Conformists and Mavericks: The Empirics of Frequency-Dependent Cultural Transmission*

Charles Efferson[1,2,4], Rafael Lalive[2,3], Peter J. Richerson[4,5], Richard McElreath[4,6], and Mark Lubell[4,5]

[1]Santa Fe Institute

[2]Institute for Empirical Research in Economics, University of Zürich

[3]Department of Econometrics and Political Economy, University of Lausanne

[4]Graduate Group in Ecology, University of California, Davis

[5]Environmental Science and Policy, University of California, Davis

[6]Anthropology, University of California, Davis

August 7, 2007

Running title: Conformists and Mavericks

Word count: ≈ 5000

***Corresponding author:** Charles Efferson, Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, *cmefferson@santafe.edu.*

Abstract: Conformity is a type of social learning that has received considerable attention among social psychologists and human evolutionary ecologists, but existing empirical research does not identify conformity cleanly. Conformity is more than just a tendency to follow the majority; it involves an exaggerated tendency to follow the majority. The "exaggerated" part of this definition ensures that conformists do not show just any bias toward the majority, but a bias sufficiently strong to increase the size of the majority through time. This definition of conformity is compelling because it is the only form of frequency-dependent social influence that produces behaviorally homogeneous social groups. We conducted an experiment to see if players were conformists by separating individual and social learners. Players chose between two technologies repeatedly. Payoffs were random, but one technology had a higher expected payoff. Individual learners knew their realized payoffs after each choice, while social learners only knew the distribution of choices among individual learners. A subset of social learners behaved according to a classic model of conformity. The remaining social learners did not respond to frequency information. They were neither conformists nor non-conformists, but mavericks. Given this heterogeneity in learning strategies, a tendency to conform increased earnings dramatically.

Key words: cultural transmission, conformity, social learning

1 Introduction

Social scientists have long recognized the importance of frequency-dependent social learning (Asch, 1956; Boyd and Richerson, 1985; Henrich and McElreath, 2003; Bowles, 2004; Boyd and Richerson, 2005; Lumsden and Wilson, 1980; Richerson and Boyd, 2005; Sherif and Murphy, 1936). Frequency-dependent social learning postulates that individuals adopt a given behavior with a probability that varies in response to how common the behavior is in a relevant social group. Conformity is a type of frequency dependence that has received considerable attention. As formally defined (Boyd and Richerson, 1982, 1985), conformity is based on the following proposition. In a simple case with two behaviors, R and B, where r_t is the frequency of R in the population, conformity means that in the near future an individual exhibits behavior R with a probability less than r_t if $r_t < 1/2$ but greater than r_t if $r_t > 1/2$. In other words, individuals do not simply follow the majority; rather they show a *disproportionate* tendency to follow the majority. They over-respond, so to speak, to frequency information. This feature of conformity is crucial because, as we show below, it homogenizes behavior within social groups. Other types of frequency dependence do not have this effect.

In spite of conformity's acknowledged importance, previous empirical research cannot identify conformity as a disproportionate tendency to follow the majority. Classic research in social psychology (Asch, 1955, 1956; Aronson *et al.*, 2002), neuroscience experiments (Berns *et al.*, 2005) in the tradition of Asch (1956), and recent experiments with chimpanzees (Whiten *et al.*, 2005) show that a focal individual is more likely to adopt a behavior as that behavior becomes more common. A simple model of non-conformity, however, makes exactly the same prediction, as do other hypotheses about positive social influences.

The distinctions, however, among different forms of positive influence are fundamental and not simply matters of definition. Below we present a general model of frequency dependence that includes both conformity and non-conformity as special cases. Although the individual psychology is different for conformity and non-conformity, in both cases the probability a focal individual adopts a given behavior increases with the frequency of the behavior in the social group. Even so, dynamics at the group level are radically different. Conformity produces multiple steady states and can lead otherwise similar societies to evolve in completely different ways in the wake of small random effects (Bowles, 2004; Efferson and Richerson, 2007). Whatever behavioral variation may exist between groups, however, conformity produces social groups that are internally homogeneous in terms of behavior. Non-conformity, in contrast, increases behavioral variation within groups and decreases variation between groups (Efferson and Richerson, 2007). Whatever names we choose to attach to these forms of social influence, such distinctions are important. To demonstrate conformity as a force that homogenizes behavior within social groups, it is not enough to show simply that individuals adopt common behaviors. Researchers must also show that this inclination is disproportionate in the way described above. In this paper we present a jointly theoretical and experimental approach to this problem.

Apart from its intrinsic interest, conformity has figured prominently in various discussions in the behavioral and evolutionary social sciences. Theoretically, conformity can be a valuable way to make good decisions in temporally and spatially variable environments (Henrich and Boyd, 1998). Imagine that R and B are two existing technologies. Individuals would like to choose the optimal technology, but returns are stochastic. The environment also varies in space and time, and so identifying the optimal technology is not easy. Assume that individuals experiment from time to time, learn individually as a result, and this produces a slight bias toward the optimal technology. As we show below, conformity exaggerates such a bias by filtering out a lot of the noise at the individual level. A powerful signal pointing toward the optimal technology is the result. By itself, however, conformity implies nothing about the optimality of individual decisions. It only exaggerates existing biases.

In addition to decision making, conformity has the interesting theoretical property that it reduces behavioral variation within populations while potentially increasing variation among populations. All else equal, this increases the strength of selective pressures at the group level. Thus, in conjunction with the punishment of norm violations and the imitation of success, conformity plays a critical role in the study of how prosocial tendencies could have evolved in humans via cultural group selection (Boyd and Richerson, 1982; Boyd *et al.*, 2003; Fehr and Fischbacher, 2003, 2004; Fehr and Gaechter, 2002; Guererk *et al.*, 2006; Guzmán *et al.*, 2006; Henrich, 2004; Henrich and Boyd, 2001). Conformity also appears to be critical in explaining aggregate patterns that characterize the diffusion of technological innovations (Rogers, 1995; Henrich, 2001).

2 How conformity works

Theoretically, conformity can be a valuable way to make decisions under uncertainty. Importantly, however, conformity is neither good nor bad by itself. It merely exaggerates existing biases in individual decision-making. To illustrate, assume technology R is optimal. Consider a group of N individuals, each of whom chooses R in a given period with probability r_t. The probability a majority of the individuals in the social group chooses the optimal technology when N is odd is simply,

$$P(\text{majority opt}) = \sum_{i=\lceil N/2 \rceil}^{N} \binom{N}{i} r_t^i (1 - r_t)^{N-i}. \tag{1}$$

Figure 1 shows how $P(\text{majority opt})$ varies as a function of r_t for 4 different values of N.

[Figure 1 about here]

Importantly, $P(\text{majority opt}) < r_t$ if $r_t < 0.5$, but $P(\text{majority opt}) > r_t$ if $r_t > 0.5$. This fact is the essence of conformity's power to reduce noise at the individual level into a useful social signal. Conformity works by identifying the optimum disproportionately if other forces, as summarized by r_t, bias choices toward the optimum. It does not work, in the sense that it disproportionately identifies the sub-optimal technology, if other forces bias choices toward the sub-optimum. Moreover, when the social group gets larger, the amount of information embedded in the group increases, and the nonlinearity intrinsic to conformity becomes more extreme. In sum, conformity exaggerates the effectiveness of other decision-making biases like individual learning in a way that depends on both the size of the social group and the strength of the other bias. If the other bias is bad, conformity is worse. If the other bias is good, conformity is even better. Additionally, this extra bias created by conformity is more extreme for larger social groups.

The electronic supplement presents additional results formalizing how conformity works. As in the experiment described shortly, the theory focuses on situations in which the payoffs associated with different behaviors are stochastic. The best behavior is not obvious because feedback is noisy. In this case, conformity can filter noisy individual feedback into a powerful signal that points clearly toward the best behavior, but only if some other force is at work.

3 The empirics of frequency dependence

As suggested in the introduction, conformity should not be defined simply as any positive social influence. Such an approach neglects important distinctions between different types of frequency dependence, some of which produce internally homogeneous groups, others of which produce social groups that are maximally heterogeneous. Here we outline the distinctions necessary to integrate theory and empiricism in the study of frequency-dependent social learning.

3.1 Individual decisions under fixed and sampled social groups

Boyd and Richerson (1982, 1985) developed a simple model of frequency dependence with the following properties. In $t+1$ each individual in the population samples N individuals from the previous period. Define $I_t \in \{0, 1, \ldots, N\}$ as a random variable with realizations i_t specifying the number of individuals choosing R in a particular sample of size N. Given a sample with a particular distribution of behaviors, the response to social information takes the following form for a focal individual,

$$P(\text{focal ind chooses R} \,|\, i_t) = \begin{cases} i_t(1-D)/N & \text{if } i_t < N/2, \\ 1/2 & \text{if } i_t = N/2, \\ i_t(1-D)/N + D & \text{otherwise.} \end{cases} \qquad (2)$$

The parameter $D \in [-1, 1]$ controls the nature of frequency dependence. When $D \in [-1, 0)$, social learning is non-conformist, when $D = 0$ social learning is linear (Boyd and Richerson, 1985), and when $D \in (0, 1]$ it is conformist.

The probabilities specified in (2) are conditional. They take the distribution of behaviors in the social group (i_t) as given and do not account for how people form social groups. Conditional probabilities may be relevant in many experimental settings where the social group is fixed (e.g. Asch (1956)), and indeed they will be relevant for the experiments presented below, but they will not always apply. In some cases, we might imagine that individuals estimate the distribution of behaviors in some larger population by sampling behavioral models randomly. In this case, if r_t is the frequency of R in the population at t, then the number who chose R in a sample of individuals will be binomially distributed. In essence, two sources of noise are present. On the one hand, given a social group with a particular mix of behaviors, individuals may exhibit positive choice probabilities for both R and B. These are the conditional probabilities of model (2). On the other hand, if individuals estimate the distribution of behaviors in the population by sampling, samples will typically differ across individuals, and these differences represent another type of noise. Accounting for both types of noise, the unconditional probability under unbiased sampling is simply,

$$P(\text{focal ind chooses R}) = r_t(1 - D) + D \sum_{i_t = \lceil N/2 \rceil}^{N} \binom{N}{i_t} (r_t)^{i_t} (1 - r_t)^{N - i_t}. \quad (3)$$

Figure 2 illustrates the basic features of models 2 and 3.

[Figure 2 about here]

Importantly, with sampling both conformity (Fig. 2c) and non-conformity (Fig. 2d) produce monotonically increasing functions. This means that both are forms of positive social influence. Using basic techniques in non-linear dynamics (Hoy *et al.*, 2001; McElreath and Boyd, 2007), however, one can show that their aggregate effects are entirely different (Efferson and Richerson, 2007). Conformity produces behavioral homogeneity within the social group or population. In a diametrically opposed fashion, non-conformity produces the maximum amount of heterogeneity.

Without sampling conformity is monotonically increasing in the number of individuals exhibiting R (Fig. 2a). Non-conformity is monotonically increasing over two restricted intervals (Fig. 2b). So long as we restrict attention to $r_t \in [0, 1/2)$ or $r_t \in (1/2, 1]$, non-conformity posits, like conformity, that the probability a focal individual adopts R increases in the number of individuals with behavior R. This assumption is compatible with the idea that individuals show relatively small biases toward the behavior in the minority. These biases are enough to move the group away from the behavior currently in the majority, thus distinguishing non-conformity from conformity, but the biases are not especially large. For this reason, we will refer to this form of non-conformity as "weak" non-conformity. With respect to empirical studies, weak non-conformity has the following important implication. Researchers cannot cleanly identify conformity by simply showing that majorities of different sizes have a positive effect on the rate of adopting the behavior in question (e.g. Asch (1955)). In essence, this is like restricting attention to the interval $r_t \in (1/2, 1]$, and both conformity and weak non-conformity make the same qualitative prediction over this interval.

In the electronic supplement, we derive a model of "strong" non-conformity in which individuals exhibit large biases toward the behavior in the current minority. This form of non-conformity produces a model in which the probability of choosing a behavior varies *inversely* with the behavior's frequency in the population, a qualitative relation very different from any form of models (2) and (3). Nonetheless, like weak non-conformity and unlike conformity, strong non-conformity cannot produce dynamically stable, behaviorally uniform social groups.

3.2 Only conformity produces internally homogeneous groups

Whether individuals sample or not when forming a group of associates, the critical feature of conformity is that it moves the group toward the behavior in the majority at any given point in time. This is formally captured by the definition presented in the introduction. In a simple two-behavior setting, if $r_t < 1/2$, the probability a focal individual adopts R in the next period is less than r_t. If $r_t > 1/2$, the probability is greater than r_t. Thus, if R is in the minority, the minority gets progressively smaller until R disappears altogether. If R is in the majority, the majority gets progressively larger until R alone is present[1].

Contrast this scenario with the following alternatives. Linear transmission (Boyd and Richerson, 1985) means the probability a focal individual chooses R is equal to r_t for all values of r_t. In this case, the dynamics of behavior are neutral with respect to frequency-dependent social influence. If the distribution of behaviors is changing through time, it is changing for some other reason and not because people are responding in a biased way to how common behaviors are. Next we consider the two forms of non-conformity. Both assume that if $r_t < 1/2$ the probability a focal individual adopts R in the next period is greater than r_t, while if $r_t > 1/2$ the probability is less than r_t. Consequently, both forms of non-conformity involve a bias *away* from the behavior currently in the majority. In the case of weak conformity, the bias is relatively small. Dynamically, this means the group moves smoothly toward a uniform distribution of behaviors, at which point non-conformity becomes irrelevant, and the system stabilizes (Efferson and Richerson, 2007). Strong non-conformity, in contrast, creates oscillations in the sense that the behavior in the majority is constantly changing (electronic supplement). Regardless of whether the dynamics are controlled by weak or strong non-conformity, groups that are stable and homogeneous cannot result. This follows precisely because of the postulated force away from any majority that might be in place at a given point in time.

Importantly, an infinite number of models exist that do not fall into the categories outlined in this section. In particular, we ignore models that meet some of the assumptions of both conformity and non-conformity. Models of this sort imply that some other bias is at work. An example would be a group of individuals who all choose R with a constant probability of 0.8. This would mean that R is intrinsically more attractive than B for reasons that have nothing to do with R's frequency in

[1]Strictly speaking, this claim about perfect homogeneity requires that the function specifying choice probabilities includes the points (0,0) and (1,1) and is continuous in the vicinity of both points. Otherwise, the result is weaker. The group is increasingly homogeneous, but perfect homogeneity does not result.

the population. We do not mean to imply that such biases do not exist. We think they do, and they probably interact with frequency dependence. But we ignore them here to focus squarely on the theoretical properties of frequency dependence without other matters clouding the issue.

3.3 Asch's study as an illustrative example

Asch (1955, 1956) initiated a tradition of conformity research in which the experiment created a conflict between what the experimental subject perceived as true and the opinions of a unanimous majority aligned against him. Unbeknownst to the experimental subject, this unanimous majority with a contradictory opinion was really composed of confederates. A typical setting involves identifying, from a set of lines, the one line that is the same length as some reference line. The experimental subject sees the right answer, but before the subject states her own opinion the confederates each choose a line that is not the same length as the reference line. The question is whether this information will influence the subject's probability of stating a wrong answer.

Figure 3 shows the results from one of Asch's experiments in which the size of the social group, and by extension the size of the unanimous majority of confederates, was varied systematically. The social group, including the experimental subject, ranged from 2 to 16, and so here we code this as unanimous majorities ranging in size from 1 to 15. As the independent variable we show the proportion of the entire group made up of confederates initially proclaiming the wrong answer, which produces variation ranging from 1/2 to 15/16. This coding assumes that, before stating their own answers, the experimental subjects all had an opinion contradictory to the stated answers of the confederates. We do not actually know this, but it is consistent with the error rate in Asch's control sessions without social influence. This type of coding also produces the most variation in the independent variable and is thus the most favorable approach to Asch's study. The response variable is the proportion of experimental subjects who, like the confederates, also stated a wrong answer. Subjects did not sample from a larger population, and conditional choice probabilities are the relevant concept. The data are superimposed on a graph showing the region of probability space compatible with conformity.

We would like to make three points. First, the experimental data lie *entirely outside the conformity region*. Second, the data only cover the interval from 1/2 to 1. Both conformity and weak non-conformity predict a monotonic increase over this interval, and thus the data are qualitatively consistent with both models. Third, the experimental protocol created a conflict between two different biases: the information provided by the subject's senses and the subject's possible susceptibility to social influence. For this reason, we do not claim, as Figure 3 might suggest, that Asch actually found non-conformity. Rather we claim that the joint effect of conflicting biases means we cannot isolate the response to frequency information, and so we cannot conclude that Asch found conformity as we define it, namely a frequency-dependent force that produces homogeneous social groups through time. The key here is the baseline rate of errors. Asch's baseline is the error rate in the absence of social influence, an error rate close to 0 (Asch, 1955, 1956). This baseline is indeed interesting and compelling, but it is not the relevant baseline for understanding the dynamical consequences of frequency-dependent social influence. As outlined above, the relevant dynamical baseline is the current distribution of behaviors. Specifically, we need to know how the probability of choosing a behavior in the future compares to the current frequency of the behavior in the group for every conceivable distribution of behaviors. This was the objective of the experimental work to which we now turn.

4 Experimental methods

With 70 students at the University of Zürich and the Swiss Federal Institute of Technology, we conducted the following experiment. In each period each player faced a choice between one of two technologies ("red" versus "blue"). Payoffs followed truncated normal distributions, but one color was optimal in that its payoff distribution had a higher expectation. Specifically, payoffs in experimental currency units for the sup-optimal technology were distributed $N(30, 12)$, and payoffs for the optimum were distributed $N(38, 12)$. Both distributions were truncated at 0 and 68, which changed the means and standard deviations slightly, and payoffs were rounded to the nearest integer. Players did not know which color was better, but they could learn through time. The basic experimental problem was thus similar to McElreath *et al.* (2005) and Efferson *et al.* (2007)

Players made choices for six blocks of 25 periods each. Each block of 25 periods had a randomly selected optimal color, but all players who played together always had the same optimal color. All of this was explained in the instructions before beginning an experimental session. The framing of the choice task was neutral, but players were explicitly told the more often they chose the optimal color the more they would earn. In addition, participants viewed an extensive graphical demonstration before the beginning of the experiment. The demonstration produced various animated histograms that gave subjects an intuition for how random payoffs would be generated even if they did not have formal training in probability theory. The entire experiment was conducted on a local computer network using z-Tree (Fischbacher, 2007). The electronic supplement provides more details[2].

Players within a session were divided into two groups that played simultaneously. In one group of 5 players, each player individually chose one of the two colors in each period and immediately received private information about her realized payoff. These players did not have any information about other players, and so we refer to them as *individual learners*. In the other group, composed of 6-7 players, each player had social information about the distribution of choices (e.g. 3 red, 2 blue) among the individual learners. These players did not have any information about their own payoffs, and thus we refer to them as *social learners*. Social information was available after all individual learners had made their choices in a period but before a given social learner had made her choice. After communicating the social information, each social learner made a choice between the two colors privately and received a payoff. Realized payoffs, however, were never communicated to players in this group, and individual learning was consequently not possible. Social learners, however, did know that players in the associated group of individual learners were

[2] Instructions in German and/or z-Tree code are available upon request.

receiving individual feedback about payoffs after each of their choices. Immediately after the experiment, all players responded to a brief questionnaire requesting basic socio-demographic information and information about how they made decisions in the experiment. The only payoff information social learners received was their total earnings after the experiment and questionnaire were completed.

Subjects were drawn from the extensive subject pool routinely used at the Institute for Empirical Research in Economics in Zürich. Psychology students were excluded. Sessions lasted about two hours, and payments were made privately after the experiment. The exchange rate was 150 experimental currency units to the Swiss Franc. Total earnings were the sum over all 150 periods. The average payment was 32.68 Swiss Francs (25.50 USD, 20.91 EUR).

5 Results and Discussion

The value of conformity in this experiment depended on the effectiveness of individual learning. Individual learning was highly effective. Individual learners exhibited a roughly uniform distribution over the two colors in period 1, but the proportion choosing the optimal color increased dramatically as the 25 periods progressed. Specifically, regressing the proportion of individual learners choosing optimally on period using the method of Newey and West (1987) to correct for heteroskedasticity and autocorrelation up to lag 3 produces a highly significant upward trend ($p < 0.01$). The estimated coefficient for period is 0.012 and the R^2 value is 0.930. Thus on average the percentage of individual learners choosing optimally in a group of five increased by roughly 30 percentage points over the course of 25 periods.

To test for conformity among social learners, we take model (2) as our theoretical framework. To evaluate the theory, we estimated the key parameter, D, using maximum likelihood under three different levels of assumed heterogeneity among social learners (electronic supplement). The simplest model posits a single value of D over all observations and all social learners. The second model divides social learners into two groups based on their response to a single questionnaire item. Specifically, to have some measure of what subjects thought they were doing in the experiment, the questionnaire asked social learners about their use of social information. They were asked whether they tended to choose (i) the same color as the majority of players in the "other group" (i.e. individual learners), (ii) the same color as the minority of players in the other group, or (iii) neither. 28 of 40 social learners claimed (i), 11 claimed (iii), and only 1 claimed (ii). Given that only one player claimed to follow the minority, we divided social learners, in an *a priori* fashion, into those who claimed they tended to follow the majority during the experiment (category (i)) and those who did not (categories (ii) and (iii)). We call these players respectively "stated conformists" and "not stated conformists." Our second model estimates a separate D value for each of these two groups. The final model estimates an individual value of D for each social learner (i.e. individual fixed effects). We used AIC$_c$, an improved form of Akaike's original criterion (Akaike, 1973; Burnham and Anderson, 2002), as a model-selection criterion.

Table 1 summarizes the results.

[Table 1 about here]

The model of individual fixed effects fits the best to an overwhelming degree, but the AIC$_c$ values also indicate that the two-parameter model is a vast improvement over the simple model that estimates a single D value. In particular, lower AIC$_c$ values indicate an improved fit, and the absolute difference in the AIC$_c$ values between models has meaning. Importantly, however, the use of information theoretic criteria like AIC$_c$ does not involve arbitrary thresholds (e.g. $\alpha \leq 0.05$) as in hypothesis testing, so there is no concept of one model being "significantly" better than another. But the differences in Table 1 are truly enormous by any standard (Burnham and Anderson, 2002). This finding means that individual variation in frequency-dependent social learning is extremely important, but nonetheless the distinction between stated conformists and those who were not stated conformists also captures important systematic variation relative to a model that simply assumes all social learners were the same.

Figure 4 compares the data and the two-parameter version of the model.

[Figure 4 about here]

The model fits poorly for social learners who were not stated conformists but quite well for those who were stated conformists. The 12 social learners who were not stated conformists, in effect, did not respond on average to information about the frequencies of alternative behaviors in any notable way. Thus the model, though it can be fit using maximum likelihood, is not based on assumptions that were generally appropriate for social learners in this group. For the 28 stated conformists, however, data and model are nearly indistinguishable for much of the function's domain. The exceptions at the boundaries show that stated conformists had a small tendency to play the absent color when all five individual learners were choosing the same color.

Figure 5 additionally shows considerable individual variation in play among both stated conformists and social learners who were not stated conformists. The graph shows results from the fixed-effects model and plots the mean earnings per period for each subject as a function of the subject's individual D estimate.

[Figure 5 about here]

Importantly, the individual D estimates are not obviously clustered into two groups. This is why the model of fixed effects fits better than the model that distinguishes simply between stated conformists and everyone else. Additionally, Figure 5 also clarifies why the distinction between stated conformists and those who were not stated conformists is better than assuming a homogeneous response to frequency information. Most of the stated conformists have positive D estimates under the fixed-effects model, while most of those who were not stated conformists have negative D estimates. In short, players were pretty good, though not perfect, at summarizing their use of frequency information. The net consequence of these results is the following. Figure 5 provides no obvious evidence for different categorical types of social learners. Nonetheless, when forced to categorize their use of frequency information, social learners did so with enough self-awareness that their claims about their use of frequency information were sufficiently accurate to provide an effective partition of the data. For these reasons, our statistical models fit progressively better as we added parameters. By adding parameters, we increasingly captured the considerable variation among social learners.

Finally, Figure 5 shows, as predicted by theory (model (1)), a strong positive relationship between conformity and earnings. In short, because the individual learners were actually learning, social learners who showed a strong inclination to follow the majority among individual learners were the social learners who made the most money. Social learners who did not respond to the available frequency-dependent information left money on the table.

These results show that individual heterogeneity is critical to understanding frequency-dependent social learning. Specifically, our data suggest a meaningful distinction between those who conform and those who largely ignore information about behavioral frequencies. Nonetheless substantial individual variation also exists within each of these two generic groups of players. The stability of this kind of heterogeneity across cultures or in different social settings is not clear, but an analogous study recently conducted among subsistence pastoralists in Bolivia produced similar patterns (Efferson, Lalive, Richerson, McElreath, and Lubell, unpublished data). Also unclear is the extent to which players might adjust their use of frequency-dependent social information according to its value. Do some players, for example, have an innate desire to conform regardless of the consequences, or do they rather recognize its practical decision-making value in appropriate situations? Our results show that some individuals do not conform even when doing so would be very much in their own interests. This conclusion is in contrast to work like that of Asch (1955, 1956), which tends to focus on how some will cave in to social influence even in direct opposition to what their senses are telling them.

Apart from the subjects who largely ignored frequency information, many did conform, and doing so paid well. To fully examine in the future how subjects adjust their tendency to follow the majority, its value would have to be systematically and exhaustively varied. The present experiment did not do so. Because the value of conformity was rooted in the performance of individual learners, conformity on average was either neutral (e.g. period 1) or valuable (e.g. subsequent periods). It was never detrimental on average. Thus we cannot say to what extent social learners who conformed recognized conformity's monetary value in this particular experiment, and to what extent they simply had a desire to match the properties of the social group provided for them regardless of the monetary consequences. More generally, the issue of how flexible biased social learning is in different settings and over short time scales remains one of the central unanswered empirical questions in the study of cultural transmission.

The heterogeneity in social learning we have documented has received little attention in the study of cultural transmission. In particular, if distinct and stable types of social learners exist, one obviously important consideration would involve how they assort into groups both within and between societies. For instance, if all else is equal and conformists form groups assortatively, conformist groups should be more productive than their less conformist counterparts as long as some basis for effective individual learning is present. More generally, the study of dynamical systems can sometimes proceed effectively by focusing on the average behavior of constituent parts, while in other cases ignoring individual variation can lead the researcher dramatically astray (Miller and Page, 2007). Which of these two scenarios holds and when is a largely unconsidered problem in cultural evolution. Even so, this kind of understanding is potentially critical when addressing aggregate behavioral dynamics and the corresponding evolutionary consequences for organisms with biased cultural transmission.

Acknowledgements

The authors would like to thank all the members in the chair of Ernst Fehr at the Institute for Empirical Research in Economics for assistance and comments. A particular thanks goes to Urs Fischbacher for his z-Tree program to produce animated histograms. We would also like to acknowledge two anonymous reviewers for much helpful input. CE thanks the U.S. Environmental Protection Agency (STAR Fellowship # U-91615601-3) for financial support. This research was funded by a U.S. National Science Foundation Grant (Award # 0340148 to PJR, RM, and ML).

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov and F. Csaki, editors, *Second International Symposium on Information Theory*, pages 267–281. Budapest: Akademiai Kiado.

Aronson, E., Wilson, T. D., and Akert, R. M. (2002). *Social Psychology*. Upper Saddle River, NJ: Prentice-Hall, 4th edition.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, **193**(5), 1–7.

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs*, **70**, 1–70.

Berns, G. S., Chappelow, J., Zink, C. F., Pagnoni, G., Martin-Shurski, M. E., and Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. *Biological Psychiatry*, **58**(3), 245–253.

Bowles, S. (2004). *Microeconomics: Behavior, Institutions, and Evolution*. New York: Russell Sage.

Boyd, R. and Richerson, P. J. (1982). Cultural transmission and the evolution of cooperative behavior. *Human Ecology*, **10**(3), 325–351.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.

Boyd, R. and Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. Oxford: Oxford University Press.

Boyd, R., Gintis, H., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences*, **100**(6), 3531–3535.

Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag, 2nd edition.

Efferson, C. and Richerson, P. J. (2007). A prolegomenon to nonlinear empiricism in the human behavioral sciences. *Biology and Philosophy*, **22**(1), 1–33.

Efferson, C., Richerson, P. J., McElreath, R., Lubell, M., Edsten, E., Waring, T. M., Paciotti, B., and Baum, W. (2007). Learning, productivity, and noise: an experimental study of cultural transmission on the Bolivian altiplano. *Evolution and Human Behavior*, **28**(1), 11–17.

Fehr, E. and Fischbacher, U. (2003). The nature of human altruism. *Nature*, **425**(23 October), 785–791.

Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, **8**(4), 185–190.

Fehr, E. and Gaechter, S. (2002). Altruistic punishment in humans. *Nature*, **415**(10 January), 137–140.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10**(2), 171–178.

Guererk, O., Irlenbusch, B., and Rockenback, B. (2006). The competitive advantage of sanctioning institutions. *Science*, **312**(7 April), 108–111.

Guzmán, R. A., Rodríguez-Sickert, C., and Rowthorn, R. (2006). When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning, and cooperation. *Evolution and Human Behavior*, **in press**.

Henrich, J. (2001). Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. *American Anthropologist*, **103**(4), 992–1013.

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization*, **53**(1), 3–35.

Henrich, J. and Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior*, **19**(4), 215–241.

Henrich, J. and Boyd, R. (2001). Why punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology*, **208**(1), 79–89.

Henrich, J. and McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology*, **12**(3), 123–135.

Hoy, M., Livernois, J., McKenna, C., Rees, R., and Stengos, T. (2001). *Mathematics for Economics*. Cambridge: The MIT Press, 2nd edition.

Lumsden, C. J. and Wilson, E. O. (1980). Translation of epigenetic rules of individual behavior into ethnographic patterns. *Proceedings of the National Academy of Sciences*, **77**(7), 4382–4386.

McElreath, R. and Boyd, R. (2007). *Mathematical Models of Social Evolution: A Guide for the Perplexed*. Chicago: University of Chicago Press.

McElreath, R., Lubell, M., Richerson, P. J., Waring, T. M., Baum, W., Edsten, E., Efferson, C., and Paciotti, B. (2005). Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior*, **26**(6), 483–508.

Miller, J. H. and Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton: Princeton University Press.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(3), 703–708.

Richerson, P. J. and Boyd, R. (2005). *Not By Genes Alone: How Culture Transformed the Evolutionary Process*. Chicago: University of Chicago Press.

Rogers, E. M. (1995). *Diffusion of Innovations*. New York: The Free Press, 4th edition.

Sherif, M. and Murphy, G. (1936). *The Psychology of Social Norms*. New York: Harper and Brothers.

Whiten, A., Horner, V., and de Waal, F. B. M. (2005). Conformity to cultural norms of tool use in chimpanzees. *Nature*, **437**(29 September), 737–740.

Table 1: Model description, the number of estimated parameters, maximized log likelihood ($\ln L^*$), Akaike value (AIC_c), and Akaike weight (w_i) for each of the three models fit to the social learners' data. Altogether the experiment produced 5000 observations for social learners, and 3749 of these could be used to estimate D and calculate AIC_c values (see electronic supplement). Akaike weights sum to 1 and summarize the proportional weight of evidence in support of each model, where larger weights indicate more support. The absolute difference between AIC_c values also has meaning (Burnham and Anderson, 2002) and is the basis for our claim that the two-parameter model is a vast improvement over the one-parameter model.

Model	Parameters	$\ln L^*$	AIC_c	w_i
Single D	1	-2159.28	4320.56	4.11×10^{-112}
Conformist, Y or N	2	-2014.39	4032.78	1.27×10^{-49}
Fixed Effects	40	-1863.36	3807.60	> 0.99

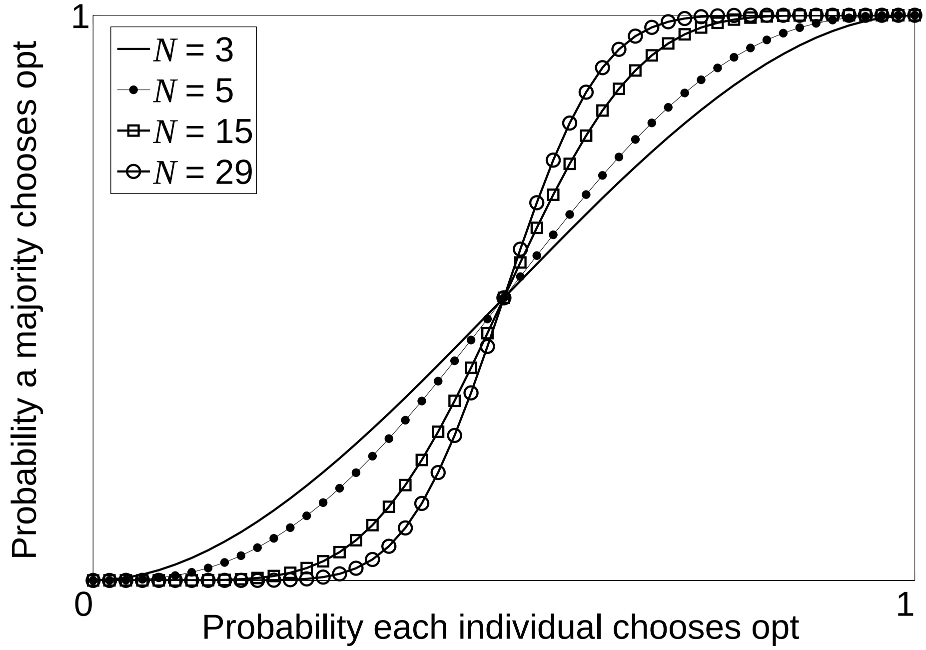

Figure 1: The probability a majority of individuals in the group, $P(\text{majority opt})$, chooses the optimal technology as a function of the probability, r_t, that each individual chooses the optimal technology for 4 different group sizes, N.

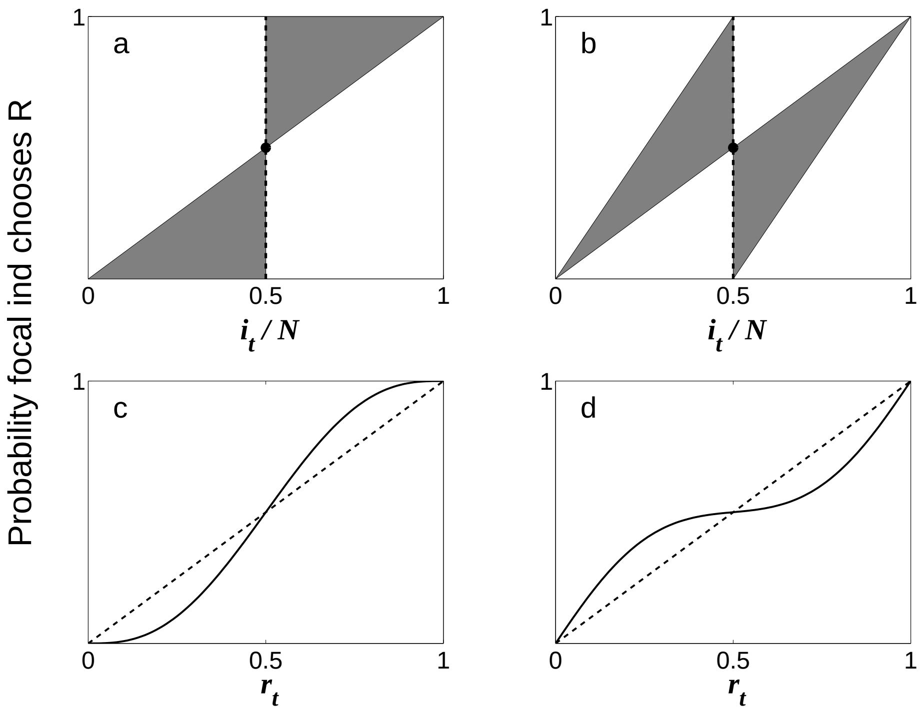

Figure 2: The top row shows in gray the entire set of possible probability functions allowed by conformity (plot a, $D \in (0, 1]$) and non-conformity (plot b, $D \in [-1, 0)$) without sampling. In both cases, choice probabilities are piece-wise linear non-decreasing functions of the frequency of behavior R, i.e. i_t/N. The 45-degree line common to both gray regions is linear transmission ($D = 0$). The bottom row shows choice probabilities under sampling for both extreme conformity (plot c, $D = 1$) and extreme non-conformity (plot d, $D = -1$). These functions depend on r_t, the proportion of individuals exhibiting R in the population from which samples are drawn. For reference the 45-degree line is also shown.

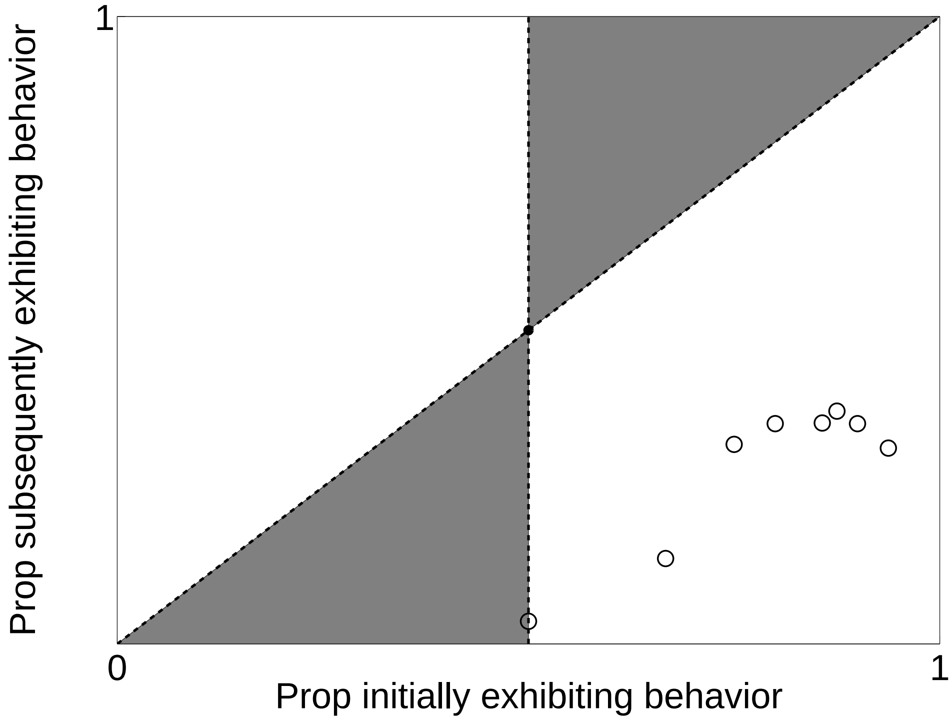

Figure 3: Functions lying within the gray area are consistent with conformity. The circles plot data from an experiment in Asch (1955, p. 6) in which the unanimous majority aligned against the experimental subject constituted a proportion of the group ranging from 1/2 to 15/16. The behavior in question is choosing a line of a different length from the reference line. The horizontal axis is the proportion of the group made up of confederates choosing a wrong line. The vertical axis shows the proportion of experimental subjects also choosing a wrong line. Only those data for which Asch (1955) provides precise numerical information are shown here.

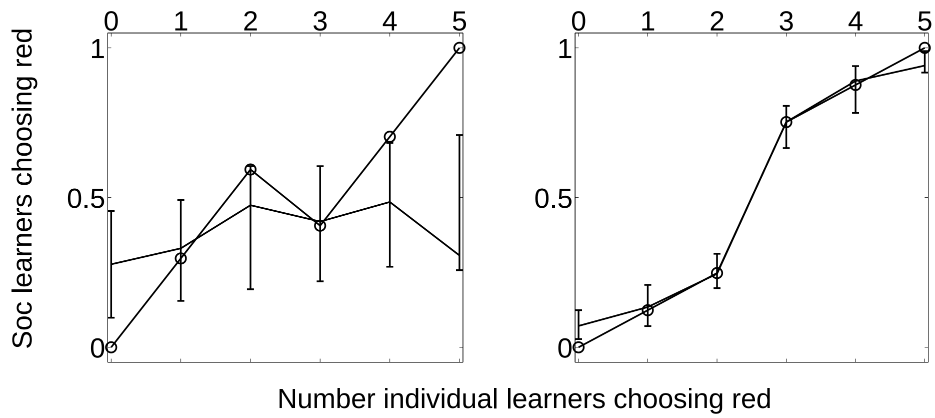

Figure 4: Data and theoretical predictions for social learners who were not stated conformists (left-hand panel) and those who were (right-hand panel). The graphs show the proportion of social learners choosing red (lines with 95% bootstrapped confidence intervals) as a function of the number of individual learners choosing red. The graphs also show (lines with circles) the theoretical probability a social learner chooses red under model (2) and the maximum likelihood estimate of D. The MLE estimate for the 12 players who were not stated conformists is -0.4843, and the standard error is 0.0438. For stated conformists the estimate and standard error are 0.3805 and 0.0250 respectively. The different point estimates of D for the two types of player account for the different theoretical predictions.

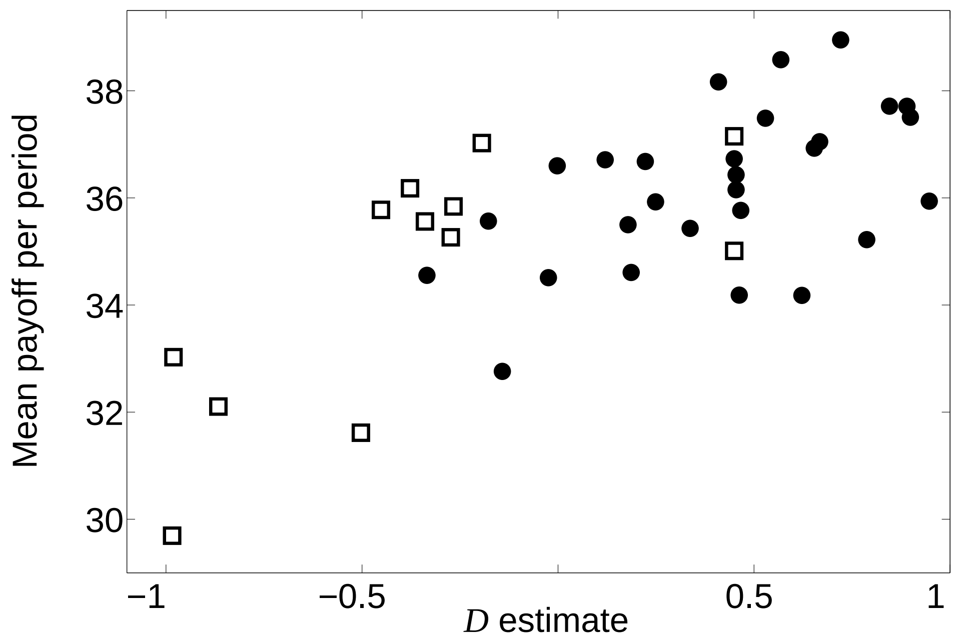

Figure 5: The mean payoff per period versus the estimated value of D for each of the 40 social learners. D estimates are based on the model of individual fixed effects. Players who were not stated conformists are shown with open squares, while stated conformists are shown in solid circles.

Experimental methods

Experiments were conducted in the laboratory of the Institute for Empirical Research in Economics at the University of Zürich and implemented entirely on a local computer network using the z-Tree software developed by Fischbacher (2007). We recruited a total of 70 undergraduate students from the University of Zürich and the Swiss Federal Institute of Technology. We ran a total of two sessions with 35 students in each session.

Students in each session were divided into 3 "worlds." Each world consisted of two groups. A group of 5 individuals received individual payoff information as described in detail below. These players were the *individual learners*. A group of 6 or 7 players who played simultaneously received no individual payoff information during the experiment. Instead, these *social learners* received information on the distribution of choices among the individual learners as described below. The only payoff information a social learner received was her total earnings after the experiment and questionnaire were finished.

Subjects were first informed they were participating in a laboratory experiment at the University of Zürich. Communication between subjects was not permitted. Students earned points in the experiment, and 150 points were worth one Swiss franc (about 0.78 USD or 0.64 EUR).

Subjects were instructed that their task was to choose either a "blue" technology or a "red" technology in each period. Technologies generated points at random according to specific probability distributions. One technology, the optimal technology, had a higher average payoff but was otherwise like the sub-optimal technology. The color of the optimal technology was chosen at random with probability 0.5.

The sub-optimal technology generated draws from a normal distribution with an expectation of 30 points and a standard deviation of 12 points. The optimal technology had a normal payoff distribution with an expectation of 38 points and a standard deviation of 12 points. Both distributions were truncated at 0 and 68. Truncation means that the truncated and untruncated distributions had slightly different expectations and standard deviations. Payoffs were rounded to integer values, and thus the set of possible payoffs for both colors was $\{0, 1, \ldots, 68\}$. We did not describe the payoff distributions to subjects in technical terms, but before they began the actual experiment we did provide them with an intuitively accessible demonstration of the random processes governing payoffs.

Specifically, the instructions before the experiment paid particular attention to the possibility that subjects may not have understood the formal concept of payoffs that follow probability distributions. Before the experiment started, subjects saw a demonstration of the two random technologies. In the demonstration, the optimal color was first determined randomly with probability 0.5. The optimal color was the same for each subject within a world but potentially different across worlds. Once an optimal color had been determined, the computer would take 250 draws from each of the two probability distributions for each subject individually. Two horizontal number lines from 0 to 68 appeared one beneath the other on the screen with one number line for each color. For each draw producing a specific value (e.g. 27 points for a red choice), the computer would place a little box (colored red or blue) along the appropriate number line (e.g. a red box at 27 for the number line being used to plot red draws). For multiple draws producing the same payoff, boxes were stacked on top of each other. As a consequence, students essentially watched a histogram being built draw by draw on the screen in front of them. This allowed them an intuitive sense of the stochastic process even if they had no training in probability theory or data analysis. Moreover, while the histograms were being built, they knew which color was optimal (but *only* for the demonstration). They could thus see, as an example, that blue was producing payoffs centered around 38, while red was producing payoffs centered around 30. The histogram was explained to them in writing, and subjects could read the explanation repeatedly while the histograms were being built. After the first demonstration, an optimal color was selected, and the demonstration was repeated. The whole demonstration phase took 5-10 minutes.

After the demonstration had been completed, subjects were informed that one repetition of the experiment would last for 25 periods. The timing was as follows. The computer first assigned the optimal color at random. The optimal color stayed the same for all 25 periods and was identical for each subject within a world. In each period, subjects chose between red and blue by indicating the desired color on the computer screen and clicking "OK."

Immediately after making a choice, individual learners were informed privately about the realized number of points received. Subjects also knew that the points from each period would be added up to yield a total payoff at the end of the entire experimental session.

Social learners were informed of the fact that in "the other part of the laboratory" a group of five individuals was facing the same two technologies with the same optimal color. Importantly, social learners also knew that subjects in the other group knew how many points they were making after each choice. In essence social learners knew that the players in the other group were receiving individual feedback.

Within each period, social learners were first informed about the number of individuals in the other group choosing red and the number choosing blue. Social learners then made their own choices. Social learners knew they would not receive information about their own payoffs until the entire session was over for the day.

The experiment was repeated six times. After the six repetitions, subjects responded to a questionnaire that recorded basic socio-demographic characteristics like gender, age, and academic major. The questionnaire also asked about learning strategies. Individual learners were questioned about the events that led them to revise their choices. Social learners were asked whether they tended to choose (i) the same color as the majority of the players in the other group, (ii) the same color as the minority of the players in the other group, or (iii) none of the two. The main paper labels social learners answering (i) as "stated conformists," while social learners answering (ii) or (iii) are labeled as "not stated conformists."

After the questionnaire, subjects received their total payoffs based on points summed over all six repetitions. The average payoff was 32.68 CHF (25.50 USD; 20.91 EUR). Sessions lasted about 2 hours.

How conformity works

To expand the theoretical discussion in the main paper, let us work with a simplified model that pertains closely to the experiment. The two technologies with random payoffs are "red" and "blue." Assume two groups, individual learners and social learners, as in the experiment. Posit a focal social learner who assumes every individual learner in period t chooses red, given that red is optimal, with probability p_t. The social learner further assumes every individual learner chooses red, given that blue is optimal, with probability v_t. One particularly simple form of conformity ignores how overwhelming the majority is in a given period and simply notes which color was chosen by a majority of individual learners. In the present experiment, individual learners were always in groups of size 5. How does noting the color in the majority among individual learners provide useful social information?

To answer this question, note that the probability a majority of individual learners chooses red (maj red) given that red is optimal (red opt), which we call $f(p_t)$, is

$$P(\text{maj red} \mid \text{red opt}) = \sum_{i=3}^{5} \binom{5}{i} (p_t)^i (1 - p_t)^{5-i} = f(p_t). \tag{1}$$

The probability a minority chooses red (min red) given red is optimal is

$$P(\text{min red} \mid \text{red opt}) = \sum_{i=0}^{2} \binom{5}{i} (p_t)^i (1-p_t)^{5-i} = 1 - f(p_t). \tag{2}$$

Similarly, the probability a majority of individual learners chooses red given blue is optimal (blue opt), which we call $h(v_t)$, is

$$P(\text{maj red} \mid \text{blue opt}) = \sum_{i=3}^{5} \binom{5}{i} (v_t)^i (1-v_t)^{5-i} = h(v_t), \tag{3}$$

while the analogous probability red is in the minority is

$$P(\text{min red} \mid \text{blue opt}) = \sum_{i=0}^{2} \binom{5}{i} (v_t)^i (1-v_t)^{5-i} = 1 - h(v_t). \tag{4}$$

Define s_t as the prior probability for a social learner that red is optimal (i.e. the probability that applies before the social learner knows the distribution of choices among individual learners in t). Bayes' rule specifies

$$P(\text{red opt} \mid \text{maj red}) = \frac{f(p_t)s_t}{f(p_t)s_t + h(v_t)(1-s_t)}, \tag{5}$$

and

$$P(\text{red opt} \mid \text{min red}) = \frac{(1-f(p_t))s_t}{(1-f(p_t))s_t + (1-h(v_t))(1-s_t)}. \tag{6}$$

To simplify matters further, assume the focal social learner believes that $p_t = 1 - v_t \Rightarrow v_t = 1 - p_t$. This assumption simply means the social learner believes individual learners have no color biases in that they are equally likely to choose the optimal color regardless of whether it is red or blue. Specifically it means that p_t is the assumed probability an individual learner chooses red given red is optimal, the assumed probability an individual learner chooses blue given blue is optimal, and thus simply the assumed probability an individual learner chooses the optimal color.

Correspondingly, if blue is the optimal technology (blue opt), we can define $g(p_t)$ by rewriting equations (3) and (4) in terms of p_t,

$$P(\text{maj red} \mid \text{blue opt}) = \sum_{i=3}^{5} \binom{5}{i} (1-p_t)^i (p_t)^{5-i} = g(p_t) = h(v_t), \tag{7}$$

and the probability of a red minority is

$$P(\text{min red} \mid \text{blue opt}) = \sum_{i=0}^{2} \binom{5}{i} (1-p_t)^i (p_t)^{5-i} = 1 - g(p_t) = 1 - h(v_t). \tag{8}$$

Further define $m_t \in \{0, 1\}$, where $m_t = 0$ means an observed minority of the individual learners chose red in t, while $m_t = 1$ means an observed majority chose red. The updated probability that red is optimal for a given social learner is then

$$s_{t+1} = \frac{m_t f(p_t) s_t}{f(p_t) s_t + g(p_t)(1-s_t)} + \frac{(1-m_t)(1-f(p_t)) s_t}{(1-f(p_t)) s_t + (1-g(p_t))(1-s_t)}. \tag{9}$$

The updating equation (9) tells us that only one of the two conditional probabilities, (5) or (6), is relevant in any given period, but which one is relevant depends on whether the social learner observed a majority or minority of red choices among the individual learners. The unconditional updated probability, s_{t+1} is a function of s_t, of course, but it also depends on p_t, a quantity that captures what the social learner thinks about how individual learners are learning. Figures S1 and S2 show the probabilities specified by (5) and (6) for five different values of p_t.

Specifically, the figures offer a simple formalization of the interplay between individual learning and conformity for social learners in the present experiment. After observing

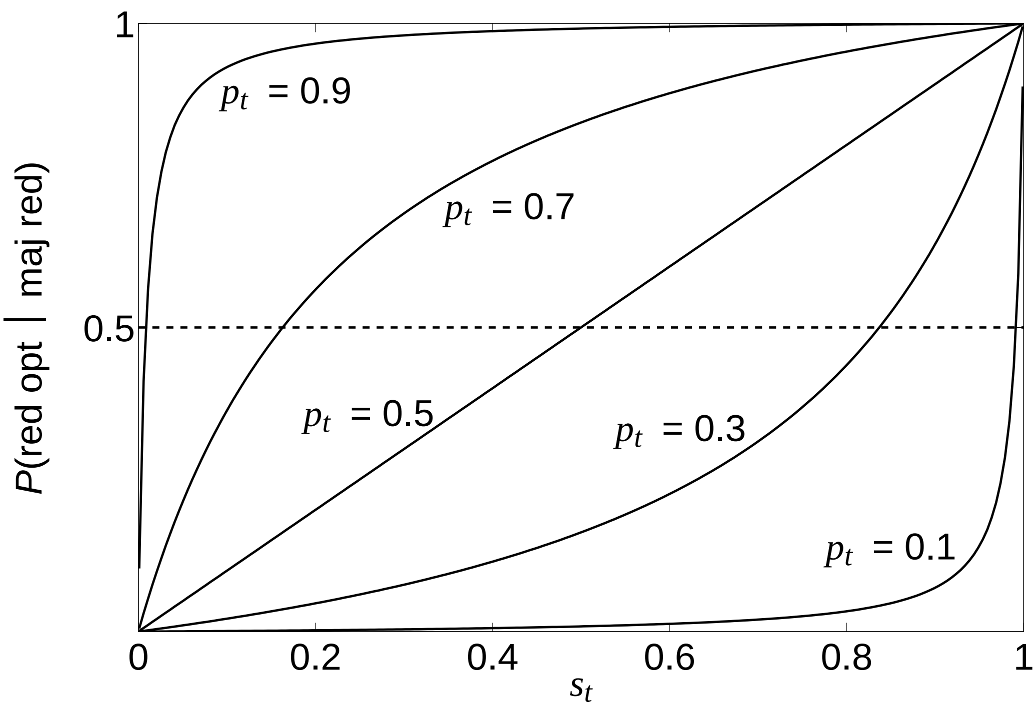

Figure S1. The updated probability that red is the optimal technology, given that a majority of individual learners chose red, as a function of s_t, the lagged unconditional probability that red is optimal. The function is shown for five different values of p_t, the social learners belief about the probability that each individual learner is choosing optimally in t.

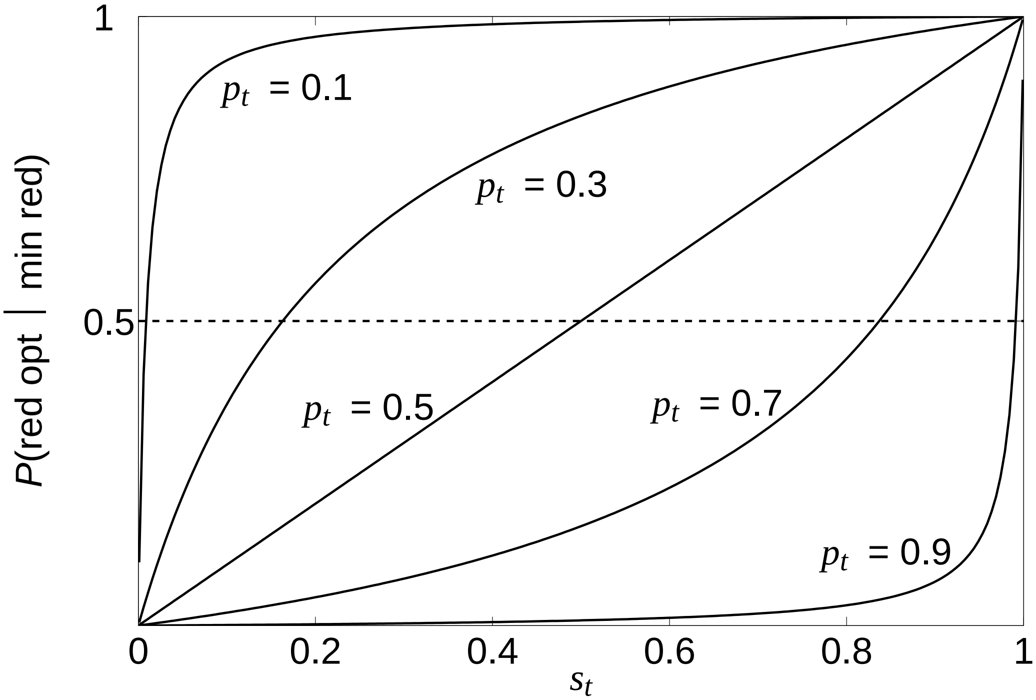

Figure S2. The updated probability that red is the optimal technology, given that a minority of individual learners chose red, as a function of s_t, the lagged unconditional probability that red is optimal. The function is shown for five different values of p_t, the social learners belief about the probability that each individual learner is choosing optimally in t.

whether red or blue was in the majority among individual learners, the Bayesian social learner we have modeled should choose the color with an unconditional updated probability of being the optimum that is greater than 0.5. This updated probability depends on the observed distribution of choices among the individual learners (specifically on whether red was in the majority or minority), the social learner's prior unconditional beliefs about the probability that red is optimal (s_t), and the social learner's beliefs about the decision-making of the individual learners as summarized by p_t.

As the figures show, if the social learner believes that individual learners are flipping a coin to make choices (i.e. $p_t = 0.5$), the information about the distribution of choices among individual learners is irrelevant. Updating essentially does not occur because the social learner's posterior probability, s_{t+1}, is equal to her prior probability, s_t. If, however, the social learner believes individual learners are biasing their choices in some way (i.e. $p_t \neq 0.5$), the social learner can use the information about the distribution of behaviors among individual learners to bias her own choice toward the optimum. How she uses the social information depends on her beliefs about the biases exhibited by individual learners.

Specifically, if the social learner believes individual learners are biasing their choices toward the optimum (i.e. $p_t > 0.5$), updating will tend to produce new beliefs that favor selecting the color in the majority among individual learners. This effect is stronger as p_t increases, s_t increases, or both. As the Figures S1 and S2 show, if p_t or s_t is close to 0 or 1, the variable with an extreme value has an overwhelming effect on updating unless the other variable has an equally extreme value with a countervailing effect.

One approach to updating would be to posit that social learners are initially neutral about which color is optimal (i.e. $s_0 = 0.5$), and they also believe individual learners are initially neutral ($p_0 = 0.5$). If social learners additionally believe individual learning is effective in the sense that $\forall t$, $p_{t+1} \geq p_t$, where the inequality is strict for some t, then these assumptions will have various implications for how social learning proceeds. In particular, as p_t increases, the updating rule implies that social learners should become more responsive to social information; the tendency to conform should become stronger through time. Moreover, if social learners vary notably with respect to p_t in any given period, some will be more responsive to changes in the social signal than others. For example, consider a situation in which red has been in the majority among individual learners. In period $t+1$, two social learners have a prior of $s_t = 0.9$, but one believes $p_t = 0.7$, while the other believes $p_t = 0.9$. Assume that, in contrast to period $t-1$, red is in the minority among individual learners in t. According to Figure S2, the social learner who believes $p_t = 0.7$

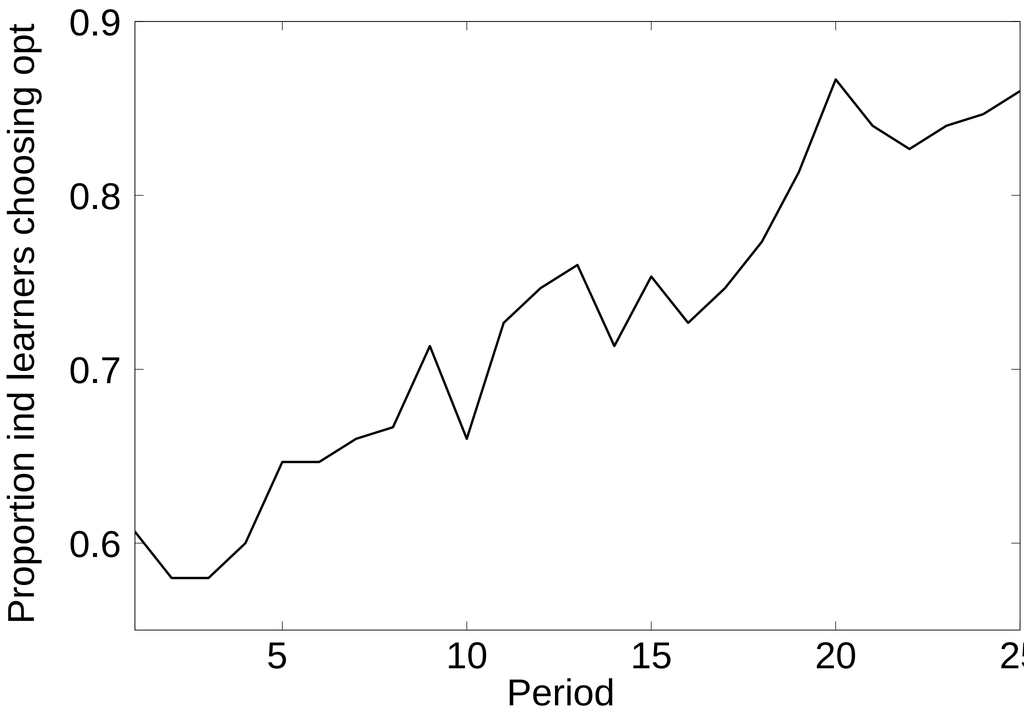

Figure S3. The proportion of individual learners choosing the optimal technology through time. The upward trend is highly significant ($p < 0.01$) when we regress the proportion choosing optimally on period using the method of Newey and West (1987) to correct for heteroskedasticity and autocorrelation up to lag 3. In this case, the estimated coefficient for period is 0.012 and the R^2 value is 0.930.

will have an updated belief that satisfies $s_{t+1} > 0.5$, while the social learner who believes $p_t = 0.9$ will update such that $s_{t+1} < 0.5$. The former will still choose red, while the latter will switch to blue. Such variation could be one source of noise in the data for social learners. In general we do not claim that social learners in our experiment were Bayesians, but the updating model we have presented provides a convenient approach to summarizing the value of conformity as it interacts with individual learning. Figure S3 shows that the individual learners in the experiment were learning effectively.

Strong non-conformity

Using the same basic notation as the main text, a model of strong non-conformity begins with the following conditional choice probabilities,

$$P(\text{focal ind chooses R} \mid i_t) = \begin{cases} 1 - M(i_t/N) & \text{if } i_t < N/2, \\ 1/2 & \text{if } i_t = N/2, \\ M(1 - i_t/N) & \text{otherwise.} \end{cases} \tag{10}$$

This model is consistent with the notion that strong non-conformists exhibit large biases away from the behavior currently in the majority. The parameter $M \in [0, 1]$ controls the strength of the non-conformity, with $M = 1$ representing the weakest form of strong non-conformity and $M = 0$ the strongest. Under unbiased sampling and N odd, the unconditional choice probability is the following,

$$\begin{aligned} P(\text{focal ind chooses R}) = &\sum_{i_t=0}^{\lfloor N/2 \rfloor} \binom{N}{i_t} (r_t)^{i_t} (1 - r_t)^{N-i_t} \\ &+ M \left\{ \sum_{i_t=\lceil N/2 \rceil}^{N} \binom{N}{i_t} (r_t)^{i_t} (1 - r_t)^{N-i_t} \right\} - M r_t. \end{aligned} \tag{11}$$

If we assume a large population, this model simply becomes the recursion specifying the value of r_{t+1} as a function of r_t. Figure 4 shows the properties of the resulting model. As panels (c) and (d) show, the dynamics oscillate, which means the behavior in the majority is constantly switching. The oscillations neither grow nor decay under $M = 1$, but they grow under $M = 0$ until the system hits the boundaries. At that point, the population literally switches completely from all exhibiting one behavior in a period to all exhibiting the other behavior in the subsequent period. This result can be easily verified

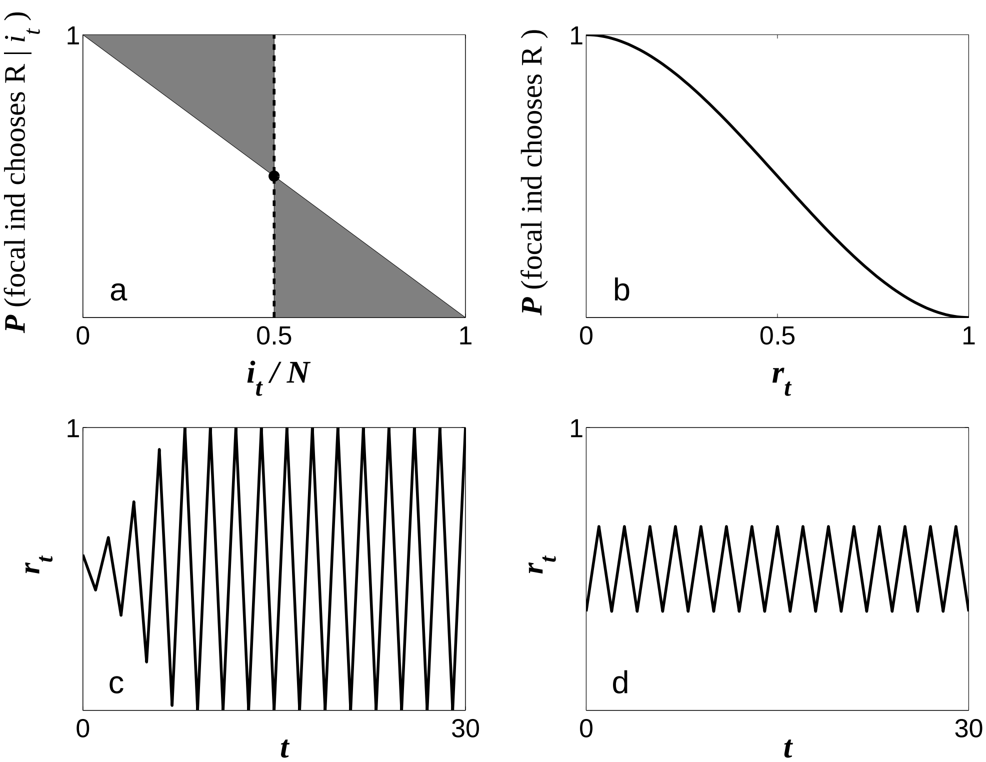

Figure S4. Panel (a) shows the set of conditional choice probabilities allowed by the model of strong non-conformity (10). Panel (b) shows the unconditional choice probabilities (11) under unbiased sampling, $M = 0$, and $N = 3$. Panel (c) shows dynamics under $M = 0$, while panel (d) shows dynamics under $M = 1$.

using standard methods like cobwebbing to analyze nonlinear difference equations (Hoy *et al.*, 2001). The oscillations here are a direct consequence of the fact that this version of non-conformity postulates a strong bias away from the current majority.

Statistical models and detailed results

To estimate conformity using data from the present experiment and the conditional probability model presented in the main paper, let $k \in K = \{1, 2, \ldots, BT\}$, where K is a set to index observations for each social learner by period combination. B is the number of social learners, and T the number of periods. Define $C_k \in \{0, 1\}$ as a random variable with realizations c_k such that $c_k = 0$ if the social learner chose blue, and $c_k = 1$ if the social learner chose red. Call the entire data set $c = \{0, 1\}^{BT}$. Further define $i_k \in \{0, 1, \ldots, 5\}$ as an associated variable that records the number of individual learners choosing red as observed by the social learner in question in the appropriate period. The probability model for a given social learner is thus

$$P(C_k = 1) = \begin{cases} (i_k/5)(1 - D) & \text{if } i_k \leq 2, \\ (i_k/5)(1 - D) + D & \text{if } i_k \geq 3. \end{cases} \tag{12}$$

To estimate D, one must remove all data points where $i_k \in \{0, 5\}$ because observations in which all individual learners chose the same option cannot be used to estimate D. This results simply from the fact that D drops out of model (12) in these cases. Accordingly, define the set $J = \{k \in K \mid i_k \in \{1, 2, 3, 4\}\}$. The log likelihood function for estimating D is thus

$$\ln\{L(D \mid c)\} = \sum_{k \in J} c_k \ln\left\{\frac{P(c_k = 1)}{P(c_k = 0)}\right\} + \ln\{P(c_k = 0)\}. \tag{13}$$

Table S1. Estimates, standard errors, and approximate 95% confidence intervals for the model that assumes a single D value for all social learners and the model that estimates a separate D for the social learners who were stated conformists (D_Y) and those who were not (D_N). Confidence intervals are the point estimates plus or minus twice the standard errors.

Model	Parameter	Estimate	Std. error	95% CI
Single D	D	0.1081	0.0005	$[0.1070, 0.1091]$
Conformist (Y or N)	D_Y	0.3805	0.0250	$[0.3305, 0.4305]$
	D_N	-0.4843	0.0438	$[-0.5720, -0.3966]$

This function is derived from a joint probability distribution over c given that observations are Bernoulli random variables under (12). We used this method to estimate the values of D reported in the main text, and approximate standard errors were calculated by inverting the Hessian of the log likelihood function evaluated at the estimated value of D.

Incorporating heterogeneity into the model (e.g. "not stated conformists" versus "stated conformists") is equivalent to splitting the sample in an appropriate way, fitting the model separately to each sub-sample, and then adding maximized log likelihood values to calculate Akaike values for the entire data set. In practice, however, one can write a routine that accommodates any desired degree of heterogeneity. Code for implementing these procedures in R (R Development Core Team, 2006) is available on request. Apart from the model-fitting results presented in the main text, Tables S1 and S2 show additional detailed results from each of the three models.

Table S2. Estimates, standard errors, and approximate 95% confidence intervals for the fixed effects model that includes a separate D for each social learner. Confidence intervals are the point estimates plus or minus twice the standard errors.

Parameter	Estimate	Std. error	95% CI
D_1	0.4542	0.1193	$[0.2155, 0.6928]$
D_2	0.6538	0.0979	$[0.4580, 0.8497]$
D_3	0.4622	0.1176	$[0.2269, 0.6975]$
D_4	-0.2736	0.1489	$[-0.5713, 0.0241]$
D_5	-0.4517	0.1404	$[-0.7325, -0.1708]$
D_6	0.1865	0.1347	$[-0.0829, 0.4559]$
D_7	-0.5028	0.1391	$[-0.7809, -0.2246]$
D_8	-0.3394	0.1550	$[-0.6495, -0.0294]$
D_9	-0.0020	0.1553	$[-0.3127, 0.3087]$
D_{10}	-0.1944	0.1607	$[-0.5158, 0.1270]$
D_{11}	-0.8665	0.1419	$[-1.1503, -0.5827]$
D_{12}	0.8456	0.0752	$[0.6953, 0.9960]$
D_{13}	0.5291	0.1249	$[0.2792, 0.7791]$
D_{14}	0.1202	0.1580	$[-0.1959, 0.4363]$
D_{15}	-0.3344	0.1458	$[-0.6260, -0.0427]$
D_{16}	-0.2667	0.1447	$[-0.5561, 0.0226]$
D_{17}	0.4542	0.1166	$[0.2210, 0.6874]$
D_{18}	0.5683	0.1071	$[0.3541, 0.7824]$
D_{19}	-0.9809	0.1083	$[-1.1977, -0.7642]$
D_{20}	0.4093	0.1220	$[0.1653, 0.6533]$

Continued on next page

Parameter	Estimate	Std. error	95% CI
D_{21}	-0.1777	0.1435	$[-0.4646, 0.1093]$
D_{22}	0.3370	0.1701	$[-0.0033, 0.6773]$
D_{23}	0.1787	0.1819	$[-0.1851, 0.5425]$
D_{24}	0.8902	0.0763	$[0.7377, 1.0428]$
D_{25}	0.6675	0.1285	$[0.4104, 0.9245]$
D_{26}	0.7211	0.1192	$[0.4826, 0.9596]$
D_{27}	0.2226	0.1726	$[-0.1225, 0.5677]$
D_{28}	-0.3775	0.1853	$[-0.7480, -0.0069]$
D_{29}	0.4493	0.1589	$[0.1314, 0.7673]$
D_{30}	-0.0246	0.1931	$[-0.4108, 0.3617]$
D_{31}	0.9471	0.0524	$[0.8422, 1.0520]$
D_{32}	0.4493	0.1589	$[0.1314, 0.7673]$
D_{33}	0.7875	0.1027	$[0.5822, 0.9929]$
D_{34}	0.4493	0.1589	$[0.1314, 0.7673]$
D_{35}	-0.9844	0.0933	$[-1.1710, -0.7978]$
D_{36}	0.4663	0.1049	$[0.2566, 0.6760]$
D_{37}	0.6220	0.0908	$[0.4403, 0.8036]$
D_{38}	0.2489	0.1196	$[0.0097, 0.4881]$
D_{39}	0.8990	0.0496	$[0.7998, 0.9982]$
D_{40}	-0.1419	0.1139	$[-0.3697, 0.0857]$

References

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, **10**(2), 171–178.

Hoy, M., Livernois, J., McKenna, C., Rees, R., and Stengos, T. (2001). *Mathematics for Economics*. Cambridge: The MIT Press, 2nd edition.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**(3), 703–708.

R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.