

# The evolution of costly displays, cooperation and religion: credibility enhancing displays and their implications for cultural evolution

Joseph Henrich

## 1. Introduction

Researchers from across the behavioral sciences have long proposed a connection between apparently costly displays — often in various ritualized forms such as firewalking, ritual scarification, animal sacrifice and subincision — and deep levels of commitment to group ideologies, religious beliefs and shared values that promote solidarity and in-group cooperation (Atran & Norenzayan, 2004; Cronk, 1994; Durkheim, 1995; Irons, 1996; Rappaport, 1999; Sosis & Alcorta, 2003). This paper provides a novel approach to understanding these observations by considering how natural selection might have shaped our cognitive processes for cultural learning so as to give salience to certain kinds of displays or actions, and what the implications of such cognitive processes are for cultural evolution. Since my goal is merely to get this approach on the table, where it can compete with alternatives, I aim to provide a *prima facie* case for considering these ideas, and not a set of conclusive tests.

The argument proceeds in three parts. Part I lays out a theory for the evolution of one particular component in the suite of cognitive adaptations that make up the human capacity for cultural learning. The core idea is that, with the evolution of substantial communicative capacities in the human lineage, cultural learners are potentially exploitable by manipulators who can convey one mental representation but actually believe something else, or at least misrepresent their depth of commitment to a particular belief. To address this adaptive challenge, I propose that learners have evolved to attend to *credibility enhancing displays* (CREDS) alongside the verbal expressions of their models (i.e., those individuals from whom people learn). These displays provide the learner with reliable measures of the model's actual degree of commitment to (or belief in) the representations that he has inexpensively expressed symbolically (e.g., verbally). Learners should use such displays in determining *how much* to commit to a particular culturally acquired mental representation such as an ideology, value, belief or preference. After laying this out, I summarize supporting findings from psychology.

Building on this, Part II explores whether such a learning bias could create interlocking sets of beliefs and costly practices that are self-stabilizing. That is, can this adaptive learning bias lead to the emergence of stable combinations of beliefs and costly

practices (displays) in a social group that could not otherwise persist (remain stable)? My formal model reveals the wide-ranging conditions under which costly practices (acting as CREDs) and associated beliefs are self-stabilizing. Such stable cultural evolutionary states are interesting because they show how particular displays or acts, which appear costly to one who does not hold the relevant corresponding belief, can be sustained by cultural evolution. Part III considers the possibility that such an interlocking system could also sustain costly practices that elevate the commitment of group members to beliefs that promote group benefits, larger-scale cooperation and solidarity, and — in particular — favor success in competition with other social groups (or institutions). This competition among stable culturally-evolved states favors social groups that are increasingly constituted by combinations of (a) beliefs that favor in-group cooperation/harmony and out-group competition, and (b) practices (e.g. rituals) that maximize participants' commitment to those beliefs.

To assess the plausibility of this account and compare it with existing approaches based on signaling, I summarize evidence indicating that (1) belief-practice (ritual) combinations are spread by cultural group selection (CGS); (2) participation in costly rituals is associated with prosocial ingroup behavior, because costly rituals *transmit* commitment to group-beneficial beliefs/goals to participants; and (3) institutions requiring costly displays are favored by cultural evolution because costly displays by members transmit higher levels of belief commitment and thereby promote cooperation and success in intergroup or interinstitution competition.

Together these three parts lay out a process, initiated by an evolved learning bias, that connects costly, even extravagant, displays to cooperation and commitment to a group's beliefs and ideology. The more costly the displays are, the potentially deeper the degree of transmitted commitment.

I close by discussing how such processes may illuminate a number of puzzling aspects of religion, including why (1) religions are often associated with prestigious paragons of virtue who make (or made) costly sacrifices; (2) martyrdom is so persuasive; (3) religions and rituals are loaded with sacrifices of various kinds; (4) gods and ancestors want costly acts; and (5) religious leaders often take costly vows, such as those involving poverty and celibacy.

## **2. The evolution of our cultural capacities**

The application of the logic of natural selection to the evolution of social learning has produced an array of novel theoretical insights, hypotheses and empirical findings (for

reviews, see Henrich & McElreath, 2006; Richerson & Boyd, 2005). One central line of inquiry arising from this research program has focused on how selection has shaped our cultural learning processes in order to more effectively acquire ideas, beliefs, values, preferences and practices from others in our social world. The set of related hypotheses about these cognitive-operational details can be partitioned into two categories, those based on *context* (e.g., cues about a model's prestige or success) and those related to mental representations' *content*. Below, I briefly review some work in this area in preparation for laying out the *CRED hypothesis*.

Contextual learning mechanisms use cues that allow learners to more effectively extract and integrate adaptive information from the range of individuals available in the learners' social world (Henrich & McElreath, 2003). One class of cognitive mechanisms, often glossed as *prestige-biased transmission* (Henrich & Gil-White, 2001), proposes that learners use model-based cues to figure out who, among their potential *models*, is most likely to possess adaptive information suitable to the learner's situation (e.g., his/her role in the social group). Theory suggests, and a wide range of empirical findings have shown, that both children and adults preferentially pay attention to and learn from others based on cues of prestige, success, skill, age, ethnicity (marked by dialect, dress, etc.) and sex (Henrich & Henrich, 2007: chapter 2). These effects influence a wide range of representations, including opinions, economic decisions, food preferences, social strategies, beliefs, technological adoptions and dialect. Moreover, these biases appear to operate across domains of expertise, as those with skill or knowledge in one field (e.g., basketball) are granted influence in other arenas (e.g., fashion or politics). Given this, and anticipating what is to come below, a highly prestigious individual motivated by self-interest could express a degree of commitment to a belief or opinion different from her own, which—once adopted by others— could yield benefits to her and costs to the learners.

Evolutionary approaches to culture also provide a rich set of cognitively informed hypotheses regarding how the *content* of representations influence their transmission (Boyd & Richerson, 1985: chapter 5; Sperber, 1996). The general insight is that learners should pay particular attention to and remember representations likely to contain adaptive information. Specifically, learners should be more likely to pay attention to and store representations when these are judged, *ceteris paribus*, more (1) fitness relevant, (2) potentially actionable and (3) plausible or compatible. Regarding the first, natural selection should favor more attention and recall for representational content of greater relevance to fitness, at least in ancestral

environments Often, such content sparks more positive or negative emotional responses, thus adaptively biasing memory storage and recall.

*Potentially actionable* means that the content of a representation leads to inferences that can readily influence subsequent actions, including additional inferences (inferential potential: Boyer, 2001). Representations, for example, in which the causes of unpleasant circumstances (e.g., storms or illnesses) are random with respect to the actions of those afflicted do not lead to useful or helpful inferences or actions, and thus are not easy to maintain. Evolutionarily nonactionable representations need not be stored because they cannot help you even if you do remember them. But, believing — for example — that illnesses are caused by the jealousy of others (e.g., the “evil eye”) can lead to inferences about who might be causing a particular illness and how one can avoid such illnesses in the future.

*The plausibility or compatibility* of a representation involves the learners' expectations about how the world works and, consequently, what is more and less likely to be true or reliable. Some such expectations of plausibility depend heavily on our evolved intuitions, including cognitive processes in such domains as mechanics and biology. For example, representations from modern physics, which involve objects (e.g., electrons) that exist only probabilistically at any point in space, violate intuitive expectations from folk mechanics and thus do not readily transmit. Such compatibility biases can also be culturally acquired, such that the possession of one mental representation biases the acquisition of others. That is, having acquired a particular idea via cultural transmission, a learner may be more likely to acquire another idea or practice, because the two “fit together” in some cognitive or psychological sense.

A variety of hypotheses generated by this approach in domains involving dangerous animals (Barrett, 2007), meat taboos (Fessler, 2003), the disgustingness of urban legends (Heath, Bell, & Sternberg, 2001) and gossip (Mesoudi, Whiten, & Dunbar, 2006) have found empirical support.

With regard to religious concepts, research has demonstrated how the presence of some counterintuitive content in concepts or narratives can bias memory in a manner that would favor such concepts or narratives in cultural evolution (Barrett & Nyhof, 2001; Lisdorf, 2004). Counterintuitive concepts or events violate our core assumptions about the nature of things in the world, usually about intentional beings, animals, inanimate objects or events (expectations from the domains of folk physics, folk psychology and folk biology). An examples of a counterintuitive concept from this literature is “a person who can be in two places at once” (Boyer & Ramble, 2001). The presence of a few counterintuitive concepts in a

narrative, even within a list of otherwise ordinary concepts, improves memory for the entire narrative or list (Norenzayan, Atran, Faulkner, & Schaller, 2006).

From the above perspective, the mnemonic advantages of counterintuitive representations arise from a mixture of plausibility, applicability and fitness relevance. Many religious beliefs, for example, would appear to be *less plausible*, more applicable and more fitness relevant than alternative nonreligious concepts or explanations. Counterintuitive concepts — by definition — make stories or beings seem less plausible (less believable and more difficult to understand) than fully intuitive concepts, which is likely part of the reason why the optimal number of such violations is small. Many counterintuitive representations are also likely to generate emotional responses, like fear or interest (see Fredrickson, 1998), as well as actionable options and additional inferences.

Heretofore, the application of ideas about counterintuitiveness to religion has not sufficiently distinguished (1) mnemonic and transmissibility effects from (2) believability of, or commitment, to the representation. While many religious concepts or narratives do have memory and transmissibility advantages, I propose that they have a believability or commitment disadvantage. Thus, the counterintuitiveness of concepts or stories can help explain the popularity of different folktales, cartoons, superheroes and myths (i.e., other people's religions), but such counterintuitiveness may actually steepen the challenge to explaining the deep commitment to the agents found in religion. Counterintuitive concepts ought to be better remembered— but not committed to or believed in — because, if true, they are important adaptively relevant information. Accepting them as true, however, should require additional learning cues not derived from representational content. Those who want to explain the ubiquity of religious belief based only on representational content need to explain why people do not adopt and commit to other people's gods as soon as they learn about them (represent their content). Below, I argue that CREDs can address this puzzle by providing a mechanism for instilling deep commitment for otherwise difficult-to-accept representations.

### **3. Part I: The emergence of an adaptive challenge**

The evolution of high-fidelity cultural learning, with all its adaptive benefits, increases the potential for exploitation by other members of one's group because cultural learners are open to modifying their behavior, and underlying mental representations, in response to others'. Models can manipulate learners by misrepresenting their (the model's) true underlying representations or commitments. Tom Sawyer famously did this when he manipulated his mates into believing that he (and they) actually *liked* painting a fence. However, prior to the

evolution of sophisticated forms of symbolic communication, of which language is the most relevant example, this potential was minimal since learners had to actually observe their models “in action” to acquire their practices, preferences, beliefs or strategies. For example, in acquiring a particular tool-making practice, learners had to watch their chosen models actually making the tools, and the final product testified—at least in part—to the effectiveness of the observed manufacturing practices. A model who wanted to deceive others about his favored technique could demonstrate a less effective technique in front of learners, but this would be costly in time and effort, and the learner may not be fooled because in the end a less effective tool would result. Similarly, in acquiring food preferences (diet choice), pre-linguistic cultural learners presumably watched what foods others actually consumed, and how this food was located, extracted and prepared. Manipulation in this case would require consuming a nonpreferred food, with all of its associated costs, not to mention the opportunity costs of the search and processing time.

With the evolution of verbal communication, in which mental representations (e.g., beliefs) can transmit at low cost, the opportunities for Machiavellian manipulators to exploit learners would have dramatically increased. These manipulators hold one mental representation but express another (e.g., state it verbally) in an effort to cause others to do things that will increase the manipulators' fitness. For example, a Sawyeresque manipulator might believe “blue mushrooms are mildly toxic” and therefore avoid eating them regularly. But, in an effort to prevent others from eating his preferred grey mushrooms (which are rarer and, he believes, delicious and nutritious), this manipulator might enthusiastically announce that “blue mushrooms are tastier and more nutritious than grey mushrooms.” An unwitting learner who has selected this prestigious Machiavellian as a model might then acquire the mental representation that “blue mushrooms are tasty and nutritious” and start eating relatively more of them, leaving more grey mushrooms for the manipulator (food preferences are heavily influenced by cultural learning). Initially, the learner experiences no ill effects, since it takes years to accumulate clinical levels of the toxin.

Since prestigious individuals can influence the beliefs (and other mental representations) of many learners, a prestigious Machiavellian could dramatically increase his fitness with well-designed culturally transmitted “mind viruses” that strategically alter others' beliefs and preferences. For example, people in many places believe “the wishes of our dead ancestors must be obeyed.” A manipulator might transmit the belief—not held by him—that he is “the mouthpiece for the ancestors, and they will talk through him; their first command is to pay the mouthpiece for his service to the ancestors with one pig from each house.”

I hypothesize that natural selection addressed the emergent problem of Machiavellian manipulators, not by suppressing the use of symbolic communication in cultural learning, but by constructing a kind of cultural immune system. This immune system is designed to assess a potential model's "degree of belief or commitment" to a symbolically communicated belief using the model's displays or actions. Cultural learners should look for displays that are most consistent with the expressed representation(s) and — more importantly — look for actions that would not be performed by a model believing something different from what the model expressed symbolically. Such diagnostic actions are evidence of commitment to the expressed belief. A model, for example, might express the view that donating to charity is important, but not donate when given the opportunity. The action, not donating, should indicate to a learner that while the model may believe in some sense that giving to charity is a good idea, he is probably not deeply committed to it. As we will see, cultural learners under such conditions would simply acquire the practice of talking about how good it is to give to charity, without actually giving. Learners imitate the model, in both actions (talking about how important charitable giving is) and in degree of commitment (little). Conversely, when a model actually gives to charity at a cost to himself, learners more readily acquire both the representation that giving to charity is good and a deeper commitment to or belief in that representation. Cultural learners are using these actions to more accurately assess the models' degree of commitment or beliefs in the expressed representation. Such diagnostic actions are *credibility-enhancing displays (CREDS)*.

CREDS will often *appear* costly to a person holding one particular belief about the world, but seem substantially less costly, neutral or even beneficial to a person holding an alternative belief about the world. In the mushroom example, the act of regularly eating the blue mushrooms would seem costly, and unlikely if the model believed that blue mushrooms were in fact toxic. However, regularly eating the blue mushrooms would not seem costly to a model who believed that blue mushrooms are tasty and nutritious. The action of regularly eating the blue mushrooms is a CRED for the verbal expression of the underlying representation that blue mushrooms are tasty and nonpoisonous because the likelihood of regularly eating such a mushroom if one actually believes they are poisonous is low. In this case, though not all cases, whether the CRED has a net fitness cost depends on the true state of the world.

This approach does not mean that learners ignore verbal statements, or other forms of communication. Such symbolic expressions can be extremely informative in a learner's efforts to replicate the underlying mental representations of a chosen model or models. Since context

and content transmission biases do not disappear in the absence of CREDs, cultural learners will still recall the verbal statements of, for example, prestigious individuals better than the statements of others (Henrich & Gil-White, 2001). The key is that, in the absence of CREDs, learners are not committed to those recalled representations in a manner that propels behavior beyond simply repeating the expression itself.

Finally, since attention to action in this approach evolved to help learners assess their models' underlying degree of belief or commitment (intrinsic motivation), costly actions that are less diagnostic (or nondiagnostic) of a model's degree of underlying commitment because of external threats or pressure to perform those actions will be relatively weaker as CREDs.

### ***3.1. Psychological findings***

The above logic proposes that learners ought to be more likely to acquire culturally transmitted representations, in the form of practices, beliefs, values or strategies, if their models perform acts that are both consistent with the possession of the underlying representation (which is expressed verbally) and inconsistent with alternative representations. Stated another way: if identical models verbally express the same belief, preference or opinion, learners should be — *ceteris paribus* — more likely to learn from models who perform accompanying CREDs. Often, the more costly a model's display would seem to someone who did *not* hold the model's expressed belief, the greater the influence of that model on the learner's subsequent commitment to, or belief in, the expressed representation.

Here I unite findings from four areas of psychology, all of which study cultural learning in one form or another. These programs focus on the transmission of (1) food preferences and consumption, (2) opinions, (3) altruism, and (4) beliefs in intangible entities and nonintuitive concepts. The acquisition of beliefs, attitudes or behaviors in the first three domains has already been shown to be influenced by cultural transmission. The question addressed here is whether learning in these areas *specifically* reveals evidence for the influence of CREDs.

#### ***3.1.1. Food preference and consumption***

Both people's preferences for certain foods and the amount of food they consume are substantially influenced by which foods those around them prefer and how much they eat. In developmental research, findings indicate that learners actually shift their intrinsic food preferences toward those of their models, especially when those models are same-sex, older children (Birch, 1980, 1987; Duncker, 1938). Work with adults demonstrates that models can



influence the quantity consumed (Herman, Roth, & Polivy, 2003; Salvy, Romero, Paluch, & Epstein, 2007).

If food choice is also influenced by CREDs, then learners should be more inclined to eat novel foods when a model is first observed to eat the food himself. As in the mushroom example, consuming something is a CRED for believing it is worthy of eating (or at least nontoxic). Harper and Sanders (1975) report experimental findings in which a female experimenter went to the homes of children (ages 14 to 48 months), spent at least 20 min playing with the child until he or she seemed comfortable, and then presented the child with a novel food. In the baseline treatment, the experimenter merely placed the novel food out (within reach of the child) and declaratively stated “something to eat” to the child. In the CRED treatment, the experimenter said the same thing as she sampled some of the food. In the baseline, only 25% of children tasted the food, while in the CRED treatment 75% sampled (pb.05). This may seem both intuitive and unsurprising, but it represents a manifestation of a tendency for learners to look for displays in models that indicate the model actually believes what she is saying.

### ***3.1.2. Opinion transmission***

Psychologists have long studied both the characteristics of effective “communicators” in the context of opinion change (Tannenbaum, 1956). From my evolutionary perspective, persuasion or opinion change is merely a kind of cultural transmission. When models express something verbally (or in writing), ostensibly their own underlying mental representations, this may cause others to alter their own mental representations in an effort to move closer to the representation inferred from the model's expression. Opinion change research shows that subjects shift their opinion substantially more when the model is more prestigious. This same work also shows evidence of CREDs, although in a more nuanced manner than with food.

Walster, Aronson and Abrahams (1966) had subjects read newspaper articles in which either a high-prestige (famed prosecutor) or a low-prestige (thug) individual expressed opinions about the need for changes in the criminal justice system. Each model called for changes that would run either *for* or *against* their own self-interest. Opinion measures from the subjects show that when models' expressed opinions that promoted their own interests, subjects' opinions shifted toward the model substantially less than when models expressed an opinion contrary to their own (the models') interests. Here, the CRED is the verbal opinion itself. It is credibility enhancing in this context because the dissemination of the expressed opinion, which was given to the mass media, runs against the self-interest of the model. It

seems unlikely that a model would argue for an opinion counter to his self-interest if he actually held an opinion consistent with his self-interest.

The evidence also suggests that the influence of highprestige individuals is damaged more when they advocate for their own interests than when low-prestige individuals advocate for their own interests. When a low-prestige individual advocates for a view that runs counter to his self-interest, his influence exceeds that of a high-prestige individual advocating for a view favoring his self-interest (see also Eagly, Wood, & Chaiken, 1978). As mentioned earlier, these findings suggest that our adaptation for using CREDs has been calibrated to recognize that high-prestige individuals have more incentives to make self-serving claims, since their opinions are more likely to spread.

### ***3.1.3. Cultural transmission of altruism requires costly acts***

Developmental research on the cultural learning of altruism shows that a model's verbal statements (“exhortations” or “preaching”) to make costly charitable donations have little or no impact on learners' donations unless such statements are accompanied by the model actually making costly donations himself. Once the model donates, cultural learning powerfully transmits altruistic behavior or charitable preferences. Actually donating is a CRED because it would be unlikely to be observed if the model held beliefs or preferences about charitable giving substantially different from those he expressed verbally.

In the paradigmatic experimental setup, from which there have been many variations, a child is brought to the experimental area to get acquainted with the experimenter. Then, the child is introduced to a miniature bowling game and shown a range of attractive prizes that can be obtained with tokens won during the bowling game. The subject is also shown the charity jar for “poor children” where they can put some of their winnings, if they want. A model, who could be a young adult or another peer, demonstrates the game by playing 10 or 20 rounds. On winning rounds the model donates (or not, depending on the treatment) to the charity jar. After the demonstration, the model departs and the child is left alone to play the bowling game (Bryan, 1971; Elliot & Vasta, 1970; Grusec, 1971; Presbie & Coiteux, 1971).

Several studies compare the effect and interaction of models who preach generosity or selfishness (“one ought to donate...”) and practice either generous or selfish giving.

Preaching alone usually has little or no effect on giving. Children's behavior seems uninfluenced by preaching when these exhortations are inconsistent with the model's actions (Bryan, Redfield, & Mader, 1971; Bryan & Walbek, 1970a, b; Rice & Grusec, 1975; Rushton, 1975). However, when a model actually donates generously, the subjects donate more

generously. Here, giving away tokens that one could use to exchange for toys is a CRED of one's commitment to the verbal claim that “one ought to donate.”

Verbal expressions are not irrelevant here. They help the learner figure out the underlying details of the model's mental representations — that is, the where, when, who and why of charitable giving. Experimental work shows that exhortations combined with CREDs allow learners to broaden the range of contexts for acquired altruism (Grusec, Saas-Kortsak, & Simutis, 1978). Thus, verbal expressions can be critical to understanding what is learned, but learners seem to “switch off” unless verbal statements about what one ought to do, when and why are accompanied by a CRED.

#### *3.1.4. Counterintuitive concepts*

Recent research suggests a similar need for CREDs in beliefs about intangible entities, such as God or germs (Harris & Koenig, 2006; Harris, Pasquini, Duke, Asscher, & Pons, 2006). This work shows that children only express beliefs in intangible entities that adults' behavior seems to “endorse.” Adults in this subculture pray to God, attend rituals and tell children to pray. Adults also refuse to eat dropped food and force children to wash their hands, while expressing a concern for germs. To the learner, these are CREDs indicating adults actually hold beliefs in God and germs. Meanwhile, entities that do not inspire CREDs in adults, such as mermaids, are not strongly believed in by children. While only suggestive, such findings are consistent with the idea that our capacities for cultural learning may have been shaped to weigh a model's CREDs in adopting and committing to culturally transmitted representations.

(...)

### **5. Part III: Cultural group selection favors interlocked belief–display combinations that increase cooperation**

Part II demonstrated that a genetically evolved reliance on CREDs can, under a wide range of conditions, yield a cultural evolutionary process with multiple stable equilibria. If this were all there were to it, the story would not be very interesting as individuals at equilibria involving costly acts would get lower payoffs than those in groups stabilized at the other equilibrium. However, showing that a reliance on CREDs can stabilize costly practices, opens the door to the possibility that such costs could be directed, in some fashion, to supply

group benefits and increase group competitiveness. There are several ways to think about this. First, the practice ( $x=1$ ) could be a cooperative or prosocial act in itself, and this could increase the success and competitiveness of the group or institution. For example, giving alms to the poor could be a CRED for a belief in Allah and a group beneficial act. Second, the practice might be an act of punishment that penalizes noncooperators (this could stabilize cooperation and similarly benefit the group). There is no first- or second-order free rider problem here, since the costly act is already stabilized by the interlocking effects of the CRED (as modeled in Part II). Third, it is possible that the costly practice in and of itself delivers nothing to the group (scarification or tattooing) but that it elevates and stabilizes a strong commitment to a group ideology ( $\theta=1$ ) that itself favors other group-beneficial contributions related to cooperation in war, self-sacrifice, bravery, etc. Costly ritual sacrifices, for example, may favor the transmission of high degrees of commitment to beliefs in a lovely afterlife. Strong commitments to beliefs in God and an afterlife could permit individuals to charge an enemy, aid the sick during a plague (Stark, 1997) or help build a community member's house after a storm. Social groups with costly acts that generate CREDs for beliefs that promote in-group cooperation and out-group competitiveness can spread more effectively—via competition among cultural groups—than those that do not.

The process of competition among social groups locked in at different stable states is a kind of Cultural Group Selection (CGS). Understanding both the importance and plausibility of CGS requires recognizing the intersection of two different lines of modeling work. First, several models including the one developed in Part II demonstrate various ways in which cultural learning gives rise to multiple stable states, including states that sustain individually costly behavior (cooperation is one type of costly behavior). Two other examples of such models come from (1) Henrich and Boyd (2001), who show how culturally transmitted forms of punishment can stabilize costly norms, and (2) Panchanathan and Boyd (2004), who show how reputation can stabilize costly norms by linking them to behavior in a dyadic helping game. Thus, the above model represents yet another means by which cultural evolution can stabilize costly behaviors, including cooperation. Each of these models reveals a range of stable equilibria involving costly practices that vary in their group payoffs, but no built-in way to determine which equilibrium eventually emerges. That is, cooperative equilibria represent only a tiny fraction of the stable states for costly behaviors, thus none of these models alone can explain the prevalence of prosocial norms or large-scale cooperation.

However, a second line of modeling work on CGS demonstrates that competition among social groups at different culturally evolved stable equilibria provides a plausible

mechanism that can favor the diffusion of cooperative, group-beneficial beliefs, practices and norms (Boyd & Richerson, 1990, 2002; Fehr & Fischbacher, 2003; Henrich, 2006). This kind of CGS, involving competition among *stable* states, suffers none of the problems typically associated with application of genetic group selection to the evolution of altruism (Henrich, 2004).

CGS can occur in several ways. First, the most straightforward form of CGS occurs when social groups — due to superior institutions for cooperation that create technological, military or economic advantages — drive out, eliminate or assimilate groups at alternative equilibria (Soltis, Boyd, & Richerson, 1995). “Institutions” here refers to the integrated sets of beliefs, values and practices that organize social interactions in groups. Second, social groups may compete demographically, with groups at some stable equilibria putting out more culture bearers than other groups or attracting more migrants than groups stuck at other inferior equilibria (Boyd & Richerson, 2009). A third form of CGS is perhaps the most subtle and important. Our evolved adaptations for cultural learning may cause people in groups stuck at less group-beneficial equilibrium to preferentially imitate the beliefs and practices of people from groups at more group-beneficial equilibrium because they show higher payoffs (Boyd & Richerson, 2002). This can cause sets of ideas, beliefs and practices to differentially spread from more successful groups to less successful groups. This can describe how institutions spread from one social group to another, or how institutions compete for membership within a social group.

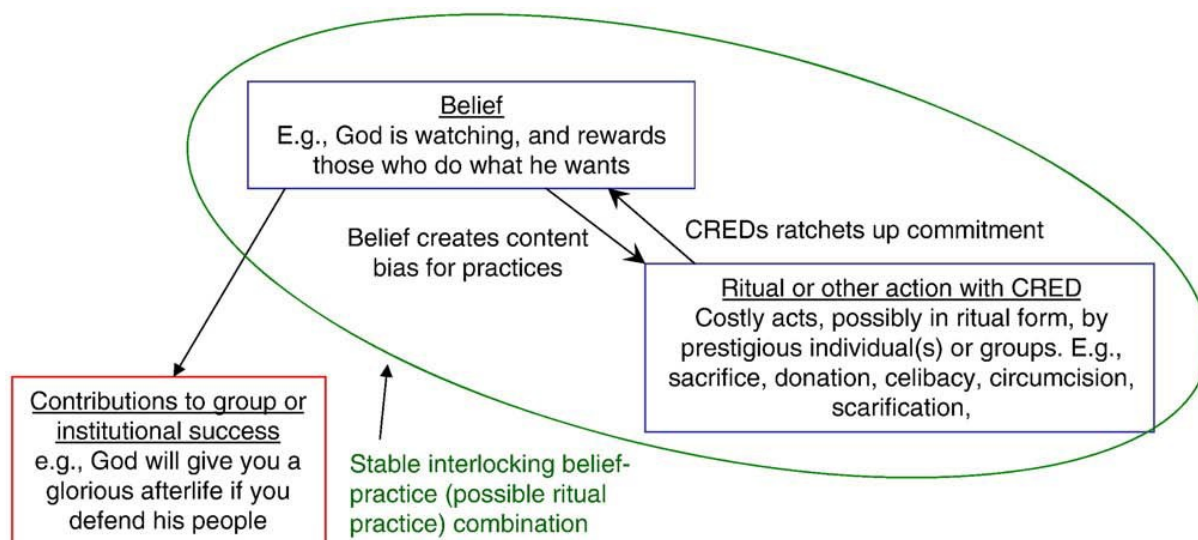
Building on this theoretical foundation, there are now numerous lines of empirical evidence supporting CGS, including data from ethnography (Atran, Medin, Ross, Lynch, Vapnarsky, Ek, et al., 2002; Soltis et al., 1995), archeology (Bettinger & Baumhoff, 1982; Flannery & Marcus, 2000; Spencer & Redmond, 2001; Young & Bettinger, 1992), ethno-history (Kelly, 1985; Sahlins, 1961) and even laboratory experiments (Gurerk, Irlenbusch, & Rockenbach, 2006).

Below, I (1) draw together insights derived above regarding CREDs with existing work on CGS and apply them to the evolution of rituals, and the relationship between rituals, costly acts, cooperation and deep commitment to group ideologies; (2) highlight some *prima facie* empirical findings indicating that packages of rituals, costly acts and group ideologies/religions do spread by CGS; and (3) interpret recent findings concerning rituals, costly acts and cooperation to illustrate their consistency with this approach.

### ***5.1. CGS favors rituals that exploit evolved learning mechanisms***

Since both religious and secular rituals have frequently been associated with costly displays — such as firewalking and scarification — and with the promotion of group solidarity, cooperation and competitiveness in warfare (Atran, 2002; Durkheim, 1995; Sosis & Alcorta, 2003; Sosis & Ruffle, 2003), I apply the above ideas to rituals, thus incorporating rituals into the discussion, and then consider empirical evidence linking rituals, cooperation, beliefs and costly acts. My goal is only to suggest how cultural evolutionary forces, rooted in our evolved cultural learning capacities, may have shaped rituals alongside other forces (Boyer & Lienard, 2006; McCauley & Lawson, 2002; Whitehouse, 2000).

Competition among groups or institutions should favor rituals that more effectively exploit our capacities for cultural learning in order to transmit deeper commitments to ideas, beliefs or values that increase in-group cooperation and solidarity (and perhaps out-group enmity). Groups with rituals that more effectively transmit commitment to groupbeneficial (self-sacrificial) beliefs will — *ceteris paribus* — outcompete groups with less effective ritual–belief combinations, causing these belief–ritual complexes to spread by the various forms of CGS discussed above. Fig. 2 illustrates the process described.



**Fig. 2. Diagram of the key relationships that give rise to the linkage between group beneficial acts (like cooperation), religious beliefs and costly acts, including rituals.**

If rituals are evolving via CGS to more effectively exploit our capacities for social learning, then we can make predictions about the nature of rituals based on our understanding of these evolved mechanisms. Effective rituals should variously make use of (1) prestige-bias transmission (Henrich & Gil-White, 2001), capturing our tendency to weight information coming from prestigious individuals more heavily than from others; (2) conformist

transmission (Henrich & Boyd, 1998), exploiting our tendency to use the frequency of others doing or professing something as a cue in adopting it; (3) folk ethnicity (Gil-White, 2001; Henrich & Henrich, 2007: chapter 9), tapping our tendencies to essentialize, preferentially interact with and differentially learn from those who share our hard-to-fake symbolic markers (dialect, dress, painful tattoos); (4) mimicry, exploiting our tendencies to both use mimicry to improve our reading of others emotions and to assess relative prestige differences; and most importantly, (5) CREDs, exploiting our reliance on diagnostic actions or displays to assess the depth of our models' commitments.

Under such selective pressures, rituals will tend to (1) put key lessons or statements of belief in the mouths of the older, more prestigious and more successful members of the community; (2) involve group professions of belief to cue conformist transmission (e.g., in prayers, chants, group public oaths); (3) make use of costly-to-acquire symbolic markers that distinguish community members from other groups; (4) include music, rhythm and synchrony to elevate solidarity (Wiltermuth & Heath, in press) via mimicry; and (5) showcase practices that only deeply committed believers would engage in, such as practices that allow prestigious members to demonstrate their degree of belief (e.g., snake handling while preaching) or practices that involve several members undergoing harsh, painful or frightening experiences. These characteristics would evolve via CGS to target participants and observers because they more effectively exploit our evolved cognitive capacities for cultural learning to convey deeper commitments. Over time, this would result in ratcheting up people's degree of commitment to some underlying beliefs.

Costly acts, particularly those found in rituals, will be more important for sustaining commitment to religious beliefs than to secular beliefs or ideologies. There are three interrelated reasons for this. First, religious beliefs often involve commitments to counterintuitive agents. Committing deeply to counterintuitive concepts may require CREDs by models because, in and of itself, counterintuitiveness violates content plausibility (Section 2). Acquiring and committing to secular ideologies often do not require accepting and committing to counterintuitive propositions and thus may not face the same uphill battle. Second, once committed to, many counterintuitive concepts — like supernatural agents (ancestors and gods) — cannot easily be falsified by real-world events or experiences in the same way or to the same degree that secular beliefs can. This means that degrees of commitment to secular ideologies will be more subject to real events and outcomes compared to religious ideologies. When religious beliefs can be directly falsified by experience, they tend not to stick around for the same reasons. For example, various groups have come to

believe that faith, or a ritual, can provide protection from bullets. Such beliefs have tended not to endure for long periods, once the shooting starts. Third, religious beliefs, *once deeply committed to*, are likely more powerful than secular beliefs at galvanizing cooperation. Supernatural agents can police (e.g., seeing all, reading minds, etc.) and motivate adherents (e.g., by bringing sickness, death, afterlife, etc.) in ways that secular agents cannot. This combination of elements means that costly acts, particularly those found in rituals, will tend to be associated with sustaining or increasing religious convictions, and any associated group-beneficial behaviors, in a manner not found for secular beliefs.

In signaling terminology (Maynard Smith & Harper, 2003), CREDs began as *cues* inadvertently or incidentally given off by individuals, according to their beliefs, that are used by learners as *indices* (more or less accurate measures) of belief commitment by learners. These indices can become true signals when (1) genetic evolution, (2) cultural evolution or (3) individual decision making favors “transmitters” strategically using these indices to influence others. Here, individuals become active transmitters or signalers as CRED cues evolve into signals. The genetic evolution of our reliance on CREDs (as cues) created an opportunity for cultural evolution to turn these *cues* into *signals* in the form of rituals and ritualized acts that exploit our learning psychology to favor deeper commitments to certain kinds of beliefs, such as those favored by CGS.

## ***5.2. Preliminary lines of evidence***

This approach makes predictions about the relationship between ritual, costly acts, cooperation and group solidarity. The three predictions addressed here ask, (1) Is there any evidence suggesting that these packages of rituals, beliefs and costly acts do spread via CGS? (2) Does ritual attendance indeed increase commitments to group ideologies? and (3) Does requiring costly acts improve a group's relative survival compared to groups demanding fewer costly acts?

### ***5.2.1. Belief-ritual packages spread by CGS***

Ethnographic, ethno-historical and comparative research indicate that belief-ritual packages are spread by CGS. I have only space to mention four studies. In New Guinea, Boyd (2001) describes how a village explicitly decides to imitate the pig-raising package of institutional practices, beliefs and rituals from their most successful and prestigious neighbors. This is prestige-biased CGS. In the East Sepik, Tuzin (1976, 2001) analyzes how the largest village in the region (five times larger than average) sustains harmony, cooperation and



solidarity using a package of costly rituals, ideologies and institutions that was copied from the Abelam, a highly successful and aggressively expanding society. In the New Guinea Highlands, Wiessner and Tumu (1998) describe belief–ritual complexes associated with painful or frightening rites, which promote “identity, welfare and unity,” as spreading by a process of emulating the more successful groups. Such rich ethnography helps us understand the cultural evolution of the observed relationship between warfare and costly rites for males (Sosis, Kress, & Boster, 2007). Increasing warfare means cultural groups with more costly rites galvanize greater cooperation and solidarity among males (more commitment to group ideals), and thus these groups survive, expand and are imitated more frequently by other groups.

### ***5.2.2. Costly rituals will elevate people's degree of belief commitment***

Participation in rituals involving costly acts will elevate people's degree of belief commitment. If the professed beliefs involve group commitment, cooperation toward fellow ingroup members, or the hatred of out-groups, then ritual attendees will trust, identify and cooperate with in-group members more than nonattendees. Demonstrating this, Sosis and Ruffle (2003, 2004) performed behavioral experiments among secular and religious members of Israeli kibbutzim to explore the relationship between ritual participation and cooperation. In these experiments, two anonymous participants from the same kibbutzim were given a monetary sum and a one-shot opportunity to contribute any portion of it to a common pot. Whatever money was contributed to this pot was increased by 50% and split equally between the pair. Pure self-interest favors contributing zero to the pot, so positive contributions are a measure of increasing cooperativeness towards the other player. Consistent with the above prediction, their results show that greater attendance at public rituals predicts higher contributions in the religious kibbutzim (controlling for a variety of other factors).

These findings also illustrate the expected link between ideological commitment, ritual and in-group favoritism. Sosis and Ruffle (2003, 2004) also used treatments in which participants knowingly interacted with either another anonymous kibbutzim member or another Israeli in general. High ritual attenders in religious kibbutzim contributed substantially more to their fellow kibbutzim members compared to nonmembers. Members of secular kibbutzim treated fellow members in the same way as other nonmember Israelis. This suggests that ritual attendance is associated with in-group favoritism.

Work by Ginges, Hansen and Norenzayan (2007) affirms this link between ritual participation and commitment for both in-group cooperation and out-group aggression. Both

survey and experimental findings from Palestinians and Jewish Israelis show that ritual participation predicts more support for suicide bomber attacks against outgroups independent of religious devotion (as measured by prayer) and a wide range of other factors. Similarly, using representative samples of Indonesian Muslims, Mexican Catholics, British Protestants, Russian Orthodox, Jewish Israelis and Indian Hindus, these researchers also showed that greater ritual attendance, independent of a person's prayer frequency and other factors, predicts both declaring a willingness to die for one's god or gods, and that other religions are responsible for much of the world's troubles.

***5.2.3. Groups that require more costly acts (CREDS) galvanize greater solidarity and cooperation because these displays effectively transmit belief commitment***

In their study of utopian communities, Sosis and Bressler (2003) assembled data on longevity, group size and costly requirements (e.g., rituals, taboos, etc.) for 83 religious and secular utopia movements in the 19th century. Costly requirements included restrictions on food, sex, material possessions, marriage and parenting rights, among other things. As predicted, the number of costly requirements strongly predicts the longevity of religious communes, though this effect does not emerge for secular communes. The authors also explored some contextual data suggesting that the driving factors for longevity were indeed related to solidarity, group commitment, and cooperation. They report that some commune members explicitly recognized that costly requirements *increased* the belief commitment and solidarity of members.

These findings, in addition to illustrating the relationship between costly displays and group success (as measured by group survival), provide a stark example of CGS in action. These communes varied in their number of costly requirements and the data show that those with the most costly requirements survived longer. Over time, the differential survival of some groups ratcheted up the mean number of costly requirements per commune by selecting out those groups unable to sustain solidarity and cooperation. It is difficult to interpret this as anything but a prime example of CGS influencing cultural evolution.

The authors, however, use these data to support a ritual signaling hypothesis, arguing that signaling predicts that those individuals who are committed to the group's ideals will be able to perform the costly requirements more cheaply than nonbelievers (the less committed) and thereby sustain more cooperation by suppressing free riders. There are several problems with this interpretation.

(1) These findings are derived from a pattern created by a historical process in which groups with more costly requirements survived longer than groups with fewer requirements. It is not clear how their signaling hypothesis actually predicts such group dynamics or historical processes. The signaling models cited by these authors are not — at this point — imbedded in a cultural evolutionary framework capable of yielding historical (nongenetic) dynamics occurring over decades.

(2) This signaling approach does not predict that costly requirements will ratchet up commitment to beliefs or ideologies. The authors, however, report that commune members believed costly requirements did increase group commitment.

(3) In contrast to most signaling applications, it is not clear why (in a fitness sense) it is more costly for nonbelievers to perform the costly requirements than for believers (more committed people). Holding a particular mental representation is not obviously parallel to possessing a physical attribute, like size, strength or stamina (as in the nonhuman literature on signaling). In nonhuman cases of signaling, it is often clear why creating a certain kind of signal is more costly for some individuals than others. Smaller animals, for example, cannot just “get big” for signaling purposes. But a human could always acquire a mental representation, if holding that representation will lead to higher fitness. Approaching this requires a theory of *belief acceptance* (i.e., a theory of cultural transmission) to explain where these ideologies come from, why people are committed to them and why humans (and not other animals) have ideologies, which can be committed to, in the first place.

(4) Lacking a theory of cultural learning, it is unclear why members do not just invent more costly requirements and thus obtain more group benefits. If this is — in fact — because the requirements are culturally transmitted or that multiple signaling equilibria exist (which is likely), then one is back to needing to embed signaling in a theory of cultural evolution.

(5) A broader problem with *ritual* signaling theory is the lack of any formal evolutionary model showing how this can solve the n-person prisoner's dilemma. Existing modeling efforts suggest that it cannot (McElreath & Boyd, 2007). And, since both signaling models (Bergstrom, Szamado, & Lachmann, 2002; Lachmann & Bergstrom, 2004; Lachmann, Szamado, & Bergstrom, 2001) and n-person models of cooperation (Boyd, 1988; Boyd & Richerson, 1992) have repeatedly yielded results (including multiple stable equilibria) that contradicted previous verbal theorizing, modeling this seems crucial.

Nevertheless, both my hypothesis and a version of the above signaling hypothesis may be important to explain the intersection of rituals, belief and cooperation. Individuals likely need to both calibrate their degree of commitment during cultural learning and assess the

degree to which their fellow group members are also committed and willing to cooperate. Norm adherence and cooperation will be maximized when (a) individuals' commitments are deepest and (b) everyone believes everyone else is also deeply committed. The problem with much existing work is that it fails to address how people get deeply committed to certain beliefs—such as those involving counterintuitive agents—in the first place.

## **6. Discussion: implications for understanding religion**

These ideas have numerous implications for understanding the cultural evolution of various religious phenomena. Here I will sketch how some of these processes may have shaped certain aspects of religion.

### ***6.1. Why are religions often associated with prestigious paragons of virtue who make (or made) costly sacrifices?***

Applying the above reasoning to this question begins by considering our evolved psychology for cultural learning. In learning how to behave and what to believe, learners give weight to both prestige and CREDs, among other things. Thus, successful cultural forms, especially those involving deep commitment to counterintuitive beliefs, will tend to begin with and be sustained by prestigious individuals performing CREDs. Cues of prestige influence who people pay attention to for learning, while CREDs convince them that the prestigious model really believes (is committed to) his or her professed beliefs. The “virtuousness” arises from these prestigious individuals' role as models. CGS will favor, over long swaths of historical time, religions with role models who effectively transmit beliefs and practices that strengthen in-group cooperation, promote intra-group harmony and increase competitiveness against out-groups.

### ***6.2. Why martyrdom is powerful***

As a corollary of the above, martyrs — be they suicide bombers or saints — can provide powerful CREDs to learners regarding their degree of commitment. Anthropologists have considered suicide bombing as a costly signal of group commitment (Atran, 2003; Sosis & Alcorta, in press), which it may be. However, this approach fails to explain the impact of these costly actions on learners' beliefs. The most important thing about martyrdom is not that everyone now knows the martyr is a committed member of the group (signaling), but that

observing this CRED increases the commitment of the (still living) learners — i.e., some moderates become radicals in the process.

Two cases help illustrate this point. First, early Christian martyrs, executed in public events, are believed by many (Stark, 1997), including observers at the time, to have substantially fueled the spread of early Christianity. Ignatius, Bishop of Antioch, after being condemned to be ripped apart by wild beasts in a Roman amphitheatre exulted in his opportunity to “imitate the passion of my God!” He then wrote letters to Christian communities along the road to Rome, who might attempt a rescue, pleading with them to allow him to go and die. A Platonist philosopher, Justin, explains that he was convinced of the divinity of Jesus and converted to Christianity, after personally witnessing the commitment demonstrated by the torture and death of some martyrs. Justin was later martyred himself (Pagels, 1989). Second, back in his hometown of Zarqa, Jordan, the death of the locally prestigious Palestinian Abu Musab al-Zarqawi at the hands of the American military ignited an epidemic of young male volunteers flowing into Iraq for martyrdom, often to die as suicide bombers. This reasoning explains why the oppression of religious minorities, or other ideologically committed groups, may actually energize the spread of these groups. Governmentdirected crackdowns, involving torture and execution, provide the faithful with opportunities for CREDs. Interested members with low commitment might not otherwise have the opportunity to observe a potent CRED from a prestigious leader, such as seeing them crucified, stoned, beheaded, eaten by wild cats, etc. Making these displays public is a really bad idea if you want to stamp out a religious movement.

### ***6.3. Why religious leaders take vows involving celibacy, fasting and poverty***

Beliefs of any kind, but especially the counterintuitive ones found in religions, will best proliferate when expressed by prestigious individuals performing CREDs. Avoiding sex, food and wealth can all act as CREDs of deep belief commitment. Individuals sticking to such vows (or appearing to) increase their potency as transmitters of the faith. Religions that prescribe the avoidance of food, sex and wealth among leaders, while effectively dealing with the obvious defection problem, will tend to proliferate because they have made their leaders better transmitters of commitment.

### ***6.4. Why are religious ideologies interlaced with ritual sacrifices of various kinds?***

Sacrifices may involve the killing of a person or nonhuman animal, or giving of money, at a public event. Such acts may arise for many reasons, but in some cases such

sacrifices are CREDs that help transmit deep commitments to participants and observers. Religions with such rituals will tend to survive and grow because these rituals instill deeper commitment than would otherwise be possible.

From this perspective, costly acts by high status leaders demonstrate — and thereby more effectively culturally transmit — the leader's professed beliefs. Atran (2002), for example, relates a scene described in Mayan glyphs in which a new ruler rises to power in Palenque. In the accession ritual, the new ruler first sacrifices a captive, by personally plunging a knife into the victim's chest, and then pierces his own penis three times, in order to pull through long strands of bark, which he then watches turn red. Such actions are likely to provide a CRED for some portion of the audience. Observing the leader's display may ratchet up the commitment to the leader's professed beliefs of his counselors, senior members of the government, the military, and perhaps even the populace.

#### ***6.5. Why counterintuitive agents (e.g., gods or ancestors) want costly acts***

The above logic proposes that religions will culturally evolve to possess counterintuitive agents, like gods, that demand or at least want CREDs. The reason for this is straightforward. Counterintuitive agents that demand CREDs can cause the transmission of deeper commitments to that agent and further spread belief in that agent. The more counterintuitive the agent, the more CREDs will be required to sustain commitment.

#### ***6.6. Why Mickey Mouse is not a god, and why people do not believe in other people's gods***

The prevailing view in evolutionary-cognitive circles is that religious representations spread because of their content (Boyer, 2001). However, many of the counterintuitive denizens of cartoons and folktales would often seem to have the “right” content to become faiths, yet no one seems ready to commit deeply to such representations. Similarly, adherents to one faith often have substantial knowledge of other faith's supernatural agents, yet they are not persuaded to commit to those gods merely by virtue of holding the same representational content as believers. This presents a problem for approaches based exclusively on content, especially when the content biases arise from innate aspects of human cognition. From the theory summarized earlier, we distinguish the effects of content on *memory* from its effects on *commitment* to, or belief in, the representation in question. Particular content may increase a representation's memorability and transmissibility, but not influence a learner's degree of commitment to that representation. To turn Mickey Mouse into God, we need CREDs,

especially by prestigious individuals or large groups (conformist transmission), and preferably by models sharing the learners' sex and ethnicity (two other evolved biases). From the perspective of a learner, the difference between Mickey and Yahweh, or Yahweh and Zeus, is that learners observe members of *their* social group, including their chosen models, performing CREDs. This makes religious commitment a cognitive, social and cultural evolutionary phenomenon.

## 7. Conclusion

I began by hypothesizing that, over the course of human evolution, cultural learners faced an adaptive challenge created by our increasing capacities for symbolic (cheap) cultural transmission. To meet this challenge natural selection favored a reliance on CREDs in determining how much to commit to, or believe in, a particular representation. Learners evolved to look for displays (often actions) that indicate a model's degree of commitment to, or belief in, verbally expressed representations. These CREDs are actions that (a) are consistent with a model's professed beliefs, and (b) a model would be unlikely to perform if he believed something different from what he expressed symbolically.

Building on this, I examined the implications of this evolved bias for cultural evolution by constructing a simple formal model. The model reveals a wide range of conditions under which this reliance on CREDs can create multiple stable states, with one of these involving an interlocking combination of a costly practice and a belief. Such situations can arise when (1) particular practices influence the transmissibility of certain belief adoptions (CREDs), (2) committing to a belief favors some practices over others (compatibility content bias) and (3) learners tend to copy more successful people (prestige-bias cultural learning).

The presence of multiple stable equilibria involving a costly practice sets up the conditions for Cultural Group Selection. Some stable practices may be only individually costly while others may also contribute benefits to the social group. Social groups that have stabilized on costly practice–belief combinations that deliver group benefits, in the form of cooperation, solidarity and group success, can spread at the expense of social groups at alternative equilibria. This leaves open the possibility that particular groups may get stuck at cultural equilibria involving interlocking belief–practice combination that are purely costly. Over the long haul of culture history, CGS will ensure these groups do not spread, though they may endure for long periods (Edgerton, 1992).

Overall, this approach suggests that the frequently observed connection between costly actions and rituals with larger-scale cooperation, solidarity and success in intergroup competition may be an emergent product of the interaction between an evolved cognitive adaptation for avoiding exploitation during social learning and larger-scale processes of cultural evolution.