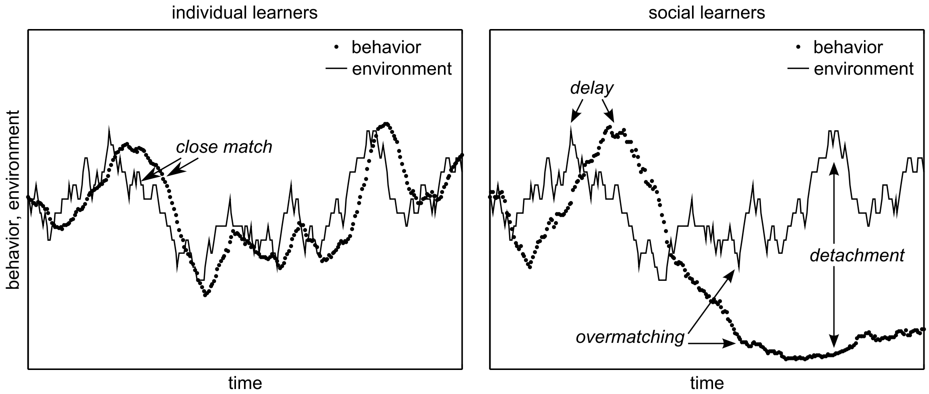

Figure 4.1: Comparison of the behavior (•) of a population of individual learners (left panel) and social learners in the form of ITW (right panel) as a function of the environment (solid line). The behavior of individual learners closely matches the environment. The behavior of social learners is characterized by higher variance caused by *overmatching*, by *delay* in response time, and by transient *detachment* from environmental changes.

2. the delay in response time to environmental changes, and
3. following from this, the transient detachment of behavior from reality.

In figure 4.1, we show an example of how individual learners (left panel) and social learners (ITW, right panel) differ in these aspects. While the behavior of individual learners (•) closely matches changes in the environment (solid line), the behavior of social learners does not. Instead, it shows more variance, caused by overmatching changes in the environment, it lags behind, and it can for some time become totally detached from the environment. We will perform several types of data analyses to study those points.

4.1.4 Social learning affects market participants

We have stressed that social learning for informational reasons should have a huge influence on decision making in modern societies, maybe more so than in former times. The importance of knowing how exactly social learning influences information processing should thus be obvious. Besides evolution and anthropology, economics is particularly interested in phenomena created by social learning. Hayek argued [85] that however beautiful and intricate macroeconomic models are, their use is limited because they assume that all necessary information is available to all participants in a society at all times. Although this assumption has been justified by assuming that learning will eventually lead to a behavior that is optimal in the aggregate, biases introduced by social learning have not been taken into account sufficiently.

In this chapter, we will have a more direct glance at how social learning affects aggregate behavior. In particular, we will be interested in how aspects of behavior that are important for markets and for societies as a whole are influenced.

As we have argued, there is strong evidence in favor of the ubiquity of social learning. For this, we do not even have to assume that our social learning propensities are shaped by natural selection, as we did in the previous chapters. Markets produce their own kind of selection [1]. For example, a bad portfolio managing strategy will sooner or later leave the market, either because the managers run out of capital or because they shift their strategy. Although this kind of evolution is not natural but cultural, it should display similar features [68, 81, 127, 175]. Therefore, we need not bother with what the exact origin of the studied social learning strategies is, as long as the more adaptive ones are finally selected.

Moreover, there is a normative argument in favor of the use of social learning by market participants [150]. As we have seen previously, adaptive social learning is characterized by a strong degree of conformity. When we look at the incentive structure of, say, investment bankers, they should also be expected to be conformist. If they imitate other market participants and their strategy goes bust, they cannot be blamed, since they did what almost everyone else did. If they follow a renegade strategy and go bust, this will result in their losing their job. And even if an investor foresees a market to go bust, it is almost impossible to predict the exact date; she might therefore be better off to be bullish, as non-conformist behavior is likely to get her fired before the crash manifests.

Conformist tendencies have presumably contributed to past bubbles [30, 117, 165, 168], from tulip mania in 1637 to the subprime crisis of 2008, and will certainly continue to do so for future bubbles. Bubbles should, however, not exist, given one of the most famous hypotheses in economics, the efficient market hypothesis. This hypothesis comes in three forms [54, 55]:

- weak form – it is impossible to consistently beat the market by predicting it's behavior,
- semi-strong form – publicly available information is instantly reflected in security prices,
- strong form – public and insider information is instantly reflected in security prices.

The semi-strong and strong form entail the weak form. If the present information is accurately reflected in the prices, it is impossible to predict how the prices will move; the only factors that will influence prices are those that are not expected, so it is impossible to predict them. Therefore, it should be impossible to beat the market. The weak form is agnostic with regard to the precise reasons why it is impossible to beat the market in a consistent fashion (i.e. excluding luck). It may be possible that this form of the efficient market hypothesis holds true but not for the reason that prices always accurately reflect the fundamental value (which is implied by the word "efficient"). We will discuss this last possibility at the end of this chapter.

There are other reasons why markets could fail to be efficient, e.g. reasons related to costs of acquiring information and to capital constraints. In our model, costs and constraints on behavior play no role, though, so we do not bother with them here.

Our understanding of social learning suggests that it may inhibit markets from being efficient. For instance, the semi-strong and the strong form of the efficient market hypothesis postulate that (public) information is integrated into security prices almost instantly. Yet the simplest social learning models predict there to be a delay in information integration [146]. Social learning by market participants is, however, not irrational but would rather be expected in equilibrium [76, 77]. So even if everyone behaved rationally, in a world of uncertainty, the semi-strong and the strong form of the efficient market hypothesis would be invalid.

There is some evidence that markets are indeed not efficient. Most famously, Shiller showed that there is excess volatility in markets [156, 158]. It was also found that stock prices tend to underreact to single new bits of information and to overreact to series of new bits of information [11, and references therein]. Furthermore, stock prices were found to have a high correlation to their fundamental value, but only in the long and not in the short run [157]. Finally, the history of financial markets is a history of bubbles and busts [30, 117, 155, 165, 168], which are hard to consolidate with the efficient market hypothesis. We will show how all these failures of market efficiency can at least be partly explained by social learning, making social learning a natural and parsimonious explanation for market inefficiencies. The empirical finding that markets are often inefficient is thus indirect evidence for the wide-spread use of social learning by market participants.

Although we use our social learning model to understand markets, we do not explicitly model markets. Markets are more complicated than the simple binary choice task we model here, and integrating a market structure into the model framework would bring about many difficulties. A major difference between markets and our model is that in markets, the behavior of the individuals affects the "environment" through the price. On the one hand, if many social learners copy one another and buy the same stock, the price for the stock should rise and demand fall, thus providing a negative feedback loop. On the other hand, if the stock price rises, this could create a self-fulfilling prophecy, as the higher price justifies the initial investment; more imitators would jump on the bandwagon, creating a positive feedback loop. It is not clear which of the two loops would prevail in what situations, but it certainly would be an interesting topic to study in the future.

Another difference to real markets are capital constraints. An investor who has been proficient in the past should have more money at her disposal and therefore a stronger influence on the market. At the other end of the spectrum, there could be unsophisticated small investors with little money who buy and sell more or less randomly, so-called noise traders [157–160]. In our model, each individual has the same overall impact on the aggregate behavior, although of course different strategies, through their frequency, can have substantially different impacts.

Despite the differences between our model and real markets, we believe that our work allows us to gain insights into market behavior. This insight should be seen more as a proof of principle. Behavior that may seem puzzling when one expects efficient markets and rational choice may be easily understood if one considers social learning to play a role. We will show that social learning can lead to behavioral patterns that deviate from efficiency in a way that is consistent with empirical findings about market inefficiencies but still cannot be easily exploited and thus corrected. Whether social learning is really the cause of inefficiences in real markets has yet to be shown, but we would argue that it should be considered a possibility.

4.2 Model description

The model we use for this chapter is the same we used in the previous chapter. No additional changes were made, so everything that was true previously is still true now. The only difference to the previous chapters is that we use the behavioral data generated by the strategies and scrutinize them in more detail. Behavioral data in the sense used here is how the strategies react to changes in the environment. More precisely, we monitored the proportion of A choices over time and made connections to the environment, which is characterized by p_A and p_B. The analyses we perform on the data will be explained in the next section.

The strategies that participate in this chapter are individual learning, Imitate The Wealthiest (ITW), and payoff-biased social learning with payoff-conformism trade-off (PC-PBSL). Individual learning consists of reinforcement learning by exponential discounting (discount factor of 0.9, as always). ITW consists of imitating the individual with the highest fitness among a sample of 7 individuals. PC-PBSL consists of comparing the average payoffs as observed in a sample of 6, and choosing the option with the highest average payoff; if averages are too close, the more frequent of the strategies is chosen instead. Thus all these strategies function exactly as they did in the previous chapter.

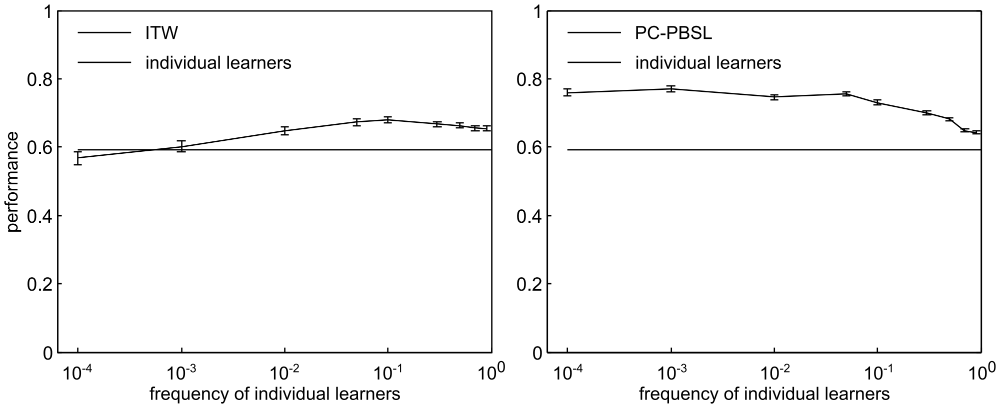

Figure 4.2: Performance of social learning strategies. Left: Performance of ITW competing with individual learners as a function of the frequency of individual learners (S.E.M. for ITW indicated). The performance of ITW is consistently higher than individual learners, except for very low frequencies of individual learners. Equilibrium frequency of individual learners is expected to lie between 10^{-3} and 10^{-4}. Right: Performance of PC-PBSL competing with individual learners as a function of the frequency of individual learners (S.E.M. for PC-PBSL indicated). PC-PBSL consistently performs better than individual learners. The equilibrium frequency of individual learners is less than 10^{-4}.

4.3 Results

4.3.1 Evolution

Before we make our analysis of the strategies' behavior, we first re-establish that social learning is expected to be the dominant form of learning. In the last chapter, we already deduced this from the simulations containing 10 different strategies. Here we compare the performance of individual learners competing either with ITW or with PC-PBSL (figure 4.2); the other strategies were not included. For this, we fixed the frequencies of social and individual learners and compared the performance as derived from 1000 simulations. On the left panel, individual learners compete with ITW. For almost the whole range of tested frequencies, ITW is superior in performance (note the log-scaled x-axis). Only for very low frequencies of individual learners do we find the performance of ITW to drop below the performance of individual learners. The equilibrium frequency of ITW can be inferred to lie between 0.999 and 0.9999.

For PC-PBSL competing with individual learners, we see a similar picture. However, even for a frequency of individual learners of 1 in 10,000, PC-PBSL still outperforms them. To gauge the equilibrium more precisely, we increased the frequency of PC-PBSL beyond the default population size of 10,000. Even if there is only 1 individual learner in a population of 500,000 did we find a high performance of PC-PBSL, reaching 75.43% (±0.92%,

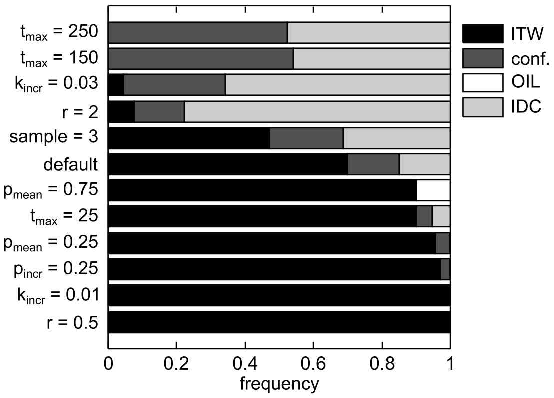

Figure 4.3: Mean frequencies of the different strategies for different conditions. Results are sorted according to the total frequency of ITW. ITW (black) reached the highest frequency on average, IDC (light gray) the second highest, conformism (dark gray) the third highest, and OIL the fourth highest. OC and individual learning always became extinct. ITW performs well except when the environment changes too quickly or when generations are too long.

S.E.M.). We know, however, that without individual learners, PC-PBSL would only perform at chance level. Therefore, their equilibrium frequency should be between 0.999998 and 1.

In the last chapter, we have already established that PC-PBSL is the dominant form of social learning and that this finding is robust to almost all changes in the parameter settings. We have further shown that in absence of PBSL strategies, ITW is the dominant form of social learning but we did not test this finding for robustness. Therefore, we checked the robustness of ITW when competing with the 5 other non-PBSL strategies: individual learning, conformism, Opportunistic Individual Learning (OIL), Opportunistic Conformism (OC), and In Doubt, Conform (IDC). 12 parameter setting were simulated 10 times for 20,000 generations. The mean frequencies at the end of the simulations are shown in figure 4.3. The parameters we tested were: generation lengths (t_{max}) of 25, 150, and 250 (default: 50); mean success rate of the environment (p_{mean}) of 0.25 and 0.75 (default: 0.5); step size of environmental change (k_{incr}) of 0.01 and 0.03 (default: 0.02); probability of environmental change (p_{incr}) of 0.25 (default: 1); reversion factor (r) of 0.5 and 2 (default: 1); reduction of the sample size of ITW to 3 (default: 7).

The first finding is that OC and individual learning became extinct in each of the 120 individual simulations. OIL reached an average of 0.1 for p_{mean} = 0.75 but became extinct otherwise. Conformism was the third most successful strategy, reaching a frequency of 0.166 on average. IDC was second most successful, reaching a frequency of 0.240 on average. ITW was the most successful, reaching a frequency of 0.585 on average. In 7 of the 12 conditions, it reached a frequency of 70% or higher.

ITW performed worse when generation length t_{max} was set to high values. The reason is that when generations last very long, a high total wealth is less likely to be informative about current behavior than it is to be about past behavior. Imitating the currently wealthiest individual is not necessarily the best strategy anymore. Furthermore, ITW's performance declined for a high reversion factor and step size. Remember that these two conditions are associated with more rapidly changing environments. ITW does not cope that well with this faster pace. Otherwise, ITW proved to reach high frequencies across conditions.

In conclusion, both social learning strategies that we study in this chapter, ITW and PC-PBSL, display good and robust performance across most conditions. They are capable of attaining extremely high equilibrium frequencies when competing with individual learners. In practical terms, with finite populations, this could well lead to informational breakdown, as discussed in the previous chapter. We will not deal with informational breakdown here but focus on how behavior of the population is affected if we assume – as we should – that social learning is extremely pervasive.

4.3.2 Over-matching and volatility

Before beginning the more technical analysis, we first have a look at the behavior of the strategies in order to get an intuition about how they behave depending on their frequency in the population. In figure 4.4, we show the behavior of the different strategies in a randomly generated environment. The proportion of A choices made by the strategies (•) is shown as compared to the environmental changes ($p_A - p_B$, solid line) over time. On the top left panel, individual learners are shown. They follow environmental changes quickly and closely, which results in environment and behavior changing almost in synchrony.

In the middle row, the behavior of ITW is shown, once for a frequency of 0.5 (left) and once for a frequency of 0.98 (right). The rest of the population is made up of individual learners but not shown. ITW tends to overmatch the environmental changes, resulting in fluctuations in behavior that are stronger than the fluctuations in the environment. On the bottom, the behavior of PC-PBSL is shown, again once for a frequency of 0.5 (left) and once for a frequency of 0.98 (right), with the rest being individual learners (not shown). PC-PBSL too shows overmatching behavior, especially when more common.

We used the terms “matching” and “overmatching” without being specific about them. These terms are borrowed from the study of animal behavior [14]. In our context, what we mean by matching is that a 1 percentage point increase in $p_A - p_B$ is matched by a roughly 1 percentage point increase in the proportion of A choices made a strategy. In contrast, overmatching would lead to a more than 1 percentage point increase in the proportion of A

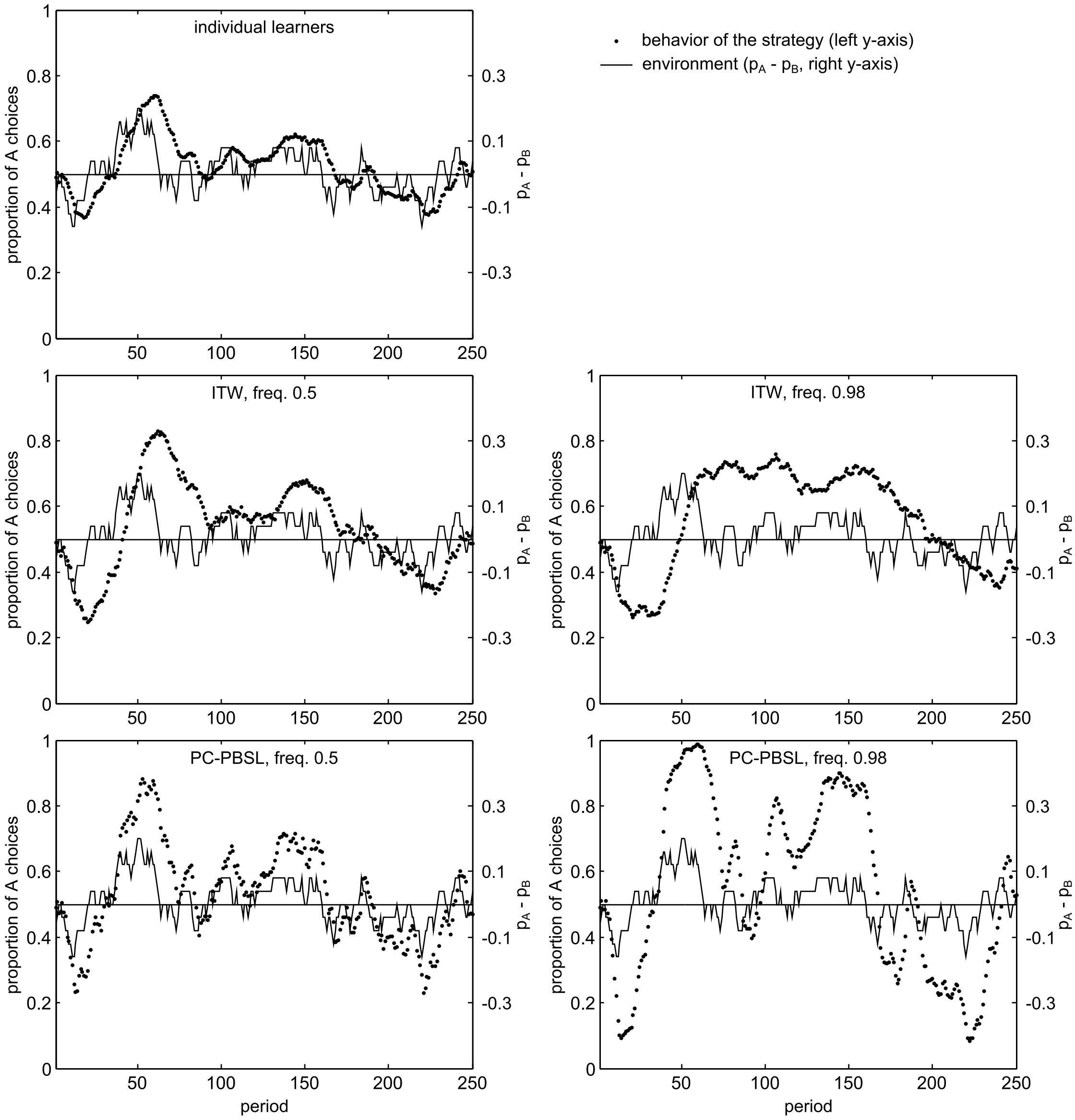

Figure 4.4: Behavior of the different strategies over time. Top: Individual learners. Middle: ITW for a frequency of 0.5 (left) and 0.98 (right); the remainder are individual learners (not shown). Bottom: PC-PBSL for a frequency of 0.5 (left) and 0.98 (right); the remainder are individual learners (not shown). The behavior of the strategies is shown as the proportion of A choices (●, left y-axis) over time; the environment is shown as $p_A - p_B$ (solid line, right y-axis). Individual learners tend to match the environmental probability, while social learners tend to overmatch, especially when very frequent.

choices. Picturing matching behavior in the fashion of figure 4.4, matching results in behavior and environment moving in parallel, as is observed for individual learners.

To quantify the tendency of strategies to match or overmatch, we simulated the behavior of the strategies for fixed frequencies over 1000 generations with 50 periods each. The resulting behavior of the different strategies is shown in figure 4.5, with proportion of A choices shown as a function of $p_A - p_B$. Optimal behavior, choosing A when $p_A > p_B$ and vice versa, would result in a step function, as shown on the top left hand side. Individual learners show a rather linear correlation between behavior and environment, as shown on the top right hand side. The social learning strategies (middle and bottom row) show a sigmoidal behavioral function.

Next we regressed the proportion of A choices onto the environment ($p_A - p_B$). Note that we corrected these results for delay, since strategies always react with a certain delay to environmental changes. To do this, we calculated the lag that was associated with the highest cross-correlation between environment and behavior and corrected for this lag. Without this correction, we get very similar results, except that the fits are worse and the slopes a little lower. (More on lags can be found in the next section.)

The results for the regression analysis are shown in table 4.1. As speculated, individual learners match probabilities. A 1 percentage point increase in $p_A - p_B$ is matched by a 1.15 percentage point increase in the proportion of A choices. In contrast, ITW shows overmatching, with a 1 percentage point increase in $p_A - p_B$ followed by a 2.14 (frequency of 0.5) or even 2.51 (frequency of 0.99) percentage point increase in the proportion of A choices. PC-PBSL also shows overmatching. For a frequency of 0.5, each 1 percentage point increase in $p_A - p_B$ is matched with a 1.96 percentage point increase in the proportion of A choices; for a frequency of 0.99, it is even matched with a 2.99 percentage point increase in the proportion of A choices. In sum, social learning strategies tend to overmatch by factor of 2:1 when not too common, and up to a factor of 3:1 when very common.

A corollary of these findings is that one should expect the social learning strategies to act in a more uniform fashion. We therefore wanted to determine the uniformity of a strategy. Uniformity is defined as exceeding value x if more than x% of the individuals with a certain strategy choose the same option. Figure 4.6 shows the proportion of periods in which the strategies have at least a certain degree of uniformity. Individual learners exhibit balanced behavior – rarely do 80% of them choose the same option. For ITW and PC-PBSL, however, more often than not do we observe 80% or more of them choosing the same option. Uniformity increases the more frequent social learners are. Higher degrees of overmatching are thus indeed associated with more uniform behavior.

It might be of interest that the uniformity in behavior apparently follows a power law. This is true for both social and individual learners. For PC-

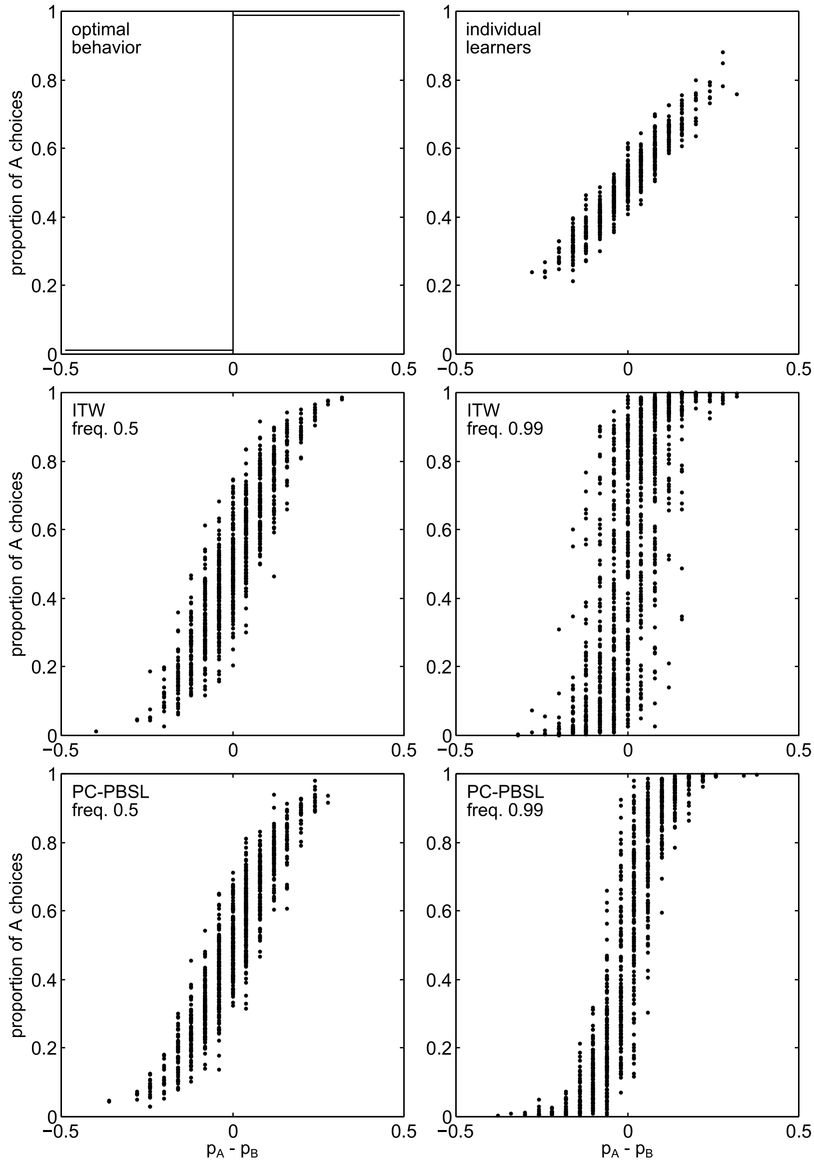

Figure 4.5: Scatter plot of the proportion of A choices as a function of $p_A - p_B$. Top left: The optimal behavior would be to choose A whenever $p_A > p_B$ and B otherwise. Top right: Behavior of individual learners; they match probabilities. Middle: Behavior of ITW (left: frequency of 0.5, right: frequency of 0.99). They overmatch probabilities, especially when more frequent. Bottom: Behavior of PC-PBSL (left: frequency of 0.5, right: frequency of 0.99). They also overmatch probabilities, especially when more frequent. The social learning strategies display a sigmoidal adoption curve. All data points were corrected for delay; only every 50th of the 50.000 data points is shown to reduce image file size.

Strategy	Frequency	Lag	Slope (95% C.I.)	R^2	S.S.E.
Individual learners	-	6	1.15 (1.14-1.15)	0.83	132
ITW	0.5	9	2.14 (2.13-2.15)	0.79	582
ITW	0.99	27	2.51 (2.49-2.53)	0.48	3214
PC-PBSL	0.5	4	1.96 (1.95-1.96)	0.88	264
PC-PBSL	0.99	7	2.99 (2.97-3.00)	0.77	1371

Table 4.1: Results of the linear regression analysis of behavior derived from the simulation of 50,000 periods, regressed onto the environment. The lag is the lag at which the cross-correlation between environment and behavior is maximized. Shown are the slope of the linear fit is shown, including 95% confidence bounds, adjusted R^2 and sums of square errors (S.S.E.).

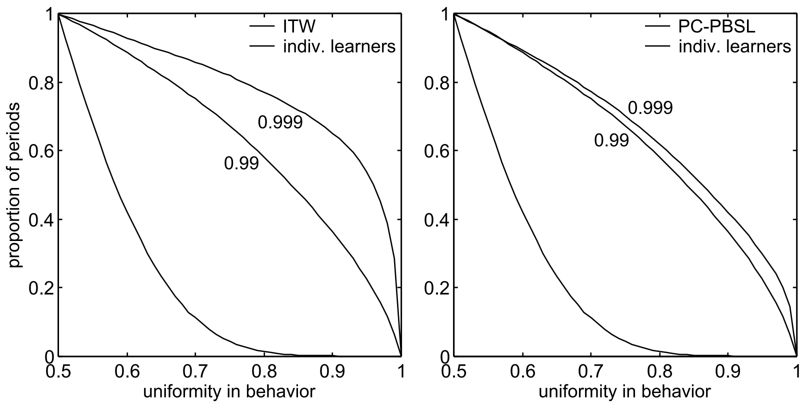

Figure 4.6: Proportion of periods in which a strategy exhibits a behavior with a certain degree of uniformity. Left: ITW (solid lines) and individual learners (dashed line); right: PC-PBSL (solid lines) and individual learners (dashed line). Frequencies of the social learning strategies as indicated. Individual learners do not exhibit very uniform behavior. ITW and PC-PBSL act more uniformly, and the more so the more frequent they are.

PBSL, we found $R^2 > 0.95$ for 9 of 9 frequencies, and $R^2 > 0.99$ for 4 of 9 frequencies. For ITW, we found $R^2 > 0.98$ for 9 of 9 frequencies, and $R^2 > 0.99$ for 7 of 9 frequencies. For individual learners, we found $R^2 > 0.98$ for 9 of 9 independent data series, and $R^2 > 0.99$ for 8 of 9 of the data series. This is strong evidence for the notion that uniformity is distributed according to a power law. Although this finding is certainly not a coincidence, we cannot think of a reason why that should be so – the environment is not distributed according to a power law – except that power law distributions are common for social phenomena [131]. Mesoudi and Lycett found that frequency-bias (which is present for ITW and PC-PBSL) disrupts power law distributions in cultural transmissions [125] but their study is made in a different context and thus results cannot be compared.

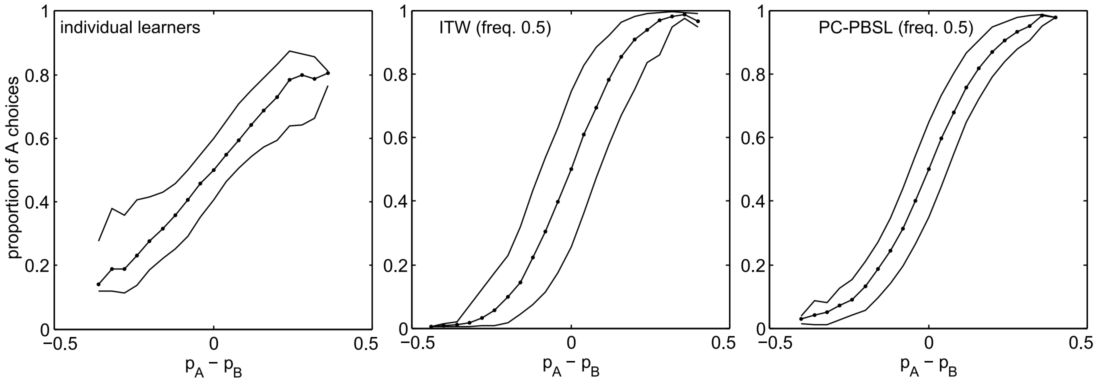

Figure 4.7: Median proportion of A choices (solid line) as a function of $p_A - p_B$, as well as 95% confidence intervals (dashed lines). As $p_A - p_B$ only takes discrete values, we could calculate the median proportion of A choices for each unique value of $p_A - p_B$. Individual learners (left) show a quite linear behavioral function, whereas ITW (middle) and PC-PBSL (right) show a sigmoidal behavioral function.

We were interested in the shape of the behavioral function of the different strategies. As there is a lot of variation in the data, we calculated, for each unique value of $p_A - p_B$, the median proportion of A choices of the different strategies. The results shown in figure 4.7 confirmed our initial guess. For individual learners, we find a linear behavioral response to environmental changes, for ITW and PC-PBSL, we find a sigmoidal behavioral response to environmental changes. Fitting a sigmoidal function of the form $y = 1/\left(1 + a \cdot \exp\left(-b \cdot x\right)\right)$ resulted in better fits for the social learning strategies but not for the individual learners.

Our findings also suggest that the behavior of social learners is more variable. When looking, e.g., at the sum of square errors (S.S.E.) of individual learners, we find it to be much lower than the S.S.E. of the social learners, especially when the latter are very frequent. However, this could be explained by the fact that the behavior of social learning strategies is not fitted as well by a simple function. To make a better comparison, we thus looked at the excess volatility that remains after subtracting the mean proportion of A choices for each singular value of $p_A - p_B$. That is, we make the assumption that each value of $p_A - p_B$ should map on a singular proportion of A choices. The variation that remains after this is the variation that is not explained after controlling for differences in the environment.

As it turns out, we indeed find more unexplained variation for social learners. In table 4.2, we show the remaining standard deviation σ after controlling for differences in the environment. In finance, this is also called "volatility". For individual learners, volatility is quite low at 0.051. For ITW at frequency 0.9, it is already almost three times as much, and at a very high frequency of 0.9999, volatility rises to 0.432. For PC-PBSL, we

Strategy	Frequency	Unexplained σ
Individual learners	-	0.051
ITW	0.9	0.141
ITW	0.99	0.250
ITW	0.999	0.361
ITW	0.9999	0.432
PC-PBSL	0.9	0.093
PC-PBSL	0.99	0.139
PC-PBSL	0.999	0.162
PC-PBSL	0.9999	0.204

Table 4.2: Unexplained standard deviation or "volatility" (σ) after controlling for differences in the environment.

also find high unexplained volatility, though less than for ITW, reaching 0.204 for a frequency of 0.9999.

4.3.3 Delay

A peak at figure 4.1 and 4.4 suggests that social learning could lead to a delayed response to environmental changes. To study this, we have a closer look at the cross-correlation between the environment in the form of $p_A - p_B$ and the proportion of A choices made by the strategies (figure 4.8). On the left panel, we see the cross-correlation of ITW for different frequencies. Cross-correlation of ITW is lower and peaks at higher lags than cross-correlation of individual learners. For example, the peak cross-correlation of ITW at frequency 0.9 can be found at 14 periods, with $\rho = 0.86$, whereas for individual learners, it can be found at 6 periods, with $\rho = 0.91$. It becomes worse for ITW at higher frequencies; for a frequency of 0.999, ITW's cross-correlation peaks at 44 periods, with $\rho = 0.49$. This means that the more social learning there is in the populatuion, the more the behavior of ITW lags behind the environment and the less environment is reflected in behavior.

For PC-PBSL, we qualitatively see a similar picture but quantitatively quite a different one. This strategy also shows higher lags and lower correlations when it becomes more frequent in the population. However, even when very frequent, the behavior of PC-PBSL still correlates reasonably well with the environment; for a frequency of 0.999, cross-correlation peaks at a lag of 8 and reaches 0.86, which is almost as good as individual learners do. For lower frequencies, PC-PBSL might even do better than individual learners.

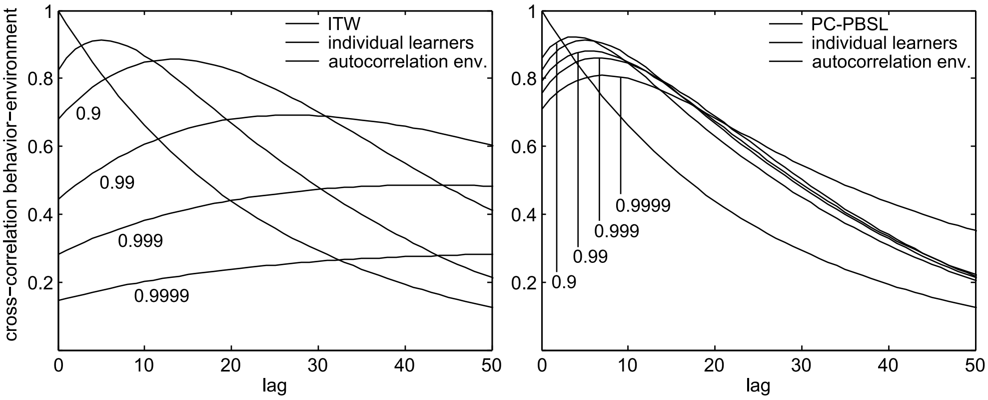

Figure 4.8: Cross-correlation between the proportion of A choices made by the strategies and the environment ($p_A - p_B$), left panel ITW, right panel PC-PBSL (solid lines). Frequencies of the social learning strategies are indicated. As a comparison, auto-correlation of the environment (dotted line) and cross-correlation of individual learners (dashed line) are shown. In general, cross-correlation of social learning is higher and lag lower the fewer social learners there are. ITW has especially low cross-correlations, whereas PC-PBSL shows good cross-correlation even for very high frequencies.

4.3.4 Detachment from reality

In this section, we want to study whether the strategies' behavior accurately reflects the environment or not. We first asked ourselves whether the strategies react to the environment in an instantaneous fashion or rather to how the environment behaves over the long run. To analyze this, we used moving average filters of variable sizes to smooth the environment. For a window size of 10, for instance, each data point of the environment corresponds to the arithmetic mean value of the environment of the last 10 periods. Next, we again determined the cross-correlation between behavior and the now smoothed environment. What we were interested in was how the highest cross-correlation we found, the peak correlation, varied with the length of the environmental trend. The results are shown in figure 4.9.

First we have individual learners. Depending on the window size of the filter, peak correlations vary, first increasing but later decreasing. The highest peak correlation can be found for a window size of approximately 15 periods. This means that the highest correlation between behavior of individual learners and environment is actually found if we look at environmental trends of 15 periods. Individual learners best track trends of this rather short duration.

Second we have PC-PBSL. At a frequency of 0.9, the highest peak correlation is found for a window size of around 10 periods. PC-PBSL thus behaves as if tracking trends of 10 periods length. When becoming more common, peak correlations of PC-PBSL shift to longer trends, but even for

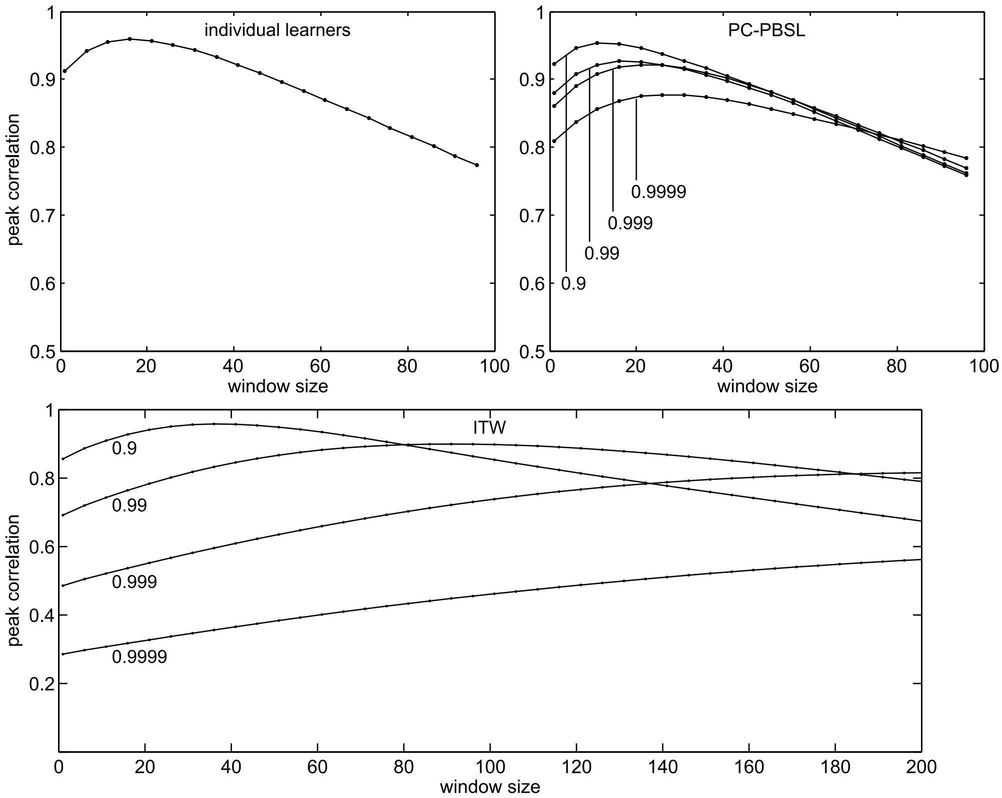

Figure 4.9: Peak correlation for different window sizes of the moving average filter applied to $p_A - p_B$. Top left: individual learners; top right: PC-PBSL (note for both the truncated y-axis); bottom: ITW. For the social learning strategies, frequencies are indicated. For individual learners, the highest peak correlation is found for a window size of about 15 periods length. For PC-PBSL, the highest peak correlations are in a similar range, and they increase the more frequent PC-PBSL is. For ITW, the highest peak correlations can be found at much higher window sizes of 40 or more.

a frequency of 0.9999, these trends are still only 20-30 periods long.

Last we have ITW. ITW too shows a shift of peak correlations towards longer trends when becoming more frequent. However, even at low frequencies, the highest peak correlations appear quite late. For a frequency of 0.9, it is at approximately 35 periods, for a frequency of 0.99 almost at 100 periods, and at frequencies of 0.999 or above, the highest peak frequencies are beyond 200 periods length. ITW thus behaves as if reacting to very long term trends, a fact that is hugely aggravated when ITW is very frequent.

Another method to study whether the behavior of the strategies reflects the environment is to analyze whether it is possible to infer the environmental state from the behavior of the strategies. To make this inference, we propose a very simple mechanism akin to conformism. According to this

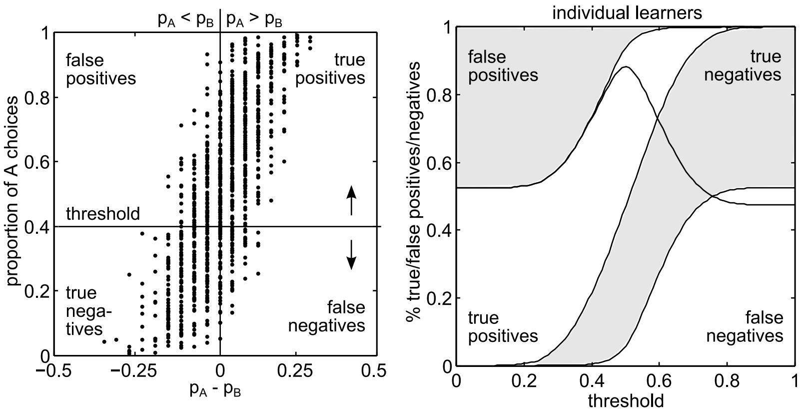

Figure 4.10: How well the environmental state can be inferred from behavior. The left panel illustrates how the choice of the prediction threshold impacts the number of data points that fall in the range of true or false positives or negatives. In the right panel, the proportion of true and false positives and negatives as derived from the simulation of individual learners is cumulatively shown. The total proportion of true estimations is shown as the dashed line. There is a high peak of total true estimations for a threshold of 0.5, reaching 88% correct estimations.

mechanism, one should infer that p_A is greater than p_B if at least x% of the population choose A (and vice versa). We illustrate this in the left panel of figure 4.10. When one infers that p_A is greater than p_B, the inference can be true, resulting in a "true positive", or false, resulting in a "false positive". In contrast, when one infers that p_A is less than p_B, a true inference results in a "true negative" and a false inference in a "false negative".

The right panel of figure 4.10 shows how the proportion of true and false positives and negatives develops when the threshold is shifted. When starting with a very low threshold, all predicitions are positives, i.e. that $p_A > p_B$. Of those predictions, approximately half will be correct and the other half false. When the threshold is increased, negative predictions (i.e. $p_A < p_B$) will join in, some of them being false and some true. The really interesting measure is the total number of correct inferences. This is shown as the dashed line. As expected, when the threshold is too low or too high, roughly half of the predictions will be correct, but when the threshold is set at around 0.5, 88% of the inferences are correct. In other words, when one observes individual learners, the best guess is to estimate that $p_A > p_B$ when more than half of them choose A and vice versa, which is of course hardly surprising. More notable is the fact that the accuracy of the inference is quite high, reaching 88%.

We were interested in whether social learning strategies also allow to make correct inferences about the environment the majority of times. We found that this depends on the strategy and how frequent it is. Our approach was for each strategy to find the best possible proportion of correct inferences by

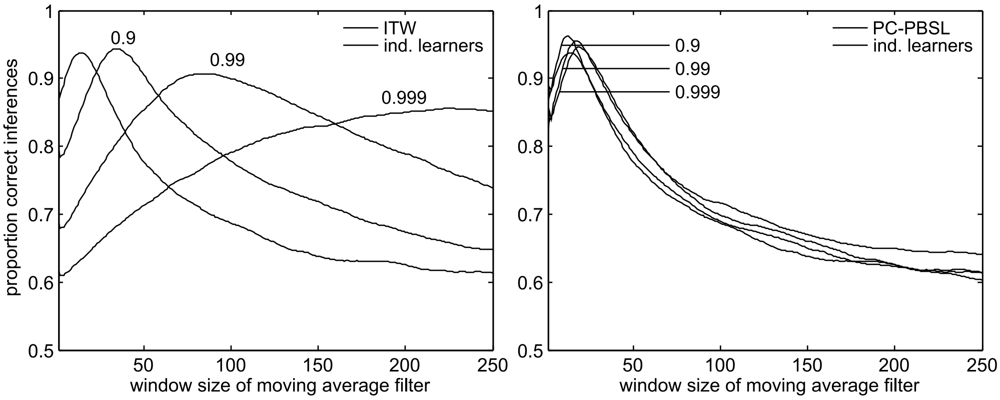

Figure 4.11: Proportion of best possible inferences from observing a strategy as a function of the window size of the moving average filter applied to $p_A - p_B$. Dashed line: individual learners; solid line: ITW (left) or PC-PBSL (right). Frequencies of the social learning strategies as indicated. From observing individual learners, one would be quite accurate in inferring the environmental state. For ITW, accuracy would be lower, especially the more frequent ITW is. Instead, ITW behaves as if fitting a smoothed environment, with filter size of 35 periods or more. PC-PBSL, in contrast, allows to make quite good predictions about the environment even if very frequent. PC-PBSL and individual learners behave as if reacting to an environment smoothed over 20 periods or less.

determining the optimal threshold ex post. For ITW at frequency 0.9, accuracy is only 79% and drops to 62% for a frequency of 0.999. For PC-PBSL, however, inferences were quite accurate, ranging from 86% for a frequency of 0.9 to 85% for a frequency of 0.999. Observing PC-PBSL would thus reveal a lot about the environment, while observing ITW, especially when very frequent, would reveal little.

Moreover, we were interested in whether the strategies react more to long term developments in the environment rather than fitting short term changes. To study this, we again applied a moving average filter to the environment and determined the highest possible proportion of correct inferences as a function of the filter size (figure 4.11). It turns out again that individual learners rather behave as if fitting a 15 period average of the environment. ITW at frequency 0.9 behaves as if reacting to a 35 period average; if more frequent, the apparent filter size ITW reacts best to can be larger than 200 periods. For PC-PBSL, we did not find this. Even at a frequency of 0.999, the strategy reacts as if fitting an environment smoothed over 20 periods only. This shows that PC-PBSL reflects the environment in a more accurate, short-term fashion.

Another important factor for keeping in touch with reality is that at all times, some contrarian views should be expressed. That is, even if, say, A is better than B, there should be some individuals who choose B. This is required because dependent social learning strategies such as ITW and PC-PBSL need role models to copy from. If all individuals choose A, there would be no “substrate” for B and a dependent social learner could never choose this option. The need for contrarians is particularly high if for a long stretch of time, one option was better, leading almost all individuals to choose that option; when there is a sudden change, a decent number of contrarians are required as models to cause a sufficiently fast switch in the population.

To quantify contrarianism, we defined the option that is chosen by less than half of the population as the minority option, regardless of whether this option is better or not. It is especially interesting to study which strategies are expected to engage in more contrarian behavior. We thus looked at what percentage of contrarians are social learners and what percentage are individual learners. Interestingly, we found that individual learners were always over-represented among contrarians. For example, when the population consists of 50% individual learners and 50% ITW, we found 57.1% of the contrarian views to be expressed by individual learners. This means that individual learners are 1.14 times more likely to express contrarian views than would be expected if contrarian views were expressed equally by all strategies.

The over-representation of contrarian views among individual learners depends on the social learning strategies that are involved and their frequencies. This is shown in figure 4.12. If contrarian views were expressed by all strategies according to their proportion, we would find a representation of 1 by individual learners (horizontal line). Instead, we find that individual learners are always over-represented among contrarians. For low frequencies of social learners, this over-representation is weak, but it rises and exceeds the factor 2 when the population consists of 99 percent social learners or more. In the extreme, if individual learners make up only 0.01% of the population, over-representation reaches a factor of 2.55 when individual learners are paired against PC-PBSL, and of 7.39 when paired against ITW.

When social learners are very frequent, despite being under-represented, they still are more common among the contrarians. But among the few contrarian social learners, many have adopted the contrarian view because they copied contrarian individual learners (but not vice versa, since individual learners do not copy). This means that an important role of individual learners is to express contrarian views in the population and “convince” others to do the same. These contrarian views are required as a substrate that allows social learners to switch options in case of an environmental change; if everyone chose the same option, there would be no one to copy the other option from. This would result in informational breakdown and thus a potentially disastrous complete detachment of behavior from reality.

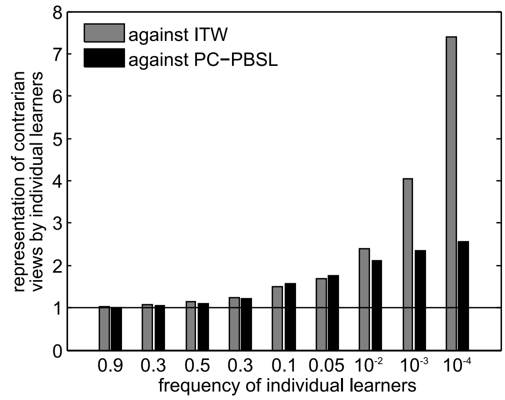

Figure 4.12: Representation of contrarian views by individual learners as a function of their frequency and of the opposing strategy (gray: ITW, black: PC-PBSL). If contrarian views were expressed equally among strategies, we would find a representation of 1 among individual learners (horizontal line). We find instead that contrarian views are more likely to be expressed by individual learners, resulting in representation exceeding 1. This is especially true the less frequent individual learners are in the population.

4.3.5 Principal findings

In this chapter, we foucs our study on three strategies, individual learning, ITW, and PC-PBSL. We chose the two latter strategies because both have been shown to have very high equilibrium frequencies under many conditions. PC-PBSL is a very dominant strategy across a wide range of parameter values but is ecologically and cognitively rather demanding. ITW, on the other hand, is the best strategy in absence of payoff-biased social learning and follows a rather simply heuristic. Therefore, these two strategies should cover a wide spectrum of social learning strategies.

First we found that ITW and PC-PBSL are characterized by overmatching – a one percentage point increase in $p_A - p_B$ leads to a 2 to 3 percentage point increase in the proportion of A choices. Individual learners match in a fashion closer to 1:1.

Overmatching by the social learners implies more uniform behavior. Most of the time, one will observe that 80% or more of them choose the same option, whereas individual learners never act in such a coordinated fashion. Moreover, since social learners behave self-referentially, more of their behavior is unexplained by differences in the environment.

Overmatching in itself should be positive; it is indeed expected if behavior were optimal. There is, however, also a negative side to the social learning strategies' behavior: We find them to lag behind environmental changes. Granted, individual learners also lag behind, but only by approximately 6 periods. PC-PBSL often shows a lower cross-correlation with the environment than individual learners, but their lag is moderate. ITW, on the other hand, shows strong tendencies to lag behind.

Interestingly, we find that the strategies fit trends in the environmental changes better than the environment itself. Individual learners show the best correlation for trends of approximately 15 periods length, while the social learning strategies seem to fit longer trends.

If social learners lag behind and overmatch, it should be expected that their behavior is often not strongly linked to the environment. For ITW, we find this to be indeed the case. When one observes them, it is difficult to infer whether A or B is currently the better choice, while for individual learners and PC-PBSL, inferences are spot on most of the time. Again, we find that if one were to infer trends in environmental changes rather than the current environmental state, the fit would be better. For individual learners and PC-PBSL, best inferences were made for trends that have a length of 15-20 periods; for ITW, the length could go into the hundreds.

Learning in varying environments always involves a trade-off between exploration and exploitation. Explore too much, and you will spend too little time with the better option; exploit too much, and you will miss when another option becomes better. In our model, the roles of explorers and exploiters seem to be well separated. Individual learners are much more likely to choose the minority option, which is often the worse option.

All the described phenomena associated with social learning depend on the frequency of social learning. If social learners are rare, they mostly imitate individual learners and thus behave in a similar fashion. Only when they are frequent do we see extreme deviations between the behavioral features. But as we noted initially, high frequencies of social learners should be expected, so that the described phenomena should be expected to occur.

4.4 Discussion

4.4.1 General discussion

Earlier theoretical work in evolution and anthropology has already studied the consequences of gene-culture coevolution, but mainly on the level of evolutionary stability. This means that it has focused on which types of social learning should be expected in equilibrium – how much to learn socially, how to copy, whom to copy, and so forth [21–23, 31, 46, 50, 61, 63, 86, 100, 101, 103, 108, 111, 120, 141–143, 146, 174, 177]. On the other hand, economic work on social learning has mainly dealt with how social learning, assuming it takes place and has certain characteristics, affects decision making in equilibrium [2, 10, 16, 17, 40, 41, 49, 96, 150, 157, 159, 183, 184].

In this work, we sought to tie together these two threads. The first chapter of this work has prepared the ground by presenting the general model framework. In the second chapter, we took a closer look at payoff-biased social learning strategies and studied dozens of strategies under many conditions. In the third chapter, we took the most promising payoff-biased social learning strategies and matched them against other strategies, most of them proposed in the literature. From this competition, two social learning strategies showed especially robust performance in the evolutionary context, one being more simplistic and one more complex. In this chapter, we took a closer look at the behavior of these two strategies compared to the behavior of individual learners, unveiling some of the consequences that should be expected from the use of social learning. This way, we connected the evolutionary and the behavioral approach and thus hope to contribute to a better understanding of social learning phenomena.

4.4.1.1 Social learning and altruism

The existence of altruistic behavior is puzzling. By definition, it involves a cost to the actor, and although the benefit to the recipient is larger than the cost, selfish natural selection should remove altruistic behavior from the population. Examples of altruistic behavior include guarding herds, game hunting with subsequent meat sharing, and risking one's life in warfare. A possible explanation for altruistic behavior could be group selection, since on the level of the group, altruistic behavior is beneficial. Group selection works only under very narrow conditions, though, and requires, among others, sufficient intergroup-variance. Empirical work has shown that group-beneficial traits are unlikely to have evolved through group-selection alone [164].

Cultural evolution has been proposed in the literature to enable the evolution of altruistic behavior between non-kin [21, 22, 25, 27, 90]. The proposed argument is roughly the following:

1. Social learning, in the form of conformism, exists because it is beneficial for reasons unrelated to cooperation.
2. Social learning promotes intergroup-variance.
3. Higher intergroup-variance facilitates group selection.
4. Group selection results in the evolution of altruistic behavior.
5. Psychological mechanisms have locked in group-beneficial behaviors.

A very simplified model may illuminate this process. Imagine there to be several groups of humans living in the same area. Each individual can either participate in helpful cooperative behavior, which has a personal cost but benefits the group, or not. Usually, such helpful behavior cannot be sustained because selfish individuals have a higher fitness. But now assume that the individuals are all conformists. Then either all or none of the group members are helpful. If a member migrates to another group, he or she, being a conformist, adopts the behavior that is currently practiced in that group. This way, disruptive selfish behavior cannot undermine cooperation through immigration.

Now there are groups that are completely cooperative and groups that are completely uncooperative. Next, one could imagine that some event, like a drought or warfare, leads to the extinction of some groups. It is most likely that the uncooperative groups vanish, since they engage in less altruistic acts. The cooperative groups, however, propagate to the vacated spots and thus cooperation becomes more common overall.

Without conformism, the level of cooperation would be more similar between groups. Therefore, the forces of group selection would be weaker than those of individual level selection. But as conformism exists, over time, cooperation dominates and defection succumbs. Psychological mechanisms such as shame evolve, making it easier to internalize the cooperative norms.

Do our findings support the notion that social learning promotes the evolution of altruistic behavior? We found earlier that pure conformism is not a very likely candidate for a social learning strategy, and even if other forms of social learning were impossible (e.g. if payoffs were not observable), we would not expect the whole population to be conformist. Therefore, the first step of the argument could be contested. However, we found that those social learning strategies that we do expect to thrive, in spite of not technically being conformists, behave in a conformist fashion. Even though pure conformism is not expected, behavior that is aligned with conformism is expected. This supports the first step of the cultural group selection argument.

Furthermore, we found social learning to result in overmatching. When the difference between the success probabilities of the two options A and B, p_A and p_B, increases by one percentage point, the behavioral response should exceed it, leading to a 2 or 3 percentage point increase in A choices. In contrast, individual learners match differences in fashion closer to 1:1. This results in more coordinated behavior of social learners. Whereas individual learners spend only 1 in 80 periods making a uniform choice (i.e. more than 80% choose the same option), ITW at frequency 99.9% acts uniformly in an average of 77 of 100 periods, and PC-PBSL at that frequency in an average of 62 of 100 periods. Therefore, social learning will indeed increase intergroup-variance towards a level that cannot be reached through individual learning. This reinforces the second step of the argument.

There is, however, a caveat. Although we should expect behavior to be very uniform, with almost all individuals choosing the same option, we also know that ITW and PC-PBSL are *noncompensatory* decision rules. This means that in principle, one single dissenter could convince them to switch options. Per definition, a selfish individual receives a higher payoff than an altruist. Thus, even if defection is the minority option, if it generates a higher payoff or wealth, it will be adopted sooner or later. Therefore, on an individual level, cultural selection, like natural selection, would favor defection over cooperation.

Certainly, this argument is purely verbal and needs to be substantiated quantitatively. It may still be possible that ITW, PC-PBSL, and other social learning strategies with a conformist bias would facilitate group-selection more than individual learning does. Yet it is doubtful that group selection would work as easily as it would if conformism was the dominant form of social learning.

Conformism will be most successful where other forms of social learning are not possible. For example, when it is impossible to link a certain behavior with its result, be it payoff or wealth, neither PBSL nor ITW could exist. But as these forms of social learning are superior to conformism, one would expect cognitive mechanisms to evolve that permit to make such links. Ultimately, it is unlikely that social learning takes purely the form of conformism.

4.4.1.2 Social learning and signal detection

The task we designed for the strategies to solve is a rather difficult one. For the default condition, the median success probability of the better option is 0.54 and the median success probability of the worse option is 0.46. How often would it be necessary to sample an option before one could be certain that it is indeed the better and not the worse option? When the usual criterion of a significance level of $p \leq 0.05$ is used, it would require on average almost 140 trials before significance is established by a binomial test. If the null hypothesis is that both options are equally good, it would take, on average, almost 490 trials before significance is reached.

During the sampling process, the values of p_A and p_B will have changed, however. After 100 periods, the autocorrelation of the environment has shrunk to a meager 0.02. The probability that if one option is currently better or equal to the other and still is so in 100 periods later is only 65.2%. The time scale at which p_A and p_B vary is thus shorter than the time scale needed to tell with certainty which of the two options is better. A “scientific approach” to the problem through use of significance tests is thus doomed to fail. The learners have to make a decision in the face not of risk but of uncertainty [105]. (By the way, taking a Bayesian approach instead of a frequentist approach does not solve the problem; as we showed in the first chapter, Bayesian individual learning, even if given a huge advantage, still leads to a modest performance only.)

A way to frame this problem is in terms of signal and noise. The signal is whether p_A or p_B is greater, but sampling from A and B is a very noisy process because p_A and p_B typically take very similar values. If, for an individual, A yielded successes in the last ten periods but then suddenly is unsuccessful twice, she has to contemplate whether these last two events are meaningful signal or just random noise. Being too conservative results in missing environmental changes, being too volatile results in fitting the noise and not the signal.

An adaptive strategy will have to encounter several failures in short order before switching away from an otherwise successful option. Therefore, it is not surprising that we find the behavior of individual learners to fit trends in the environment rather than instantaneous changes. In fact, the behavior of individual learners was best described if assuming that they fit trends of 15 periods length. If individual learners had no memory, i.e. if they used win-stay lose-shift, they would react instantaneously to the environment. In the first chapter, we already saw, however, that such immediate switches in behavior result in a lower performance. Having a longer memory, and thus fitting trends rather than instantaneous changes, is the adaptive thing to do.

Social learning strategies, by drawing on the experience of several individuals, could in principle react faster to environmental changes. Yet this advantage vanishes the more social learning occurs in the population [146]. Our findings support the notion that social learning does not lead to faster adoption but instead to delays. Furthermore, we found social learning to create excess volatility in the behavior, making the observations derived from other social learners more noisy.

At the end of the day, social learners face the same trade-off as individual learners in terms of sampling. They, too, indirectly sample over several periods by imitating old behavior, thus avoiding overfitting; their behavior consequently fits trends in the environment and not instantaneous changes. This is supported by our findings that PC-PBSL behaved as if fitting trends of length 10 to 30 and ITW as if fitting even longer trends – depending on the frequency, these trends could be several hundreds of periods long. If anything, social learning should eventually lead to less instantaneous behavior.

4.4.2 Social learning and finance

4.4.2.1 Excess volatility

Robert Shiller showed that the volatility of a stock price forecast has to be lower than volatility of the realized stock price [156, 158]. To see this, let $p_t = E(p_t) + U_t$, with p_t being the real stock price, $E(p_t)$ the expected value of the stock price (the forecast), and U_t being the forecasting error. It must true be that the covariance of $E(p_t)$ and U_t is zero because else, there would be a systematic error that could be exploited to beat the market, contradicting the efficient market hypothesis. From the equation, it follows that $\mathrm{Var}(p_t) = \mathrm{Var}(E(p_t)) + \mathrm{Var}(U_t)$, and thus, as variance is never negative, that $\mathrm{Var}(p_t) \geq \mathrm{Var}(E(p_t))$, implying that $\sigma(p_t) \geq \sigma(E(p_t))$ (σ being the standard deviation or “volatility”).

In terms of finance, $E(p_t)$ is the stock price and p_t is the present value of all future dividends. Whether the inequation holds true is thus a matter of simple statistical analysis of the two data sets, stock prices and dividends, and Shiller found that the inequality is undoubtedly violated – stock prices vary much more than present value of dividends as determined ex post. The standard deviation that is not explained by changes in the present value of dividends is called excess volatility.

Is this excess volatility in behavior mirrored in the behavior our social learning strategies? For this, we do not simply take the variation of $p_A - p_B$, since the optimal behavior is not to track this difference as accurately as possible, but to choose A when $p_A > p_B$ and vice versa. Therefore, we took the standard deviation of the sign of $p_A - p_B$, mapped onto $[0; 1]$, because we deal with probabilities. This is called the "effective environment", as fitting this environment exactly would really lead to optimal behavior. It turns out that variation in behavior is not greater than variation in the effective environment. Even for ITW at frequency 0.9999, standard deviation of behavior was 0.451, while standard deviation of the environment is 0.458. One could thus conclude that the findings from finance that forecasts of stock prices show excess volatility are not replicated by our model.

There is a more sophisticated way to look for excess volatility in our model, though. Because of how we generate the environment, $p_A - p_B$ can only take discrete values. It is therefore easy to compute for each singular value of $p_A - p_B$ the median proportion of A choices made by a strategy. Next, one can calculate the standard deviation of the behavior corrected for the median behavior for the given value of $p_A - p_B$. We also corrected for delay in behavior. The remaining standard deviation is the volatility of the behavior that cannot be explained by reactions to the environment. There are two possible sources for this remaining volatility:

- historical contingencies and
- self- and other-referential behavior (social learning).

For individual learners, clearly the first point applies – they decide based on the past experience, which is partly subject to random outcomes. The second point, however, does not apply, since they cannot learn socially. For social learners, the second point applies but the first does not (except, indirectly, if they copy an individual learner); this is because the two social learning strategies we studied only sample from the last period and have no memory of earlier events. The reason why social learning generates volatility is mostly because sampling is random and small initial differences may be exaggerated.

Our results suggest that excess volatility generated by social learning is much higher than excess volatility generated by individual learning. For a frequency of, say, 99%, ITW has five times the excess volatility of individual learners and PC-PBSL has almost three times the excess volatility. When social learning is more ubiquitous, the differences become even larger. This means that volatility generated by social learning is greater than volatility generated by the contingencies of individual learning. For finance, this means that social learning behavior could be the main source of excess volatility.

In behavioral finance models, similar results were found. For example, Shiller [157] assumed that in addition to “smart money”, there are feedback traders in the market. Feedback traders observe the movement of prices, which a rational actor knows are random in efficient markets, and buy or sell as a consequence of these movements. As prices are determined by demand and supply, which are in turn determined by the market participants, feedback trading is a form of social learning. Shiller found that feedback traders can generate excess volatility, which lends support to our finding that social learning is the main source of volatility.

In Shiller’s model, there is no explanation why feedback trading, which is irrational, should exist in the first place. Although, according to Shiller, casual observation suggests that feedback trading exists and data are consistent with it, one should expect that irrational behavior, which in the long run leads to monetary losses, should vanish from the market. This is a common criticism raised against behavioral finance models [11, 76, 77]. Our model, though not a market model, has the advantage of explaining why social learning should be an important factor, instead of presupposing its existence. Future models that merge our evolutionary approach while taking market mechanisms into account should be a welcome addition to the study of anomalies in financial markets.

4.4.2.2 Reaction to short and long term trends

Barberis, Shleifer, and Vishny found that stock prices underreact to singular new bits of information but overreact to series of new bits of information [11, and references therein]. In the short term, which according to the authors is of a length between one and twelve months, stock prices thus underreact to changes in fundamentals, whereas in the long term, if a consistent series of information is revealed (all good or all bad), they overreact. In their model, the authors explain this by the existence of “investor sentiment”. These sentimental investors are assumed to either believe that the environment is mean-reverting or trending when really it is a random walk. The investors thus react to the prices, which is a form of social learning. The real model, random walk, is not supported in the boundedly rational behavior of the investors, which explains why they never fully learn to behave rationally. Therefore, the proposed model could be criticized for precluding fully rational behavior.

In our model, social learning is rational in the sense that one would expect it to evolve. When we look at the cross-correlation between behavior and environment in our model, we find that it can be improved when not looking at the environment instantaneously but instead at the long term trend of the environment. For individual learners, these trends are of the length of about 15 periods but for social learners, these trends are longer. For PC-PBSL at frequency 0.999, these trends are rather 20 to 25 periods long; at frequency 0.9999, 25 to 30 periods. For ITW, this finding is exacerbated. At frequency 0.99, the best cross-correlation is found for trends of roughly 100 periods length; at a frequency of 0.999 and higher, the best cross-correlation is found for trends of more than 200 periods length.

Interpreting our results in terms of finance, they suggest that investors should be expected to react more weakly to short term trends in the fundamentals and more strongly to long term trends in the fundamentals. Are these reactions under- and overreactions, respectively? When optimal behavior is fitted to the environment, we find a correlation of 0.860. Then we would indeed find that ITW, especially when frequent but not too frequent, underreacts to short-term trends and overreacts to long term trends. We would, however, also find that both individual learners and PC-PBSL almost always overreact, so this interpretation is a little shaky. What remains, though, is that the strategies, especially the social learning strategies, react more strongly to long than to short term trends; whether we deal with over- and underreactions is hard to tell, especially since there is no objective way to relate periods in our model to real time.

4.4.2.3 Relation between prices and fundamentals

The weak form of the efficient market hypothesis states that markets are unpredictable. The impossibility to make accurate forecasts of markets having been supported by abundant evidence [54, 55], this is seen as evidence for the notion that market participants are rational, because:

If all available information is accurately reflected in prices, price movements should be unpredictable.

From here, it is a short step to conclude the converse, namely that:

If price movements are unpredictable, all available information is accurately reflected in prices.

This is stated by the strong and semi-strong form of the efficient market hypothesis [54, 55]. But the first statement being true does not imply its converse to be true. Yet this converse statement is what proponents of efficient markets often imply – that is why it is called *efficient* market hypothesis. But a simple example should illustrate why the converse is false. If all buyers and sellers would base their decisions on a random number generator, price movements would be unpredictable but prices would still not accurately reflect all the available information, or any.

Certainly, nobody suggests that all market participants just act randomly.[1] Another reason for unpredictability could be social influence. Very small differences in initial values, unmeasurable under most circumstances, can lead

[1] If the efficient market hypothesis were true, though, market participants could as well act completely randomly, since there would be nothing to gain or to lose, on average.

to completely opposite outcomes when social influence is involved [75]. When informational cascades occur, one new piece of information can completely reverse them and lead to the opposite result [16]. Social learning could thus be the reason why price movements are unpredictable [159].

Whether markets accurately reflect fundamental values is, however, also dependent on the observed time scale. It has been found that in the long run, real stock prices have very high correlation with real dividends, while in the short run, correlation is low. Using S&P stock prices for the time of 1871 to 1987, Shiller found an R^2 of 0.637 on an annual basis, and for the time of 1951 to 1987, of 0.819 [157]. Yet when he regressed stock prices on contemporaneous changes in real dividends, R^2 fell to 0.301 for the period of 1871 to 1987 and to 0.231 for the period of 1951 to 1987. Prices thus seem to well reflect long term trends of fundamentals but not short term trends.

In our model, we tested how well it is possible to infer the environmental state from observing the behavior. We found that for ITW, especially at high frequencies, inferences were quite bad. If, however, inferences are not made about the contemporaneous environment but rather about long term trends in the environment, this was not necessarily the case. For example, for ITW at frequency 0.99, instantaneous inferences were only right about 69% of the times; when looking at long term trends of about 85 periods length, however, inferences were right more than 90% of the times. More extremely, for ITW at a frequency of 0.999, instantaneous inferences were only right about 61% of the times but inferences of trends with a length of 230 periods were right more than 85% of the times. When ITW is dominant in the population, one should thus expect that in the short run, there is little relation between behavior and environment, whereas in the long run, there is a lot of relation between behavior and environment.

In finance, behavior is reflected in prices and our model's environment corresponds to the fundamental value of a security. Therefore, our findings have similarities with the findings from finance reported above. There are some caveats, though. Apart from not modeling real markets with all their feedback mechanisms, our model's time scale is not explicit. Therefore, we cannot know whether long term trends in our model are long term trends in finance. Moreover, with PC-PBSL, we did not find the phenomena related to short and long term trends. Therefore, it crucially depends on which social learning strategy is played whether phenomena as those we described should be expected.

4.4.2.4 Overmatching

Social learning in general cannot be condemned. Indeed, in our model, some forms of social learning lead to an increase in performance – this is why they are expected to rise in frequency. At least transiently, social learning can improve upon the possibilities of individual learning. Depending on which strategy is involved, one may find higher performance even in equilibrium, as we showed in the previous chapter. But when could social learning actually being harmful?

As we found in the first chapter, if the main advantage of social learning is saving the costs of individual learning, performance will be worse when social learners reach equilibrium. Furthermore, the evolution of the frequency of social learning strategies takes place on a similar time scale as the adaptations to the changing environment. Therefore, when circumstances are right, one may transiently find too much social learning. For example, when one option is better than the other for an unusually long stretch of time, imitation, even pushed to the extreme, may perform better than individual learning.

To illustrate this, we simulated a population of conformists and individual learners that has reached equilibrium. The resulting frequencies show a decent amount of fluctuation around the equilibrium, as can be seen in the left panel of figure 4.13. These fluctuations are, however, not purely random. Instead, conformists tend to do better when the environment greatly favors one option over the other during a generation. This is illustrated in the right panel of figure 4.13. When the mean absolute difference between p_A and p_B is especially high during one generation, meaning that one option was much better, fitness of conformists is also especially high. In fact, a 0.1 percentage point increase in the mean absolute difference of p_A and p_B leads to an increase of 1.89 in fitness (95% confidence bounds: 1.60 - 2.19). In contrast, individual learners gain less from constancy, only a 1.21 (0.95 - 1.45) increase in fitness for a 0.1 percentage point increase in absolute mean $p_A - p_B$. Conformists thus tend to become more frequent in time of constancy and vice versa, which is consistent with earlier findings [177].

Another risk posed by excessive social learning is complete loss of feedback from the environment. In the last chapter, we discussed the possibility of such informational breakdowns when equilibrium frequencies of dependent social learning strategies are extremely high. Our results suggested that for finite populations, informational breakdown should be expected under most circumstances. Therefore, the existence of social learning, while generally positive for performance, is sometimes predicted to cause peril.

History is ripe with examples of financial bubbles [30, 117, 155, 165, 168], ranging from tulip mania in 17th century Holland to the recent subprime crisis.[2] Bubbles are marked by a constant rise in the price of some asset, be it tulip bulbs or stocks of dotcom companies. For a long stretch of time, being bullish is the right thing to do, as imitating those who made a fortune indeed yields the best returns. All market participants act in a coordinated fashion, even if this means that tulip bulbs become worth a month's wage, but this detachment from fundamental values did not seem to have an influence on

[2]Proponents of the efficient market hypothesis obviously deny the existence of bubbles.

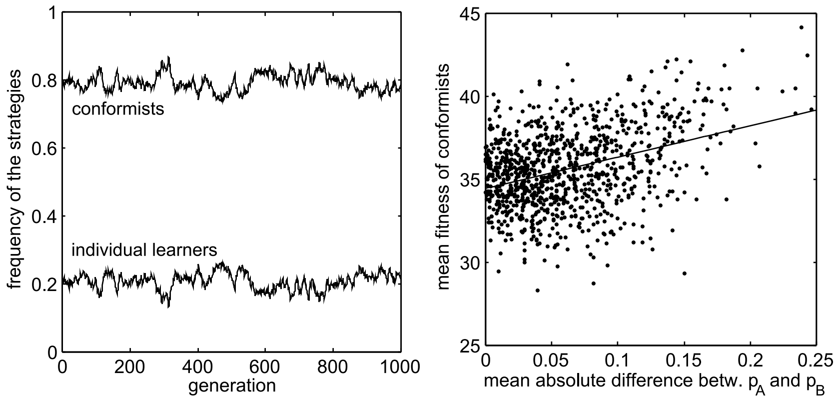

Figure 4.13: Left: Frequency of conformists and individual learners as derived from simulating 1000 generations. Frequencies fluctuate around the mean. Right: Mean fitness of conformists as a function of the absolute mean value of $p_A - p_B$ during a generation. Data from the same simulation. As the linear fit suggests, there is a positive relationship. This means that when, during a generation, one option is constantly better than the other, conformists perform especially well; if the options perform about equally well during a generation, conformists perform less well.

decision making of the market participants.

We found that in our model, contrarian views are especially likely to be expressed by individual learners but are crowded out if social learning is the dominant form of learning. Only when the detachment from the fundamental values of the asset becomes unsustainable do bubbles collapse. It could be argued that the typical downswings after bubbles burst are also bubbles, albeit bearish ones.

4.4.2.5 Social learning and the subprime crisis

Many a word has been written about the putative causes of the recent subprime crisis, which culminated in 2008. The truth is probably that many causes were at work and that the absence of any one of the causes may have prevented the crisis or would at least have mitigated the grave consequences. We will focus on one of the culprits of the crisis that is especially related to imitation but has been given little attention.

A new financial instrument, an asset-backed security going by the name of Collateralized Debt Obligations, was at the center of the crisis. CDOs mostly consist of bundles of mortgages and are divided in tranches of different risks. Even if mortgages were very risky ("subprime"), as CDOs consisted of many mortgages, the risk of all of them defaulting at the same time was thought to be low. Especially the most secure CDO tranches, which are paid first in case of default, should bear very little risk. However, the exact default risk of a CDO tranche depends crucially on the correlation between default

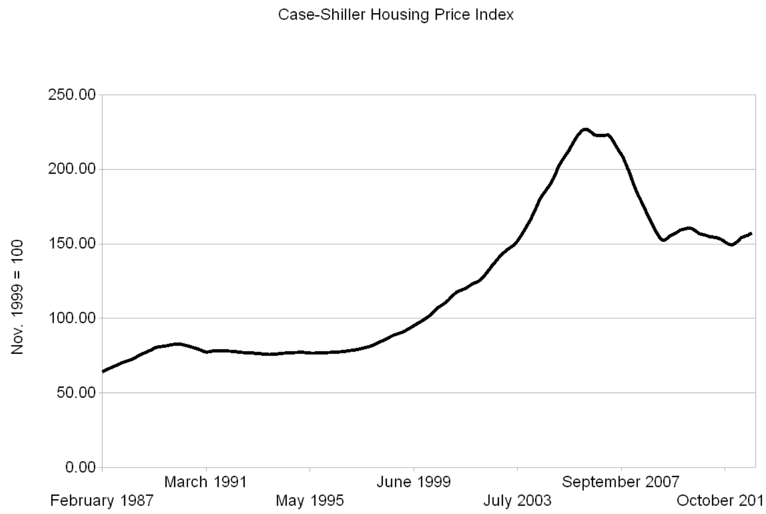

Figure 4.14: Case-Shiller Housing Price Index, Nov. 1999 = 100. Housing prices show signs of a bubble that burst in 2007. Data from Standard and Poor's, http://bit.ly/rH42Ep (as of Feb. 2013).

risks. If mortgage defaults were completely uncorrelated, risks would almost be non-existent, but if they were completely correlated, risks would be much higher than expected, especially for the safest CDO tranches.

CDOs were responsible for most of the write-downs by the affected banks. In 2007, Citigroup and Merrill Lynch had to write down a combined $48.5 billion, mostly because of CDOs on their accounts [169]; until February 2009, total write-downs due to CDOs accumulated to $218.2 billion [15]. These write-downs were caused by CDOs defaulting completely or losing almost all of their market value.

Why were the values of CDOs so blatantly miscalculated? Of course, there was the housing bubble (figure 4.14). But this would not have caused any trouble if the default risk of mortgages had been accurately reflected in the pricing of CDOs. We have to look elsewhere to understand the causes of the misjudgments.

To estimate default risks and their correlations, derivatives by the name of Credit Default Swaps (CDSs) were used as a measure. CDSs can be seen as insurances on the default risk, or as bets on the probability of default. The cost of a CDS for a certain mortgage should reflect the probability of a default of that mortgage, and the correlation between CDS prices should reflect the correlation between default risks. Using CDS prices, one can therefore hypothetically infer the correct pricing of CDOs without ever having to look at historical default rates [114]. If default risks rise, CDS prices should anticipate this and therefore the risk on CDOs could be correctly re-assessed. This approach to calculating the values of CDOs has been excessively practiced on Wall Street [149].

The twist of the story is of course that the calculations were totally wrong. In fact, default rates of the safest, AAA rated CDO tranches were 200 times higher than predicted by the models [161]. What links this story to social learning is that instead of doing their own research about default risks, bankers tasked with pricing CDOs relied on the judgment of others, as reflected in the CDS prices. This is a convenient form of social learning, since neither does it require bankers to do the tedious work of studying historical data on default risks, nor does it involve researching market fundamentals.

As this chapter has shown, a high amount of social learning is accompanied by exaggeration, delayed responses, and, transiently, detachment from reality. During the subprime crisis, statistical input from real world data has scarcely been present; instead market participants relied upon each other to make accurate judgments. This should be seen as one of the causes for the financial crisis.

4.4.2.6 Conclusion

In the first three chapters of this work, we studied social learning mainly from an evolutionary perspective. Social learning is crucial to explain cultural practices, which in turn allowed humans to successfully spread to almost all terrestrial habitats on Earth. Theoretical findings do, however, suggest that simple forms of social learning should not contribute to human adaptedness. Although solutions to this problem have been proposed, our detailed analyses showed that they do not actually solve it. Although we found that some strategies that we proposed were sometimes able to explain human adaptedness, this was only true in narrow cases. The research quest is thus still going on.

It could be argued that all these question have are interesting from an academic standpoint but have no consequences for our modern societies. However, understanding what makes us human requires us to take an evolutionary perspective. In the last chapter, we showed that if we take the evolution of social learning seriously, we should expect different behaviors in populations than if individual learning dominated. These differences could parsimoniously explain many phenomena that are hard to explain otherwise, for example market failures in the domain of finance, which we know to have grave consequences for economies and societies. Studying social learning and its origins could therefore be more necessary, and the lessons we draw more valuable, than ever in human history.

Acknowledgments

I would like to thank all who contributed to this work. Discussions with Pantelis Analytis, Victor Anaya, Ruben Arslan, Robert Boyd, Sofia Figueiredo, Christa Finkenwirth, Matthias Flor, Victor Greiff, Hanspeter Herzel, Arnulf Köhncke, Martina Michalikova, Bernhard Völkl, Adam Wilkins, Roman Zug and probably others whom I shamefully forgot to mention helped me in thinking about and writing my work. My father helped me with the German abstract. I thank Ngoc Schneeweiß and Jan Weber for proofreading. I am grateful to Peter Hammerstein and Ole Jann for collaborating with me on this topic.

Bibliography

[1] A.A. Alchian. Uncertainty, evolution, and economic theory. *The Journal of Political Economy*, 58(3):211–221, 1950.

[2] J. Apesteguia, S. Huck, and J. Oechssler. Imitation–theory and experimental evidence. *Journal of Economic Theory*, 136(1):217–235, 2007.

[3] K.J. Arrow. *Social Choice And Invidual Values*, volume 12. Yale University Press, 1963.

[4] S.E. Asch. Effects of group pressure upon the modification and distortion of judgments. *Groups, leadership, and men*, pages 177–190, 1951.

[5] S.E. Asch. Opinions and social pressure. *Scientific American*, 193:31–35, 1955.

[6] R. Axelrod. *The Evolution Of Cooperation*. Basic Book, 1984.

[7] R. Axelrod and W.D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390, 1981.

[8] R. Baldini. Success-biased social learning: Cultural and evolutionary dynamics. *Theoretical Population Biology*, 82:222–228, 2012.

[9] A. Bandura. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of personality and social psychology*, 1(6):589–595, 1965.

[10] A.V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[11] N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of Financial Economics*, 49(3):307–343, 1998.

[12] J.A. Bargh and T.L. Chartrand. The unbearable automaticity of being. *American psychologist*, 54(7):462, 1999.

[13] M. Bateson, D. Nettle, and G. Roberts. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters*, 2(3):412–414, 2006.

[14] W.M. Baum. Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior*, 32(2):269–281, 1979.

[15] E. Benmelech and J.L. Dlugosz. The credit rating crisis. *NBER Macroeconomic Annual 2009*, pages 161–207, 2009.

[16] S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy*, 100(5):992–1026, 1992.

[17] S. Bikhchandani, D. Hirshleifer, and I. Welch. Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12(3):151–170, 1998.

[18] S.A.J. Birch, S.A. Vauthier, and P. Bloom. Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, 107(3):1018–1034, 2008.

[19] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 3):7280–7287, 2002.

[20] J.C. Borda. Mémoire sur les élections au scrutin. *Histoire de l'académie royale des sciences*, 2, 1781.

[21] R. Boyd and P.J. Richerson. Cultural transmission and the evolution of cooperative behavior. *Human Ecology*, 10(3):325–351, 1982.

[22] R. Boyd and P.J. Richerson. *Culture and the evolutionary process.* University of Chicago Press, 1985.

[23] R. Boyd and P.J. Richerson. Why does culture increase human adaptability? *Ethology and Sociobiology*, 16(2):125–143, 1995.

[24] R. Boyd and P.J. Richerson. Why culture is common, but cultural evolution is rare. In *Proceedings - British Academy*, volume 88, pages 77–94. Oxford University Press, 1996.

[25] 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.

[26] R. Boyd, P.J. Richerson, and J. Henrich. The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10918, 2011.

[27] R. Boyd, P.J. Richerson, and J. Henrich. Rapid cultural adaptation can facilitate the evolution of large-scale cooperation. *Behavioral ecology and sociobiology*, 65(3):431–444, 2011.

[28] C. Camerer and R.H. Thaler. Anomalies: Ultimatums, dictators and manners. *The Journal of Economic Perspectives*, 9(2):209–219, 1995.

[29] A. Caplin, M. Dean, and D. Martin. Search and satisficing. Technical report, NYU working paper, 2009.

[30] J. Carswell. *The South Sea Bubble.* Cresset Press, 1960.

[31] L.L. Cavalli-Sforza and M.W. Feldman. *Cultural transmission and evolution: A quantitative approach.* Princeton University Press, 1981.

[32] Y. Chen. Herd behavior in purchasing books online. *Computers in Human Behavior*, 24(5):1977–1992, 2008.

[33] M. Chudek, S. Heller, S. Birch, and J. Henrich. Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior*, 2011.

[34] N. Claidière, M. Bowler, and A. Whiten. Evidence for weak or linear conformity but not for hyper-conformity in an everyday social learning context. *PloS one*, 7(2):e30970, 2012.

[35] G. Clark. *A farewell to alms: a brief economic history of the world.* Princeton University Press, 2008.

[36] M. Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix.* L'imprimerie royale, 1785.

[37] L. Corazzini and B. Greiner. Herding, social preferences and (non-) conformity. *Economics Letters*, 97(1):74–80, 2007.

[38] C. Darwin. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. *New York: D. Appleton*, 1859.

[39] R. Dawkins. *The blind watchmaker: Why the evidence of evolution reveals a universe without design.* WW Norton & Company, 1986.

[40] J.B. De Long, A. Shleifer, L.H. Summers, and R.J. Waldmann. Noise trader risk in financial markets. *The Journal of Political Economy*, 98(4):703–738, 1990.

[41] J.B. De Long, A. Shleifer, L.H. Summers, and R.J. Waldmann. Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance*, 45(2), 1990.

[42] M. Deutsch and H.B. Gerard. A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, 51(3):629, 1955.

[43] J.M. Diamond. *Collapse: How societies choose to fail or succeed*. Penguin Group USA, 2006.

[44] W. Doyle. *The Oxford History of the French Revolution*. OUP Oxford, 2002.

[45] A. Dreber, D.G. Rand, D. Fudenberg, and M.A. Nowak. Winners don't punish. *Nature*, 452(7185):348–351, 2008.

[46] C. Efferson, R. Lalive, P.J. Richerson, R. McElreath, and M. Lubell. Conformists and mavericks: the empirics of frequency-dependent cultural transmission. *Evolution and Human Behavior*, 29(1):56–64, 2008.

[47] C. Efferson, P.J. Richerson, R. McElreath, M. Lubell, E. Edsten, T.M. Waring, B. Paciotti, and W. Baum. Learning, productivity, and noise: an experimental study of cultural transmission on the bolivian altiplano. *Evolution and Human Behavior*, 28(1):11–17, 2007.

[48] M. Ehn and K.N. Laland. Adaptive strategies for cumulative cultural learning. *Journal of theoretical biology*, 2012.

[49] G. Ellison and D. Fudenberg. Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, 110(1):93, 1995.

[50] M. Enquist, K. Eriksson, and S. Ghirlanda. Critical social learning: A solution to rogers's paradox of nonadaptive culture. *American Anthropologist*, 109(4):727–734, 2007.

[51] M. Enquist and S. Ghirlanda. Evolution of social learning does not explain the origin of human cumulative culture. *Journal of Theoretical Biology*, 246(1):129–135, 2007.

[52] K. Eriksson, M. Enquist, and S. Ghirlanda. Critical points in current theory of conformist social learning. *Journal of Evolutionary Psychology*, 5(1):67–87, 2007.

[53] A. Falk and U. Fischbacher. A theory of reciprocity. *Games and Economic Behavior*, 54(2):293–315, 2006.

[54] E.F. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of finance*, 25(2):383–417, 1970.

[55] E.F. Fama. Efficient capital markets: Ii. *Journal of finance*, pages 1575–1617, 1991.

[56] J.J. Faria, S. Krause, and J. Krause. Collective behavior in road crossing pedestrians: the role of social information. *Behavioral Ecology*, 21(6):1236–1242, 2010.

[57] E. Fehr and C.F. Camerer. Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11(10):419–427, 2007.

[58] E. Fehr and S. Gächter. Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4):980–994, 2000.

[59] E. Fehr and S. Gächter. Cooperation and punishment in public goods experiments. *American Economic Review*, 90(4):980–994, 2000.

[60] E. Fehr and K.M. Schmidt. A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868, 1999.

[61] M.W. Feldman, K. Aoki, and J. Kumm. Individual versus social learning: evolutionary analysis in a fluctuating environment. *Anthropological Science*, 104:209–232, 1996.

[62] R.A. Fisher. 1930the genetical theory of natural selection, 1930.

[63] M. Franz and C.L. Nunn. Rapid evolution of social learning. *Journal of Evolutionary Biology*, 22(9):1914–1922, 2009.

[64] M. Friedman. *A Theory of the Consumption.* princeton university press Princeton, 1957.

[65] S. Gächter and B. Herrmann. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):791–806, 2009.

[66] B.G. Galef, L.M. McQuoid, and E.E. Whiskin. Further evidence that norway rats do not socially transmit learned aversions to toxic baits. *Learning & behavior*, 18(2):199–205, 1990.

[67] G. Gigerenzer. Fast and frugal heuristics: The tools of bounded rationality. *Blackwell handbook of judgment and decision making*, pages 62–88, 2004.

[68] H. Gintis. *Game theory evolving: A problem-centered introduction to modeling strategic behavior.* Princeton University Press, 2000.

[69] L.A. Giraldeau and T. Caraco. *Social foraging theory.* Princeton Univ Pr, 2000.

[70] L.A. Giraldeau, T.J. Valone, and J.J. Templeton. Potential disadvantages of using socially acquired information. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1427):1559–1566, 2002.

[71] J.K. Goeree and L. Yariv. Conformity in the lab. *California Institute of Technology Working Paper*, 2007.

[72] D.E. Goldberg. Probability matching, the magnitude of reinforcement, and classifier system bidding. *Machine Learning*, 5(4):407–425, 1990.

[73] D.G. Goldstein and G. Gigerenzer. *The recognition heuristic: How ignorance makes us smart.* Oxford University Press, 1999.

[74] D.G. Goldstein and G. Gigerenzer. Models of ecological rationality: The recognition heuristic. *Psychological review*, 109(1):75–90, 2002.

[75] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, pages 1420–1443, 1978.

[76] S.J. Grossman and J.E. Stiglitz. Information and competitive price systems. *The American Economic Review*, 66(2):246–253, 1976.

[77] S.J. Grossman and J.E. Stiglitz. On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3):393–408, 1980.

[78] J. Haidt. Morality. *Perspectives on Psychological Science*, 3(1):65–72, 2008.

[79] W.D. Hamilton. The genetical evolution of social behaviour. i. *Journal of theoretical biology*, 7(1):1–16, 1964.

[80] P. Hammerstein. *Why is reciprocity so rare in social animals? A protestant appeal.* MIT Press, 2003.

[81] P. Hammerstein and E.H. Hagen. The second wave of evolutionary economics in biology. *Trends in ecology & evolution*, 20(11):604–609, 2005.

[82] G. Hardin. The tragedy of the commons. *Journal of Natural Resources Policy Research*, 1(3):243–253, 1968.

[83] R.M. Harding, SM Fullerton, RC Griffiths, J. Bond, M.J. Cox, J.A. Schneider, D.S. Moulin, and JB Clegg. Archaic african and asian lineages in the genetic ancestry of modern humans. *American journal of human genetics*, 60(4):772, 1997.

[84] D. Haun, Y. Rekers, and M. Tomasello. Majority-biased transmission in chimpanzees and human children, but not orangutans. *Current Biology*, 2012.

[85] F.A. Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945.

[86] J. Henrich and R. Boyd. The evolution of conformist transmission and the emergence of between-group differences. *Evolution and human behavior*, 19(4):215–241, 1998.

[87] 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.

[88] J. Henrich, S.J. Heine, and A. Norenzayan. Most people are not weird. *Nature*, 466(7302):29–29, 2010.

[89] J. Henrich, S.J. Heine, and A. Norenzayan. The weirdest people in the world. *Behavioral and Brain Sciences*, 33(2-3):61–83, 2010.

[90] Jo. Henrich and R. McElreath. The evolution of cultural evolution. *Evolutionary Anthropology: Issues, News, and Reviews*, 12(3):123–135, 2003.

[91] R. Hertwig, G. Barron, E.U. Weber, and I. Erev. Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, 15(8):534–539, 2004.

[92] R. Hertwig and I. Erev. The description-experience gap in risky choice. *Trends in Cognitive Sciences*, 13(12):517–523, 2009.

[93] R. Hertwig and T.J. Pleskac. Decisions from experience: Why small samples? *Cognition*, 115(2):225–237, 2010.

[94] B.S. Hewlett, H.N. Fouts, A.H. Boyette, and B.L. Hewlett. Social learning among congo basin hunter–gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1168–1178, 2011.

[95] S.E. Hill, D.J. DelPriore, and P.W. Vaughan. The cognitive consequences of envy: Attention, memory, and self-regulatory depletion. *Journal of personality and social psychology*, 101(4):653, 2011.

[96] S. Huck, H.T. Normann, and J. Oechssler. Learning in Cournot oligopoly–An experiment. *The Economic Journal*, 109(454):80–95, 1999.

[97] L.P. Kaelbling, M.L. Littman, and A.W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[98] D. Kahneman, J.L. Knetsch, and R.H. Thaler. Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives*, 5(1):193–206, 1991.

[99] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

[100] T. Kameda and D. Nakanishi. Cost-benefit analysis of social/cultural learning in a nonstationary uncertain environment: An evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior*, 23(5):373–393, 2002.

[101] T. Kameda and D. Nakanishi. Does social/cultural learning increase human adaptability? Rogers's question revisited. *Evolution and Human Behavior*, 24(4):242–260, 2003.

[102] Y. Kareev. Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review*, 107(2):397–402, 2000.

[103] J. Kendal, L.A. Giraldeau, and K. Laland. The evolution of social learning rules: Payoff-biased and frequency-dependent biased transmission. *Journal of theoretical biology*, 260(2):210–219, 2009.

[104] V. Klucharev, K. Hytoenen, M. Rijpkema, A. Smidts, and G. Fernández. Reinforcement learning signal predicts social conformity. *Neuron*, 61(1):140–151, 2009.

[105] F.H. Knight. *Risk, Uncertainty, and Profit.* Houghton Mifflin Company, 1921.

[106] J.R. Krebs and N.B. Davies. *Behavioral ecology.* Wiley-Blackwell, 1997.

[107] Z. Kunda. The case for motivated reasoning. *Psychological Bulletin*, 108(3):480–498, 1990.

[108] K.N. Laland. Social learning strategies. *Learning & Behavior*, 32(1):4–14, 2004.

[109] D.F. Lancy. Learning'from nobody': The limited role of teaching in folk models of children's development. *Childhood in the Past*, 3(1):79–106, 2010.

[110] W. Lee. *Decision theory and human behavior.* John Wiley & Sons, 1971.

[111] L. Lehmann, MW Feldman, and R. Kaeuffer. Cumulative cultural dynamics and the coevolution of cultural innovation and transmission: an ess model for panmictic and structured populations. *Journal of Evolutionary Biology*, 23(11):2356–2369, 2010.

[112] S.D. Levitt and J.A. List. What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 21(2):153–174, 2007.

[113] S. Levy. *In the plex: How Google thinks, works, and shapes our lives.* Simon and Schuster, 2011.

[114] D.X. Li. On default correlation: a copula function approach. *Journal of Fixed Income*, 9(4):43–54, 2000.

[115] C.G. Lord, L. Ross, and M.R. Lepper. Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11):2098–2101, 1979.

[116] R.D. Luce and H. Raiffa. *Games and decisions: Introduction and critical survey.* Dover Publications, 1989 (1957).

[117] C. Mackay. *Extraordinary Popular Delusions and the Madness of Crowds.* Barnes & Noble, 2004 (1841).

[118] N.J. Mackintosh. *The psychology of animal learning.* Academic Press, 1974.

[119] R. McElreath. The coevolution of genes, innovation and culture in human evolution. In Joan Silk and Peter Kappeler, editors, *Mind The Gap: Tracing the Origins of Human Universals*, pages 451–474. Springer, 2010.

[120] R. McElreath, A.V. Bell, C. Efferson, M. Lubell, P.J. Richerson, and T. Waring. Beyond existence and aiming outside the laboratory: estimating frequency-dependent and pay-off-biased social learning strategies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1509):3515–3528, 2008.

[121] R. McElreath, B. Fasolo, and A. Wallin. The evolutionary rationality of social learning. In Ralph Hertwig, Ulrich Hoffrage, and the ABC Research Group, editors, *Simple Heuristics in a Social World*, pages 381–408. Oxford University Press, 2013.

[122] R. McElreath, M. Lubell, P.J. Richerson, T.M. Waring, W. Baum, E. Edsten, C. Efferson, and B. Paciotti. Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior*, 26(6):483–508, 2005.

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

[124] A. Mesoudi. *Cultural evolution: how Darwinian theory can explain human culture and synthesize the social sciences*. University of Chicago Press, 2011.

[125] A. Mesoudi and S.J. Lycett. Random copying, frequency-dependent copying and culture change. *Evolution and Human Behavior*, 30(1):41–48, 2009.

[126] A. Mesoudi and M.J. O'Brien. The cultural transmission of great basin projectile-point technology i: an experimental simulation. *American Antiquity*, pages 3–28, 2008.

[127] A. Mesoudi, A. Whiten, and K.N. Laland. Towards a unified science of cultural evolution. *Behavioral and Brain Sciences*, 29(4):329–346, 2006.

[128] G.A. Miller. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2):81, 1956.

[129] T.J.H. Morgan, L.E. Rendell, M. Ehn, W. Hoppitt, and K.N. Laland. The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*, 2011.

[130] J. von Neumann and O. Morgenstern. Theory of games and economic behavior. *Princeton University, Princeton*, 1947.

[131] M.E.J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[132] R.E. Nisbett and T.D. Wilson. Telling more than we can know: Verbal reports on mental processes. *Psychological review*, 84(3):231–259, 1977.

[133] M. Nowak and K. Sigmund. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature*, 364(6432):56–58, 1993.

[134] M.A. Nowak, C.E. Tarnita, and E.O. Wilson. The evolution of eusociality. *Nature*, 466(7310):1057–1062, 2010.

[135] T. Offerman and J. Sonnemans. Learning by experience and learning by imitating successful others. *Journal of Economic Behavior & Organization*, 34(4):559–575, 1998.

[136] T. Ohta. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, 23:263–286, 1992.

[137] E. Ostrom. Coping with tragedies of the commons. *Annual review of political science*, 2(1):493–535, 1999.

[138] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[139] C. Perreault, C. Moya, and R. Boyd. A bayesian approach to the evolution of social learning. *Evolution and Human Behavior*, 2012.

[140] K.L. Powell, G. Roberts, and D. Nettle. Eye images increase charitable donations: Evidence from an opportunistic field experiment in a supermarket. *Ethology*, 2012.

[141] L. Rendell, R. Boyd, D. Cownden, M. Enquist, K. Eriksson, MW Feldman, L. Fogarty, S. Ghirlanda, T. Lillicrap, and KN Laland. Why copy others? insights from the social learning strategies tournament. *Science*, 328(5975):208, 2010.

[142] L. Rendell, L. Fogarty, W.J.E. Hoppitt, T.J.H. Morgan, M.M. Webster, and K.N. Laland. Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences*, pages 68–76, 2011.

[143] L. Rendell, L. Fogarty, and K.N. Laland. Rogers' paradox recast and resolved: population structure and the evolution of social learning strategies. *Evolution*, 64(2):534–548, 2010.

[144] P.J. Richerson and R. Boyd. *Not by genes alone: How culture transformed human evolution.* University of Chicago Press, 2005.

[145] H. Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[146] A.R. Rogers. Does biology constrain culture? *American Anthropologist*, 90(4):819–831, 1988.

[147] E.M. Rogers. *Diffusion of innovations.* Free Press, 1995 (1962).

[148] B. Ryan and N.C. Gross. The diffusion of hybrid seed corn in two iowa communities. *Rural sociology*, 8(1):15–24, 1943.

[149] F. Salmon. Recipe for disaster: The formula that killed Wall Street. *Wired Magazine*, 17(3):17–03, 2009.

[150] D.S. Scharfstein and J.C. Stein. Herd behavior and investment. *The American Economic Review*, pages 465–479, 1990.

[151] K.H. Schlag. Why imitate, and if so, how? a bounded rational approach to multi-armed bandits. *Journal of economic theory*, pages 130–156, 1996.

[152] K.H. Schlag. Which one should I imitate? *Journal of Mathematical Economics*, 31(4):493–522, 1999.

[153] E. Shafir and A. Tversky. Thinking through uncertainty: Nonconsequential reasoning and choice* 1. *Cognitive Psychology*, 24(4):449–474, 1992.

[154] M. Sherif. A study of some social factors in perception. *Archives of Psychology (Columbia University)*, pages 1–60, 1935.

[155] R. Shiller. *Irrational Exuberance*. Princeton University Press, 2000.

[156] R.J. Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? *American Economic Review*, 71(3):421–436, 1981.

[157] R.J. Shiller. Market volatility and investor behavior. *The American Economic Review*, 80(2):58–62, 1990.

[158] R.J. Shiller. From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, 17(1):83–104, 2003.

[159] R.J. Shiller, S. Fischer, and B.M. Friedman. Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 1984(2):457–510, 1984.

[160] R.J. Shiller and J. Pound. Survey evidence on diffusion of interest and information among investors. *Journal of Economic Behavior & Organization*, 12(1):47–66, 1989.

[161] N. Silver. *The Signal and the Noise: Why Most Predictions Fail–But Some Don't*. Penguin Press, 2012.

[162] J.M. Smith. *The evolution of sex*. Cambridge University Press, 1978.

[163] J.M. Smith and E. Szathmary. *The major transitions in evolution*. Oxford University Press, 1997.

[164] J. Soltis, R. Boyd, and P.J. Richerson. Can group-functional behaviors evolve by cultural group selection?: An empirical test. *Current Anthropology*, 36(3):473–494, 1995.

[165] J.B. Stewart. *Den of Thieves.* Touchstone, 1991.

[166] J. Surowiecki. *The wisdom of crowds.* Anchor, 2005.

[167] N. Takahata. Allelic genealogy and human evolution. *Molecular Biology and Evolution*, 10(1):2–22, 1993.

[168] G. Tett. *Fool's gold: how unrestrained greed corrupted a dream, shattered global markets and unleashed a catastrophe.* Little Brown GBR, 2009.

[169] The Financial Crisis Inquiry Commission. The financial crisis inquiry report. 2011.

[170] R.L. Trivers. The evolution of reciprocal altruism. *Quarterly review of biology*, pages 35–57, 1971.

[171] A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

[172] C. Ungemach, N. Chater, and N. Stewart. Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, 20(4):473–479, 2009.

[173] F. Vega-Redondo. The evolution of Walrasian behavior. *Econometrica: Journal of the Econometric Society*, 65(2):375–384, 1997.

[174] J.Y. Wakano, K. Aoki, and M.W. Feldman. Evolution of social learning: a mathematical analysis. *Theoretical population biology*, 66(3):249–258, 2004.

[175] J.W. Weibull. *Evolutionary game theory.* The MIT press, 1995.

[176] G. Weizsäcker. Do we follow others when we should? a simple test of rational expectations. *American Economic Review*, pages 2340–2360, 2010.

[177] H. Whitehead and P.J. Richerson. The evolution of conformist social learning can cause population collapse in realistically variable environments. *Evolution and Human Behavior*, 30(4):261–273, 2009.

[178] A. Whiten, N. McGuigan, S. Marshall-Pescini, and L.M. Hopper. Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1528):2417–2428, 2009.

[179] R.A. Williamson and A.N. Meltzoff. Own and others' prior experiences influence children's imitation of causal acts. *Cognitive Development*, 26(3):260–268, 2011.

[180] R.A. Williamson, A.N. Meltzoff, and E.M. Markman. Prior experiences and perceived efficacy influence 3-year-olds' imitation. *Developmental Psychology*, 44(1):275–285, 2008.

[181] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931.

[182] H.P. Young. Social choice scoring functions. *SIAM Journal on Applied Mathematics*, pages 824–838, 1975.

[183] H.P. Young. The diffusion of innovations in social networks. *Economy as an Evolving Complex System. Proceedings Volume in the Santa Fe Institute Studies in the Sciences of Complexity*, 3:267–282, 2002.

[184] H.P. Young. Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *The American economic review*, 99(5):1899–1924, 2009.

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den Benjamin Vinh Bossan