APA NLM
tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

Journal of Experimental Psychology: General

0096-3445/15/$12.00 http://dx.doi.org/10.1037/xge0000071

AQ: 1

Mechanisms of Social Avoidance Learning Can Explain the Emergence of Adaptive and Arbitrary Behavioral Traditions in Humans

AQ: au
AQ: 2

Björn Lindström and Andreas Olsson
Karolinska Institutet

Many nonhuman animals preferentially copy the actions of others when the environment contains predation risk or other types of danger. In humans, the role of social learning in avoidance of danger is still unknown, despite the fundamental importance of social learning for complex social behaviors. Critically, many social behaviors, such as cooperation and adherence to religious taboos, are maintained by threat of punishment. However, the psychological mechanisms allowing threat of punishment to generate such behaviors, even when actual punishment is rare or absent, are largely unknown. To address this, we used both computer simulations and behavioral experiments. First, we constructed a model where simulated agents interacted under threat of punishment and showed that mechanisms' (a) tendency to copy the actions of others through social learning, together with (b) the rewarding properties of avoiding a threatening punishment, could explain the emergence, maintenance, and transmission of large-scale behavioral traditions, both when punishment is common and when it is rare or nonexistent. To provide empirical support for our model, including the 2 mechanisms, we conducted 4 experiments, showing that humans, if threatened with punishment, are exceptionally prone to copy and transmit the behavior observed in others. Our results show that humans, similar to many nonhuman animals, use social learning if the environment is perceived as dangerous. We provide a novel psychological and computational basis for a range of human behaviors characterized by the threat of punishment, such as the adherence to cultural norms and religious taboos.

Keywords: Social learning, reinforcement learning, punishment, avoidance, traditions

Supplemental materials: http://dx.doi.org/10.1037/xge0000071.supp

The fundamental importance of social learning for humans and other animals has received wide recognition during the last decades (Boyd, Richerson, & Henrich, 2011; Meltzoff, Kuhl, Movellan, & Sejnowski, 2009; Rendell et al., 2010). Among social organisms, the behavior of others can provide critical information for survival by alerting to danger. Accordingly, many nonhuman animals preferentially copy the actions of others when the environment involves predation risk or other types of uncertainty (Cook & Mineka, 1990; Curio, Ernst, & Vieth, 1978; Griffin, 2004; Hoppitt & Laland, 2013). In spite of a growing interest in the social transmission of threat-relevant information (Olsson & Phelps, 2007) and nonsocial avoidance learning (Lawson et al., 2014; Roy et al., 2014), there has been little connections between these lines of research in humans, and the role of social learning in avoidance of danger is still unknown. Humans, like many other primates (Cook & Mineka, 1990), reliably learn to predict danger by observing the reactions of a conspecific undergoing fear conditioning (Olsson & Phelps, 2007). Because these studies have been modeled on the classical conditioning paradigm, in which the learner remains a passive observer, they do, however, not speak directly to the wide range of human behaviors involving active decisions between different options or courses of action. It remains unknown how humans use social learning to avoid danger (avoidance learning). Because social learning takes particularly complex forms in humans, and commonly is assumed to be at the root of what makes humans as a species successful (Boyd et al., 2011; Dean, Kendal, Schapiro, Thierry, & Laland, 2012; Tomasello, Kruger, & Ratner, 2010), characterizing the interaction of avoidance learning and social learning could have important implications for understanding both adaptive and nonadaptive human behavior. The goal of the present study was to directly investigate whether basic learning mechanisms involved in the avoidance of punishment could help to understand the emergence and maintenance of large-scale behavioral patterns in humans. To this end, we used a combination of theoretical simulation studies and experiments with human subjects.

The Role of Social Learning and Punishment in the Emergence of Norms and Traditions

On a general level, it is clear that many complex social behaviors in humans involve both social learning and the threat of punishment. Social learning is crucial for creating and propagating

Björn Lindström and Andreas Olsson, Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet, Sweden.

This research was supported by a grant from the Swedish Science Council (Vetenskapsrådet; 421-2010–2084) and an Independent Starting Grant (284366; Emotional Learning in Social Interaction project) from the European Research Council to Andreas Olsson. We thank Olof Hjorth and Paulina Wiktor for assistance with the data collection, and Johan Lundström, Bo Melin, and Magnus Enquist for valuable comments on previous draft of the article.

Correspondence concerning this article should be addressed to Björn Lindström, Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet (KI), Institutionen för klinisk neurovetenskap (CNS), K8, Psykologi, Nobels Väg 9 171 77 Stockholm, Sweden. E-mail:
AQ: 26 bjorn.lindstrom@ki

APA NLM
tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

large-scale, intergenerational patterns of behavior, such as those manifested in the adherence to social norms and traditions in humans (Boyd et al., 2011) and nonhuman primates (van de Waal, Borgeaud, Whiten, & Waal, 2013). It is important that many of these large-scale behavioral patterns involve the threat of danger or punishment, although the specific currency (e.g., physical, monetary, social) of the punishment varies between contexts. For example, social norms, such as those prescribing cooperation between unrelated individuals, can be maintained by the threat of punishment (Boyd, Gintis, & Bowles, 2010; Fehr & Fischbacher, 2004). Although such norms can be maintained by the swift punishment dealt to violators, the objective probability of punishment appears less important than the subjective *expectation* that punishment will follow norm-violating behaviors (Bicchieri, 2005). Norms can emerge and be maintained even if the actual probability of punishment is close to, or even, zero. For example, the phenomenon pluralistic ignorance describes that a norm can be sustained in the absence of any punishment for transgressions through each individual's false interpretation of why others' comply with the norm. In this way, pluralistic ignorance can lead to a state where most individuals believe that everyone else supports, and is willing to punish violations, of the norm (Bicchieri, 2005). Examples of norms under pluralistic ignorance include attitudes about drinking behaviors among college students, where most respondents mistakenly believed that they themselves were more uncomfortable with campus alcohol practices than was the average student, which can have detrimental consequences by promoting exaggerated drinking behavior to comply with the imagined norm (Prentice & Miller, 1993). Similar dynamics resulting from social learning have been described in models of information cascades in economics (Bikhchandani, 1998). Information cascades form when every individual base his or her choices on what others do under the (false) assumption that others' actions are based on private information about, for example, what alternative has the highest risk of punishment, while simultaneously disregarding their own private information. More direct evidence for how threat of, rather than actual, punishment can generate large-scale behavioral patterns comes from research on taboos (Aunger, 1994; Henrich & Henrich, 2010). Transgressions of religious taboos are often expected to result in supernatural punishment by the disapproving deity, providing further support for the suggestion that behavioral traditions can be maintained by threat of rare or nonexistent punishment (Aunger, 1994; Johnson & Krüger, 2004).[1] Even in the case of local food taboos that are on average adaptive (i.e., avoided foods are reliably associated with aversive outcomes), many nondangerous food types are subject to the same taboo (Henrich & Henrich, 2010), indicating that taboos can be maintained even when the objective probability of punishment is low. Such avoidance behaviors might involve aspects of magical thinking or superstition, which from an evolutionary perspective might represent rational responses to unlikely but highly dangerous outcomes (e.g., the reverse of the famous Pascal's wager; Foster & Kokko, 2009; Johnson, Blumstein, Fowler, & Haselton, 2013), but from a mental health perspective might be individually detrimental (Dymond, Schlund, Roche, De Houwer, & Freegard, 2012).

The Structure of Avoidance Traditions

Collectively, large-scale behavioral patterns involving threat of punishment can be described as "avoidance traditions." As described earlier, many avoidance traditions are likely to be adaptive, such as cooperative norms maintained by threat of punishment mentioned previously (Fehr & Fischbacher, 2004) and taboos prohibiting dangerous food (Henrich & Henrich, 2010). Other avoidance traditions are arbitrary or even destructive (Aunger, 1994; Mani, 1987), such as when behaviors that were adaptive in the past has become maladaptive in modern societies (Boyd et al., 2011).

In social organisms, avoidance traditions share a common underlying structure: (a) the organism believes punishment is possible in a particular context, and (b) the actions of others can be observed.[2] In addition, (c) most avoidance traditions are characterized by an uncertainty about what constitutes an optimal choice of action. Uncertainty can, for example, be caused by novelty of the situation or actions having delayed consequences and is known to amplify social influences on individual behavior (Claidière & Whiten, 2012). Furthermore, as discussed previously, avoidance traditions include both situations where the probability of punishment following transgressions is high and situations where it is low or zero. Examples of avoidance traditions meeting criteria (a)–(c) include everyday social interactions; for example, when an individual is placed in an unfamiliar social context, such as when starting a new job, joining a club, or as a freshman in college. The individual is aware that inappropriate action (e.g., certain topics of conversation when interacting with the new boss or peers) can be punished by social sanctions (e.g., scorn or social exclusion from the new boss or colleagues). Although others' behaviors are observable, there is an uncertainty in terms of which behaviors are appropriately displayed, for example, because the outcomes of others behaviors are ambiguous or delayed. Similarly, but placed in the nonsocial domain, an individual gathering or hunting food might be aware that certain fruits and animals in the surrounding can be toxic and dangerous. The individual can monitor others' behavior, but similar to the previous example, the effects of toxins might be ambiguous or delayed, leading to uncertainty to what actions are optimal (see Social Learning About Rare Punishments section for elaboration of this point).

Avoidance traditions have primarily been analyzed in evolutionary game theory, which has shown that punishment of deviant behaviors can be a possible explanation for both the emergence and maintenance of cooperation and for the maintenance of arbitrary social equilibria (Boyd & Gintis, 2003; Boyd & Richerson, 1992; Fehr & Gächter, 2002). In this research tradition, punishment is typically defined by a fitness cost incurred by the *punisher*, and thus requires an explanation at the ultimate, evolutionary level of analysis (Boyd & Gintis, 2003; Jensen, 2010). In the present study, we address a little studied but crucial aspect for understanding how punishment shapes avoidance traditions: the motivational

[1] It is of course possible that violation of religions taboos is associated with social forms of punishment.

[2] Note that "belief" does not imply an explicit or conscious representation. For example, many nonhuman animals behave as if they have beliefs about predation risk (Lima & Dill, 1990). The cognitive basis for such behaviors are, however, unknown.

APA NLM
tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

influence of punishment on the *punishee* and how this influence together with social learning can support the emergence, maintenance, and transmission of avoidance traditions. Despite the large theoretical and empirical literature investigating the influence of punishment on cooperation and other behavioral traditions, surprisingly little is known about the psychological and computational mechanisms of learning and decision-making underpinning this influence. What is known mainly relates to high-level constructs, such as social emotions (e.g., greed; Yamagishi & Sato, 1986) or individual differences in social motives (e.g., social value orientation; Van Lange, Joireman, Parks, & Van Dijk, 2013).

Learning Mechanisms Involved in Avoidance Traditions

A growing consensus in computational and cognitive neuroscience indicate that reinforcement learning (RL) constitutes a basic computational mechanism for human and animal learning and decision-making in both the asocial and social domain (Dayan & Balleine, 2002; Dunne & O'Doherty, 2013; Glimcher, 2011; Rangel, Camerer, & Montague, 2008; Ruff & Fehr, 2014; Seymour, Singer, & Dolan, 2007). RL, in brief, describes how the expected value of an action (e.g., to consume a food object) is determined by the previous negative or positive experiences resulting from that action. The future expected value of that action is then updated based on the difference between the organism's prediction and the actual outcome, the prediction error. From this learning theoretical perspective, it is predicted that avoidance traditions should form if the objective risk of punishment is high (because individuals can reliably learn from the punishment following their own actions), but it is unclear how rare or nonexisting punishment can generate avoidance traditions. In the typical game theoretical approach to avoidance traditions, the value of avoiding punishment is taken as given (Boyd, Gintis, Bowles, & Richerson, 2003). However, in humans and many other organisms with complex behavioral repertoires, it is likely that this value needs to be learned through experience to subsequently guide decision-making (Erev & Roth, 1998; Rangel et al., 2008; Seymour et al., 2007). Acquisition of such values might be relatively simple when the relationship between one's own behavior and resulting punishment is clear and consistent, but difficult if punishment is rare or nonexistent because of the lack of a direct punishment experience. Thus, the documented existence of avoidance traditions with low probability of punishment (see The Role of Social Learning and Punishment in the Emergence of Norms and Traditions section for examples) constitutes a puzzle. In the present study, we hypothesized that social learning might be a necessary complementary mechanism to threat of punishment for the emergence, maintenance, and transmission of avoidance traditions when punishment has a low objective probability. To decompose avoidance traditions into basic psychological mechanisms, we adopted a learning theoretical perspective defining punishment as any behavior-contingent cost (in any currency) incurred by the punishee that reduces the probability of repeating the punished behavior (Seymour et al., 2007; Solomon, 1964). By this definition, punishment includes negative consequences resulting from interactions with both animate and inanimate aspects of our environment, including aggressive conspecifics, predators, and toxic foods. We hypothesized that the interaction of two learning mechanisms: (a) the rewarding properties of avoiding an possible, threatening punishment, and (b) a tendency to copy the actions of others, which is amplified if the environment is perceived as dangerous and uncertain (Hoppitt & Laland, 2013; Webster & Laland, 2008a), can help to explain how avoidance traditions emergence, are maintained, and transmitted. We discuss these two mechanisms in turn.

Avoiding a Possible Punishment Is Rewarding

As outlined previously, threat of punishment exerts a potent influence on both human (Fehr & Gächter, 2002) and animal behavior (Cant, 2011; Jensen, 2010). It is important, and less appreciated, that threat of punishment instills successful avoidance of the possible punishment with reward value. Classical experiments in animal avoidance learning have shown that active avoidance of a punished behavior can be maintained long after the punishment is removed (Solomon, 1980). Behavioral avoidance in the absence of any external reinforcement initially presented a problem for learning theory, a conundrum that was later resolved by showing that the omission of an possible punishment functions as an internal reward (Solomon, 1980). This finding has subsequently been corroborated by human and nonhuman neuroscience research, which has demonstrated a symmetric response profile for procuring rewards and avoiding punishments in cortical valuation areas (Kim, Shimojo, & O'Doherty, 2006) and the involvement of the dopaminergic neurotransmitter in both processes (Ilango, Shumake, Wetzel, Scheich, & Ohl, 2012). Thus, in both humans and nonhuman animals, procuring rewards and avoiding punishments appear to be functionally equivalent in terms of their motivational value (Eder & Dignath, 2014; Solomon, 1980). We refer to this reinforcement mechanism as *rewarding punishment omission* (RPO, also known as an aversive inhibitor; Seymour et al., 2007). RPO causes a "sigh of relief," a reduction of anxiety or fear that can reinforce behavior (Boureau & Dayan, 2011; Gerber et al., 2014; Kim et al., 2006; Solomon, 1980). The consequence of this reduction in anxiety is that avoidance behaviors continue to be emitted even in the absence of any extrinsic aversive consequence. For example, a socially anxious individual tends to avoid new social situations because of the possible aversive consequences that might follow, which leads to an associated reduction of anxiety that reinforces continued avoidance. Crucially, continued avoidance prevents any opportunity to confirm the expectation that a specific action risks being punished (Dymond et al., 2012; Mineka & Zinbarg, 2006). Such chronic avoidance behaviors are believed to be a maintaining factor in many anxiety disorders (Borkovec, Alcaine, & Behar, 2004; Dymond et al., 2012; Mineka & Zinbarg, 2006). However, the role of RPO in behavioral avoidance traditions has hitherto been unexplored. We hypothesize that RPO can also function as a motivational basis for avoidance when punishment is rare or nonexistent, and that such avoidance tendencies can be transmitted between individuals by social learning. We discuss this possibility next.

Social Learning About Rare Punishments

When punishment is rare, such as when avoidance traditions involving actual dangers are well established or when the objective

probability of punishment is low, it is unlikely to support learning based on direct experience or even by observing the punishment administered to others (Seymour et al., 2007; Seymour, Yoshida, & Dolan, 2009). In computational terms, social learning of avoidance behaviors constitutes an "inverse avoidance problem" (where inverse refers to the estimation of a value function given observed actions, rather than the opposite which is the standard domain of RL; Seymour et al., 2007): How does an individual that observes others consistently avoid an action infer that this action might be punished, also after overt signs of emotional distress have extinguished? There are several possible solutions to this problem (see the Discussion section for analysis of alternatives), but the simplest solution is to directly assign the observed actions an "action value" that can serve as the basis for RL (Seymour et al., 2007). In psychological terms, basing one's behavior on the behavior of others, in the absence of any observable consequences, is termed *behavior copying* (BC; Giraldeau, Valone, & Templeton, 2002;
Fn3 Seymour et al., 2007).[3] Whereas several forms of BC have been demonstrated in nonhuman animals (Giraldeau et al., 2002; Matthews, Paukner, & Suomi, 2010; Rieucau & Giraldeau, 2011), human research in the aversive domain has focused on learning from observing punishing consequences to others (Olsson & Phelps, 2007; Selbing, Lindström, & Olsson, 2014). Behavior copying, even when ignorant of its consequences, is, however, likely to be adaptive also in humans, because most individuals will rationally try to procure rewards and avoid punishments (Rendell et al., 2010).

The Present Study

Here, we sought to understand how these two basic psychological mechanisms, RPO and BC, interact to shape behavior. We hypothesized that they together can provide building blocks to help explain the emergence, maintenance, and transmission of avoidance traditions, also when the objective probability of punishment is low or zero. We hypothesized that BC based on observational action values can influence behavior, which subsequently is reinforced by RPO (if no punishment follows) or actual punishment. The joint impact of BC and RPO on behavior would thereby present an important complement to game theoretic analyses of how punishment affects large scale behavior patterns. To address our hypothesis, we first developed and analyzed an agent-based simulation model that mimicked the core features of avoidance traditions that we discuss previously: (a) The organism believes punishment is possible in a particular context, (b) others' actions can be observed, and (c) there is an uncertainty about the optimal course of action. The use of agent-based simulation modeling, crucially, allowed us to investigate the large-scale consequences of the interaction of BC and RPO in avoidance traditions, which would be difficult or even impossible, to do in a laboratory (as it would involve prohibitively large groups and long time-spans). To provide empirical support for our agent-based simulation model of avoidance traditions, we implemented our hypothesis about BC and RPO in a serious of four behavioral experiments that each addressed specific aspects of the agent-based simulation model. In this way, we confirmed the realism of the key psychological assumptions underlying our simulation model of avoidance traditions. Similar to well-established models of social interaction, such as the prisoner's dilemma (Rapoport, 1965), our theoretical and experimental model is an abstraction aimed at capturing the essential structure of a complex phenomenon. We make no claim about the completeness of our theoretical account for explaining all (aspects) of avoidance traditions, but view it as a basic, and expandable, framework for understanding some of the key underlying mechanisms.

Method: Modeling Avoidance Traditions

We used agent-based simulation to model the large scale interaction of BC and RPO. Agent-based simulation has proven to be a valuable tool for understanding phenomena that emerge from the interaction of individual agents in complex and stochastic systems (Bonabeau, 2002). Agent-based simulations are widely used as both as "proof of principle" and as tools for analysis and prediction throughout biology and parts of the social sciences (Smith & Conrey, 2007), but have of yet been rarely used in psychology (but see Gray et al., 2014, for a notable recent exception). The model was implemented in NetLogo 5.0.2 (Wilensky, 1999; see Appendix for additional details).

The simulation model was intended to capture the essential structure common to avoidance traditions in both social and nonsocial domains. Individual agents made repeated decisions between two possible actions [*A*, *B*] under threat of punishment. Threat of punishment was not explicitly represented in the model but is a necessary auxiliary assumption for the modeling results. In natural settings, threat of punishment can, for example, be induced by previous individual learning (if the situation resemblances previous situations where punishment was possible), social information (e.g., verbal information about the dangers associated with specific foods, sacred texts prescribing certain taboos, news media reports about the likelihood of violent crime in an area), or, as for many nonhuman animals, by any cue signaling predation risk. In the model, none of the two actions were associated with external reward. Instead, we varied the probability that action *B* was punished, i.e., P(Punishment|B), to explore the role of objective punishment probability in avoidance traditions. To model a wide variety of avoidance traditions with the structure we outline above (see The Structure of Avoidance Traditions section) in both the social and asocial domains, the model had a simple structure without strategic interaction between individual agents. Instead, each agent first (Observation phase) observed *M* number of choices by another, randomly selected, agent (the "Demonstrator" agent), and subsequently chose individually *M* times (Choice phase; see the Appendix for details). To generate a simple "birth-death" process, the agents were probabilistically removed and replaced with completely naïve individuals. This structure was intended to model any situation where individuals enter a novel situation but can observe the actions of a (possibly) knowledgeable other before taking action. The model did not include any heritable traits.

[3] In humans, information transfer by social learning is often based on verbal or symbolic communication apart from direct observation of other (Laland & Rendell, 2013). Our focus on direct observation of the actions of others is motivated by the generality of this process both across situations and species (Laland & Rendell, 2013). The model thus represents a low-level version of social learning, which can be augmented by other means of information transfer.

Each individual agent was based on a simple RL model (*individual-level model*; see the Appendix for details) controlled by three learning parameters, regulating the individual learning rate (αI), the social learning rate (αO), and the psychological impact of RPO (Ω) (Dayan & Balleine, 2002). The individual learning rate (αI) regulates how sensitive the agent is to prediction errors, the difference between the actual and expected value of an action, which is used to update the future expected value of performing the same action. A high value (e.g., approaching 1) of this parameter result in a high sensitivity to unexpected outcomes and generally rapid learning. Similarly, the social learning rate (αO) determines how strongly social prediction errors, the difference between the observed and the expected behavior of another agent, affects the future expected value of the observed action for the observing agent. If this parameter is zero, the agents do not learn socially. The psychological impact of RPO (Ω) parameter affects how sensitive the decisions of the agent are to RPO, by scaling the reinforcement value of successfully avoiding a possible punishment. Low values (approaching zero) of the Ω parameter result in decisions that are strongly influenced by RPO, which in our model leads to a strong tendency for the agent to repeat unpunished actions. Psychologically, lower values of Ω would typically be associated with the successful avoidance of more severe punishments. In sum, the individual-level model gives a simple account of how BC and RPO interact. If the individual agent learns about the value of actions by observing another agent ($\alpha O > 0$), then the probability of choosing the action most often observed at the beginning of the Choice phase is >0.5, resulting in BC. The chosen action is subsequently reinforced by the (internal) reward elicited by RPO. RL has received extensive support in both human and nonhuman research and provides a coherent account of both behavioral and neural aspects of learning from rewards and punishments (Dayan & Balleine, 2002; Glimcher, 2011).

Results: Simulated Avoidance Traditions

We first explored how the presence of rare punishment affected the majority behavior in groups of $N = 100$ agents. To keep our report of the results as clear as possible, we focus on a specific parameterization in which most parameters were held constant

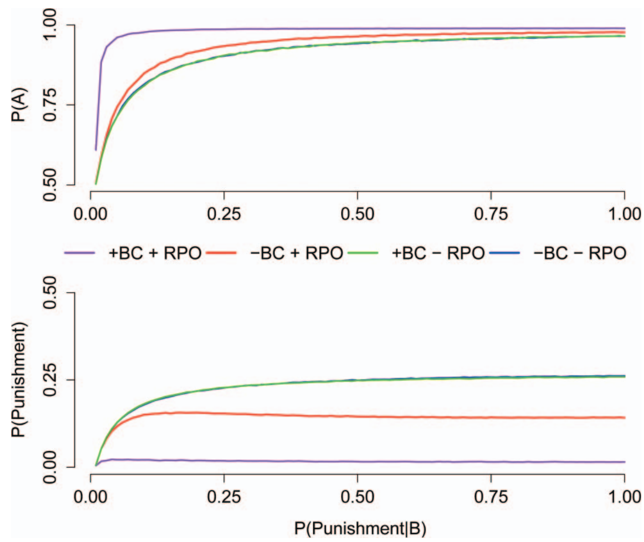
Fn4 ($\alpha O = 0.5$, $\alpha O = 0.5$, $\Omega = 0.03$, $M = 20$, $N = 100$).[4] In the supplemental material available online (see Sensitivity Analysis), we explored the effects of varying parameters that were kept constant in the main analyses.

We used two measures of the effect of RPO and BC on avoidance traditions: the proportion of individuals exhibiting the *A* behavior (which never was punished), *P*(*A*), and the average lifetime probability of (one) punishment. The latter was calculated by assuming a constant rate of individual punishment, converted to probability across the agents behaving life span (i.e., the Choice phase involving *M* choices):

$$P(Punishment) = 1 - e^{-rM}$$

where *r* is the average individual rate of punishment.

F1 Figure 1 shows that the objective probability of punishment, as expected, had a strong effect on both the proportion of individuals displaying the *A* behavior (i.e., the avoidance tradition) and the individual risk of punishment. In concordance with our hypothesis about the interaction between BC and RPO, reliable avoidance



ONLINE COLOR

Figure 1. The effect of objective punishment probability on the majority behavior (upper) and the average individual lifetime risk of punishment (lower). BC = behavior copying; RPO = rewarding punishment omission. $\Omega = 0.03$, $\alpha I = 0.5$, $\alpha O = 0.5$, $M = 20$, $N = 100$. + indicates the presence of the learning mechanism and − indicates absence of the mechanism. Each line represents averages across 100,000 time steps and 10 runs of the simulation. See the online article for the color version of this figure. AQ: 27

traditions emerged also when the objective probability of punishment was very low. It is important that these mechanisms also resulted in a very low individual risk of punishment across all levels of objective punishment probability. These results show that these mechanisms jointly can result in behavioral avoidance traditions at low (e.g., P(Punishment|B) < 0.01), but nonzero, punishment probabilities, and that the combination of BC and RPO adaptively decreased the individual risk of punishment, also when the *B* option was objectively highly dangerous. We discuss the explanation for these results next. AQ: 4

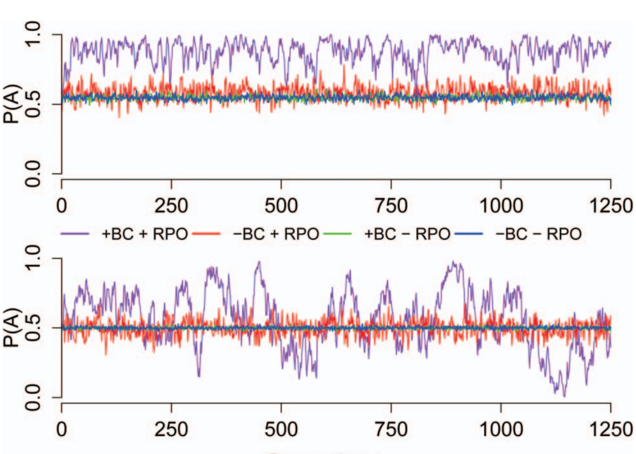
While it is intuitive that individuals will learn to avoid objectively dangerous options (with high probability of punishment) on their own, the pronounced avoidance traditions exhibited by populations with BC and RPO also at very low punishment probabilities show the potency of these mechanisms in interaction (which also is demonstrated by stark reduction in individual punishment risk across all levels of objective punishment probability; Figure 1). Therefore, we ran further simulations at zero, and near zero, probability of punishment to explore this effect further. The maximum difference in the proportion of individuals exhibiting the *A* behavior (i.e., the strength of the avoidance tradition) in groups with RPO and BC, relative to groups without both, could be seen when the probability of punishment was ~1% (Figure S4, available as supplemental material online). The explanation for these results is that at such low objective punishment probabilities, nonsocial learners ($\alpha O = 0$) failed to avoid *B* because the pun-

[4] Preliminary simulations showed that these parameter settings resulted in representative results. The value of the Ω parameter was arbitrarily set to a value at the lower end of the parameter ranger that reliably created avoidance traditions when the probability of punishment was zero or close to zero (Supplemental Figure 9).

APA NLM
tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

ishment was too rare to support direct individual learning. In contrast, social learners ($\alpha O > 0$) can profit from the punishment incurred by of earlier generations, and their resulting avoidance behaviors, through BC. Agents without RPO ($R = 0$) failed to maintain the tradition across time, because in the absence of external punishment, the behavior of these agents quickly became random. Crucially, these results show that rare punishment in itself is not sufficient for generating avoidance traditions. Instead, avoidance traditions require both social learning and a representation of the punishment threat that can generate RPO. In situations where the objective probability of punishment is high, RPO and BC together sharply reduces the risk of incurring a punishment (Figure 1, lower) by allowing the agent to make adaptive decision even in the absence of any direct prior experience of the situation.

F2 Arbitrary avoidance traditions also emerged with BC and RPO, but not without either (Figure 2, lower). In the absence of any punishment, i.e., P(Punishment|B) = 0), or initial bias for either action, these traditions lasted across hundreds of "generations." However, traditions in the absence of punishment were unstable: The average behavior tended to periodically switch, resembling how norms collapse following norm transgressions in the pluralistic ignorance phenomenon described in the introduction (see Social Learning and Punishment in Norms and Traditions). Similar dynamics have been described in economic models of information cascades (Bikhchandani, 1998). Such cascades easily form among rational decision-makers, but can be overturned if just one individual behaves differently. In our model, accumulated random fluctuations (e.g., mistakes) had the same effect; once a majority of agents prefer one option, a tipping point was reached. The existence of this tipping point depended jointly on BC and RPO; the oscillations between a majority of *A* or *B* choices suggest that RPO creates two "attractor states" (all *A* or all *B* choices rather than a mixture) on the individual level, which were propagated to naïve individuals through BC. However, if the objective probability of punishment is above zero, then the model suggests this will on average drive the majority behavior away from the punished action, although considerable stochasticity can be evident in the short term (Supplemental Figure 6).



ONLINE COLOR

Figure 2. Adaptive and arbitrary avoidance traditions. Two typical simulation runs, with (upper) and without (lower) punishment. In the upper simulation, behavior B had a 0.5% risk of being punished. BC = behavior copying; RPO = rewarding punishment omission. $\Omega = 0.03$, $\alpha I = 0.5$, $\alpha O = 0.5$, $M = 20$, $N = 100$. One generation = $2M$. + indicates the presence of the learning mechanism and − indicates absence of the mechanism. See the online article for the color version of this figure.

These results were stable for groups above ~50 agents. In small groups ($N < 30$ agents), single model runs involving rare punishment showed occasional temporary switches in the dominant behavior, from *A* to *B* (Supplemental Figures 5–6). Naturally, the average behavior is less sensitive to individual variability in larger, compared with smaller, groups (Franz & Matthews, 2010). Our results unequivocally demonstrate that BC and RPO jointly can create, maintain, and transmit both adaptive, P(Punishment|B) > 0, and arbitrary, P(Punishment|B) = 0, avoidance traditions. The initial emergence of avoidance traditions in the model depend on BC, and their maintenance depends on RPO. The transmission of avoidance traditions is logically contingent on their emergence and maintenance. In addition, transmission requires a degree of transmission fidelity that mainly is determined by the strength of RPO, which is regulated by the parameter Ω (Supplemental Figure 9). Taken together, our analysis of the agent-based simulation model showed that the interaction of BC and RPO can provide a mechanistic psychological underpinning for the emergence, maintenance, and transmission of avoidance traditions. In the presence of both common and rare punishment, these mechanisms can lead to adaptive avoidance traditions. In the absence of punishment, the same mechanisms can cause arbitrary traditions. Arbitrary behavioral traditions can thus be viewed as a byproduct of two basic, and generally adaptive, mechanisms (Franz & Matthews, 2010).

All results reported previously are based on groups composed of completely homogenous individual agents. In real world settings, it is known that individuals differ in their propensity to be influenced by the behavior of others (Claidière & Whiten, 2012). We assessed the robustness of the basic results (Figure 1) for such individual differences by varying the proportion of individuals using BC in the population. These simulations showed (Supplemental Figure 8) that avoidance traditions readily emerged also if a sizable minority of individuals never learned from others. Naturally, these traditions were quantitatively weaker if a large minority did not learn socially.

How do the psychological assumptions underlying the agent-based simulation model correspond to empirical human behavior? We have previously demonstrated that humans use social learning to take advantage of both observed actions (i.e., BC) and the observed outcomes of these actions when certain options were probabilistically punished by mild electric shocks, and that this behavior can be well-explained by computational RL models (Selbing et al., 2014). In that study, however, the probability of punishment was relatively high ($p = .8$), which thus corresponds to the higher range of objective punishment probabilities explored in our agent-based simulation model (Figure 1). It is unclear how humans behave at the very low end of objective punishment probabilities, where the joint effect of BC and RPO was most pronounced in our model simulations (Figures 1–2). Thus, we next sought to provide empirical support for the psychological realisms of the simulation-based findings that BC and RPO together can underpin the emergence, maintenance and transmission of avoidance traditions when punishment is rare or nonexistent. This was done by empirically testing the assumptions underlying the agent-based simulations. These assumptions can be reformulated as two simple predictions about human behavior: First, humans will tend to faithfully copy the actions of others (the Demonstrator) under threat of punishment, but in the absence of any actual punishment (Experiments 1–4). Second, behavior can in this way be transmitted between individuals to form avoidance traditions (Experiment 4). We evaluated these predictions by conducting four experiments with human participants.

Method: Experiments

Each of the four experiments was based on the same general design and addressed closely related questions and are therefore presented and discussed together.

Participants

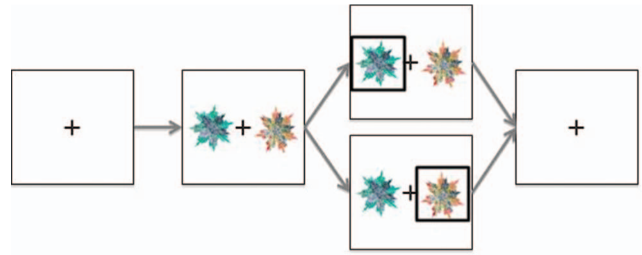
A total of 120 volunteers participated: 25 (18 women) in Experiment1, 20 (9 women) in Experiment 2, 25 (17 women) in Experiment 3, and 50 (25 women) in Experiment 4. All participants provided written informed consent and were rewarded with one movie voucher (value ≈ 11.5 USD). No individual participated in more than one experiment. All experimental procedures were approved by the local ethics committee at Karolinska Institutet.

Task and Stimuli

Prior to the experimental procedure, subjects were attached to the shock electrodes (all experiments except Experiment 2), and the shock amplitude was determined individually to be "uncomfortable but not painful." The electric shock stimulus was a monopolar 100 ms DC-pulse electric stimulation (STM200; Biopac Systems Inc, www.biopac.com) applied to the participant's nondominant forearm.

The participants in all experiments, except Experiment 2, received identical instructions. They were informed that they should choose between two images, that they first (Observation phase: 20 trials) would observe the choices, but not their consequences, of a previous participant in the experiment, and that they thereafter (Choice phase: 20 trials) would do the same task (i.e., choose between the same images) themselves during which they might receive shocks based on their choices. To provide unambiguous and clearly defined social information, we used a computer-controlled Demonstrator that consistently chose option *A* across all 20 Observation trials in Experiment 1–3. In Experiment 4, we relaxed this premise, and allowed the participants to observe the choices of real former experimental participants (see "Experiment F5 4"; Figure 5). In Experiment 2, the subjects received the same information as in the other experiments but were told that their choices could be awarded by points. These points would increase their chance of winning a lottery for two extra movie vouchers. They were instructed that delivery of points would be signaled with a "$" symbol. The "$" symbol was never shown, as an analogue to the omission of punishment in the other experiments.

The same stimuli were used for all experiments, and consisted of two abstract fractals with predominantly red or green color hue F3 (Figure 3). The stimulus corresponding to *A*, the stimulus chosen by the Demonstrator, was randomized for each subject. All trials began with a fixation cross (1,500–2,500 ms), followed by presentation of the two choice stimuli (2,000 ms). The choice (the Demonstrator's or subject's depending on experimental phase) was indicated by a black frame surrounding the stimulus (2,000 ms), followed by a blank white screen (3,000–5000 ms). In all experiments, all events were identical in appearance and timing for the Observation phase and the Choice phase. In Experiment 3, participants could randomly ($p = .5$) receive an electric shock during the blank white screen. In all experiments except Experiment 2, nonresponding had a 0.5 probability of being followed by a shock.



ONLINE COLOR

Figure 3. The experimental task. The design of the experimental task was identical in all experiments. Participants first observed the choices by the Demonstrator (Observation phase: 20 trials), and subsequently chose themselves (Choice phase: 20 trials). The color of the stimulus chosen by the Demonstrator (referred to as *A* in the text) was randomized across participants. The arrows indicate the temporal order of events. See the online article for the color version of this figure.

Statistical Analyses

The statistical analyses were conducted with Generalized Linear Mixed Models (GLMMs), using a logit link-function and random intercept for each participant (Jaeger, 2008). The reported tests of the difference of *P*(*A*) from 0.5 (i.e., the predicted value in the absence of BC) were conducted by testing if the fixed effects intercept differed from zero. The *P*(*A*) reported in Figure 4 is the F4 fixed effects intercept converted to the probability scale (from log-odds). These estimates are conditional on the distribution of individual behaviors in the sample. The consequence of this conditional weighting is that rare behaviors, e.g., a low *P*(*A*), has less influence on the estimate than if raw averaging is used. The point estimate thereby accounts for the distribution of the (experimental) population, which reduces the risk that it is unrepresentative of any given individual (i.e., Simpsons paradox; Dixon, 2008; Lee & Nelder, 2004). All results remain unchanged if summary statistics are analyzed instead using *t* tests or the nonparametric equivalent (supplemental material available online; see Statistical Analyses), but the point estimates of *P*(*A*) are generally lower. For the GLMM analysis, we report the simple effect size (i.e., unstandardized β estimate). Presently, no consensus methodology for effect sizes in GLMM exists, but the simple effect size has the desirable property of being easily interpreted (Baguley, 2009). Estimation of the individual-level RL model was done using maximum-likelihood optimization, which finds the set of parameter values that maximize the probability of the participant's trial-by-trial choices given the model. The model was fit individually to each participant's data (see Parameter Estimation, available online as supplemental material, for additional information).

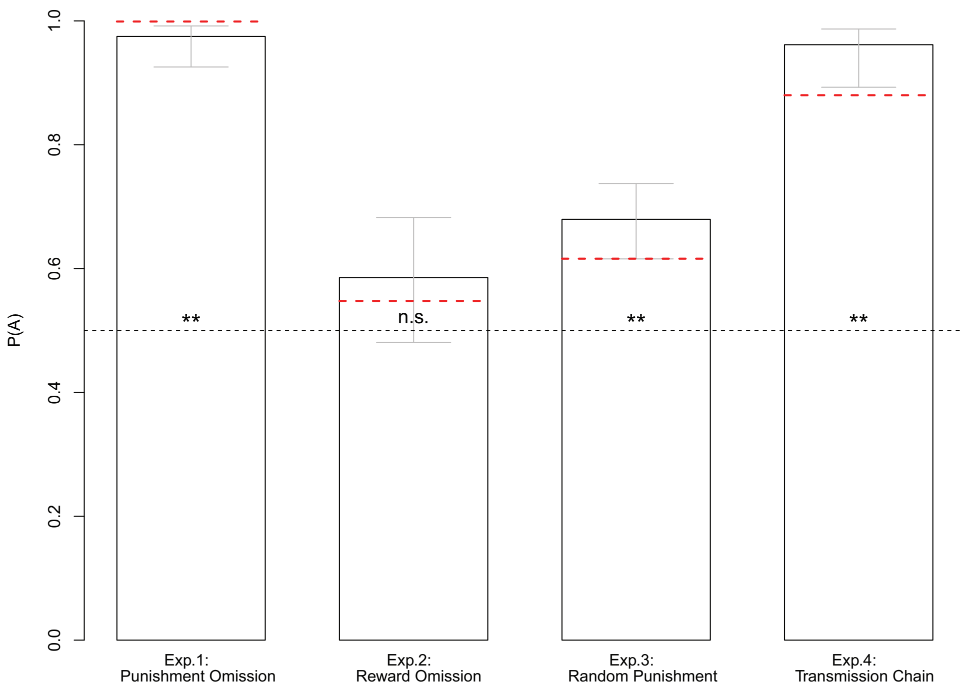


Figure 4. The probability of copying the Demonstrator choice (A) in Experiments 1–4. ** = $p < .01$ compared with 0.5. The red (dark gray) dotted lines show the simulated *P(A)* from the individual-level model with parameter values estimated from the empirical data. The error bars represent 1 *SE*. See the online article for the color version of this figure.

Results: Experimental Avoidance Traditions

The experiments mimicked the structure of the agent-based model, and had a computerized setup involving repeated choices between the two options [*A*, *B*] (represented by colored fractals; Figure 3). To induce the experience of threat of punishment, participants were informed that they might receive mild electric shocks depending on how they chose.

Experiment 1

The first experiment was directly designed to test the first prediction, which stated that humans will faithfully copy the behavior of others under threat of punishment. The participants believed their actions risked be punished, but they never received any punishment regardless of their choices. This setup mimicked conditions with very low, or zero, objective probability of punishment, but where the threat of punishment is constant (Gächter, Renner, & Sefton, 2008; Henrich & Boyd, 2001). Incorporating actual punishment at the low probabilities ($p < .01$) shown by the agent-based simulation to be effective in generating stable avoidance traditions was not experimentally possible because it would require an unfeasibly large number of experimental trials.

As predicted from our agent-based simulations, the social influence was strong under threat of punishment (logistic random effects regression against the predicted value in absence of BC [0.5]: $\beta = 3.66$, $SE = 1.14$, $p = .001$, 95% confidence interval (CI) [1.46, 5.89], see Figure 4 and Method for description of the regression approach). A large majority of participants chose *A* on all choice trials (Supplemental Figure S1), as predicted from the interaction of BC and RPO (Figure 4).[5] When interviewed about the reason for their choices after the experiment, many of the participants reported that they had continued to choose the first unpunished action throughout, and that they had wanted to test the other action, but did not dare for risk of punishment. These reports are qualitatively consistent with the assumption about RPO implemented in the individual-level model (see Discussion for alternative interpretations).

We fitted the individual level-model (Appendix: Equations 1–5) to each participant's choices to estimate the empirical parameter values, in order to verify that it could accurately describe human behavior (see Parameter Estimation, available online as supplemental material, and Supplemental Table 1 for details). The model-derived *P(A)* (based 1,000 runs of the individual-level model with parameter values based on the group median) were highly similar to the empirical *P(A)* (Figure 4). This indicates that the model, on average, did not misrepresent human behavior. This measure is, however, coarse because it does not take individual differences in the best fitting parameter values into account. To address this, we assessed the correlations between the estimated parameters and proportion of *A* responses, i.e., the individual *P(A)*, across the sample to verify that the model captured the qualitative patterns in the data. Both the estimated Ω, $r(23) = -0.49$, $p = .012$, and the αO, $r(23) = 0.59$, $p = .002$, parameters were

[5] There were large individual differences in the propensity to copy the actions of the Demonstrator in all experiments (Supplemental Figure 1). Note that the basic results from the agent-based simulation are robust for such differences (Supplemental Figure 8).

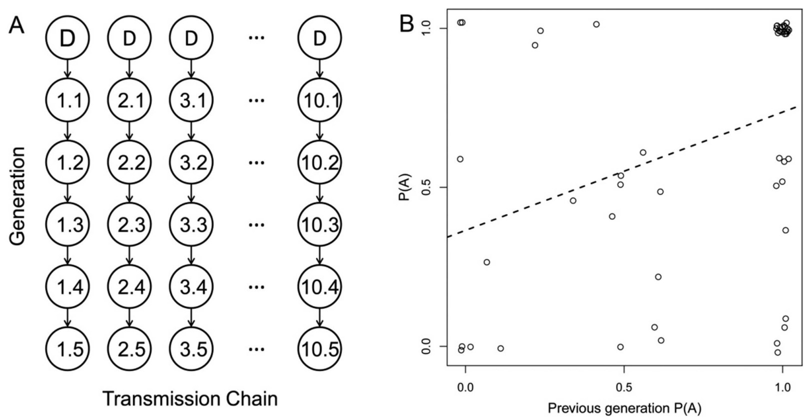


Figure 5. Experiment 4. (A) The transmission chain—design. Ten independent chains of participants observed the choices of the previous generation in the chain. The first generation observed the computerized "Demonstrator" (D), which consistently chose option *A*. (B) The average correlation in *P(A)* across generations. Note that the large cluster in the upper right corner represents individuals who faithfully copied the A behavior of the previous generation. The data points are slightly jittered for visibility.

correlated with the individual *P(A)* in the anticipated directions. Thus, stronger RPO (i.e., low value of the Ω parameter) and BC (higher values of the αO parameter) was associated with a stronger tendency to consistently copy the demonstrated action, in clear correspondence with the analysis of the agent-based simulation (see Sensitivity Analysis, available as supplemental material online). Next, we sought to establish the role of RPO for promoting BC by conducting a second experiment. In this experiment, we wanted to rule out any possibility that the lack of an external reinforcer in itself led to the strong tendency to copy the *A* action in Experiment 1, for instance, by retarding individual learning.

Experiment 2

Experiment 2 was identical to Experiment 1, except that the new sample of participants believed their choices could be rewarded. If the results of Experiment 1 did not involve RPO but rather reflected the lack of learning or other factors unrelated to threat of punishment, Experiment 2 should result in an identical pattern of results as Experiment 1. This was not the case. As predicted by our RPO and BC account (Figure 4), Experiment 2 resulted in a very different behavior pattern. The allocation of choices followed a random distribution (β = 0.35, *SE* = 0.42, *p* = .41, 95% CI [−0.48, 1.17], corresponding to ~ [0.38, 0.73] on the probability scale), a result our model explained by a simple mechanism: every *A* choice *not* resulting in reward (R = 0) caused a negative prediction error, and therefore more explorative future behavior. The joint action of the learning mechanisms, RPO and BC, in social learning of avoidance is potent because it motivates highly imitative and nonexplorative behavior. In contrast, omission of reward results in explorative behavior that overrides the social influence.

Experiment 3

Next, we wanted to (a) asses the qualitative strength of social learning during threat of punishment, and (b) establish that behavior was influenced also by external punishment in concordance with the predictions from the individual-level model. Experiment 3 was identical to Experiment 1, except that the participants choices were randomly (p = .5) punished by electric shocks. Consistent with the individual-level RL model simulation (Figure 4), external punishment and RPO together resulted in less faithful BC than in the absence of external punishment, but still with a clear social influence on behavior (Figure 4, difference against 0.5: β = 0.75, *SE* = 0.28, *p* = .008, 95% CI [0.2, 1.3]). In fact, when the demonstrated action actually was punished, it was still on average preferred. These results demonstrate how potent social learning is when people believe their actions risk being punished, and show the validity of the individual-level model for explaining behavior driven by both internal (e.g., RPO) and external reinforcement.

Experiment 4

Experiments 1–3 confirmed our first prediction, that humans faithfully copy the actions of others under threat of punishment, but in the absence of any actual punishment. Next, we conducted an experiment to empirically test prediction two; that avoidance behavior can be transmitted across individuals to form traditions. We used a linear "transmission chain" design (Figure 5) that closely resembled the structure of the agent-based simulation model. Five generations of human participants who were organized into 10 separate "chains" observed the decisions, but not their consequences, of the prior participant in the chain. As in Experiments 1–3, the first generation was exposed to a computerized Demonstrator choosing option *A* across all choice trials (Figure 5). Thereafter, we measured the proportion of choice *A* in each of the five generations, averaged over the 10 separate chains, to assess social transmission (Mesoudi & Whiten, 2008). Across all generations, *P(A)* was again above chance (β = 3.22, *SE* = 1.1, *p* = .003, 95% CI [1.06, 5.36]; Figure 4). On average, the proportion of *A* choices in each generation of the transmission chain significantly predicted the proportion of *A* choices in the subsequent generation, $r(48)$ = .35, *p* = .01, demonstrating that the behavior was transmitted across generations (Figure 5 and Addi-

tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

tional Analyses of Experiment 4, available as supplemental online material).

In qualitative resemblance to simulated groups composed of 10 agents, where the group-average behavior occasionally switched to *B* (Supplemental Figure 6), the majority behavior was stochastic: a random effects logistic regression (Additional analyses of Experiment 4 and Supplemental Figure 2, available online) showed a significant quadratic (but not linear) effect of generation (β = 17.74, *SE* = 5.37, $p < .001$) that indicated a decrease in the proportion of *A* choices in the third generation, followed by a subsequent increase (this was also reflected by high estimates for the Ω parameter in the third generation). The simulation model explains this pattern as because of two "attractor" states on the individual level; all *A* or all *B* choices, which result from RPO (see Results: Simulated Avoidance Traditions). Put differently, most people were likely to stick to the first unpunished response rather than mix responses. Considering the pattern of responses by generation clarifies this reasoning: in the third generation, four out of 10 participants actually showed mixed responses (i.e., individual *P(A)* ~0.5). In the fourth generation, however, only two individuals did so, whereas two individuals had a *P*(*A*) of ~0.2. Finally, in the fifth generation, only two individuals diverged from either *P*(*A*) = 1 or *P(A)* = 0. Thus, the "system" converges on either of the two extreme points. It should be apparent from this reasoning that the original majority behavior (*P(A)* = 1) can be both lost and regained in arbitrary avoidance traditions, but that the initial condition (e.g., the computerized demonstrator in Experiment 4) is important for the majority behavior (as shown by the on average above chance *P(A*) and correlation between generations). If the objective probability of punishment is above zero, our agent-based simulation model suggests this will on average drive the majority behavior away from the punished action, although considerable stochasticity can be evident in the short term. Notably, and as described above (see Results: Simulated Avoidance Traditions), larger groups are less sensitive to the behavior of any given individual.

Additional Analyses of Experiment 1–4

To provide a model-independent verification that RL is an appropriate framework for explaining the trial-by-trial data, we analyzed how the probability of choosing an action (e.g., *A*) on trial *t* was affected by successful avoidance of punishment on the previous trial, $t-1$. In other words, we tested the reinforcement effect of successful avoidance of punishment. In Experiment 1, where no external punishment could be received, the choice on *t* was predicted by the choice on $t-1$ (logistic random effects regression with by-subject random slopes for the effect of $t-1$: β = 7.94, *SE* = 1.36, z = 5.86, $p < .001$), indicating a strong tendency to stick with unpunished choices. As the individual-level model predicted that the first choice ($t = 1$) in Choice phase would be especially important for determining the subsequent behavioral trajectory, we calculated the proportion of *A* and *B* choices on $t = 2$, conditional on the choice on $t = 1$. The probability of choosing *A* on $t = 2$ was 1 if the choice on $t = 1$ was *A*, i.e., $P(A(t = 2)|A(t = 1)) = 1$. The indices for trial are henceforth dropped for clarity), and 0.2 if the choice on Trial 1 was B (i.e., $P(A|B) = 0.2$). For Experiment 4, which had the identical reinforcement structure as Experiment 1, the effects were highly similar, both averaged over trials (β = 9.99, *SE* = 1.34, z = 7.44, $p < .001$), and for the choices on Trial 2 ($P(A|A) = 0.97$, $P(A|B) = 0.17$). For Experiment 3, we took the external punishment into consideration by including the reception of a shock on $t-1$ as a predictor. Here, there was a main effect of the choice on $t-1$, $\chi^2(1) = 6.48$, p = .01, and an interaction between the choice on $t-1$ and if a shock was received on $t-1$, $\chi^2(1) = 34.33$, $p < .001$, which showed a strong tendency to repeat the same choice if unpunished and switch to the other choice if punished (too few response were available for directly calculating all four conditional probabilities on $t = 2$). Finally, in Experiment 2, where reward was omitted, there was also a tendency to repeat prior actions (β = 1.84, *SE* = 0.67, z = 2.71, p = .007), likely reflecting the general and well-known stickiness of decision-making (Erev & Barron, 2005; Samuelson & Zeckhauser, 1988), but as expected, this tendency was considerably weaker than when punishment was possible but avoided (tested by adding an interaction between the effect of choice on $t-1$ and experiment (Experiment 1 vs. Experiment 2: β = 5.30, *SE* = 1.71, z = 4.09, $p < .001$. Experiment 4 vs. Experiment 2: β = 4.63, *SE* = 1.44, z = 3.22, $p < .001$). In accordance with these differences, the conditional probabilities on $t = 2$ indicated explorative behavior (as discussed earlier): $P(A|A) = 0.6$, $P(A|B) = 0.37$. These results confirm that the RL framework was appropriate for modeling individual-level behavior.

Taken together, Experiments 1–4 provide strong support for our two predictions about human behavior: (a) humans faithfully copy the actions of others under threat of punishment, and (b) behavior can in this way be transmitted between individuals to form avoidance traditions. The experiments thereby confirm the key psychological assumptions underlying the agent-based model.

Discussion

Collectively, the results from our agent-based simulation model and the four behavioral experiments suggest that social learning has a strong effect on human behavior when the environment is perceived as dangerous, in striking analogue to the well-documented tendency of nonhuman animals to copy the behavior others during predation threat (Hoppitt & Laland, 2013; Webster & Laland, 2008b). To our knowledge, our experimental results represent the first comprehensive investigation of social influences on avoidance behaviors in humans, and thereby extend the literature on social learning of fear through classical conditioning to voluntary behavior and decision-making (Olsson & Phelps, 2007). Critically, together with the results from our agent-based simulation model, these results show how social learning (BC) in combination with the rewarding property of avoiding a threatening punishment (RPO) can underlie the emergence, maintenance, and transmission of avoidance traditions in humans. These findings show how low-level psychological mechanisms can underlie complex social behaviors (Franz & Matthews, 2010; Mesoudi, 2009) and offer an important complement to traditional game theoretic analyses of avoidance traditions by providing a parsimonious and mechanistic motivational basis for how individuals are affected by the threat of both rare and common punishments when the actions of others can be observed. These findings can thereby help explain how human (and perhaps some nonhuman) avoidance traditions take shape and spread when the environment is experienced as dangerous or risky, as exemplified by different types of social norms and taboos. Furthermore, our agent-based simulation showed that when punishment is rare, avoiding the dangerous option is a problem that only can be solved in a cumulative manner, across generations of individuals, for which social learning is indispensable (Boyd et al., 2011). Depending on the consequences of punishment, the importance of this solution may range from trivial to paramount, but is likely to have consequences on both the individual, and the group, level.

However, there might also be important downsides of social learning of avoidance behaviors. As outlined in the introduction, chronic and excessive avoidance behaviors are believed to be a maintaining factor in many anxiety related problems, such as social phobia (Borkovec et al., 2004; Dymond et al., 2012; Lissek et al., 2005; Mineka & Zinbarg, 2006). The joint results of the agent-based simulations and experiments suggest that avoidance behaviors are likely to be socially transmitted and maintained if individuals *believe* their actions risk being punished, regardless of the objective probability of punishment, hinting that magical thinking and superstitious behaviors can spread in groups of susceptible individuals, even in the absence of any actual danger (Foster & Kokko, 2009)

The role of the proposed psychological mechanisms, BC and RPO, are consistent with current knowledge about the computational and neurobiological basis of RL and value-based decision-making (Rangel et al., 2008; Seymour et al., 2007), and thereby provides a plausible explanation for how rare punishment can motivate behavior and how this behavior can spread in groups. As demonstrated by the agent-based simulation model, the combination of BC and RPO adaptively decreased the individual risk of punishment both when punishment had high and low probabilities, underscoring the generally adaptive nature of these mechanisms. It is important, as described earlier, that we also showed that arbitrary traditions can arise also in the absence of the effects of the same psychological mechanisms, BC and RPO. This finding provides a plausible psychological account that serve as a complement to the game theoretic demonstration that punishment in strategic interaction can make any behavior a Nash equilibrium, where no single individual can profit by changing their behavior, if alternative behaviors are punished (Boyd & Richerson, 1992).

Directly exploring how BC and RPO contribute to strategic social interactions represents an important future direction of research. For example, a certain variant of the serotonin receptor gene is associated with increased cooperative behavior in economic games in conditions with, but not without, punishment (Schroeder, McElreath, & Nettle, 2013), hinting at a serotonergic contribution to RPO in social interaction (Crişan et al., 2009). The combination of BC and RPO might be especially important in situations where either the payoff matrix governing the outcomes of an interaction or the characteristics of the other players are unknown (Seymour et al., 2009). For example, RPO might modify the objective payoff matrix by adding a positive value to any action not resulting in punishment, thereby bias behavior toward repeating this action. Together with BC, this might generate convergence on cooperative traditions.

Collectively, our findings also have relevance for animal social learning theory. Theoretical and empirical work on animal social learning has supported the existence of social learning strategies or heuristics, such as “copy when individual learning is costly” (Laland, 2004) that operate under similar conditions as those described here, namely that the environment is perceived as dangerous, the optimal course of action is uncertain, and the behavior of other individuals can be observed. Our findings show how two low-level psychological mechanisms, BC and RPO, together can result in the emergence of behaviors consistent with such social learning strategies (Heyes & Pearce, 2015).

Caveats and Future Directions

Clarifying the exact nature of the psychological representation of threat of punishment underlying the behavior in our experiments represents an important focus for future research. As described above (see Social Learning About Rare Punishments), directly assigning a positive action value to the observed actions of the demonstrator represent the simplest computational solution (i.e., model-free RL; Dolan & Dayan, 2013) to the “inverse avoidance problem” (Seymour et al., 2007). Alternatively, this problem could be solved by a more complex “model-based” approach where the agent infers that observed choices of *A* imply a negative value of *B* (Liljeholm, Molloy, & O’Doherty, 2012). Similarly, although we have assumed that a positive value of the *A* action (because of RPO) motivates continued avoidance of the *B* action, it is possible that humans by inference assign a negative value to the *B* action rather than a positive value to the *A* action (based on model-based RL; Dolan & Dayan, 2013). However, the fact that a simple RL model based on RPO successfully could account for the experimental results lend support for our assumptions about the motivational basis of behavior, substantiated by the well-established importance of the RPO mechanism in nonsocial avoidance learning is well-established both in humans (Kim et al., 2006; Solomon, 1980) and nonhuman animals (Solomon, 1980). It is important that these alternative representational accounts result in the same *behavioral* predictions as our model (because the difference in values between *A* and *B* are identical). The computational implementation of both the “inverse avoidance problem” and the motivating factor underlying continued avoidance in our model likely represent the simplest, and thus hopefully most general, form.

Our agent-based model demonstrated how large-scale complex behavioral traditions can emerge from a combination of two psychological mechanisms, BC and RPO, in a way not directly evident from viewing individual behavior in isolation (Bonabeau, 2002; Smith & Conrey, 2007). The use of agent-based simulation models as a way to explore situations and environments not directly available for experimental scrutiny (e.g., because of unfeasibly large experimental groups and time-spans) is an important aspect of the present study, and a possible methodological contribution to the field. Specifically, and in contrast to some recent examples of agent-based modeling in psychology (Gray et al., 2014), we verified the key psychological assumptions underlying the model in a series of experiments. The agent-based simulation model used in our study can be viewed as a bridge between the microlevel of individual behavior and the macrolevel of real-world avoidance traditions. The simple modeling framework presented in the present study can be elaborated and extended in several ways, and thereby hopefully applied to an even wider range of human avoidance traditions. We discuss some of these possible extensions, together with some important limitations and caveats associated with the current version of the model.

tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

Beyond the current focus on the interaction of BC and RPO, the findings from our agent-based simulation might inform about situations where RPO is *not* present, but where agents instead chose between two options that both are externally rewarded while one option is associated with punishment risk (because procuring rewards and avoiding punishment are both rewarding). Thus, the scope of the simulation-based findings extends to situations where external reward is possible. For example, a series of classical studies from Galef and colleagues showed that rats reliably can develop and sustain behavioral traditions involving choices between two equivalent food types over several generations if one of the food types were associated with punishment (a nausea-inducing injection) at the outset of the experiment (e.g., Galef & Aleen, 1995). Such patterns can readily be reproduced and explained by our agent-based simulation model. However, our experimental investigation of the psychological assumptions underlying the model where focused on the interaction of BC and RPO, which therefore precludes strong conclusions about situations with external rewards. Further verification and generalization of the assumptions underlying the model represent important routes for future empirical work, for example, by combining external rewards with threat of punishment or assessing the efficiency of other types of social learning (e.g., verbal or symbolical information) apart from direct observation in generating avoidance traditions.

Our simulations and experiments focused on situations where only the actions of one Demonstrator were observable. In reality, many sources of social information can concurrently be available, although there might be natural limits to how well attention can be allocated to multiple sources. The interaction of RPO and BC underlying our model is, however, unlikely to be contingent on the assumption of a single Demonstrator. To directly address this possible limitation, we performed additional analyses of a modified version of our agent-based model, where a new Demonstrator was chosen randomly on each on each point in time, thereby simulating situations where the actions of many individuals can be observed. These analyses confirmed that our results generalize to a multi-Demonstrator setting (see Multidemonstrator Model, available as online supplemental material). Both the behavior of the original and this modified agent-based model resembled the dynamic of informational cascades (Bikhchandani, 1998), where the actions of a few individuals come to influence the subsequent actions of many others through a "snowball" process. It is important that we also confirmed that our basic results hold in settings where more than two actions are available (see Three-Action Model, available online as supplemental material), which presumably is the case in many real world environments. Exploring how avoidance traditions emerge in such more complex environments represents an, in general, focus for future research.

Another factor that was excluded from the our agent-based simulation is the fact that avoidance traditions are likely to depend on the identity of the Demonstrator (Laland, 2004). Both humans and nonhumans are embedded in social networks, in which certain individuals are more likely to serve as demonstrators than others (Henrich & Broesch, 2011; Henrich & Henrich, 2010; Hoppitt, Boogert, & Laland, 2010). It is possible that such network structures will enhance the convergence and stability of avoidance traditions, for example if high status individuals reliably avoid certain actions or food sources (Henrich & Henrich, 2010). Such network based transmission biases represent another important future venue for research on avoidance traditions.

An important question arising from our results is how general the findings are in terms of different time scales and punishment currencies. Our agent-based model makes no assumptions about either of these important factors. The general pattern of results is predicted to generalize to any situations with the same underlying structure, and where the psychological impact of avoiding threatening punishment is sufficiently potent. Such simplicity in the modeling formulation is commonly seen as a virtue, because the model might reveal explanatory generalities across many situations (Evans et al., 2013). It can, however, also be criticized for not actually corresponding to any natural system (Evans et al., 2013). Our experimental tests of the model assumptions, however, are naturally limited to a specific timescale (brief) and a punishment currency (electric shocks), and mainly pertain to arbitrary avoidance traditions since sufficiently rare punishment for adaptive avoidance traditions cannot be experimentally administered on the time scale of an experimental session. This situation is common to many experimental tests of theoretical models, such as the Prisoner's Dilemma. The Prisoner's Dilemma model is considered to have explanatory power for human behaviors at different time scales, and in a variety of payoff currencies, but experimental tests are normally restricted to the behavioral laboratory with monetary payoffs (Guala, 2012). The discrepancy between theoretical scope and experimental investigations is an important topic of current debate (Guala, 2012). Nevertheless, there is a general agreement that lab experiments are important for model and theory development, and that simple models do contribute to understanding of complex phenomena (Bowles & Gintis, 2011; Fehr & Gächter, 2002; Janssen, Holahan, Lee, & Ostrom, 2010).

Conclusion

In sum, we have demonstrated empirically, across four behavioral experiments, that social learning (BC) has a strong influence on human behavior when the environment is perceived as threatening. We have shown theoretically, using agent-based simulations, that this social influence, together with the rewarding effect of avoiding expected punishment (RPO), might provide a novel mechanistic psychological basis for the generation, maintenance, and transmission of both adaptive and arbitrary avoidance traditions in humans. Because the combined workings of the BC and RPO mechanisms require no intrinsic tendency for conformity or personal experience of punishment to maintain avoidance traditions, it is plausible that they can help explain many different large scale behavioral patterns involving threat and avoidance of punishment, such as norms and traditions, in humans and nonhuman animals alike.

References

Aunger, R. (1994). Are food avoidances maladaptive in the Ituri Forest of Zaire? *Journal of Anthropological Research, 50,* 277–310.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100,* 603–617. http://dx.doi.org/10.1348/000712608X377117

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms.* New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511616037

Bikhchandani, S. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives, 12,* 151–170.

Bonabeau, E. (2002). 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.

Borkovec, T. D., Alcaine, O. M., & Behar, E. (2004). Avoidance theory of worry and generalized anxiety disorder. In R. G. Heimberg, C. L. Turk, & D. S. Mennin (Eds.), *Generalized anxiety disorder: Advances in research and practice* (pp. 77–108). New York, NY: Guilford Press.

Boureau, Y.-L., & Dayan, P. (2011). Opponency revisited: Competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36,* 74–97. http://dx.doi.org/10.1038/npp.2010.151

Bowles, S., & Gintis, H. (2011). *A cooperative species: Human reciprocity and its evolution* (p. 280). Princeton, NJ: Princeton University Press. Retrieved from http://www.amazon.com/Cooperative-Species-Human-Reciprocity-Evolution/dp/0691151253. http://dx.doi.org/10.1515/9781400838837

Boyd, R., & Gintis, H. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 3531–3535.

Boyd, R., Gintis, H., & Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science, 328*, 617–620. http://dx.doi.org/10.1126/science.1183665

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 3531–3535. http://dx.doi.org/10.1073/pnas.0630443100

Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*, 171–195.

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences, of the United States of America, 108* (Suppl. 2), 10918–10925. http://dx.doi.org/10.1073/pnas.1100290108

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 14431–14436. http://dx.doi.org/10.1073/pnas.1003111107

Cant, M. A. (2011). The role of threats in animal cooperation. *Proceedings Biological Sciences/The Royal Society*, *278,* 170–178. http://dx.doi.org/10.1098/rspb.2010.1241

Claidière, N., & Whiten, A. (2012). Integrating the study of conformity and culture in humans and nonhuman animals. *Psychological Bulletin, 138,* 126–145. http://dx.doi.org/10.1037/a0025868

Cook, M., & Mineka, S. (1990). Selective associations in the observational conditioning of fear in rhesus monkeys. *Journal of Experimental Psychology: Animal Behavior Processes, 16,* 372–389. http://dx.doi.org/10.1037/0097-7403.16.4.372

Crişan, L. G., Pana, S., Vulturar, R., Heilman, R. M., Szekely, R., Druğa, B., . . . Miu, A. C. (2009). Genetic contributions of the serotonin transporter to social learning of fear and economic decision making. *Social Cognitive and Affective Neuroscience*, *4,* 399–408. http://dx.doi.org/10.1093/scan/nsp019

Curio, E., Ernst, U., & Vieth, W. (1978). Cultural transmission of enemy recognition: One function of mobbing. *Science, 202,* 899–901. http://dx.doi.org/10.1126/science.202.4370.899

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron, 36,* 285–298. http://dx.doi.org/10.1016/S0896-6273(02)00963-7

Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science*, *335,* 1114–1118. http://dx.doi.org/10.1126/science.1213969

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language*, *59,* 447–456. http://dx.doi.org/10.1016/j.jml.2007.11.004

Dolan, R. J. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron, 80,* 312–325. http://dx.doi.org/10.1016/j.neuron.2013.09.007

Dunne, S., & O'Doherty, J. P. (2013). Insights from the application of computational neuroimaging to social neuroscience. *Current Opinion in Neurobiology*, *23*, 387–392. http://dx.doi.org/10.1016/j.conb.2013.02.007

Dymond, S., Schlund, M. W., Roche, B., De Houwer, J., & Freegard, G. P. (2012). Safe from harm: Learned, instructed, and symbolic generalization pathways of human threat-avoidance. *PLoS ONE*, *7,* e47539. http://dx.doi.org/10.1371/journal.pone.0047539

Eder, A. B., & Dignath, D. (2014). I like to get nothing: Implicit and explicit evaluation of avoided negative outcomes. *Journal of Experimental Psychology: Animal Learning and Cognition*, *40,* 55–62. http://dx.doi.org/10.1037/xan0000005

Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist simulation of attitude learning: Asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin*, *29,* 1221–1235. http://dx.doi.org/10.1177/0146167203254605

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, *88,* 848–881.

Evans, M. R., Grimm, V., Johst, K., Knuuttila, T., de Langhe, R., Lessells, C. M., . . . Benton, T. G. (2013). Do simple models lead to generality in ecology? *Trends in Ecology & Evolution*, *28,* 578–583. http://dx.doi.org/10.1016/j.tree.2013.05.022

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25,* 63–87.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415,* 137–140.

Foster, K. R., & Kokko, H. (2009). The evolution of superstitious and superstition-like behaviour. *Proceedings of the Royal Society B: Biological Sciences, 276,* 31–37.

Franz, M., & Matthews, L. J. (2010). Social enhancement can create adaptive, arbitrary and maladaptive cultural traditions. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 3363–3372. http://dx.doi.org/10.1098/rspb.2010.0705

Gächter, S., Renner, E., & Sefton, M. (2008). The long-run benefits of punishment. *Science*, *322*, 1510.http://dx.doi.org/10.1126/science.1164744

Galef, B. G., Jr., & Aleen, C. (1995). A new model system for studying behavioural traditions in animals. *Animal Behaviour, 50,* 705–717. http://dx.doi.org/10.1016/0003-3472(95)80131-6

Gerber, B., Yarali, A., Diegelmann, S., Wotjak, C. T., Pauli, P., & Fendt, M. (2014). Pain-relief learning in flies, rats, and man: Basic research and applied perspectives. *Learning & Memory, 21,* 232–252. http://dx.doi.org/10.1101/lm.032995.113

Giraldeau, L.-A., Valone, T. J., & Templeton, J. J. (2002). Potential disadvantages of using socially acquired information. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 357,* 1559–1566. http://dx.doi.org/10.1098/rstb.2002.1065

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 108* (Suppl. 3), 15647–15654. http://dx.doi.org/10.1073/pnas.1014269108

Gray, K., Rand, D. G., Ert, E., Lewis, K., Hershman, S., & Norton, M. I. (2014). The emergence of "us and them" in 80 lines of code: Modeling group genesis in homogeneous populations. *Psychological Science*, *25*, 982–990. http://dx.doi.org/10.1177/0956797614521816

Griffin, A. (2004). Social learning about predators: A review and prospectus. *Animal Learning & Behavior, 32,* 131–140.

Guala, F. (2012). Reciprocity: Weak or strong? What punishment experiments do (and do not) demonstrate. *Behavioral and Brain Sciences, 35,* 1–15. http://dx.doi.org/10.1017/S0140525X11000069

Henrich, J., & Boyd, R. (2001). Why people punish defectors. Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208,* 79–89. http://dx.doi.org/10.1006/jtbi.2000.2202

Henrich, J., & Broesch, J. (2011). On the nature of cultural transmission networks: Evidence from Fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 366*(1567), 1139–1148.

Henrich, J., & Henrich, N. (2010). The evolution of cultural adaptations: Fijian food taboos protect against dangerous marine toxins. *Proceedings Biological Sciences/The Royal Society, 277,* 3715–3724. http://dx.doi.org/10.1098/rspb.2010.1191

Heyes, C., & Pearce, J. M. (2015). Not-so-social learning strategies. *Proceedings of the Royal Society B: Biological Sciences,* Advance online publication. http://dx.doi.org/10.1098/rspb.2014.1709

Hoppitt, W., Boogert, N., & Laland, K. (2010). Detecting social transmission in networks. *Journal of Theoretical Biology, 263,* 544–555.

Hoppitt, W., & Laland, K. (2013). *Social learning: An introduction to mechanisms, methods, and models*. Princeton, NJ: Princeton University Press.

Huys, Q. J. M., & Dayan, P. (2009). A Bayesian formulation of behavioral control. *Cognition, 113,* 314–328. http://dx.doi.org/10.1016/j.cognition.2009.01.008

Ilango, A., Shumake, J., Wetzel, W., Scheich, H., & Ohl, F. W. (2012). The role of dopamine in the context of aversive stimuli with particular reference to acoustically signaled avoidance learning. *Frontiers in Neuroscience, 6,* 132.http://dx.doi.org/10.3389/fnins.2012.00132

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Janssen, M. A., Holahan, R., Lee, A., & Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. *Science, 328,* 613–617. http://dx.doi.org/10.1126/science.1183532

Jensen, K. (2010). Punishment and spite, the dark side of cooperation. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 365,* 2635–2650. http://dx.doi.org/10.1098/rstb.2010.0146

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology & Evolution, 28,* 474–481. http://dx.doi.org/10.1016/j.tree.2013.05.014

Johnson, D. D. P., & Krüger, O. (2004). The good of wrath: Supernatural punishment and the evolution of cooperation. *Political Theology, 5,* 159–176. http://dx.doi.org/10.1558/poth.2004.5.2.159

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology, 4,* e233. http://dx.doi.org/10.1371/journal.pbio.0040233

Laland, K. N. (2004). Social learning strategies. *Learning & Behavior, 32,* 4–14. http://www.ncbi.nlm.nih.gov/pubmed/15161136. http://dx.doi.org/10.3758/BF03196002

Laland, K. N., & Rendell, L. (2013). Cultural memory. *Current Biology, 23,* R736–R740. http://dx.doi.org/10.1016/j.cub.2013.07.071

Lawson, R. P., Seymour, B., Loh, E., Lutti, A., Dolan, R. J., & Dayan, P., . . . Roiser, J. P. (2014). The habenula encodes negative motivational value associated with primary punishment in humans. *Proceedings of the National Academy of Sciences of the United States of America, 111,* 11858–11863. http://dx.doi.org/10.1073/pnas.1323586111

Lee, Y., & Nelder, J. A. (2004). Conditional and marginal models: Another view. *Statistical Science, 19,* 219–238. Retrieved from http://projecteuclid.org/euclid.ss/1105714159. http://dx.doi.org/10.1214/088342304000000305

Liljeholm, M., Molloy, C. J., & O'Doherty, J. P. (2012). Dissociable brain systems mediate vicarious learning of stimulus-response and action-outcome contingencies. *The Journal of Neuroscience, 32,* 9878–9886. http://dx.doi.org/10.1523/JNEUROSCI.0548-12.2012

Lima, S., & Dill, L. (1990). Behavioral decisions made under the risk of predation: A review and prospectus. *Canadian Journal of Zoology, 68,* 619–640. http://dx.doi.org/10.1139/z90-092

Lissek, S., Powers, A. S., McClure, E. B., Phelps, E. A., Woldehawariat, G., Grillon, C., & Pine, D. S. (2005). Classical fear conditioning in the anxiety disorders: A meta-analysis. *Behaviour Research and Therapy, 43,* 1391–1424. http://dx.doi.org/10.1016/j.brat.2004.10.007

Mani, L. (1987). Contentious traditions: The debate on sati in colonial India. *Cultural Critique, 7,* 119–156. http://dx.doi.org/10.2307/1354153

Matthews, L. J., Paukner, A., & Suomi, S. J. (2010). Can traditions emerge from the interaction of stimulus enhancement and reinforcement learning? An experimental model. *American Anthropologist, 112,* 257–269. http://dx.doi.org/10.1111/j.1548-1433.2010.01224.x

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325,* 284–288. http://dx.doi.org/10.1126/science.1175626

Mesoudi, A. (2009). How cultural evolutionary theory can inform social psychology and vice versa. *Psychological Review, 116,* 929–952. http://dx.doi.org/10.1037/a0017062

Mesoudi, A., & Whiten, A. (2008). Review: The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 363,* 3489–3501. http://dx.doi.org/10.1098/rstb.2008.0129

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: It's not what you thought it was. *American Psychologist, 61,* 10–26. http://dx.doi.org/10.1037/0003-066X.61.1.10

Olsson, A., & Phelps, E. A. (2007). Social learning of fear. *Nature Neuroscience, 10,* 1095–1102. http://dx.doi.org/10.1038/nn1968

Prentice, D. A., & Miller, D. T. (1993). Pluralistic ignorance and alcohol use on campus: Some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology, 64,* 243–256. http://dx.doi.org/10.1037/0022-3514.64.2.243

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience, 9,* 545–556. http://dx.doi.org/10.1038/nrn2357

Rapoport, A. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., . . . Laland, K. N. (2010). Why copy others? Insights from the social learning strategies tournament. *Science, 328,* 208–213. http://dx.doi.org/10.1126/science.1184719

Rieucau, G., & Giraldeau, L. (2011). Exploring the costs and benefits of social information use: An appraisal of current experimental evidence. *Proceedings of the Royal Society B: Biological Sciences, 366,* 949–957.

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature Neuroscience, 17,* 1607–1612. http://dx.doi.org/10.1038/nn.3832 AQ: 21

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience, 15,* 549–562. http://dx.doi.org/10.1038/nrn3776

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty, 1,* 7–59. http://dx.doi.org/10.1007/BF00055564

tapraid5/zfr-xge/zfr-xge/zfr00315/zfr2639d15z | xppws | S=1 | 3/27/15 | 10:30 | Art: 2014-0653

Schroeder, K. B., McElreath, R., & Nettle, D. (2013). Variants at serotonin transporter and 2A receptor genes predict cooperative behavior differentially according to presence of punishment. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 3955–3960. http://dx.doi.org/10.1073/pnas.1216841110

Selbing, I., Lindström, B., & Olsson, A. (2014). Demonstrator skill modulates observational aversive learning. *Cognition, 133,* 128–139. http://dx.doi.org/10.1016/j.cognition.2014.06.010

Seymour, B., Singer, T., & Dolan, R. (2007). The neurobiology of punishment. *Nature Reviews Neuroscience, 8,* 300–311. http://dx.doi.org/10.1038/nrn2119

Seymour, B., Yoshida, W., & Dolan, R. (2009). Altruistic learning. *Frontiers in Behavioral Neuroscience, 3,* 23. http://dx.doi.org/10.3389/neuro.08.023.2009

Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review, 11,* 87–104.

Solomon, R. L. (1964). Punishment. *American Psychologist, 19,* 239–253.

Solomon, R. L. (1980). The opponent-process theory of acquired motivation: The costs of pleasure and the benefits of pain. *American Psychologist, 35,* 691–712. http://www.ncbi.nlm.nih.gov/pubmed/7416563. http://dx.doi.org/10.1037/0003-066X.35.8.691

Tomasello, M., Kruger, A. C., & Ratner, H. H. (2010). Cultural learning. *Behavioral and Brain Sciences, 16,* 495. http://dx.doi.org/10.1017/S0140525X0003123X

van de Waal, E., Borgeaud, C., Whiten, A., & van de Waal, E. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science, 340,* 483–485. http://dx.doi.org/10.1126/science.1232769

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120,* 125–141. http://dx.doi.org/10.1016/j.obhdp.2012.11.003

Webster, M., & Laland, K. (2008a). Social learning strategies and predation risk: Minnows copy only when using private information would be costly. *Proceedings of the Royal Society B: Biological Sciences, 275*(1653), 2869–2876. http://dx.doi.org/10.1098/rspb.2008.0817

Webster, M. M., & Laland, K. N. (2008b). Social learning strategies and predation risk: Minnows copy only when using private information would be costly. *Proceedings of the Royal Society B: Biological Sciences, 275,* 2869–2876. http://dx.doi.org/10.1098/rspb.2008.0817

Wilensky, U. (1999). *Netlogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. Retrieved from http://ccl.northwestern.edu/netlogo/

Yamagishi, T., & Sato, K. (1986). Motivational bases of the public goods problem. *Journal of Personality and Social Psychology, 50,* 67–73. http://dx.doi.org/10.1037//0022-3514.50.1.67

(*Appendix follows*)

Appendix

Agent-Based Simulations

Scheduling

The agent-based simulations were implemented in NetLogo 5.0.2 (Wilensky, 1999). The simulation scheduling is summarized in pseudocode. The simulation proceeded in discrete time steps, where each agent either observed the action of another agent, or chose themselves between *A* and *B* on each time-step.

For each agent [pick one other agent at random and set as Demonstrator. Set own Age to 0]

For each Time-step to total number of time-steps

Set Time-step = Time-step + 1;

For each agent

If Demonstrator = none [pick another agent at random and set as Demonstrator]

If Age < = M(Observation-trials) and Time-steps > M [do observational learning (Appendix: Equations 3–4)]

Else [do individual learning and decision-making (Appendix: Equations 1–2 and Equation 5)]

If random number between 0 and 1 < P(death) [remove agent and introduce a completely naïve new agent]

End

End

End

The removal rate, P(death) was set to 1/(Observation-trials * 2), resulting in average lifetime 2M.

Individual-Level Model

Each agent represented the expected value of the actions as $Q_i \in [Q_A, Q_B]$. The individual level model was specified as:

$$Q_i(t+1) = Q_i(t) + \alpha I * \delta(t) \tag{1}$$

$$\delta(t) = R(t) - Q_i(t) \tag{2}$$

where αI is the asocial learning rate, which regulates how the difference between the actual and expected value, the prediction error $\delta_i(t)$, affects the future expected value of that action at $t + 1$. The outcome, R, was set to 1 to represent RPO and 0 to represent the absence of RPO (Eiser, Fazio, Stafford, & Prescott, 2003; Huys & Dayan, 2009). Punishment was represented by setting $R = -1$. Thus, RPO was rewarding if punishment was expected.

Social learning (Observation phase) was modeled identically as individual learning, but with a separate learning rate (αO) that determined the rate of the social influence (Burke, Tobler, Baddeley, & Schultz, 2010). The observed action was assumed to have a positive action value ($R = 1$), leading to a positive Q value associated with the action observed with the highest frequency at the outset of the Choice phase if $\alpha O > 0$.

$$Q_i(t+1) = Q_i(t) + \alpha O * \delta(t) \tag{3}$$

$$\delta(t) = 1 - Q_i(t) \tag{4}$$

An equivalent formulation of observational learning would be to treat R as a parameter and αO as a positive constant. The probability that an individual would chose A was proportional to the difference in value between the n alternatives (weighted by the parameter Ω, which determines the psychological magnitude of RPO. Note that this parameter in the Softmax function is commonly interpreted as determining the exploration/exploitation tendency of the decision-maker. In the current context we prefer to view the parameter as scaling the difference in value between the stimuli) according to the Softmax function:

$$P(A)(t) = \frac{e^{Q_A(t)\Omega}}{\sum_{i=1}^{n} e^{Q_i(t)\Omega}} \tag{5}$$

In sum, the individual-level model gives a simple account of how BC and RPO interact. If the individual assign a positive value to the actions of another agent ($\alpha O > 0$), the probability of choosing the action most often observed at the beginning of the Choice phase is >0.5, resulting in BC. The chosen action is thereafter maintained by the (internal) reward elicited by RPO. It's important to note that the individual-level model did not include any intrinsic bias toward conformity (Claidière & Whiten, 2012).

Received November 8, 2014
Revision received February 16, 2015
Accepted February 23, 2015 ■