

© 2022 American Psychological Association ISSN: 0022-3514 2022, Vol. 123, No. 2, 249-271 https://doi.org/10.1037/pspa0000301

# INNOVATIONS IN SOCIAL PSYCHOLOGY

The Importance of Being Unearnest: Opportunists and the Making of Culture

Ivan Hernandez<sup>1</sup>, Dov Cohen<sup>2</sup>, Karl Gruschow<sup>3</sup>, Andrzej Nowak<sup>4, 5</sup>,

Michele J. Gelfand<sup>6</sup>, and Wojciech Borkowski<sup>4</sup>

<sup>1</sup> Department of Psychology, Virginia Polytechnic Institute and State University

<sup>2</sup> Department of Psychology, University of Illinois

<sup>3</sup> Illinois Criminal Justice Information Authority, Chicago, Illinois, United States

<sup>4</sup> The Robert B. Zajonc Institute for Social Studies, University of Warsaw

<sup>5</sup> Department of Psychology, Florida Atlantic University

<sup>6</sup> Graduate School of Business, Stanford University

Opportunistic actors—who behave expediently, cheating when they can and offering minimal cooperation only when they have to—play an important role in producing some puzzling phenomena, including the flourishing of strong reciprocity, the peculiar correlation between positive and negative reciprocity within cultures of honor, and low levels of social capital within tight and collectivist cultures (that one might naively assume would produce high levels of social capital). Using agent-based models and an experiment, we show how Opportunistic actors enable the growth of Strong Reciprocators, whose strategy is the exact opposite of the Opportunists. Additionally, previous research has shown how the threat of punishment can sustain cooperation within a group. However, the present studies illustrate how stringent demands for cooperation and severe punishments for noncooperation can also backfire and reduce the amount of *voluntary, uncoerced* cooperation in a society. The studies illuminate the role Opportunists play in producing these backfire effects. In addition to highlighting other features shaping culture (e.g., risk and reward in the environment, "founder effects" requiring a critical mass of certain strategies at a culture's initial stage), the studies help illustrate how Opportunists create aspects of culture that otherwise seem paradoxical, are dismissed as "error," or produce unintended consequences.

Keywords: honor, culture, social capital, individualism-collectivism, tightness

Supplemental materials: https://doi.org/10.1037/pspa0000301.supp

Psychologists studying culture often think about the way cultures socialize their members to behave appropriately (Kim & Lawrie, 2019; Kitayama et al., 2019; Wrong, 1961). Psychologists studying the evolution of prosocial behavior often think about the strong predispositions humans have to behave fairly and to punish those who transgress social norms (Bian et al., 2018; Dawkins et al., 2019; Margoni et al., 2018; Mikulincer & Shaver, 2010; Sloane et al., 2012; Ting et al., 2019). Well-behaved actors—whether socialized into or adapted for their good behavior—are crucial for thinking

about the development of social systems. Psychologists have paid much less attention to opportunistic actors, but they, too, are important for understanding the emergence of social systems.

By opportunists we mean people who have no strong predisposition to behave in one way or another. They act expediently cheating when they can and cooperating only when they have to. Intuitively, such opportunists might seem likely to overrun a system and make it dysfunctional. However, there seems to be a crucial balancing act that occurs between opportunists and well-socialized

Michele J. Gelfand https://orcid.org/0000-0002-9780-9230

Wojciech Borkowski D https://orcid.org/0000-0002-5793-6884

Correspondence concerning this article should be addressed to Dov Cohen, Department of Psychology, University of Illinois, 603 East Daniel, Champaign, IL 61820, United States. Email: dovcohen@illinois.edu

This article was published Online First March 24, 2022. Ivan Hernandez https://orcid.org/0000-0002-3141-7525 Dov Cohen https://orcid.org/0000-0002-0520-8527

Karl Gruschow D https://orcid.org/0000-0003-3905-5810

Andrzej Nowak b https://orcid.org/0000-0001-9965-0684

Dov Cohen acknowledges Grants SES080002N, SES100003 from National Center for Supercomputing Applications (NCSA) TeraGrid and Grant SES140010 from Extreme Science and Engineering Discovery Environment (XSEDE). Andrzej Nowak acknowledges support from National

Science Center (Poland), decision number: DEC-2011/02/A/HS6/00231.

Data, materials, code, and output may be found at https://osf.io/pf2ny.

Ivan Hernandez, Dov Cohen, and Karl Gruschow played an equal role in contributing at all stages. Andrzej Nowak played a supporting role in conceptualization, writing of original draft, and writing of review and editing. Michele J. Gelfand played a supporting role in conceptualization, writing of original draft, and writing of review and editing. Wojciech Borkowski played a supporting role in conceptualization, writing of original draft, and writing of review and editing.

members of a society that keeps systems functioning in a way often observed in the world.

In this article, we model a game among players who have very high or very low inherent predispositions to cooperate with or punish others. We conduct two simulation studies (in silico), examining the role that opportunists play in promoting the prevalence of their opposites-Strong Reciprocators. We then conduct an experiment with human participants (in vivo) to test two of the more striking implications of the model. Together, the studies show the importance of opportunists for producing some real-world cultural phenomena that otherwise seem puzzling. Specifically, as noted below, we describe how the presence of opportunists: (a) allows for the growth of a very unopportunistic strategy of strong reciprocity (Henrich et al., 2006, 2010); (b) creates the peculiar correlation between virtue and violence found in honor cultures (Cohen & Leung, 2012; Leung & Cohen, 2011); and (c) produces effects in which tight or collectivistic social norms crowd out volunteerism and civic virtue (Allik & Realo, 2004), as opportunistic actors displace more altruistic ones.

## **Cultural Puzzles**

# Puzzle 1: The Flourishing of Strong Reciprocity

Why cooperate when it is more advantageous to cheat? Why take vengeance on (punish) another at a cost to oneself? Strong reciprocity is defined by the willingness to make costly personal sacrifices to cooperate with others (positive reciprocity) and to punish wrongdoers (negative reciprocity; Fehr et al., 2002). The puzzle is how such "irrational" behavior survived evolution. Researchers have offered various answers that have run the gamut from kin selection theory and reciprocal altruism to social norms and group selectionist arguments that groups with Strong Reciprocators outcompete other groups (Fehr & Gächter, 2002; Frank, 1988; Gavrilets & Richerson, 2017; Ghate et al., 2013; Henrich et al., 2006, 2010; Jansson, 2013; Norenzayan, 2013; Ostrom, 2000; Santos et al., 2008; Trivers, 1971; Yamagishi et al., 2012). The many perspectives highlight the importance of understanding strong reciprocity, as well as the still open nature of explanations for the emergence of this multiply determined behavior.<sup>1</sup>

# **Puzzle 2: The Correlation Between Virtue and Violence**

In many contemporary cultures, tendencies to cooperate and tendencies toward vengeance are seen as opposed. Agreeableness, a hallmark trait of the Big 5 and Big 6, is defined on the high end by altruism and cooperation and on the low end by revenge taking and quarrelsomeness (Ashton & Lee, 2007; McCrae & Costa, 1987).<sup>2</sup> From this perspective, the pattern of behavior seen in honor cultures represents a puzzle because cooperativeness and vengefulness are *positively* correlated. In laboratory experiments, among participants from an honor culture (Hispanic/Latinx persons and southern Anglos), those most endorsing honor-related violence showed the greatest prosocial reciprocity, traipsing longer distances to pay back a confederate who had done them a favor (Leung & Cohen, 2011).

Honor cultures entwine virtue and virility in a way that nonhonor cultures do not (Cohen & Leung, 2012; Peristiany, 1966). In archival studies of U.S. political elites (Presidents and Congresspersons), among southerners, those scoring highest on martial honor (as indicated by their military service or lack thereof<sup>3</sup>) also scored highest on ratings of their character, integrity, and incorruptibility (Cohen & Leung, 2012).

In these archival studies (Cohen & Leung, 2012) and experiments (Leung & Cohen, 2011), correlations between tendencies toward violence and tendencies toward virtue ranged from about .2 to .6 for those from an honor culture. Among those not from an honor culture, there was either a negative correlation between violence and virtue—consistent with psychology's understanding of "Agreeableness"—or no correlation (Cohen & Leung, 2012; Leung & Cohen, 2011; see also Egloff et al., 2013; Yamagishi et al., 2012).<sup>4</sup>

# Puzzle 3: Low Social Capital in Cultures That Could Socialize for High Social Capital

Most social scientists regard social capital, defined by the network of relationships allowing individuals to work together to achieve common ends, as leading to positive outcomes: effective government, peaceful neighborhoods, and communities high in health and well-being (Coleman, 1990; Putnam, 2001; Putnam et al., 1994). Volunteerism, neighborliness, and the ability to organize for the common good bring large benefits. Why is it then that cultures with a strong capacity to socialize for high social capital do not do so? Tight cultures have welldefined standards of appropriate behavior that people are expected to comply with (Eriksson et al., 2021; Gelfand et al., 2011; Pelto, 1968). Collectivist cultures have cohesive families (and other in-groups) in which individuals need to prioritize group goals over the more selfish goals of the individual (Hong et al., 2001; Miller et al., 1990; Oyserman et al., 2002; Triandis, 1988; see also Oishi, 2010; Oishi & Kisling, 2009). Yet despite their clearly defined standards, requirements to sacrifice individual desires and whims for the good of the group, and the potential threat of exclusion and other strong punishments for deviance, researchers find that tight and collectivist cultures usually

<sup>3</sup> There are many reasons one might choose to engage in military service. However, scholars of the South have long seen the actual and mythical military traditions of the South—its high rates of military volunteerism, support for wars, dominance of the military elite, feats of heroism and persistence in the face of long odds, its storied military schools, and displays of soldierly pageantry—as interwoven with ideals of masculine honor in a land once characterized as America's "Sparta" (Franklin, 1956, p. 2; Napier, 1989).

<sup>4</sup> Results vary from study to study, of course. However, most studies do not focus on honor cultures, and there also seems to be a disciplinary/ methodological split. Conventional psychology studies are likely to use questionnaires to assess tendencies toward prosocial behavior and revenge. As noted, agreeableness scales sometimes put positive reciprocity and (reversed) negative reciprocity items together. Sometimes separate scales for prosocial behavior and negative reciprocity are used and these are often found to be negatively correlated (Brown, 2004; Hoyt et al., 2005; Lee & Ashton, 2012; McCullough et al., 2001; McCullough & Hoyt, 2002; Sheppard & Boon, 2012; but cf. Eisenberger et al., 2004, finding a .1 correlation between positive and negative reciprocity). In the world of behavioral economics, however, studies often find that tendencies toward positive reciprocity and negative reciprocity are orthogonal (e.g., Brethel-Haurwitz et al., 2016; Weber et al., 2018; Yamagishi et al., 2012). As with most broad constructs, the particulars of a study matter a great deal.

<sup>&</sup>lt;sup>1</sup> Though occasionally people will operationalize strong reciprocity in terms of being willing to sacrifice resources to cooperate (positive reciprocity) *or* to punish another (negative reciprocity), we follow the convention that defines strong reciprocity as people's willingness to sacrifice resources to engage in *both* positive *and* negative reciprocity (Bowles & Gintis, 2004; Fehr et al., 2002; Yamagishi et al., 2012).

<sup>&</sup>lt;sup>2</sup> Sample agreeableness items from John et al. (2008): Respondents characterize themselves as: "likes to cooperate with others," "is helpful and unselfish with others," "has a forgiving nature," and (reverse scored) "starts quarrels with others." Sample items from International Personality Item Pool (2020): "respect others," "have a good word for everyone," and (reverse scored) "get back at others" and "hold a grudge."

have low social capital (see Allik & Realo, 2004; Luria et al., 2015). Table 1 shows correlations between indicators of tightness and collectivism and levels of social capital; and whereas the correlations range in size, they are all negative.

# Proposed Solution to Puzzles: The Role of Opportunistic Agents

These three puzzles—how does strong reciprocity survive and even thrive, why are virtue and violence correlated in honor cultures in a way not seen elsewhere, and why do cultures that would seemingly be most able to produce high levels of social capital actually have very low levels of it—would seem quite disparate. However, we argue that these three puzzles have a common solution (though it is not the only solution for explaining these multiply determined phenomena). More specifically, understanding the role of Opportunistic actors is key to all three puzzles.

We use simulations from agent-based models to explore these issues. Such modeling is well suited to understanding how these processes unfold over time. It provides analytic leverage by stripping behavior down to simple preprogrammed rules and examining how manipulations of situational parameters can lead to different long-term equilibria. Conversely, simplification and reductionism can also be disadvantages because what works in theory does not always work in practice. We thus follow-up the simulations with experiments involving actual people to verify some of the simulations' more striking findings.

#### An Agent-Based Modeling Approach

Agent-based models (ABMs) are computer simulations in which a collection of heterogeneous agents interact with other agents in a virtual environment. Depending on how agents' preprogrammed strategies match up, agents either gain or lose resources. If agents gain enough resources, they reproduce. If they lose enough

#### Table 1

Correlations Between Tightness, Collectivism, and Indices of Social Capital

Cultural syndrome	Within 48 states of U.S.	Cross-country analysis	Average 1
Tightness	$r =43^{*}$	r =14	$R =32^{\circ}$
Collectivism	$r =76^{*}$	$r =67^*$	$R =72^{\circ}$

Note. For states of the U.S., indicators of collectivism, tightness, and social capital come from, respectively, Vandello and Cohen's (1999) individualism-collectivism index, Harrington and Gelfand's (2014) tightness index, and Putnam's (2001) social capital index that includes measures of volunteerism, social trust, and participation in community organizations. For countries, indicators of collectivism, tightness, and social capital come from, respectively, Hofstede (2003) for collectivism; Gelfand et al. (2011) for tightness–looseness; and a three-item ( $\alpha = .86$ ) social capital index of (a) a generalized trust indicator from the World Values Survey (often used as a proxy for social capital; Norris, 2017), (b) the Johns Hopkins Civil Society index (highly relevant for gaging voluntary/optional cooperation; Salamon et al., 2004), and (c) the social capital subindex from the Legatum Institute (2016; measuring social cohesion, network ties, and institutional trust). The -.76 correlation of collectivism with social capital is the same as that of Allik and Realo (2004). Average r is computed from Fischer's Z and weighted by the degrees of freedom.  $p \leq .01.$ 

resources, they die. This differential reproduction creates the natural selection that lets some types of agents proliferate as other types dwindle. Importantly, the final state of the society is not necessarily characterized by a single strategy, but rather a set of strategies. In practice, the equilibrium of a society is quantified as the ultimate stable proportions of strategies to which a population evolves (Bergstrom & Godfrey-Smith, 1998). Some equilibrium states will robustly emerge across a variety of assumptions and conditions. Other equilibrium states will emerge only under a restrictive set of conditions.

#### The Agents and Their Interactions in This Context

The game in this article models a simple interaction. The interaction begins as one party can invest in cooperation (or not) and the other party can respond by either investing in cooperation or cheating (taking the other's investment). Agents not liking the response of the other party can choose to punish (or not) that other person, who may then retaliate (or not) for this punishment.

For the purpose of modeling, we have to put numbers on the "goods" that are gained or lost. The goods at stake can be economic (you invest in cooperation and I can match your investment vs. cheat by taking your investment), status (you politely treat me as a person of honor and I can politely treat you as a person of honor vs. accept your deference without giving you reciprocal honor and courtesy), comity (you offer peaceful coexistence and I can offer peaceful coexistence vs. act aggressively toward you), alliance (you become my ally and I can become your ally vs. remain uncommitted or exploit your loyalty), and so on.

Regardless of the "goods" at stake, we can create four types of agents in the model, crossing a high versus low tendency toward positive reciprocity (investment in cooperation) and a high versus low tendency toward negative reciprocity (punishment).

Strong Reciprocators (high in predisposition to positive reciprocity, high in predisposition to negative reciprocity): Strong Reciprocators believe in behaving appropriately. They cooperate with others and will punish others who do not cooperate with them. Because Strong Reciprocators punish those who they believe have crossed them, they can end up in costly spirals of retaliation (feuds) if they get in conflicts with other agents who are also high in negative reciprocity (Exploitive Egoists or other Strong Reciprocators).

*Trusters* (high in predisposition to positive reciprocity, low in predisposition to negative reciprocity): These agreeable "nice guys" want to cooperate and will not retaliate if others behave badly toward them.

*Exploitive Egoists* (low in predisposition to positive reciprocity, high in predisposition to negative reciprocity): These disagreeable agents exploit others' positive behavior, not reciprocating with positive behavior of their own. They punish others who do not behave positively toward them, and they also can end up in spirals of retaliation if they get in a conflict with others who are high in negative reciprocity (Strong Reciprocators and other Exploitive Egoists).

*Opportunists* (low in predisposition to positive reciprocity, low in predisposition to negative reciprocity): These agents will cheat if they can, invest in cooperation minimally and only when they have to, and will not retaliate. Unlike Strong Reciprocators and Trusters, they have no strong predisposition to cooperate. Unlike Exploitive Egoists, they have no willingness to retaliate or defend themselves, fearing conflict with anyone who might punish them. Opportunists try to avoid punishment, appeasing Strong Reciprocators and

Exploitive Egoists by offering only the minimal amount of investment that will allow them to escape being punished (see also Geniole et al., 2017). Opportunists do not trust and essentially follow the Minimax strategy (minimizing the maximum damage that can be done to them). Depending on one's perspective, the Opportunists can be seen as "unearnest," sycophantic "weasels"; as gametheoretic Minimaxers; as cautious isolates not wanting to risk being cheated or being punished; or as apathetic slackers with no strong commitments, just trying to get over or get by. Either way, these agents are low in predispositions toward positive and negative reciprocity, offering only the minimum investment required to avoid punishment by others.

# **Payoffs Between Agents**

The skeletal payoff matrix in Table 2 recaps what happens when various agents meet. When two agents who are both high in positive reciprocity (Strong Reciprocators or Trusters) meet, they receive the payoff to cooperation (abbreviated as Coop gain). When Agent 1 is high in positive reciprocity but Agent 2 is low, Agent 1 will be cheated (-Sucker payment) and Agent 2 will receive the cheater's payoff (+Sucker payment)—unless Agent 1 is high in negative reciprocity and chooses to retaliate, administering punishment (-Punish Amount). If Agent 2 is also high in negative reciprocity, he will retaliate for Agent 1's retaliation and this will escalate into a feud (-Feud rounds × Punish Amount, where "Feud rounds" is the number of rounds of retaliation).

As noted, Opportunists have low predispositions to cooperate or retaliate, meaning they do not reciprocate with Trusters and do not gain much with other Opportunists. With Strong Reciprocators and Exploitive Egoists, the Opportunists offer only enough investment to appease the other, meaning they gain only a fraction of what they could gain from full cooperation with Strong Reciprocators and lose only a fraction of what they could lose from being cheated by Exploitive Egoists (+Appease offer × Coop gain with the Strong Reciprocators and –Appease offer × Sucker payment with the Exploitive Egoists, with "Appease offer" being the proportion of the investment offered to appease the partner).

To derive these payoffs, we have thus made a few simplifying assumptions about behavior. (a) There is no need to coerce cooperation from those who wholeheartedly offer it (Strong Reciprocators and Trusters). (b) Among people with a low predisposition to cooperate, we assume that appeasement is possible—that people who might not want to trust another can be induced to offer at least partial cooperation out of the fear of being punished (see also Krasnow et al., 2015). Their behavior can be "good enough" to appease Strong Reciprocators and Exploitive Egoists, even if it is not fully compliant. (c) People with a high predisposition toward negative reciprocity are unlikely to be successfully coerced. Thus, among those with a low inclination to cooperate, Opportunists may be coerced into partial cooperation, but "tough guy" Exploitive Egoists cannot be. When two people high in negative reciprocity meet, attempts at coercion will result in escalating aggression (Daly & Wilson, 2017).

As seen in Table 2, as long as (a) the benefit from cheating exceeds the benefit from cooperating (otherwise why cheat?) and (b) the punishment from an angry retaliator is great enough (otherwise why not cheat?), no strategy will dominate in the sense of being always at least as good and sometimes better than another strategy. To this skeletal outline (Table 2), we added the following to further flesh out the game: a minimal "base" payoff that is the same for every cell as an incentive to engage with other agents (this might be thought of as a basic "gains from trade" or "benefits from interaction" term) and two terms that model the possibility of misperception-Type I errors (the probability of believing someone has cheated you when they have not) and Type II errors (the probability of not noticing when someone has actually cheated you). The mathematical formulas for this fully fleshed-out payoff matrix are given in the Supplemental Materials (Table S1), along with the range of values that parameters may take.

# Study 1: Modeling an Honor Culture

In Study 1, we modeled an honor culture (Barnes et al., 2012; Brown, 2016; Cross et al., 2014; Pitt-Rivers, 1968; Rodriguez Mosquera, 2016; Üskül et al., 2012), adding one additional constraint to the model. Because honor cultures tend to spring up in lawless environments where people must take justice into their own hands rather than rely on the state, we imposed a common principle of retributive justice on the game. That is, we required that punishment exceeds the benefit one could gain by cheating (Punish Amount > Sucker payment). The principle of "an eye for an eye" is a common one in honor cultures, and here we assume that those high in negative reciprocity would retaliate at at least that level.

Schematic of Payoffs for an	Agent (Shown by the	Column) When It Encounters	Another Agent (Shown by the Row)

Agent	Strong Reciprocator	Truster	Exploitive Egoist	Opportunist
Meets a				
Strong Reciprocator	Coop gain	Coop gain	-Feud rounds × Punish Amount	Appease offer $\times$ Coop gain
Truster	Coop gain	Coop gain	Sucker payment	Sucker payment
Exploitive Egoist	$-$ Feud rounds $\times$ Punish Amount	-Sucker payment	-Feud rounds $\times$ Punish Amount	-Appease offer × Sucker payment
Opportunist	Appease offer × Coop gain	-Sucker payment	Appease offer $\times$ Sucker payment	0

*Note.* Payoffs are to the column. Coop gain is the reward gained from cooperation. Sucker payment is the amount received/lost from exploiting/being exploited. Appease offer is the fraction offered to appease another player. Punish Amount is the penalty from being punished. Feud rounds is the multiplier associated with a punishment escalating into a feud. The constraint is that Coop gain < Sucker payment < Punish Amount in the honor culture simulation of Study 1. With the appeasement fraction (Appease offer) being less than 1 and the Feud multiplier being greater than 1, the payoffs thus follow the rank order: Cheating (Sucker payment) > Cooperating (Coop gain) > Appeasement met with cooperation (Appease offer × Sucker payment) > Being cheated (–Sucker payment) > Feud (–Feud rounds × Punish Amount). In the social capital simulation (Study 2), the constraints that Punish Amount > Sucker payment and that Punish Amount > Coop gain are removed.

Lex Talionis-or the Law of Talion-is a principle of strict retribution that *limits* retaliation (an eye for an eye, not two eyes for an eye). However, negative retribution tends to escalate, rather than remain at a level of strict equivalence. Rage tends to blind one's judgment. Beyond this, perpetrators and victims tend to have asymmetric perceptions of harm, with victims seeing more harm in a perpetrator's action than the perpetrator does. Thus, even when victims aim at strict equivalence, perpetrators see them as overreacting and going on the offensive (on asymmetric perceptions and the tendency to escalate, see Baumeister et al., 1990; Halali et al., 2013; Kearns & Fincham, 2005; Keysar et al., 2008; Kowalski, 2000; Stillwell et al., 2008; Simantov-Nachlieli & Shnabel, 2014; Vandermeer et al., 2019; but cf. Miller, 2007 on the fine levels of calibration that can occur in honor cultures). Further, even if one was able to keep rage in check and aim for strict equivalence, there are potentially rational reasons to make sure one punishes offences severely: In a world where the only protection is self-protection, a person may want to retaliate harshly to deter others or to end a conflict quickly through the "shock and awe" of punishing with overwhelming force. "You start it, I'll finish it" reflects this latter principle.

The tendency of negative reciprocity to escalate is demonstrated when disputes over what seem like "trivial incidents"—a dinged car door, stepped-on shoes, a rude bump—escalate into violence in honor cultures (Cohen & Nisbett, 1994; Cohen et al., 1996, 1999; Daly & Wilson, 2017; see also Cohen et al., 1998, 2018, and Daly & Wilson, 2017, on the rational irrationality of "overreacting" in honor cultures). Fights and feuds can develop when one side feels the other has not just gotten "even" but has gone one up or made one look like a chump (Cohen et al., 1996, Study 3). In our modeling of honor culture, we thus make sure our Strong Reciprocators and Exploitive Egoists retaliate at least at the level of what the Law of Talion would imply—and perhaps more importantly, run very serious risks in order to do so. As Anderson (1994) notes in describing inner-city "street" culture,

Not to be afraid to die is by implication to have few compunctions about taking another's life. Not to be afraid to die is the quid pro quo of being able to take somebody else's life—for the right reasons, if the situation demands it.

Courage (the willingness to risk) and the pursuit of vengeance are twinned in the honor mentality. And more generally, as the anthropologist David Gilmore (1990, p. 223) noted, "the acceptance of expendability constitutes the basis of the manly pose everywhere it is encountered"; "honorable" men do not back down from the risk of conflict.<sup>5</sup>

Because they show positive reciprocity and run such high risks in pursuit of vengeance, the Strong Reciprocators in this simulation can be thought of as the Honorable people in an honor culture. In Study 1 and Study 3, we thus refer to Strong Reciprocators and Honorables interchangeably.

Exploitive Egoists may engage in risky behavior as well, but they do not show the positive reciprocity that would make them paragons of an honor culture. Trusters show positive reciprocity but do not retaliate when someone crosses them, so they do not embody the honor ideal. Opportunists are the opposite of Honorable, seeking to minimize their risks of being taken advantage of or getting punished and caught in a feud.

# Game of Chicken

Because retaliation is so risky, the payoffs when modeling an honor culture most resemble those of the game of Chicken. The Chicken game—named after the teenage contest in which cars drive toward each other on a collision course until one player "chickens" out—is structured such that the worst possible outcome occurs when neither player backs down (the game is also called "Hawks and Doves" with the Hawk strategy being aggressive, the Dove strategy being pacific, and the worst outcome occurring when both players are Hawks and go to war).

Colman and Wilson (1997) argued that, when considering the properties of all strategically distinct two-person, two-strategy games (Rapoport & Guyer, 1966), the Chicken game may be particularly well suited to modeling certain types of risk taking and aggression: it occurs frequently; is compulsory, in that "declining a challenge to play Chicken amounts to playing a version of it and losing" (p. 27); and rewards a reputation for toughness and irrational vengeance (see also Jankowski, 1990).

Chicken (or Hawks and Doves) is different from the Prisoner's Dilemma (PD) in that the worst outcome in the PD is the "sucker's payoff" (that occurs when I cooperate and the other person defects). In Chicken, the worst outcome results when both players act aggressively—a situation we think better models the risks in a contest over honor, whether this occurs with two cars going at each other on the blacktop, two nations playing brinksmanship in their foreign policy, two pioneers playing for keeps on the frontier, or even two headstrong persons walking toward each other in a narrow hallway where there is room for only one of them to comfortably pass (for the last example, see Experiment 3 of Cohen et al., 1996; Pasley, 2008; Schelling, 1980; Snyder, 1971 on brinksmanship in international relations).

It should be noted, however, that the situation we model is different from a traditional Chicken game in at least two ways. First, instead of simply having an aggressive or a peaceful strategy, the strategies here have two dimensions underlying them—the dimensions of positive reciprocity (a desire to cooperate) and negative reciprocity (a desire to retaliate) as they might unfold in a given interaction. Second, the presence of Opportunists—who will cheat but will also offer partial cooperation if they must—changes the game. Indeed, as we show later, if Opportunists are removed, the situation essentially does boil down to a game of Hawks (Exploitive Egoists) and Doves (Trusters), with Strong Reciprocators making up only a small fraction of the population.

# **Key Hypotheses**

Relevant to the first two puzzles described in the introduction, we examine two hypotheses: (1) The presence of Opportunistic

<sup>&</sup>lt;sup>5</sup> Punish Amount > Sucker payment is an oversimplification. For present purposes, it does not matter whether this inequality holds every time or reflects an expected value. We also add that while conflicts over "trivial" causes occur in honor cultures, some honor cultures are more hair-trigger than others. That is, some encourage reactions to even small perceived slights and accidents; others are content to let small matters slide until they build to an explosion (Anderson, 1994; Cohen et al., 1999; Daly & Wilson, 2017; Peristiany, 1966). We do not know how to estimate the number of "trivial incidents" that result in full-scale blow-ups versus those that are skillfully avoided. Suffice it to say that the relatively heavy punishments here probably resemble a fairly stern honor culture.

strategies will help Strong Reciprocators flourish and (2) Over time, the tendencies toward positive reciprocity and negative reciprocity will go from uncorrelated to positively correlated, as Opportunists and Strong Reciprocators grow in number during the honor culture simulation.

# **Exploratory Hypotheses**

Though not directly related to the role of Opportunists, this simulation also allows us to explore other hypotheses relevant to the flourishing of Strong Reciprocators (Honorable people). These hypotheses derive from research on cultures of honor and other ethnographic data: (Exploratory Research Question 1) Greater vulnerability to predation will lead to more Honorable/Strong Reciprocity strategies. This is because deterrence is more important where predation is particularly tempting-that is, where taking advantage of another person is particularly rewarding and being taken advantage of is particularly costly. In terms of ecological conditions, this is illustrated in the contrast between farmers and herders (Cao et al., 2021; Edgerton, 1971; O'Kelly & Carney, 1986). Farmers face little danger of criminals stealing their entire crop. Herders, on the other hand, can have their wealth instantly stolen from them. Grazing animals represent portable, easily stolen wealth. More generally, where one's valuable "goods" are portablewhether those goods are animals (Peristiany, 1966; Schilling et al., 2012; Sweet, 1965), people (Chagnon, 2012; Lacey, 2013; Witsenburg & Adano, 2009), or contraband (including, e.g., drugs, diamonds, and other smuggled goods; Anderson, 1994, 2000; Prasad, 2012)-the risks and costs of being exploited are high, and thus, the need to cultivate a tough reputation is high as well.

Finally, we expect there to be sensitive dependence on initial conditions, such that there will be "founder effects": (Exploratory Research Question 2) The initial population in a place will have an important influence on how the culture evolves. In particular, having an initial critical mass of Honorables/Strong Reciprocators is likely important for the strategy to become popular. Historians have argued this was true of the U.S. South, where people from the war-torn border between Scotland and England came to settle the Appalachians and other "back country" areas (see, e.g., Fischer, 1988; McWhiney, 1988; Nisbett & Cohen, 1996; see also Grosjean, 2014, predicting violence in the South from the presence of settlers from the Scottish–English border in U.S. regions where institutions were weak).

#### Study 1 Method

Using NetLogo software (Wilensky, 1999), we began each simulation by creating 4,800 agents in two-dimensional space. We assumed tendencies toward positive and negative reciprocity were uncorrelated and so Honorables, Trusters, Exploitive Egoists, and Opportunists were always exactly 25% of the starting population. Each agent was assigned a random number of starting "resources" that varied uniformly between 0 and 15. These resources are a proxy for the goods (either material or social) that affect an agent's ability to live and reproduce. The goods increase or decrease via interactions with others. At 0 goods, the agent dies; at 15, the agent reproduces with his/her "child" starting at 5 goods and the parent losing 10 goods. Note we use the phrase "parent" and "child" to indicate that one originating agent (the "parent") produces as the

parent. How much these behavioral tendencies are actively socialized (vs. passively passed on through genetics) and who the originating agent is—whether it is a parent, aunt, village elder, or friend—is irrelevant for present purposes.

### **Match-Up Rules**

On every turn, an agent moved one step forward within a random  $45^{\circ}$  radius in an attempt to find a partner. If two or more agents occupied the same location in two-dimensional space, they selected one of the other agents to interact with. Once that agent was selected, they were unavailable to be chosen by another agent for that turn.

If an agent's resources increased in an interaction, the agent stayed in the same location. If the agent's resources decreased, the agent moved three steps away in a random direction. Agents in the same location for 10 turns moved forward three steps in a random direction for their next turn. However, the particulars of these movement rules do not matter much. We also ran simulations allowing agents to stay in the same place twice as long (20 turns), half as long (5 turns), or requiring them to move randomly after every turn. Results for those simulations are extremely similar to those below.

# **Population Carrying Capacity**

To account for natural limits on an ecology's ability to sustain a population, we placed a limit on how many agents could exist at once. We set this limit to 5,000 agents arbitrarily; however, we ran other simulations with twice as much carrying capacity (10,000 agents) and half as much carrying capacity (2,500 agents), and the pattern of results did not change. Because of the reduced computational complexity of simulating 5,000 concurrent agents, we left the carrying capacity at the original amount. After each round, if the population exceeded 5,000, the simulation killed a random selection of agents until the number of agents was reduced to 5,000.

# **Running Time**

We gave agents a fixed amount of turns before the simulation stopped. One turn occurs when all agents have completed the process of movement, interaction, and reproduction/death if applicable. We ran the simulation to 30,000 turns, which as will be seen, was long enough to ensure stability.<sup>6</sup>

### **Model Assumptions**

The model thus incorporates a number of assumptions—about movement rules, initial composition of the population, population carrying capacity and initial density, and so on. As may be seen in Supplemental Materials (Table S2) and at

<sup>&</sup>lt;sup>6</sup> Results presented below were verified in a "replication." We asked an independent coder to write an agent-based model using NETLOGO with the parameters described in the Method section and Supplemental Materials. This programmer followed the same payoff structures and routines of the original program, but without access to the original code. Results of the independently coded program show a great deal of consistency with the original (percent Honorables, Trusters, Exploitive Egoists, Opportunists in model = 41, 19, 7, 33; Corresponding percents in replication model = 40, 22, 3, and 35). This replication minimizes the chance that the current results are due to a programming error and increases the likelihood they are robust in different implementations.

Hernandez et al. (2021; https://osf.io/pf2ny), the model is quite robust to violations of these assumptions, though a few important exceptions are noted below. As noted, parameter values were generated randomly, and in this first simulation, were accepted when Coop gain < Sucker payment < Punishment.

#### Study 1 Results

We ran the simulations 1,000 times. (An *n* of 788 will have 80% power to detect a "small" effect of d = .20 and an *n* of 1,396 will have 80% power if d = .15. One thousand iterations is approximately the midpoint.) Results after each interaction opportunity were averaged and presented in Figure 1, thus lines in the figure represent the means of 1,000 runs. The percentages, standard deviations, and 99% confidence intervals for this baseline condition as well as for some variations from the baseline are shown in Table 3. Standard deviations should be considered cautiously, however, as run outcomes are not normally distributed, with strategies (especially among Exploitive Egoists and Trusters) getting eliminated in many runs. The eight parameters (in Table 2 and described above) were allowed to vary widely over the runs, again, with the constraint that gains to cooperation (Coop pay) are smaller than gains from cheating (Sucker pay), which in turn are smaller than the punishment (Punish Amount) exacted by Honorables or Exploitive Egoists if they think they have been cheated.

# Research Question 1: Prevalence of Honorables and Their Symbiosis With Opportunists

Figure 1

**Baseline Simulation** 

As may be seen (Figure 1; Table 3, line 1), when collapsing across all runs, agents of all types continue to persist at some level.

However, consistent with honor cultures developing in environments of self-help justice (Nisbett & Cohen, 1996; Peristiany, 1966; see also Nowak et al., 2016), the most prevalent behavior type is that of the Honorables (40.5%), who show high positive reciprocity and high negative reciprocity.

The second most prevalent type (32.7%) are the Opportunists (low positive reciprocity, low negative reciprocity)-the exact opposite of Honorables. Indeed, Honorables and Opportunists create a space for each other-they are in some ways symbiotic. As may be seen in the skeletal payoff matrix of Table 2, if Opportunists did not exist, the Truster strategy would dominate the Honorable strategy. The "niche" for high positive reciprocity agents would get filled almost exclusively by Trusters, crowding out the Honorables. Indeed, if we run the simulations without the Opportunists, there would in fact be very few Honorables. The Truster strategy would become much more popular (78.5%), the Exploitive Egoists would become the second most popular strategy (16.8%), and the Honorables would fall to very low levels (3.5%; Figure 2; Table 3, line 2); effect of the presence/ absence of Opportunists on percent Honorables, Trusters, and Exploitive Egoists was  $r(1,998) = .56 (p \le .001); r(1,998) = -.66 (p \le .001);$ and  $r(1,998) = -.18 (p \le .001)$ , respectively. The world would mostly boil down to a version of a simple "Hawks" (Exploitive Egoists) and "Doves" (Trusters) game, with few agents showing strong reciprocity (Honorables). Thus, it is the presence of the Opportunists that enables the popularity of Honorable/Strong reciprocity agents.

# The Presence of Honorables and the Long-Term Viability of the Society

Interestingly, to the extent that viable societies require some level of positive reciprocity to survive, the Opportunists also depend on

*Note.* Lines represent means collapsed across 1,000 runs of the simulation. Each run has 30,000 turns. See the online article for the color version of this figure.



Table 3

the Final Percentage of Different Strategies in the Population Under Various Simulation Conditions (SD in Parentheses)									
	Hono	orables	Tru	Trusters		Exploitive Egoists		Opportunists	
Simulation Condition	Percent	99% CI	Percent	99% CI	Percent	99% CI	Percent	99% CI	
Baseline	40.5 (37.3)	[37.4, 43.5]	19.1 (36.0)	[16.2, 22.1]	7.3 (22.5)	[5.5, 9.1]	32.7 (34.9)	[29.9, 35.6]	
No Opportunists	3.5 (9.1)	[2.8, 4.2]	78.6 (30.7)	[76.0, 81.1]	16.8 (30.1)	[14.3, 19.2]	_	_	
No Honorables		_	0.0(0.0)	[0.0, 0.0]	27.6 (32.7)	[25.0, 30.3]	70.6 (33.9)	[67.8, 73.3]	

**r**.

Note. Total percents do not always add up to 100 because of rounding and because in some cases, the entire population died. In the latter case, we assigned a percentage of 0 to each agent type.

the existence of the Honorables. If the Honorables did not exist, the Trusters would be dominated by the Opportunists and the world would then (temporarily) exist with only the Opportunists and the Exploitive Egoists. Indeed, if we run a simulation without the Honorables, Opportunists come to comprise 70.6% of the population, Exploitive Egoists comprise 27.6%, and Trusters are basically eliminated (Figure 3; Table 3, line 3); effect of the presence of Honorables on percent Opportunists, Exploitive Egoists, and Trusters was r(1,998) = -.48,  $p \le .001$ ; r(1,998) = -.34,  $p \le .001$ ;  $r(1,998) = .35, p \le .001$ , respectively. In such a world, no one shows any positive reciprocity, and the ultimate prospects for this society-especially if it was in competition with other societies or were to receive some exogenous shock to the system-seem quite limited (Diamond, 2004; Norenzayan, 2013). Thus, Honorables are likely necessary for the long-term functioning and growth of a society.

The point above about the necessity of Honorables for sustaining cooperation is complemented by another variation of the simulation in which the sufficiency of Honorables for sustaining cooperation was shown. In that variation, we removed the other type of cooperating agents; specifically, we ran simulations in which Trusters did not exist. Whereas Trusters are basically eliminated in a world without Honorables (as indicated in the paragraph above), the reverse is not true. Honorables do just fine without Trusters-and in fact prosper (as seen in Table S2). In a simulation where the initial starting composition is varied to have no Trusters, Honorables become 65.7% of the population, completely filling the positive reciprocity niche left by the removal of Trusters. Exploitive Egoists slip a bit to 2.8% of the population, and Opportunists remain steady at about 31.5% of the population. Positive reciprocity survives without Trusters; it is basically eliminated in a world without Honorables.7

Honorables punish. Trusters do not. That Honorables matter for sustaining cooperation is consistent with empirical work showing the necessity of punishment for maintaining cooperation within a group (Fehr & Gächter, 2000, 2002).

# **Research Question 2: The Peculiar Correlation Between Positive and Negative Reciprocity Emerges**

At the start of baseline simulations, all four strategy types were equally represented and hence there was no correlation between predispositions toward positive reciprocity and predispositions toward negative reciprocity ( $\varphi = 0$ ). However, as indicated in Table 3, by the end of the simulation, we note that Honorables (40.5%); high on both positive and negative reciprocity) and Opportunists (32.7%; low on both positive and negative reciprocity) combine to form about three fourths of the society, with Trusters (19.1%) and Exploitive Egoists (7.3%) making up the other one fourth. By the end of the simulation, predispositions toward positive reciprocity and toward negative reciprocity have become correlated ( $\varphi = .49, p \le .001$ ). The correlation between positive and negative reciprocity that is often observed in honor cultures-admittedly, with a large range of effect sizes, .2-.6-has been recreated in this simulation (Cohen & Leung, 2012; Leung & Cohen, 2011; cf. Krasnow et al., 2015).

This contrasts with the large negative correlation between positive and negative reciprocity observed when Opportunists are artificially removed (and hence almost all Honorables die off). The correlation would then be an unreasonably high  $\varphi = -.89$  $(p \leq .001)$ , as the population would consist almost entirely of agreeable Trusters and disagreeable Exploitive Egoists.<sup>8</sup>

# Exploratory Research Question 1: High Costs of Exploitation and the Prevalence of Honorables

The baseline results-showing that Honorables are the most prevalent strategy, followed by Opportunists-collapse over a great range of parameters. However, it is worth examining how strategies fare under some theoretically relevant variations. For example, as noted, culture-of-honor research and ethnographic data suggest that an Honorable strategy that deters predators is likely especially important where predation is particularly tempting and being taken advantage of is particularly costly.

To examine the effect of the cost of exploitation in our simulations, we predicted the percentage of each of the four strategy types from the Sucker's payment (the variable representing the gain for exploiting someone and conversely, the cost of being exploited), controlling for all other independent variables.<sup>9</sup> Obviously, the larger the Sucker's payment was, the more Exploitive Egoists

<sup>&</sup>lt;sup>7</sup> The final type one can remove are the Exploitive Egoists. Simulations without them look reasonably similar to the baseline condition (percent Honorable = 42.7, Trusters = 20.1, Opportunists = 37.2).

<sup>&</sup>lt;sup>5</sup> Obviously, when Opportunists are *completely* absent, the formula for phi guarantees that phi cannot be positive. It can be undefined (if Trusters or Exploitive Egoists get wiped out) or negative; but even if negative, phi can be vanishingly small (unlike the sizeable correlation in the text). Also, numbers will differ slightly depending on whether phi is based on percentages or counts. In the latter case, simulations are implicitly weighted by the size of the surviving population in a simulation. Results will differ to the extent that small populations are present in many runs. In most cases, differences are small. When phi is calculated based on population counts,  $\phi$  would be .6 rather than .49 in runs with Opportunists present; it would be -.76 rather than –.89 in runs without.

<sup>&</sup>lt;sup>9</sup> The eight independent variables in the regression were as follows: sucker's payment, cooperation gain, punishment amount, appeasement amount, rounds of feuding, minimal "base" payoff, Type I errors ( $\alpha$ ), and Type II errors ( $\beta$ ).

Simulation With No Opportunists in the Population 0.8 0.7 Strategy 0.6 H- Honorables Proportion of the Population 0.5 Trusters Exploitive Egoists 0.4 0.3 0.2 0.1 0.0 0 5000 10000 15000 20000 25000 30000 Turn

*Note.* Lines represent means collapsed across 1,000 runs of the simulation. Each run has 30,000 turns. See the online article for the color version of this figure.

benefitted, t(991) = 12.16,  $p \le .001$ . But, consistent with the idea that honor norms are strong where the cost of being exploited is steep, Honorables also benefitted in environments where exploitation was *more* (rather than less) costly, b = .036, t(991) = 2.23,

Figure 3

Figure 2

p = .026. Overall, such risky environments benefitted the "tough guys"—Exploitive Egoists and Honorables—who were willing to punish those who crossed them, even though doing so might lead them into a costly spiral of retaliation (a feud).



*Note.* Lines represent means collapsed across 1,000 runs of the simulation. Each run has 30,000 turns. See the online article for the color version of this figure.

Another way to analyze the data is to examine the success of strategies when the difference between the Sucker's payment and Cooperation gains is low (the gains to the cheater and hence cost to the cheated are small, relative to the gains realized from simply cooperating) versus when the difference is high (the gains to the cheater and hence cost to the cheated are large, relative to the gains from cooperating). Figure 4 shows the final percentage of each strategy as a function of Sucker's payment minus Coop gain (broken into deciles). As may be seen, the Honorables, Trusters, and Opportunists were all relatively close together through the first four deciles (when the gains to the cheater and cost to the cheated were small). For deciles 5 through 9-that is, for worlds in which there was moderate to high risk-Honorables became the most popular group. Only at the very end, where gains to cooperation (Coop gain) were quite small, risks to exploitation (Sucker's payment) were great, and hence the costs of escalating feuds were also extremely high (since Feud × Punish Amount > Sucker's payment) were Honorables not the most popular group.

Note that this effect—Honorables' success in risky environments does *not* operate through Honorables directly benefitting from moderate-to-large values of the Sucker's payment. Honorables never cheat anyone so they never gain the Sucker's payment. Rather, the mechanism is one of selection as exploitation diminishes the population of Trusters, and the "niche" for high positive reciprocity agents increasingly gets filled by Honorables.

# Exploratory Research Question 2: Sensitive Dependence on Initial Conditions

As may be seen in the Supplemental Materials (Table S2), the model was robust under a variety of assumptions. In all cases, the

group that was the most common was the Honorables, who were at least 6 percentage points above the next most frequent group, the Opportunists. The salient exception was the case where the starting percentage of agents was completely random. In this case, Honorables were still the most populous, but their lead over the Opportunists went down to 2% (39% Honorables vs. 37% Opportunists). A closer look revealed that this difference was driven by cases where the initial percent of Honorables was small. Figure 5 breaks up the initial starting proportion of Honorables into deciles. As may be seen, Honorables ended the simulations behind Opportunists through the first two deciles (when the initial percentage of Honorables was between 1% and 12% of the population) and only surpassed them when they got to the fourth decile (the initial percentage of Honorables was between 17% and 21%). Thus, Honorables can become the modal "type" in the society, even if they are initially underrepresented-but they cannot do so if they lack a critical mass and are too small a percent of the initial population. These data are thus consistent with "founder effects" for honor cultures (e.g., Fischer, 1988; Grosjean, 2014; Nisbett & Cohen, 1996).

However, perhaps the biggest demonstration of sensitive dependence on initial conditions was that there were huge differences between having a strategy not exist at all versus having it exist even at a very low level of the initial population. This can be seen dramatically in the co-evolution of Honorables and Opportunists. As shown in Table 3, in the baseline condition, if there are no Opportunists at all, the number of Honorables becomes extremely small (3.5%). However, if Opportunists exist in the initial population, even at extremely low levels, Honorables can establish a foothold. As another variation on the baseline condition, we ran 1,000 simulations in which Opportunists began as 1% of the population. The results were extremely close to the baseline

#### Figure 4





Note. See the online article for the color version of this figure.

Final Number of Honorables, Trusters, Exploitive Egoists, and Opportunists as a Function of the Initial Number of Honorables in the Population



Note. The initial number of Honorables is broken into deciles. See the online article for the color version of this figure.

condition: Honorables, Trusters, Exploitive Egoists, and Opportunists ended up as 40%, 22%, 5%, and 33% of the population, respectively. In sum, Opportunists need to exist (even if they do not need to be plentiful initially) for Honorables to thrive.

### Summary

In this game modeling an honor culture, Opportunists are symbiotic with their opposites, Honorables. In a moderate- to high-risk environment and with a relatively small number of Honorable compatriots, the presence of at least some Opportunists leads to Honorables becoming a plurality of the population. The Honorables are the Strong Reciprocators and paragons of an honor culture. But, without Opportunists, the game mostly boils down to a variation of a one-dimensional Hawk-Dove game between the aggressive versus pacific, rather than a two-dimensional honor culture where Honorables are both cooperative and vengeful while "peaceful" nonretaliators can be either high in cooperation (Trusters), or more frequently, low in cooperation (Opportunists; Cohen & Leung, 2012; Leung & Cohen, 2011). Further, the large number of Opportunists and Honorables that come to emerge in these simulations produces the peculiar correlation between predispositions toward positive reciprocity and predispositions toward negative reciprocity found in honor cultures.

# Study 2: Modeling Social Capital in Cultures With High Standards for Cooperation and Severe Punishment for Noncooperation

Civil society is that layer of society that exists between citizens' private life and the government (Putnam, 2001; Putnam et al., 1994;

Tocqueville, 2002). Civil society consists of *voluntary* organizations, activities, and institutions that exist in a society to promote some common good (Salamon et al., 2004). "Civil society" is highly related to and sometimes used interchangeably with the concept of social capital (loosely defined as social relations based on trust that allow people to work with and through others to effectively achieve their goals).

Research Question 3: We use our model to investigate the puzzle of why cultures with high standards for cooperation and strong punishments for noncooperation—such as tight and collectivistic societies—seemingly could have high levels of social capital or a thriving civil society but usually do not (Table 1). Thus, our independent variables are as follows: (a) Appease offer (the standard for the amount of cooperation that is considered "good enough" to avoid punishment) and (b) Punish Amount (the severity with which noncooperation is punished by Strong Reciprocators and Exploitive Egoists). We expect both variables to negatively predict the amount of *voluntary* cooperation that exists in our society (the dependent variable) and examine the role Opportunists play in producing these effects.

# **Removing Constraints on Punishment**

There is no reason to believe that the relationships described above are limited to honor cultures. Thus, we remove the constraint that offenses must be paid back at least "an eye for an eye, a tooth for a tooth" (Punish Amount > Sucker's payment). We let Punish Amount vary freely, running higher or (mostly) lower than the Sucker's payment. This means that agents who punish can be considerably more forgiving—and take far less risk—than their counterparts in Study 1. On the one hand, this expands the generalizability of the model. On the other hand, it blurs the interpretation of the Strong Reciprocators, who no longer run such large risks nor pursue "eye for an eye, tooth for a tooth" vengeance. In Study 2, we do not refer to Strong Reciprocators as "Honorables."

Note that allowing Punish Amount to vary freely changes the structure of the game. The payoffs no longer necessarily resemble those of a Chicken game, in which the worst possible outcome for an individual is the feud. Removing constraints on Punish Amount opens up the possibility of payoffs instead resembling those of a Prisoner's Dilemma, in which the worst possible outcome for the individual is the Sucker's payment. Because feuds allow for multiple rounds of retaliation, in about half of the cases where Punish Amount < Sucker's payment, the feud is still the worst possible outcome (similar to a game of Chicken); in the other half, the "Sucker's payoff" is the worst outcome (similar to a PD).

Again, however, the game is not a traditional one-shot PD nor a traditional Chicken game for two reasons: (a) the strategies have two dimensions (positive and negative reciprocity) baked into them when they interact and (b) as will be seen below, the presence of Opportunists changes the game. Our concern, though, is not so much whether the payoffs resemble those of Chicken or PD. Rather, we simply let levels of Punishment and Appeasement amount vary greatly and observe whether higher levels of these variables drive out voluntary cooperation.

### Study 2 Method

We use the same basic model from Study 1, but remove the constraints that Punishment > Sucker's payment and Punishment > Coop gain. We retain only the constraint that Sucker's payment > Coop gain as this constraint is necessary for creating any sort of dilemma. (If cooperation pays better than cheating, then there is no reason for an agent to consider another option and everyone would simply cooperate.)

Our conceptual dependent variable is social capital, operationalized here as the total amount of *voluntary* cooperation manifested in the society. Total voluntary cooperation is indicated by the total number of interactions between two people with a high predisposition to cooperate. It is based on the population in the last round of the simulation and is equivalent to the number of interactions between Strong Reciprocators with Strong Reciprocators, Trusters with Strong Reciprocators, and Trusters with Trusters. Coerced cooperation—that occurs between Strong Reciprocators and Opportunists only because of the threat of punishment—does not count as voluntary.

We ran 3,000 simulations. After the main analysis, we would then have about 1,000 cases where (a) revenge was at least at the level of "an eye for an eye" (Punish Amount > Sucker's payoff) as in Study 1; (b) revenge was at a very low level such that the payoffs resembled a PD (Feud rounds  $\times$  Punish Amount < Sucker's payoff); and (c) revenge was at a mild level such that (Punish Amount < Sucker's payoff) but the Feud was still the worst outcome (Feud rounds  $\times$  Punish Amount > Sucker's payoff).

### Study 2 Results

We ran a regression predicting the total amount of voluntary cooperation in the society from our eight variables (Punish Amount, Sucker's payment, Coop gain, Appease offer, rounds of feuding, base payoff, Type I errors, and Type II errors). The main predictions were that more severe punishments (Punish Amount) and higher standards for what is "good enough" behavior to avoid punishment (Appease offer) would predict lower levels of voluntary cooperation. Table 4 presents the relevant beta weights as well as breaks down effects on each strategy.

# Severity of Punishment for Selfish Behaviors

As seen in the first line of Table 4, greater punishment for selfish behavior is associated with less voluntary cooperation in the society, b = -258.32,  $\beta = -.25$ , t(2,991) = -20.29, p = .001. This is because higher levels of Punishment escalate the costliness of feuds, driving down the number of Strong Reciprocators, t(2,991) = -12.29,  $p \le .001$ ; and to a lesser extent, Exploitive Egoists, t(2,991) = -3.53,  $p \le .001$ , and making room for a greater number of Opportunists, t(2,991) = 23.06,  $p \le .001$ .

Punishing noncooperation more severely thus has the ironic effect of reducing the number of voluntary cooperators. Even if we add in cooperation that is coerced from Opportunists, the effect of Punishment on total cooperation (voluntary cooperation + coerced

#### Table 4

Beta Weights for Severity of Punishment and Level of Standards for Appropriate Behavior in Predicting Cooperation and the Proportions of Strategy Types

	Dependent variables						
Beta weight from regressions	Amount of voluntary cooperation	Percent Strong Reciprocators	Percent Trusters	Percent Exploitive Egoists	Percent Opportunists	Amount of coerced cooperation	Total cooperation (voluntary and coerced cooperation)
Severity of punishment (Punish) Level of standards for what is acceptable or "good enough" (Appease offer)	25** 15**	21** .10**	.03 23**	06** .17**	.32** 04*	.31** .26**	14** 05**

*Note.* Dependent variables for each of the different regressions are listed across the top. In addition to the punishment (Punish) and standards for what is good enough (Appease offer) variables, independent variables for each regression include benefits/costs of cheating (Sucker payment), cooperation benefits (Coop gain), levels of feuding (Feud rounds), minimal "base" payoff (Base), Type I errors ( $\alpha$ ), and Type II errors ( $\beta$ ). \*  $p \le .05$ . \*\*  $p \le 001$ . v

Composition	of Pa	opulation	Under	Various	Conditions
	./	1			

	Simulations with Opportunists present				Simulations without Opportunists present		
Parameter constraints within simulation	% Strong Reciprocators	% Trusters	% Exploitive Egoists	% Opportunists	% Strong Reciprocators	% Trusters	% Exploitive Egoists
Vhen Punish > Sucker payment (first simulation)	40.5	19.1	7.3	32.7	3.5	78.6	16.8
Vhen Punish < Sucker payment and the worst outcome is the feud (from second simulation)	49.0	23.6	10.6	15.3	3.8	56.4	34.4
Vhen Punish < Sucker payment and the worst outcome is the "sucker's payoff" (from second simulation)	64.5	20.6	11.4	2.8	38.6	51.0	9.4

cooperation) is also negative, b = -137.60,  $\beta = -.14$ , t(2,991) = -10.55, p = .001.

Many argue that punishment is necessary to maintain cooperation in a group (Fehr & Gächter, 2000). Nothing here contradicts this. Indeed, recall the section on the "Presence of Honorables and the Long-Term Viability of the Society" in Study 1, showing Honorables as necessary and sufficient for cooperation. The point we are making here is that *severity* of punishment reduces *voluntary cooperation offered without the specter of punishment*. In this case, the effect is so large that total (coerced and voluntary) cooperation is reduced.

# High Standards for Cooperative Behavior

As seen in line 2 of Table 4, as standards for what is a "good enough" level of cooperation become more demanding (as Appease offer increases), levels of voluntary cooperation drop, b = -154.46,  $\beta = -.15$ , t(2,991) = -12.10, p = .001. Even total cooperation (voluntary + coerced cooperation) drops, b = -51.62,  $\beta = -.05$ , t(2,991) = -3.95, p = .001. If Opportunists must do more to appease Strong Reciprocators and Exploitive Egoists, both of those groups gain; however, the largest effect is that Trusters get pushed out, b = -.08,  $\beta = -.23$ , t(2,991) = -15.60, p = .001. The number of Opportunists decreases slightly, b = -.01,  $\beta = -.04$ , t(2,991) = -2.57, p = .010.

The main importance of Opportunists here is that raising or lowering their level of appeasement does not much affect their own proliferation, rather it affects the proliferation of all the *other* groups. In addition, whereas raising Punishment reduces voluntary cooperation primarily by reducing Strong Reciprocators and increasing Opportunists, raising the Appeasement offer reduces voluntary cooperation primarily by affecting the other two groups: reducing Trusters and increasing Exploitive Egoists.

As we note later, there are good psychological reasons for why demanding higher levels of cooperation to avoid punishment and increasing the severity of punishment might crowd out voluntary cooperative behavior. However, for the findings here, we do not have to appeal to any internal psychological explanation. The structure of rewards in the social ecosystem is sufficient for producing the effect.

# Comparing Simulations Where Punishment Can Vary Freely Versus Must Be Greater Than the Sucker's Payment

To explore the effect of constraints on Punishment and examine robustness, we can also compare results in this simulation (where there are no constraints on Punishment) to those from Study 1 (where Punishment must be greater than the Sucker's payment). We note a few salient similarities and differences:

- 1. Table 5 shows the results of Study 1 along with a subset of cases from the present simulation: (a) cases where Punishment is less than the Sucker's payment but the feud is still the worst possible outcome (resembling payoffs in a game of "Chicken"; n = 923) and (b) cases where Punishment is very low and the "Sucker's payoff" is the worst possible outcome (resembling payoffs in a Prisoner's Dilemma; n = 1,095). In the latter case, Opportunists shrink dramatically, whereas Strong Reciprocators become not simply a plurality but a majority. Strong Reciprocators can thus achieve majority statusbut only by abandoning the principle of paying back at least "an eye for an eye, a tooth for a tooth" and living in a world where a feud is not the worst outcome. That is, they can achieve this status-but only by dramatically softening what made them "Honorable" in Study 1.<sup>10</sup>
- 2. Even though levels of Punishment move the fortunes of Strong Reciprocators and Opportunists in different directions and even though Opportunists are only a small portion of the population when the "Sucker's payoff" replaces the feud as the worst outcome, the relationship between Opportunists and Strong Reciprocators is still in some ways symbiotic. The right side of Table 5 shows what happens under various conditions when we artificially remove Opportunists from the simulation: Strong Reciprocators become far less prevalent. Under such conditions, Trusters and Exploitive Egoists comprise the majority of the population. Even when the "Sucker's payoff" is the worst outcome, Trusters are the majority of the population as Strong Reciprocators shrink from

<sup>&</sup>lt;sup>10</sup> There were 1,000 cases in the baseline condition of Study 1 with Opportunists present, and we ran an additional 1,000 cases with Opportunists removed. In Study 2, we also had cases where Punishment amount > Sucker's payment (as in Study 1). There were 968 of such cases with Opportunists present and 330 cases with Opportunists removed. For greater stability and ease of comparison, we use the Study 1 figures in Table 5. The Study 2 figures were extremely similar: When Opportunists were present, the percent of Strong Reciprocators, Trusters, Exploitive Egoists, and Opportunists were 43.3%, 18.6%, 5.8%, and 32.1%, respectively. With Opportunists removed, the figures for Strong Reciprocators, Trusters, and Exploitive Egoists were 3.7%, 80.5%, and 15.2%, respectively.

64.5% of the population (when Opportunists are present) to 38.6% (when they are not). When Opportunists are initially present—even when they ultimately become a small fraction of the population—they shift the balance from a population in which Trusters would be most numerous to one in which Strong Reciprocators are most numerous.

3. When Opportunists are not present (right side of Table 5), the composition of the population makes it so that the correlation between positive and negative reciprocity is negative ( $\varphi = -.73$  when P < R but feud is the worst outcome;  $\varphi = -.16$  when the Sucker's payoff is). Even when Opportunists are present, in situations where the "Sucker's payoff" is worse than the feud, the correlation between positive and negative reciprocity is near zero ( $\varphi = -.04$ ). Thus, the positive correlation between positive and negative reciprocity (often found in honor cultures) does not emerge (a) when feud is *not* the worst possible outcome or (b) when Opportunists are absent.<sup>11</sup>

# Summary

One oft-found "puzzle" is that social capital tends to be low in both collectivistic and tight cultures-cultures that typically demand high levels of cooperation with in-group members and thus might erroneously be expected to have high levels of social capital. Consistent with this "puzzle," the present simulations showed that levels of voluntary cooperation in a society were reduced when (a) standards for cooperation with in-group members were made more demanding and (b) punishments for noncooperation were made more severe. In the latter case, voluntary cooperation is lessened because Strong Reciprocators decrease and Opportunists increase. In the former case, voluntary cooperation is lessened because Trusters decrease and Exploitive Egoists increase. Opportunists are key to understanding both these findings, either because they displace Strong Reciprocators (when Punishment is raised) or because their degree of appeasement affects the fortunes of all the other groups (when Appeasement offer is raised).

Whereas the purpose of lifting the Punishment amount > Sucker's payment constraint was to expand the generalizability of findings past those in which the dictates of honor require payback at least at the level of an eye for an eye, there were also other findings of note. First, Strong Reciprocators can grow to a very large portion of the population—but only by losing that which made them most honorable: the principle of seeking retribution that equals or exceeds the affront, and their willingness to risk escalating negative reciprocity (over even "trivial" incidents) in a world where the feud is the worst possible outcome.

Second, we note that in worlds where the "Sucker's payoff" is worse than the feud, the correlation between positive and negative reciprocity moves toward zero (unlike the positive correlation typically found in honor cultures). This positive correlation is also absent when Opportunists are removed from the population. Finally, Opportunists and Strong Reciprocators remain symbiotic in some ways, regardless of constraints on Punishment levels. Without Opportunists present, Trusters became the most popular strategy in the population. With Opportunists present (even in small numbers), Strong Reciprocators surpassed Trusters.

# Study 3: Empirical Replication of Key Simulation Findings

The simulations' results indicate how the presence of Opportunists is important for increasing the prevalence of Strong Reciprocity/Honorable types (Research Question 1, Study 1). They also indicate how high levels of punishment can crowd out voluntary cooperation, replacing Strong Reciprocity/Honorables with Opportunistic agents (Research Question 3, Study 2). Study 3 offers an independent experimental test of these two hypotheses using real participants. The experiment here involved a "multiplayer," multiround economic game that followed the parameters of the simulation. "Multiplayer" is in quotes because—unbeknownst to real participants—the other players were robot confederates, allowing us tighter experimental control of various factors, mirroring the conditions of the simulation more closely, and making the experiment more feasible to run.

Although simulations offer a high degree of precision and can model evolutionary dynamics, simulations in silico only indicate what could happen, not what actually happens with real human beings. Humans are not dispassionate agents seeking only to maximize their payoffs but have a multitude of factors that drive them. These factors include various individual differences or "traits," cognitive biases, aspirations of altruism, desire for revenge, empathy, aversion to betrayal, ideals of fairness and justice, selfimage needs, and social norms that mix with acute episodes of greed, fear, wrath, guilt, and apathy to produce behavior that can be both surprisingly impetuous or stubborn (e.g., Bohnet et al., 2008; Frank, 1988; Ketelaar & Au, 2003).

In simulations, agents behave according to their preprogrammed rules—always. The prevalence of different strategies reflects their success over time as some strategies thrive, while others do not. Reproduction and death operate according to mathematical rules. Between those boundaries, the rules do not make losses loom larger than gains; memory gets wiped clear with every new partner, meaning there are no spillover effects from the warm glow of cooperation and no one becomes "once bitten, twice shy"; agents never copy their neighbors or get influenced by social norms; and they have no need to look themselves in the mirror the next morning (Andreoni, 1995; De Heus et al., 2010; Ketelaar & Au, 2003; Liebrand et al., 1986; McCusker & Carnevale, 1995). Humans have a different operating system. Thus, it is useful to see if what could happen actually does.

# Study 3 Method

#### **Participants**

Participants in this multiplayer game were from Amazon's Mturk crowdsourcing platform. All were U.S. adults. We placed no other restriction on participation, though we included various attention check questions to eliminate bots and ensure that participants understood the rules of the game, as noted below. The final sample contained 418 participants with an average age of 42.36 (SD = 13.03) and a fairly even gender balance (44.7% females). Participants' ethnicities were similar to the U.S. as a

<sup>&</sup>lt;sup>11</sup> Again, calculations for phi will differ somewhat depending on whether counts or percentages are used. When counts are used, the  $\varphi$ s given in the text are as follows: -.92, -.33, and .09.

whole, with a slight underrepresentation of Hispanic/Latinx persons and slight overrepresentation of Asian Americans (10.50% Asian American, 9.80% Black, 5.5% Hispanic/Latinx, 1.2% Other, 72.5% White).

### **Study Conditions**

The study followed the parameters of the honor culture simulations, in which retaliation demands at least an eye for an eye, a tooth for a tooth (Punishment amount > Sucker's payment > Cooperation gain). The behavior of Strong Reciprocators and Exploitive Egoists is thus more extreme and riskier; and because the purpose of the experiment was to see if results hold with real human beings—with all their idiosyncrasies, biases, emotions, and values—we thought examining the more extreme behavior was appropriate. As in Study 1, Strong Reciprocators can be seen as the Honorables of an honor culture, and we refer to them as such in this study.

# Validating the Punishment $\rightarrow$ Voluntary Cooperation Hypothesis

First, we manipulated the level of punishment used in the calculation of the payoff matrix values. Participants either saw a payoff matrix where the payoffs were calculated using a low value of punishment (1.2 *SD*s below the mean punishment level in Study 1), the median level of punishment, or a high level (1.2 *SD*s above the mean in Study 1). The choice of the number of standard deviations is arbitrary; we chose 1.2 to create a difference large enough to observe effects on voluntary cooperation and to stay true to the parameter constraints in the simulation.

#### Validating the Presence of Opportunists Hypothesis

The second factor we manipulated was the presence or absence of Opportunists. In the absence of Opportunists, we expect Trusters to dominate, Honorables to diminish, and the situation to mostly boil down to a game of Hawks and Doves between Trusters and Exploitive Egoists.

### **Dependent Variable**

For each of 50 rounds, participants chose which of the four strategies (or which of the three strategies in games without Opportunists) they wanted to play. Our dependent variables were the percent of rounds on which each strategy was chosen and the amount of voluntary cooperation resulting from those choices. The amount of voluntary cooperation—interpretable as the proportion of all interactions that would involve both parties voluntarily cooperating—was computed similarly to the way it was for the simulation:  $P(\text{Honorables}) \times P(\text{Honorables}) + P(\text{Trusters}) \times P(\text{Trusters}) + 2 \times P(\text{Honorables}) \times P(\text{Trusters}).^{12}$  All proportions were, of course, derived from the responses of real participants and not the confederates they played against.

#### Sample Size and Power

The primary effect sizes of interest were correlations from the simulations involving the effect of (a) severity of punishment on the

amount of voluntary cooperation ( $\beta = -.19$ , partial r = -.26) and on the prevalence of Opportunists ( $\beta = .25$ , partial r = .30) and (b) the presence or absence of Opportunists on the prevalence of Honorables (r = .56) and Trusters (r = -.66). With the *ns* in this study, we would have 99% power to detect effects of this size.

### Procedure

Participants were recruited from MTurk for an "Economic Multiplayer Game." Participants who signed up were taken to an outside website that was ostensibly an online game involving other real-life players. However, as noted, for purposes of experimental control over various factors and for technical feasibility, the participant was the only real player in the game, with the others being computer agents.

On the initial page, participants were told that the game offers them the chance to win additional money, depending on the choices they and the other players make. To create the perception of playing with other actual participants, participants were told that there will be a brief waiting period to obtain the necessary minimum number of players for the game. They were then redirected to the waiting area, where they could see how many other players were taking part. The sign ups slowly populated until they reached 34 participants. We chose 34 as we wanted enough confederates to reflect the distributions in the simulation, while also making the game believable in that it seems feasible that 33 other Mturk players could log into the game within a few minutes of waiting. Participants were then taken to the next screen, which provided an overview of the game and the different strategies.

Participants were told that they will be playing a game with 34 players including themselves; that they and another randomly selected player will decide on a particular strategy to use during their interaction; and that this process would be repeated over multiple rounds, each involving a randomly selected player.

We described the interaction strategies as involving decisions about offering full, partial, or no cooperation and decisions about whether or not to punish someone who has crossed you. To maintain consistency with the simulation and to reduce noise (increase statistical power), players were told they can choose from four strategies when playing against another participant: Strategy 1 offers full cooperation but retaliates against those who cross it. Strategy 2 offers full cooperation and never retaliates. Strategy 3 does not cooperate but retaliates against those who cross it. Strategy 4 offers partial cooperation to avoid punishment and never retaliates. Obviously, the four strategies matched the simulation's Honorable, Truster, Exploitive Egoist, and Opportunist strategies, respectively. However, strategies were only referred to by their number to minimize biases associated with any particular label (Liberman et al., 2004). For games without Opportunists, all mention of Strategy 4 (and corresponding payoffs) were omitted.

After describing the strategies, we confirmed participants' understanding using four multiple-choice items that served as screening

<sup>&</sup>lt;sup>12</sup> Levels of cooperation in the simulation used the proportion above times the size of the population divided by two because a population with 3,000 voluntary cooperators has more voluntary cooperation than a population with 300 cooperators. In the simulation, agents died off and reproduced as a function of their success. During the experimental sessions with participants, no one died or gave birth.

questions. The questions only allowed one wrong response. If participants provided a wrong answer, they were told the answer was incorrect and needed to reanswer the question. Participants who got more than one question wrong were not allowed to continue with the study.

Following these attention checks, participants received further information about the strategies and how the nature of each leads to different results when they interact. The specific instructions are given in Appendices A and B. After being provided with a narrative instruction of how the various strategies fare when they play each other, participants were shown the payoff matrix on the next screen. They were told that the way each person is rewarded or punished depends on how their and the other person's chosen strategies interact. We presented a  $4 \times 4$ table showing what all players earn or lose depending on the combination of strategies. This payoff matrix was the same throughout and was displayed in every round of the game.

Subsequent slides highlighted each of the 16 possible strategy combinations and what payoff the participant and the other player would receive under that combination. To ensure that participants understood how to read the matrix, the game then asked participants to report how many points they would receive/lose under various strategy combinations. Participants were allowed no more than two wrong responses to these open-ended questions and were informed if an answer was incorrect. A total of 82 persons failed to pass the understanding questions (16.4%).

For purposes of examining robustness, we varied two other factors in the procedure: (a) the distribution of strategies in the game and (b) whether those distributions were shown to participants or not. As noted in Supplemental Materials, these factors had little effect on the findings reported below-at least within the range we examined. Participants assigned to see the distribution of strategies in the population were told this prior to the game starting. As an explanation for why participants would be able to see the distribution of other players' choices, they were told that because they were the first one to sign up, they would be the last one to choose their strategy. They were told that no other player received distribution information. All the other "players" would simultaneously have up to 45 s to make their choice. After that period, the distributions of their strategies appeared and remained on the screen. (The information had a small jitter each round to create the perception of variability.) All participants then had up to 45 s to make a choice for that round by clicking one of four buttons, labeled "Strategy 1," "Strategy 2," "Strategy 3," and "Strategy 4." For participants not assigned to see distribution information, they were merely told that as the first to sign up, they would be the last to choose. They also had to wait for the other players to pick their strategies, but saw no distribution information.

After selecting a strategy for the round, participants saw what their "randomly selected partner" chose and how many points each gained or lost in that round. The next round began automatically after giving participants 10 s to view the results. Each round involved a new randomly selected "player" and the game repeated for 50 total rounds. We used 50 rounds so that gameplay would take roughly 30 min and not fatigue participants. Following the 50 rounds, participants completed a few demographic items as well as two open-ended questions asking them to describe what they thought the study was about and to provide any other questions or comments. On these questions, only three participants reported any suspicion regarding the manipulations or other players not being actual people. Results without these three participants look extremely similar to the results (presented below) that include them.

### Compensation

Participants were paid a base rate of \$2 USD. Values in the payoff matrix were equal to those specified in the original simulation. We told participants that each point earned is equal to 50 cents USD. Therefore, in most cases, participants could earn a maximum of nearly \$9.00 USD in the game (based on the distribution of moves the computer agents would choose), and this payoff structure meant that the majority of payment would come from in-game performance, heightening the importance of the player's choices in determining their financial outcomes. Participants who finished the game with a negative point total earned no additional payment. The average amount of additional compensation earned was \$3.28.

# Study 3 Results

### Hypothesis 1: Level of Punishment

We examined the hypothesis that greater punishment levels would decrease the amount of voluntary cooperation in the society. Similar to the simulations, voluntary cooperation was measured as the proportion of interactions in which voluntary cooperation between two players would be achieved (according to the formula derived from the data of actual participants, not confederates). The correlation of punishment with voluntary cooperation in the experiment was r(340) = -.17 (p = .002, 95% CI [-.27, -.06]). This was very close to that predicted by the simulations in Study 1:  $\beta = -.19$ , n = 1,000, 95% CI [-.24, -.15], partial  $r = -.26, p \le .001$ . (Because payoff matrices in the experiment were constructed with Punishment amount > Sucker's payment, we compared Study 3 results with those of the simulation showed the negative effect that severity of punishment had on levels of voluntary cooperation.

In predicting the various subtypes, increasing punishment decreased the percent of Honorable strategies chosen, r(340) =-.18, p = .001, 95% CI [-.28, -.08], and increased the percent of Opportunist strategies, r(340) = .12, p = .027, 95% CI [.01, .22], in the experiment. The effects on Trusters, r(340) = -.01, p = .86,95%CI [-.12, .10], and Exploitive Egoists, r(340) = .08, p = .16, 95% CI [-.03, .18], were not significantly different from zero. These correlations were similar to those predicted by the simulation in Study 1. The 95% CIs of the experimental effect sizes for the effect of punishment at least partly overlap with their simulation counterparts, with both studies appropriately powered; in simulations, the effect of punishment severity on proportion of Honorables,  $\beta = -.09, 95\%$  CI [-.17, -.02], partial r = -.08, t(991) = -2.44,  $p \le .015$ ; Trusters,  $\beta = -.10$ , 95% CI [-.16, -.04], partial r = -.11, t(991) = -3.44,  $p \le .001$ ; Exploitive Egoists,  $\beta = -.07$ , 95% CI [-.14, -.01], partial r = -.07,  $t(991) = -2.13, p \le .033$ ; Opportunists  $\beta = .25, 95\%$  CI [.20, .30], partial r = .30, t(991) = 9.76,  $p \le .001$ . We thus "replicate" the significant effect of increasing the severity of punishment on decreasing the amount of voluntary cooperation, reducing the prevalence of Honorables, and raising the prevalence of Opportunists.

Readers may compare the distributions of response frequencies from the experiment with the predicted responses from the simulation (in parentheses) in Table 6. In terms of raw numbers, both the

 Table 6

 Participants' Choice of Strategies as a Function of Severity of

 Punishment

Punishment	Honorable	Truster	Exploitive Egoist	Opportunist
Low	.30 (.45)	.14 (.23)	.07 (.09)	.48 (.22)
Medium	.26 (.41)	.15 (.19)	.08 (.07)	.51 (.33)
High	.18 (.36)	.14 (.15)	.10 (.05)	.58 (.43)

Note. Numbers in parentheses are predictions from the simulation in Study 1.

experiment and the model predicted that Honorables and Opportunists would make up between two thirds and three fourths of the population. The biggest difference, however, was that people in the experiment were more likely to be Opportunistic and less likely to be Honorable than the simulation predicted. Predictions for the raw numbers of Trusters and Exploitive Egoists were similar in the experiment and simulation. Using the proportions of the various strategies to compute the correlation between positive and negative reciprocity,  $\varphi$  would be approximately .52 ( $p \le .001$ ) in the experiment, compared to the simulation's  $\varphi = .49$  ( $p \le .001$ ).

# Hypothesis 2: Presence of Opportunists

Testing the hypothesis that the presence of Opportunists is important for Honorables to become prominent, we can compare conditions where Opportunists were absent versus present and punishment was at the medium level. (Punishment levels in experimental games without Opportunists were always set at the medium level.) As expected, without Opportunists, the game turned primarily into one of Hawks and Doves as Honorables decreased in number while Exploitive Egoists and especially Trusters increased. The presence or absence of Opportunists in the experiment had statistically significant effects on the percent of the population that were Honorables, F(1, 192) = 8.67, p = .004; the percent that were Trusters, F(1, 192) = 142.52, p = .001; and the percent that were Exploitive Egoists, F(1, 192) = 28.78, p =.001. Readers may see Table 7 to examine the percent of each group in games with and without Opportunists for both the experiment and (in parentheses) the simulation.

As predicted by the simulation, removing Opportunists led to a significant drop in the prevalence of Honorables; however, the drop was not as large as the simulation predicted. The drop for Honorables was 11%, r(192) = -.21, p = .004, rather than the 37% in the simulation (r = -.56). The gain for Trusters was 46%, r(192) = .65, p = .001, in the experiment and 60% (r = .66,  $p \le .001$ ) in the simulation. The gain for Exploitive Egoists was 16%, r(192) = .36,

Table 7

Participants' Choice of Strategies as a Function of the Presence or Absence of Opportunists

Experimental condition	Honorable	Truster	Exploitive Egoist	Opportunist
Opportunists absent	.15 (.035)	.61 (.79)	.24 (.17)	.51 (.33)
Opportunists present	.26 (.41)	.15 (.19)	.08 (.07)	

Note. Numbers in parentheses are predictions from the simulation in Study 1.

p = .001, in the experiment and 10% (r = .18, p = .001) in the simulation. Together, Hawks (Exploitive Egoists) and Doves (Trusters) made up 85% of the experiment when Opportunists were not present (compared to 96% in the simulation). They made up 23% of the experiment when Opportunists were present (compared to 26% in the simulation). Using the proportions of the various strategies to compute the correlation between positive and negative reciprocity,  $\varphi$  would be approximately -.70 ( $p \le .001$ ) in the experiment, compared to the simulation's  $\varphi = -.89$  ( $p \le .001$ ).

# Summary

Overall, the experiment and the simulation produced similar conclusions regarding the central hypotheses. We found that increasing the severity of punishment reduced the levels of voluntary cooperation, decreasing the percent of Honorables, and increasing the percent of Opportunists. The experiment and simulation also concurred that the presence of Opportunists is important for producing a relatively large number of Honorables in a population. Removing the Opportunists reduced the number of Honorables and caused the population to be dominated by Trusters and Exploitive Egoists.

The simulations, however, (a) generally overestimated the attractiveness of the Honorable strategy while underestimating the appeal of the Opportunistic strategy and (b) underestimated the degree to which hard-core Honorables would "stick to their guns" in the absence of Opportunists. We suspect that (a) occurred partly because the simulations fail to account for people's risk aversion (Opportunism is the Minimax strategy). We suspect that (b) occurred because the simulations fail to account for the hard-core Honorables' values or their emotions or identity concerns related to reciprocity and paying back measure for measure. We return to these discrepancies later. Setting them aside, however, we think there was generally a reasonably high concurrence between model predictions and experimental results regarding (a) more severe punishments lessening the amount of voluntary cooperation and (b) the presence of Opportunists being important for Honorables to prosper.

# **General Discussion**

Psychologists tend to think about culture in terms of the way cultures socialize their citizens to behave appropriately and be attuned to the rewards of being a good member of the cultural group (Wrong, 1961). Those who are not paragons of the culture are often regarded as "error" or "noise." Thus, psychologists often neglect the crucial role that Opportunists play in the functioning of a culture. In the research above, we saw how Opportunists are important for understanding some puzzling cultural phenomena. For example, we saw the importance of Opportunists for the emergence of strong reciprocity. In silico and in vivo, Opportunists were symbiotic with their exact opposites, the Strong Reciprocators/ Honorables. Additionally, in both the simulations and experiment, Opportunists and Honorables created the peculiar correlation between positive reciprocity and negative reciprocity found in honor cultures. In simulations, when being a Strong Reciprocator involved being less punitive and less risky (as when retaliation was less than an eye for an eye and being a sucker was worse than getting into a feud), the correlation between positive and negative reciprocity shrunk close to zero. In simulations and the experiment, when Opportunists were removed, the correlation went sharply negative.

In Study 2, we saw how Opportunists play a key role in explaining the level of development of civil society. More particularly, Opportunists play a key role in explaining the low levels of social capital often found in collectivistic and tight cultures. Such cultures exert strong socializing pressure by effectively monitoring and sanctioning selfish behavior and by raising the standards of what constitutes cooperative behavior that is "good enough" to avoid punishment. However, the strong pressure to cooperate crowds out voluntary cooperation: In both the simulations and experiment, when punishments for noncooperation were made more severe, the level of voluntary cooperation in a society dropped, with Strong Reciprocators/Honorables decreasing and Opportunists increasing as a percent of the population. In simulations, raising the standards for what is "good enough" (i.e., increasing the amount Opportunists had to offer to appease Strong Reciprocators and Exploitive Egoists) lessened voluntary cooperation as Exploitive Egoists gained and altruistic Trusters were crowded out. Ironically-but congruent with the "puzzle" of social capital being lower in tight and collectivistic cultures-raising the standards for what is "good enough" cooperation and increasing the punishment for noncooperation reduces the amount of cooperation voluntarily given, without the specter of possible punishment, in a society. The dynamics observed in silico and in vivo shed light on this "puzzle" and the role Opportunists play in producing it.

# Mental Models

Two other points are important to consider. Both are relevant to how these findings fit with psychology's disciplinary biases.

First, because this article is written for a psychology journal, we have emphasized how the presence of Opportunistic agents explains some phenomena that seem puzzling at first glance or are often dismissed as "error." The role of Opportunists is underappreciated in psychology because our mental model of people is usually not one of Minimaxers or unearnest, sycophantic "weasels," or apathetic individuals. However, had this article been written for a different audience, there would likely be a different takeaway. Many economists do model people as habitual Minimaxers. What might stand out for them are the presence of devoted actors—those who have strong predispositions to trust though they risk betrayal or those who have a strong predisposition to punish even if that leads them into extremely costly feuds.

Obviously, this is a caricature of many psychologists' and economists' views. But it seems generally safe to say that the basic mental model for psychologists is different than that for economists. Regardless of the focal type of interest, however, the point here is that the *mix* of these types—devoted actors behaving in accord with their predispositions or Opportunistic actors trying to get over or get by—is useful for understanding a cultural system. Cultures consist of a mix of individuals—not just paragons but their opposites, as well as partial deviants. In Konner's (2007) metaphor, cultures are more symphonies than solos: articulation of its various personality types ... rather than from fundamental tendencies shared by a majority—a sort of symphony orchestra model of culture and personality, in which each culture provides a series of scores (p. 79).

# **Asocial Psychology**

# **Between-the-Ears**

A second point: Some of the "puzzling" cultural phenomena noted here can be understood purely at the psychological level, without any need for evolutionary game theory or ABM. For example:

- a. in terms of harsh punishments and strong external demands crowding out good behavior, we can understand such phenomena in terms of overjustification effects, with highly salient situational demands undermining internal attributions for our own behavior (Aronson & Carlsmith, 1963; Tenbrunsel & Messick, 1999). Harsh punishments and strong external demands also may crowd out trust. Seeing a person behave properly when there is tremendous external pressure to do so, we attribute their good behavior to those pressures rather than the person's "true" desires (Morris and Larrick, 1995). And so, not convinced that others are truly trustworthy, we are less likely to trust them and hesitant to be made a fool by attempting to cooperate.
- b. Relatedly, one can explain collectivism's negative relation with social capital through its creation of an ideology that regards sacrifice for close in-group members as necessary but sacrifices for anyone else as foolish (e.g., Banfield's, 1967 "amoral familism" and the sentiment behind the aphorism, "charity begins at home.")
- c. Concerning honor, one can explain the way an honor ideology ties together positive and negative reciprocity in terms of overarching ideals of toughness and virtue— and can note that people can either accept this ideology or reject it. Thus, people may not always be of a culture, but they are always in a culture—and they react toward or against the major cultural templates that organize behavior (Cohen & Leung, 2012; Leung & Cohen, 2011). Additionally, such an ideology can also explain why Honorables prosper in moderate- to high-risk environments (where it is advantageous to cheat and costly to be cheated). Honorables are predisposed to cooperate and never cheat anyone because cheating—even if it is highly rewarding—is just not something an honorable person does.

Thus, these "puzzling" phenomena—the positive correlation between tendencies toward vengeance and prosocial reciprocity in honor cultures, the crowding out of social capital by strong demands for cooperation, and so on—*can* be understood at a purely psychological, between-the-ears level. They can be explained in terms of ideologies that people either believe in or reject—and hence, they can be understood without any recourse to agent-based models and the differential selection of strategies. However, what a purely psychological/ideological/between-the-ears account omits are the contingencies in the environment that maintain certain

individual variation in personality and character is great in every known culture ... at best there is perhaps a 'modal personality' shared by a substantial minority of a culture's members ... and in any case a culture must derive its distinctiveness from the particular mutual

strategies and make them differentially viable (Harton & Bullock, 2007; Heine et al., 1999; Sato et al., 2014; Savani et al., 2011; Schug et al., 2010; Üskül & Oishi, 2018). That is, giving a purely psychological account ignores the real rewards and punishments that other people provide for us; and ignoring what other people do for and to us leads to a very asocial psychology. ABM involves simplifying agents' psychological richness, it (partly) makes up for in describing aspects of the *social ecosystem* that sustain certain behavioral strategies at certain levels.

We think it useful to not only understand culture as a set of beliefs but to understand it as a social system that—in some nonobvious ways—makes some beliefs, ideologies, and strategies more viable than others (Heine, 2015; Markus & Hamedani, 2019; Talhelm & Oishi, 2019). Doing so helps us understand how—at a system level—cultures emerge, maintain themselves, and change.

### Limitations

ABM suggests where a social system may find some quasistationary equilibrium, as evolutionary game theoretic principles influence how big the niche is for strategies of different types. However, there are principles beyond cold, rational calculations that influence people's choices. Based on the in vivo experiment, it seems the simulations may have underestimated the appeal of the risk-reducing strategy of Opportunists, who minimize the maximum damage that can be done to them. Regarding Honorables, the simulations both overestimated how many people were willing to engage in this punitive, risky strategy and underestimated the number of hard-core Honorables, who stuck with this strategy even when the environment became less favorable (such as when Opportunists were absent).

One question is how disturbed one should be about the discrepancies. In terms of simulations' overestimating the absolute number of Honorables and underestimating the number of Opportunists in experiments where we varied punishment, we are of two minds about this. On the one hand, the absolute number is not terribly important because these numbers can probably be pushed around by the framing of the game—as in Liberman et al. (2004) who moved around levels of cooperation in a Prisoner's Dilemma by approximately 30% merely by calling the game either the "Community" game or the "Wall Street" game. This suggests the profound importance of the framing of the game-a point that has nothing to do with the mathematics of the game but one that social psychologists have no trouble recognizing as true. Accepting that point, the key issue is not so much about the absolute numbers of people who choose a certain strategy; it is more about how our independent variables-such as severity of punishment, standards for what is good enough, the presence or absence of Opportunists, and so forth-push around those choices, regardless of the "mean" level of those choices in a particular environment.

On the other hand, absolute numbers do matter. If one is talking about a culture of honor, for example, then one should have some significant fraction of the population holding to honor norms. Whether the fraction engaging in this high-risk strategy is 10%, 25%, or 50% is less clear to us. Whether we should talk about that fraction in terms of the number of *people* embracing an honor strategy (as we have done in this article) versus the number of settings and situations calling for an honorable response is also less clear, and this second approach makes it especially apparent how absurd the quest for a specific number may be.

### Conclusion

There are other aspects of the present model that seem worthy of exploring—including implementing different rules for reproduction or considering the evolution of strategies within the context of both within-group and between-group competition (Norenzayan, 2013; Norenzayan et al., 2016). Further, some dynamics of the present model produce instances similar to the cycling found in Nowak et al. (2016), replicating findings that illustrate Lotka–Volterra predator–prey cycles in cultural dynamics.

Leaving this aside, however, the current model shows promise for contributing to our understanding of some aspects of culture that are worthy of explanation (e.g., the emergence of strong reciprocity, the correlation between positive and negative reciprocity shown in honor cultures, the symbiosis between opposite agent types, and the crowding out of civic virtue in tight and collectivistic societies). While the model is not the only explanation for these phenomena, it contributes novel explanations for their occurrence. Further, we hope the article helps illuminate for psychologists an understudied "type" that plays a major role in the workings of culture. Psychologists often focus either on well-socialized individuals or on individuals with strong internal predispositions to act one way or another (Wrong, 1961). This article highlights the crucial role of the Opportunist for producing some of the unintended consequences and surprising phenomena in a social system.

### References

- Allik, J., & Realo, A. (2004). Individualism-collectivism and social capital. Journal of Cross-Cultural Psychology, 35(1), 29–49. https://doi.org/10 .1177/0022022103260381
- Anderson, E. (1994). The code of the streets. Atlantic Monthly, 273(5), 81– 94. Retrieved May 15, 2021, from https://www.theatlantic.com/magazine/ archive/1994/05/the-code-of-the-streets/306601/
- Anderson, E. (2000). Code of the streets. W.W. Norton.
- Andreoni, J. (1995). Warm-glow versus cold-prickle: The effects of positive and negative framing on cooperation in experiments. *The Quarterly Journal of Economics*, 110(1), 1–21. https://doi.org/10.2307/2118508
- Aronson, E., & Carlsmith, M. (1963). Effect of the severity of threat on the devaluation of forbidden behavior. *Journal of Personality and Social Psychology*, 66(6), 584–588. https://doi.org/10.1037/h0039901
- Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality* and Social Psychology Review, 11(2), 150–166. https://doi.org/10.1177/ 1088868306294907
- Banfield, E. (1967). Moral basis of a backward society. Free Press.
- Barnes, C. D., Brown, R. P., & Osterman, L. L. (2012). Don't tread on me: Masculine honor ideology in the U.S. and militant responses to terrorism. *Personality and Social Psychology Bulletin*, 38(8), 1018–1029. https:// doi.org/10.1177/0146167212443383
- Baumeister, R. F., Stillwell, A., & Wotman, S. R. (1990). Victim and perpetrator accounts of interpersonal conflict: Autobiographical narratives about anger. *Journal of Personality and Social Psychology*, 59(5), 994– 1005. https://doi.org/10.1037/0022-3514.59.5.994
- Bergstrom, C. T., & Godfrey-Smith, P. (1998). On the evolution of behavioral heterogeneity in individuals and populations. *Biology & Philosophy*, 13(2), 205–231. https://doi.org/10.1023/A:1006588918909
- Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the*

National Academy of Sciences, 115(11), 2705–2710. https://doi.org/10 .1073/pnas.1719445115

- Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *The American Economic Review*, 98(1), 294–310. https://doi.org/10.1257/aer.98.1.294
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1), 17–28. https://doi.org/10.1016/j.tpb.2003.07.001
- Brethel-Haurwitz, K. M., Stoycos, S. A., Cardinale, E. M., Huebner, B., & Marsh, A. A. (2016). Is costly punishment altruistic? Exploring rejection of unfair offers in the Ultimatum Game in real-world altruists. *Scientific Reports Nature*, 6(1), Article 18974. https://doi.org/10.1038/srep18974
- Brown, R. P. (2004). Vengeance is mine: Narcissism, vengeance, and the tendency to forgive. *Journal of Research in Personality*, 38(6), 576–584. https://doi.org/10.1016/j.jrp.2003.10.003
- Brown, R. P. (2016). *Honor bound: How a cultural ideal has shaped the American psyche*. Oxford University Press.
- Cao, Y., Enke, B., Falk, A., Giuliano, P., & Nunn, N. (2021). Herding, warfare, and a culture of honor: Global evidence (Working Paper No. 29250). National Bureau of Economic Research.
- Chagnon, N. (2012). The Yanomamo. Cengage.
- Cohen, D., Hernandez, I., Gruschow, K., Nowak, A., Gelfand, M., & Borkowski, W. (2018). Rationally irrational? The ecologies and economics of honor. In A. Uskul & S. Oishi (Eds.), Socioeconomic environment and human psychology (pp. 77–104). Oxford University Press.
- Cohen, D., & Leung, A. K.-Y. (2012). Virtue and virility: Governing with honor and the association or dissociation between martial honor and moral character of U.S. presidents, legislators, and justices. *Social Psychological & Personality Science*, 3(2), 162–171. https://doi.org/10.1177/1948550611 412792
- Cohen, D., & Nisbett, R. E. (1994). Self-protection and the culture of honor: Explaining southern violence. *Personality and Social Psychology Bulletin*, 20(5), 551–567. https://doi.org/10.1177/0146167294205012
- Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An "experimental ethnography". *Journal of Personality and Social Psychology*, 70(5), 945–960. https://doi.org/10.1037/0022-3514.70.5.945
- Cohen, D., Vandello, J., Puente, S., & Rantilla, A. (1999). "When you call me that, smile!" How norms for politeness, interaction styles, and aggression work together in southern culture. *Social Psychology Quarterly*, 62(3), 257–275. https://doi.org/10.2307/2695863
- Cohen, D., Vandello, J., & Rantilla, A. (1998). The sacred and the social: Cultures of honor and violence. In P. Gilbert and B. Andrews (Eds.), *Shame: Interpersonal behavior, psychopathology, and culture* (pp. 261– 282). Oxford University Press.
- Coleman, J. S. (1990). Foundations of social theory. Harvard University Press.
- Colman, A. M., & Wilson, J. C. (1997). Antisocial personality disorder: An evolutionary game theory analysis. *Legal and Criminological Psychology*, 2(1), 23–34. https://doi.org/10.1111/j.2044-8333.1997.tb00330.x
- Cross, S. E., Uskul, A. K., Gerçek-Swing, B., Sunbay, Z., Alözkan, C., Günsoy, C., Ataca, B., & Karakitapoglu-Aygün, Z. (2014). Cultural prototypes and dimensions of honor. *Personality and Social Psychology Bulletin*, 40(2), 232–249. https://doi.org/10.1177/0146167213510323
- Daly, M., & Wilson, M. (2017). *Homicide*. Routledge. https://doi.org/10 .4324/9780203789872
- Dawkins, M. B., Sloane, S., & Baillargeon, R. (2019). Do infants in the first year of life expect equal resource allocations? *Frontiers in Psychology*, 10, Article 116. https://doi.org/10.3389/fpsyg.2019.00116
- De Heus, P., Hoogervorst, N., & Van Dijk, E. (2010). Framing prisoners and chickens: Valence effects in the prisoner's dilemma and the chicken game. *Journal of Experimental Social Psychology*, 46(5), 736–742. https:// doi.org/10.1016/j.jesp.2010.04.013

Diamond, J. (2004). Collapse. Penguin.

- Edgerton, R. (1971). *The individual in cultural adaptation*. University of California Press.
- Egloff, B., Richter, D., & Schmukle, S. C. (2013). Need for conclusive evidence that positive and negative reciprocity are unrelated. *Proceedings* of the National Academy of Sciences, 110(9), E786–E786. https://doi.org/ 10.1073/pnas.1221451110
- Eisenberger, R., Lynch, P., Aselage, J., & Rohdieck, S. (2004). Who takes the most revenge? Individual differences in negative reciprocity norm endorsement. *Personality and Social Psychology Bulletin*, 30(6), 787– 799. https://doi.org/10.1177/0146167204264047
- Eriksson, K., Strimling, P., Gelfand, M., Wu, J., Abernathy, J., Akotia, C. S., Aldashev, A., Andersson, P. A., Andrighetto, G., Anum, A., Arikan, G., Aycan, Z., Bagherian, F., Barrera, D., Basnight-Brown, D., Batkeyev, B., Belaus, A., Berezina, E., Björnstjerna, M., . . . Van Lange, P. A. (2021). Perceptions of the appropriate response to norm violation in 57 societies. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-21602-9
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, 13(1), 1–25. https://doi.org/10.1007/s12110-002-1012-7
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *The American Economic Review*, 90(4), 980–994. https:// doi.org/10.1257/aer.90.4.980
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140. https://doi.org/10.1038/415137a
- Fischer, D. (1988). Albion's seed. Oxford University Press.
- Frank, R. H. (1988). Passions within reason. W.W. Norton.
- Franklin, J. (1956). The militant South. Beacon Press.
- Gavrilets, S., & Richerson, P. J. (2017). Collective action and the evolution of social norm internalization. *Proceedings of the National Academy of Sciences*, 114(23), 6068–6073. https://doi.org/10.1073/ pnas.1703857114
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliach, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., ... Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104. https://doi.org/10.1126/science.1197754
- Geniole, S. N., MacDonell, E. T., & McCormick, C. M. (2017). The threat premium in economic bargaining. *Evolution and Human Behavior*, 38(5), 572–582. https://doi.org/10.1016/j.evolhumbehav.2016.12.004
- Ghate, R., Ghate, S., & Ostrom, E. (2013). Cultural norms, cooperation, and communication: Taking experiments to the field in indigenous communities. *The International Journal of the Commons*, 7(2), 498–520. https:// doi.org/10.18352/ijc.376
- Gilmore, D. D. (1990). Manhood in the making: Cultural concepts of masculinity. Yale University Press.
- Grosjean, P. (2014). A history of violence: The culture of honor and homicide in the U.S. South. *Journal of the European Economic Association*, 12(5), 1285–1316. https://doi.org/10.1111/jeea.12096
- Halali, E., Bereby-Meyer, Y., & Ockenfels, A. (2013). Is it all about the self? The effect of self-control depletion on ultimatum game proposers. *Frontiers in Human Neuroscience*, 7. Article 240 https://doi.org/10.3389/ fnhum.2013.00240
- Harrington, J. R., & Gelfand, M. J. (2014). Tightness-looseness across the 50 United States. *Proceedings of the National Academy of Sciences*, 111(22), 7990–7995. https://doi.org/10.1073/pnas.1317937111
- Harton, H. C., & Bullock, M. (2007). Dynamic social impact: A theory of the origins and evolution of culture. *Social and Personality Psychology Compass*, 1(1), 521–540. https://doi.org/10.1111/j.1751-9004.2007.00022.x
- Heine, S. J. (2015). Cultural psychology. W.W. Norton.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review*, 106(4), 766–794. https://doi.org/10.1037/0033-295X.106.4.766

- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2010). Markets, religion, community size, and the evolution of fairness and punishment. *Science*, 327(5972), 1480–1484. https://doi.org/10.1126/science.1182238
- Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., & Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–1770. https://doi.org/10.1126/ science.1127333
- Hernandez, I., Cohen, D., Gruschow, K., Nowak, A., Gelfand, M. J., & Borkowski, W. (2021). *The importance of being unearnest: Opportunists* and the making of culture [Data set]. https://osf.io/pf2ny
- Hofstede, G. (2003). Culture's consequences. Sage Publications.
- Hong, Y. Y., Ip, G., Chiu, C. Y., Morris, W., & Menon, T. (2001). Cultural identity and dynamic construction of the self: Collective duties and individual rights in Chinese and American cultures. *Social Cognition*, 19(3), 251–268. https://doi.org/10.1521/soco.19.3 .251.21473
- Hoyt, W. T., Fincham, F. D., McCullough, M. E., Maio, G., & Davila, J. (2005). Responses to interpersonal transgressions in families: Forgivingness, forgivability, and relationship-specific effects. *Journal of Personality and Social Psychology*, 89(3), 375–394. https://doi.org/10.1037/0022-3514.89.3.375
- International Personality Item Pool. (2020). Agreeableness. Retrieved January 7, 2020, from https://ipip.ori.org/newNEOKey.htm#Agreea bleness
- Jankowski, R. (1990). Punishment in iterated chicken and prisoner's dilemma games. *Rationality and Society*, 2(4), 449–470. https://doi.org/ 10.1177/1043463190002004004
- Jansson, F. (2013). Pitfalls in spatial modelling of ethnocentrism. Journal of Artificial Societies and Social Simulation, 16(3), Article 2. https://doi.org/ 10.18564/jasss.2163
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big-5 trait taxonomy. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality*. Guilford Press.
- Kearns, J. N., & Fincham, F. D. (2005). Victim and perpetrator accounts of interpersonal transgressions: Self-serving or relationship-serving biases? *Personality and Social Psychology Bulletin*, 31(3), 321–333. https:// doi.org/10.1177/0146167204271594
- Ketelaar, T., & Au, W. T. (2003). The effects of feelings of guilt on the behaviour of uncooperative individuals in repeated social bargaining games: An affect-as-information interpretation of the role of emotion in social interaction. *Cognition and Emotion*, 17(3), 429–453. https://doi.org/ 10.1080/02699930143000662
- Keysar, B., Converse, B. A., Wang, J., & Epley, N. (2008). Reciprocity is not give and take: Asymmetric reciprocity to positive and negative acts. *Psychological Science*, 19(12), 1280–1286. https://doi.org/10.1111/j .1467-9280.2008.02223.x
- Kim, H., & Lawrie, S. (2019). Culture and motivation. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed., pp. 268– 291). Guilford Press.
- Kitayama, S., Varnum, M., & Salvador, C. (2019). Cultural neuroscience. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed., pp. 79–118). Guilford Press.
- Konner, M. (2007). Evolutionary foundations of cultural psychology. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (1st ed., pp. 77–108). Guilford Press.
- Kowalski, R. M. (2000). "I was only kidding!": Victims' and perpetrators' perceptions of teasing. *Personality and Social Psychology Bulletin*, 26(2), 231–241. https://doi.org/10.1177/0146167200264009
- Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2015). Group cooperation without group selection: Modest punishment can recruit much

cooperation. PLOS One, 10(4), Article e0124561. https://doi.org/10.1371/journal.pone.0124561

- Lacey, L. (2013). Women for cows: An analysis of abductions of women in South Sudan. Agenda (Durban, South Africa), 27(4), 91–108. https:// doi.org/10.1080/10130950.2013.861685
- Lee, K., & Ashton, M. C. (2012). Getting mad and getting even: Agreeableness and Honesty-Humility as predictors of revenge intentions. *Personality and Individual Differences*, 52(5), 596–600. https://doi.org/10.1016/j .paid.2011.12.004
- Legatum Institute. (2016). Prosperity index.
- Leung, A. K. Y., & Cohen, D. (2011). Within- and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of Personality and Social Psychology*, 100(3), 507–526. https://doi.org/10.1037/a0022151
- Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychol*ogy Bulletin, 30(9), 1175–1185. https://doi.org/10.1177/0146167 204264004
- Liebrand, W. B., Wilke, H. A., Vogel, R., & Wolters, F. J. (1986). Value orientation and conformity: A study using three types of social dilemma games. *The Journal of Conflict Resolution*, 30(1), 77–97. https://doi.org/ 10.1177/0022002786030001006
- Luria, G., Cnaan, R. A., & Boehm, A. (2015). National culture and prosocial behaviors: Results from 66 countries. *Nonprofit and Voluntary Sector Quarterly*, 44(5), 1041–1065. https://doi.org/10.1177/ 0899764014554456
- Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy* of Sciences, 115(38), E8835–E8843. https://doi.org/10.1073/pnas .1801677115
- Markus, H., & Hamedani, M. (2019). People are culturally-shaped shapers. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed., pp. 11–52). Guilford Press.
- McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality* and Social Psychology, 52(1), 81–90. https://doi.org/10.1037/0022-3514 .52.1.81
- McCullough, M. E., Bellah, C. G., Kilpatrick, S. D., & Johnson, J. L. (2001). Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin*, 27(5), 601–610. https://doi.org/10.1177/0146167201275008
- McCullough, M. E., & Hoyt, W. T. (2002). Transgression-related motivational dispositions: Personality substrates of forgiveness and their links to the Big Five. *Personality and Social Psychology Bulletin*, 28(11), 1556– 1573. https://doi.org/10.1177/014616702237583
- McCusker, C., & Carnevale, P. J. (1995). Framing in resource dilemmas: Loss aversion and the moderating effects of sanctions. *Organizational Behavior and Human Decision Processes*, 61(2), 190–201. https://doi.org/ 10.1006/obhd.1995.1015
- McWhiney, G. (1988). Cracker culture: Celtic ways in the Old South. University of Alabama Press.
- Mikulincer, M., & Shaver, P. (2010). *Prosocial motives, emotions, and behavior*. American Psychological Association.
- Miller, W. (2007). An eye for an eye. Cambridge University Press.
- Miller, J. G., Bersoff, D. M., & Harwood, R. L. (1990). Perceptions of social responsibilities in India and in the United States: Moral imperatives or personal decisions? *Journal of Personality and Social Psychology*, 58(1), 33–47. https://doi.org/10.1037/0022-3514.58.1.33
- Morris, M. W., & Larrick, R. P. (1995). When one cause casts doubt on another: A normative analysis of discounting in causal attribution. *Psychological Review*, 102(2), 331–355. https://doi.org/10.1037/0033-295X .102.2.331

- Napier, J. (1989). Military tradition. In C. Wilson & W. Ferris (Eds.), *Encyclopedia of southern culture* (pp. 641–643). University of North Carolina Press.
- Nisbett, R., & Cohen, D. (1996). Culture of honor. Perseus.
- Norenzayan, A. (2013). Big gods: How religion transformed cooperation and conflict. Princeton University Press.
- Norenzayan, A., Shariff, A. F., Gervais, W. M., Willard, A. K., McNamara, R. A., Slingerland, E., & Henrich, J. (2016). The cultural evolution of prosocial religions. *Behavioral and Brain Sciences*, 39, Article e1. https:// doi.org/10.1017/S0140525X14001356
- Norris, P. (2017). Democracy cross-national data. Retrieved July 26, 2017, from https://sites.google.com/site/pippanorris3/research/data#TOC-The-World-Values-Study
- Nowak, A., Gelfand, M. J., Borkowski, W., Cohen, D., & Hernandez, I. (2016). The evolutionary basis of honor cultures. *Psychological Science*, 27(1), 12–24. https://doi.org/10.1177/0956797615602860
- O'Kelly, C., & Carney, L. (1986). Women and men in society. Wadsworth.
- Oishi, S. (2010). The psychology of residential mobility: Implications for the self, social relationships, and well-being. *Perspectives on Psychological Science*, 5(1), 5–21. https://doi.org/10.1177/1745691609356781
- Oishi, S., & Kisling, J. (2009). The mutual constitution of residential mobility and individualism. In R. Wyer, C. Y. Chiu, & Y.-y. Hong (Eds.), Understanding culture: Theory, research, and application (pp. 223–234). Psychology Press.
- Ostrom, E. (2000). Collective action and the evolution of social norms. *The Journal of Economic Perspectives*, *14*(3), 137–158. https://doi.org/10 .1257/jep.14.3.137
- Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin*, 128(1), 3–72. https:// doi.org/10.1037/0033-2909.128.1.3
- Pasley, J. F. (2008). Chicken pax atomica: The cold war stability of nuclear deterrence. Journal of International and Area Studies, 15(2), 21–39.
- Pelto, P. (1968). The differences between "tight" and "loose" societies. Society, 5(5), 37–40. https://doi.org/10.1007/BF03180447
- Peristiany, J. (1966). Honor and shame. University of Chicago Press.
- Pitt-Rivers, J. (1968). Honor. In D. Sills (Ed.), International encyclopedia of the social sciences (Vol. 6, p. 506). Macmillan.
- Prasad, K. (2012). Economic liberalization and violent crime. *The Journal of Law & Economics*, 55(4), 925–948. https://doi.org/10.1086/666487
- Putnam, R. D. (2001). Bowling alone: The collapse and revival of American community. Simon and Schuster.
- Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). Making democracy work. Princeton University Press.
- Rapoport, A., & Guyer, M. (1966). A taxonomy of 2 x 2 games. General Systems, 11, 203–214.
- Rodriguez Mosquera, P. M. (2016). On the importance of family, morality, masculine, and feminine honor for theory and research. *Social and Personality Psychology Compass*, 10(8), 431–442. https://doi.org/10.1111/spc3.12262
- Salamon, L., Sokolowski, S., & associates. (2004). Global civil society (Vol. 2). Kumarian Press.
- Santos, F. C., Santos, M. D., & Pacheco, J. M. (2008). Social diversity promotes the emergence of cooperation in public goods games. *Nature*, 454(7201), 213–216. https://doi.org/10.1038/nature06940
- Sato, K., Yuki, M., & Norasakkunkit, V. (2014). A socio-ecological approach to cross-cultural differences in the sensitivity to social rejection: The partially mediating role of relational mobility. *Journal of Cross-Cultural Psychology*, 45(10), 1549–1560. https://doi.org/10.1177/0022022114544320
- Savani, K., Morris, M. W., Naidu, N. V. R., Kumar, S., & Berlia, N. V. (2011). Cultural conditioning: Understanding interpersonal accommodation in India and the United States in terms of the modal characteristics of interpersonal influence situations. *Journal of Personality and Social Psychology*, 100(1), 84–102. https://doi.org/10.1037/a0021083
- Schelling, T. C. (1980). The strategy of conflict. Harvard University Press.

- Schilling, J., Opiyo, F. E., & Scheffran, J. (2012). Raiding pastoral livelihoods: Motives and effects of violent conflict in north-western Kenya. *Pastoralism: Research. Policy & Practice*, 2(1), 1–16. https://doi.org/10 .1186/2041-7136-2-25
- Schug, J., Yuki, M., & Maddux, W. (2010). Relational mobility explains between- and within-culture differences in self-disclosure to close friends. *Psychological Science*, 21(10), 1471–1478. https://doi.org/10.1177/ 0956797610382786
- Sheppard, K. E., & Boon, S. D. (2012). Predicting appraisals of romantic revenge: The roles of honesty–humility, agreeableness, and vengefulness. *Personality and Individual Differences*, 52(2), 128–132. https://doi.org/10 .1016/j.paid.2011.09.014
- Simantov-Nachlieli, I., & Shnabel, N. (2014). Feeling both victim and perpetrator: Investigating duality within the needs-based model. *Personality and Social Psychology Bulletin*, 40(3), 301–314. https://doi.org/10 .1177/0146167213510746
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, 23(2), 196–204. https://doi.org/10.1177/ 0956797611422072
- Snyder, G. H. (1971). "Prisoner's dilemma" and "chicken" models in international politics. *International Studies Quarterly*, 15(1), 66–103. https://doi.org/10.2307/3013593
- Stillwell, A. M., Baumeister, R. F., & Del Priore, R. E. (2008). We're all victims here: Toward a psychology of revenge. *Basic and Applied Social Psychology*, 30(3), 253–263. https://doi.org/10.1080/01973530802375094
- Sweet, L. E. (1965). Camel raiding of North Arabian Bedouin: A mechanism of ecological adaptation. *American Anthropologist*, 67(5), 1132–1150. https://doi.org/10.1525/aa.1965.67.5.02a00030
- Talhelm, T., & Oishi, S. (2019). Culture and ecology. In D. Cohen & S. Kitayama (Eds.), *Handbook of cultural psychology* (2nd ed., pp. 119– 143). Guilford Press.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. Administrative Science Quarterly, 44(4), 684– 707. https://doi.org/10.2307/2667052
- Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences*, *116*(13), 6025–6034. https://doi.org/10.1073/pnas.1817849116

Tocqueville, A. (2002). Democracy in America. Chicago.

- Triandis, H. (1988). Collectivism v. individualism. In C. Bagley & G. Verma (Eds.), Cross-cultural studies of personality, attitudes and cognition (pp. 60–95). Palgrave. https://doi.org/10.1007/978-1-349-08120-2\_3
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46(1), 35–57. https://doi.org/10.1086/406755
- Üskül, A. K., Cross, S. E., Sunbay, Z., Gercek-Swing, B., & Ataca, B. (2012). Honor bound: The cultural construction of honor in Turkey and the Northern United States. *Journal of Cross-Cultural Psychology*, 43(7), 1131–1151. https://doi.org/10.1177/0022022111422258
- Üskül, A. K., & Oishi, S. (Eds.). (2018). Socio-economic environment and human psychology: Social, ecological, and cultural perspectives. Oxford University Press. https://doi.org/10.1093/oso/9780190492908.001.0001
- Vandello, J. A., & Cohen, D. (1999). Patterns of individualism and collectivism across the United States. *Journal of Personality and Social Psychology*, 77(2), 279–292. https://doi.org/10.1037/0022-3514.77.2.279
- Vandermeer, J., Hosey, C., Epley, N., & Keysar, B. (2019). Escalation of negative social exchange: Reflexive punishment or deliberative deterrence? *Journal of Experimental Social Psychology*, 84, Article 103823. https://doi.org/10.1016/j.jesp.2019.103823
- Weber, T. O., Weisel, O., & Gächter, S. (2018). Dispositional free riders do not free ride on punishment. *Nature Communications*, 9(1), Article 2390. https://doi.org/10.1038/s41467-018-04775-8
- Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo/

- Witsenburg, K. M., & Adano, W. R. (2009). Of rain and raids: Violent livestock raiding in northern Kenya. *Civil Wars*, 11(4), 514–538. https:// doi.org/10.1080/13698240903403915
- Wrong, D. H. (1961). The oversocialized conception of man in modern sociology. American Sociological Review, 26(2), 183–193. https://doi.org/ 10.2307/2089854
- Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., Miura, A., Inukai, K., Takagishi, H., & Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences*, 109(50), 20364–20368. https://doi.org/10.1073/pnas .1212126109

# Appendix A

# **Overview Instructions for In Vivo Experiment**

#### **Overview of Game**

You will be playing a game with 34 players including you. During the game, you and another randomly selected player will decide on a particular strategy. You will play multiple rounds. The person you will play is newly and randomly selected in each round.

Strategies involve decisions about offering full, partial, or no cooperation, and they involve decisions about whether or not to punish someone who has crossed you.

To simplify things, we will let you and the other players choose from four strategies to use when playing against another player:

 Strategy 1 offers full cooperation but retaliates against those who cross it.

- Strategy 2 offers full cooperation and never retaliates.
- Strategy 3 does not cooperate but retaliates against those who cross it.
- Strategy 4 offers partial cooperation to avoid punishment and never retaliates.

The payoff you receive depends on how the strategy you choose interacts with the strategy the other player chose.

The more points you get, the more additional money you will earn at the end of the game.

Each point is worth 50 cents.

# Appendix B

# **Overview Description for In Vivo Experiment**

What you receive depends on what you choose and the choice of a player randomly paired with you in that round.

Strategies 1 and 2 are fully cooperating strategies. The basic principle with fully cooperating is that cooperating is beneficial to you and the other person, but there is risk because you can lose more than you can gain if you try to cooperate but the other person does not want to.

Some strategies—Strategies 1 and 3—can punish other players. The basic principle with punishment is that you can hurt someone who you think is taking advantage of you by not cooperating, but you risk getting badly hurt yourself if the other person is also a punisher. (Note: We said you punish those who YOU THINK are taking advantage of you. There is miscommunication here—just like in the real world, and we have built that into the payoff matrix you will see.)

Strategy 4 tries to avoid risk by offering partial cooperation to avoid punishment from Strategies 1 and 3 and not cooperating with nonpunishing Strategies 2 and 4. The basic principle with partial cooperation is minimizing (as much as possible) the risk of being punished and being taken advantage of.

> Received May 25, 2019 Revision received November 29, 2021 Accepted December 17, 2021