2015-04-20

Are Smiles Reliable Due to Costs of Social Punishment?

Daniel Forster
University of Miami, dforster@psy.miami.edu

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

UNIVERSITY OF MIAMI

ARE SMILES RELIABLE DUE TO COSTS OF SOCIAL PUNISHMENT?

By

Daniel E. Forster

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

May 2015

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

ARE SMILES RELIABLE DUE TO COSTS OF SOCIAL PUNISHMENT?

Daniel E. Forster

Approved:

_____
Michael E. McCullough, Ph.D.
Professor of Psychology

_____
Debra L. Lieberman, Ph.D.
Associate Professor of Psychology

_____
William A. Searcy, Ph.D.
Professor of Biology

_____
M. Brian Blake, Ph.D.
Dean of the Graduate School

FORSTER, DANIEL E.   (M.S., Psychology)
<u>Are Smiles Reliable Due to Costs of</u>   (May 2015)
<u>Social Punishment?</u>

Abstract of a thesis at the University of Miami.

Thesis supervised by Professor Michael E. McCullough.
No. of pages in text. (40)

Evidence suggests that smiles can function as signals of cooperative intent—by producing a smile, the smiler can expect to receive benefits from perceivers and perceivers can expect a return from smilers. However, this type of signal is seemingly susceptible to the evolution of cheats who smile in the absence of cooperative intent, thereby receiving the benefits of smiling without paying the costs of cooperation. If smiles were to maintain reliability over evolutionary time, some mechanism(s) must have prevented the evolution of cheats. Researchers have not yet developed a paradigm to directly test *why* smiles might have maintained reliability over evolutionary time. This experiment was the first to assess whether smiles might have maintained reliability due to receiver-dependent costs associated with smiling without cooperative intent. 385 participants played a Trust Game with a confederate who was either smiling or not smiling, and who then behaved either fairly or unfairly. Subsequently, participants provided self-report levels of anger and happiness. Finally, participants had an opportunity to punish, reward, or do nothing to the confederate. Results showed that those treated unfairly by a smiler did not become angrier toward, nor did they inflict more costs on, the confederate than did those treated unfairly by a non-smiler. There are some methodological concerns worth addressing before dismissing the hypothesis; most notably, it needs to be addressed why participants in this study were not more likely to trust smilers than non-smilers, as has been shown in prior research.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Smiles have been the subject of scientific research since the late 18th century, when physiologists began to study the neural and muscular systems underlying facial displays (Duchenne, 1862). Darwin (1872), in *The Expression of the Emotions in Man and Animals*, drew on the work of Duchenne and others to theorize about the evolved functions of smiling and other facial expressions, thereby inaugurating the functional approach to the study of human expressions of emotion. Darwin theorized that humans' species-typical facial displays of emotions evolved for expressive functions—conveying their bearers' emotional states—and that those displays, along with receivers' abilities to understand their fixed meaning, are evolved instincts possessed by all humans. Nearly a century later, Ekman, Sorenson, and Friesen (1969) demonstrated that facial expressions of emotion were in fact displayed in identical patterns, and recognized as conveying the same information across cultures, supporting Darwin's theory of emotion expression.

The view that particular facial expressions are species-typical displays that function to reveal particular emotional states (termed the Emotions View of facial expression; Fridlund, 1997) appears to dominate the scholarly literature, although it is not universally accepted (e.g., see Jack, Garrod, Yu, Caldara, & Schyns, 2012), and is not without its critics (e.g., Fridlund, 1991, 1997; Hinde, 1985). In what Fridlund (1997) has termed the Behavioral Ecology View of facial expression, he calls into question the theoretical claims made by those advocating the Emotions View. Specifically, Fridlund observed that postulating a system that functions to display a representation of its bearer's internal state specifies no selection pressure that favors the evolution of such a system by natural selection. In other words, the Emotions View has not clearly articulated

the advantage that might come from displaying an accurate representation of one's

internal state. More strongly, to the extent that displaying one's inner state provides

receivers with information that can be put to strategic advantage over the individuals

making the displays (e.g., in the context of competition over resources, mating, or some

other conflict of interest), natural selection would strongly penalize displayers, and a

system for displaying one's inner state would be unlikely to evolve. Though it might still

be plausible, even though it is not necessary, that some or all facial displays are

representative of an individual's internal states, this hypothesis is, in its best form, an

incomplete explanation for the evolution of the displays. Therefore, it is likely that facial

displays convey information, or have some effect, beyond merely reflecting their bearers'

inner states.

In contrast to the Emotions View, the Behavioral Ecology View of facial

expressions situates facial displays in the category of *signals*, which can evolve by

natural selection because of their ability to manipulate observers' behavior in ways that

are favorable to the signaler (Dawkins, 1999; Krebs & Dawkins, 1984; Maynard Smith &

Harper, 2003; Searcy & Nowicki, 2005). On this view, if smiles do not confer any

advantage on the individual who smiles, then natural selection would not favor the use of

smiles and the expression would be unlikely to evolve. This latter view grants the

possibility that smiles evolved initially as reflexive actions coinciding with emotions,

thereby making smiles candidates for cues (i.e., properties of individuals that convey

information that can, in turn, be used by receivers to predict the cue-emitting individual's

future behavior, but which did not evolve for that purpose; Scott-Phillips, Blythe,

Gardner, & West, 2012). Under this scenario, the availability of smiles as potential

sources of strategic information about the cue-emitter's internal state would create a selection pressure for the evolution of perceptual mechanisms that enabled consumers of those cues to strategically alter their behavior in response to them (Guilford & Dawkins, 1991; Scott-Phillips et al., 2012). The evolution of such a smile-using perceptual system would, in turn, create a selection pressure for the modification of smile control so that smilers could produce the signal (unconsciously or not) when receivers' evolved responses to them were beneficial to the individual who is smiling. Over evolutionary time, this arms race between the evolution of perceptual systems for processing smiles combined with motivational systems acting strategically upon the social information they contain (for receivers) and the evolution of mechanisms for increasingly finer strategic control over the production of the smiles themselves (for signalers), would be expected to lead to ever-finer adaptive responses to signals (so-called "mind-reading") and ever better use of smiles to influence smile-consumer behavior (so-called "manipulation"; Dawkins, 1999; Krebs & Dawkins, 1984).

The Behavioral Ecology View of facial expressions does not imply that facial displays lack corresponding internal states, nor does this view imply that these internal states are unimportant when considering the utility of the displays; rather, this view merely suggests that the expression of the internal state is an incomplete explanation for the evolved function of the display (Fridlund, 1997). Instead, from a view grounded in the theory of natural selection, the function of a signal is better understood in terms of its effects on the behavior of the recipient and the second-order effects of the recipient's changed behavior on the signaler (Fridlund, 1997; Krebs & Dawkins, 1984). For this reason, when attempting to understand the function of smiles, it may be fruitful to

consider the effects of smiles on the behavior of those who observe them, and how their altered behavior might benefit the individuals who display the smiles.

Smiles as Cooperative Signals

It is possible that smiles, and any variants of what we call smiles (e.g., "Duchenne" smiles, smirks, grins, etc.), have different effects in different contexts (Fridlund, 1997). Researchers have found, for instance, that smiles can convey amusement, motivation to appease an adversary, and dominance, though not all at once, but rather at different times and in different contexts (Hess, Beaupré, & Cheung, 2002). The context-dependence of the information that particular smiles might convey in different social situations underlines the importance of specifying a clear social context in which one might be trying to find the function of a particular signal (Fridlund, 1997). Several studies have found that ratings of smiling faces, compared to the same faces when not smiling, are greater on measures of perceived intelligence, competence, generosity, attractiveness, kindness, trustworthiness, and honesty (Gueguen & De Gail, 2003; Mehu, Little, & Dunbar, 2007; Otta, Abrosio, & Hoshino, 1996). This constellation of perceptual responses to smiles suggests one route by which signalers might benefit from smiling at unrelated conspecifics: By causing smile-consumers to perceive the smiler as cooperatively disposed, smilers might convince smile-consumers that it is in their own interest to confer benefits on the smiler (e.g., so as to induce the smiler, in turn, to produce benefits for the smile-consumer). On this view, a smile is (among the many other things it might be) a signal whose function is to communicate one's readiness to engage in cooperation for mutual benefit with a previously unacquainted interaction partner (West, El Mouden, & Gardner, 2011; West, Griffin, & Gardner, 2007b).

If this functional account is correct, then smiles should be expected to elicit cooperation. In fact, studies have shown that smiling occurs more frequently in dyads who are sharing than in dyads who are not sharing (Mehu, Grammer, & Dunbar, 2007), and that people who smile are trusted more in cooperative dilemmas (Centorrino, Djemai, Hopfensitz, Milinski, & Seabright, 2015; Krumhuber et al., 2007) and receive more help (Gueguen & De Gail, 2003) than do people who do not smile. Additionally, people who smile tend to be more trustworthy (Centorrino et al., 2015) and more cooperative (Reed, Zeglen, & Schmidt, 2012) in cooperative dilemmas than people who do not smile. Taken together, these findings are consistent with the possibility that smiles indeed function as signals of cooperative intent in the context of social exchange: Not only do smile-consumers perceive smiles as cues to the smiler's cooperative disposition, but they also do appear to be correlated with the propensity to cooperate.

The Problem of Cooperation and the Problem of Smile Evolution

To understand exactly why selection might favor the signaling and detection of a cooperative disposition, it is useful to consider the evolution of cooperation. Given the reasonable evidence that cooperation has played a crucial role in human evolution, natural selection has likely favored individuals who successfully cooperated with kin and non-kin to acquire, pool, and defend resources that are difficult to obtain (Delton & Robertson, 2012; Fuentes, Wyczalkowski, & MacKinnon, 2010). Despite these likely advantages, non-discriminate cooperative systems are vulnerable to invasion by free riders (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012; Price, 2006; Price, Cosmides, & Tooby, 2002). The reasoning is as follows: as cooperative dispositions begin to evolve, natural selection would favor individuals who readily consumed the

benefits associated with cooperation without paying the costs required to produce those benefits (i.e., the free-rider problem; Price et al., 2002; see also Maynard Smith, 1982). As one solution to the so-called problem of cooperation in humans (among several others; West, Griffin, & Gardner, 2007a), natural selection appears to favor discriminating cooperation strategies—for example, strategies that cause individuals to cooperate with individuals who were cooperative in prior interactions and to abstain from cooperating with individuals who were not cooperative in prior interactions (e.g., reciprocity, tit-for-tat strategy; Axelrod & Hamilton, 1981; Delton & Robertson, 2012; Trivers, 1971).

However, even with such a conditional cooperative strategy in place, individuals can still suffer from cooperating indiscriminately on initial interactions. This vulnerability might create selection pressures for individuals to reliably detect and avoid cheaters (as well as to reliably detect and interact with cooperators) *even before any interaction has taken place* (Price, 2006). In keeping with this proposition, research has revealed evidence that human social cognition is adaptively structured to facilitate the recognition of both cheaters (Cosmides, 1989) and cooperators (Brown & Moore, 2000) in the context of social exchange.

Here, it is tempting to conclude summarily that smiles are likely to be signals of cooperative intent, but if this were the case, natural selection would favor cheaters who were also able to display the smiles that caused smile-perceivers to confer benefits upon the smiler. As natural selection favored the evolution of such "cheater-smilers," smiles would lose their informational content: they would cease to be a reliable indicator of cooperative dispositions. With this loss of useful information, natural selection would

favor mutants that caused individuals to ignore smiles rather than to attend to them. Thus, the claim that smiles might be a signal of cooperative intent raises a potentially more interesting evolutionary question: If the claim is true, what is the evolutionary mechanism that has maintained their reliability through hominid evolution? One possibility that has been discussed in the literature is that the reliability of smiles has been preserved over evolutionary time by virtue of their intrinsic costs.

Smiles as Handicaps?

Signaling theorists have shown that signals can maintain reliability through costs imposed on the signaler (Maynard Smith & Harper, 2003; Searcy & Nowicki, 2005). Signals that evolve in this fashion are called handicaps (Grafen, 1990; Zahavi, 1975). For handicaps to evolve by natural selection, the benefits of signaling must outweigh the costs for higher-quality signalers, whereas the costs of signaling must outweigh the benefits for lower-quality signalers (Maynard Smith & Harper, 2003; Scott-Phillips, 2008; Searcy & Nowicki, 2005). The iconic example of a handicap is the peacock's tail (Scott-Phillips, 2008). The peacock's tail epitomizes three types of costs: development, maintenance, and production (Searcy & Nowicki, 2005). Theorists have proposed that the costs associated with *developing* the tail must come at the cost of not developing other vital organs that could be used for survival (Grafen, 1990; Searcy & Nowicki, 2005; Zahavi, 1975). Individuals can only pay the costs associated with *maintaining* the tail by having access to ample metabolic resources to prevent the tail from withering away. Additionally, only highly agile individuals can *maintain* efficient maneuverability to escape predators while carrying the burden of a massive tail. Finally, individuals can only pay the costs associated with *production* by being capable of sustaining the display for an

effective amount of time while also being cautious of potential predators that might be drawn in by the allure of the vibrant display. The costs of building, bearing, and displaying the tail at the expense of survivorship are what maintain the signal as an honest indicator of a peacock's quality. These costs, of course, are repaid by virtue of the fact that peafowl use information about tail quality to guide their mate choice, thereby yielding increased reproductive success for males with high-quality tails.

Boone and Buck (2003) argued that the reliability of facial displays could also have been maintained over evolutionary time through handicap-related costs. They claimed that emotional expressiveness (i.e., one's propensity to make facial displays) constitutes a signal of trustworthiness in the context of social exchange inasmuch as expressive individuals are subject to the costs associated with exposing their true intentions. In other words, a person who is highly expressive is signaling trustworthiness and this signal can be trusted because if expressive individuals were not trustworthy, it would show in their facial expression (called "leakage"; Ekman & Friesen, 1969). This begs the question as to why the facial displays themselves are considered to be honest. Their explanation is that facial displays are inherently honest due to the fact that most people have difficulties producing them when asked to (Boone & Buck, 2003). Similar reasoning has been used by other researchers who argue that facial expressions are reliable due to their relative difficulty to be produced voluntarily (Brown & Moore, 2000, 2002; Brown, Palameta, & Moore, 2003; Mehu, Grammer, et al., 2007; Mehu, Little, et al., 2007; Reed et al., 2012).

This explanation, however, appears to equate involuntariness with honesty—a misleading (if seductive) idea. Consider, for example, cases of Batesian mimicry in

butterflies (Jeffords, Sternburg, & Waldbauer, 1979), whereby non-toxic butterflies of one species share the same markings as toxic butterflies of a separate species. Both the toxic and non-toxic butterflies share the advantage of not being eaten by potential predators that recognize the signal. The production of these markings in non-toxic butterflies is involuntary (i.e., the non-toxic butterflies do not "intend" to look toxic), but they are at the same time dishonest (i.e., the display conveys information suggesting the bearer is toxic when in fact the bearer is not). This example illustrates that involuntariness of a signal's display is neither necessary nor sufficient to render a signal honest.

The second shortcoming of Boone and Buck's (2003) account is that it does not specify a causal pathway by which genes responsible for cost-based honesty in facial displays would have been favored by natural selection, which is a critical requirement for a plausible account for behavior based on natural selection (Scott-Phillips, Dickins, & West, 2011). In particular, what is needed for cost-based honest signaling to be favored by natural selection is a causal pathway by which the costs (relative to benefits) of facial displays are lower for higher-quality signalers than for lower-quality signalers (Maynard Smith & Harper, 2003; Searcy & Nowicki, 2005)—as is the case, for example, with the peacock's tail.

A third shortcoming with cost-based explanations for smiles is that it is difficult to identify any intrinsic costs associated with them at all. For starters, smiles do not appear to have costs associated with development or maintenance. With regard to developmental costs, the muscles used to form smiles do not appear to be produced at a great expense to other, more vital tissues, nor are these muscles used specifically *for*

smiles (suggesting utility for some other, potentially important, function that is independent of smiling). With regard to maintenance costs, smiles are produced sporadically and for short periods of time, so it does not seem reasonable to think that humans have to maintain a feature that is not a permanent phenotype. With regard to production costs, the sporadic and ephemeral nature of smiles likewise seems to imply that the metabolic resources required to produce a smile are marginal at most (though this question has evidently never been researched). Granted, Centorrino, Djemai, Hopfensitz, Milinski, and Seabright (2010) have proposed that smiles are possibly reliable because, via the contraction of muscles around the eyes, smiles have the cost of reducing an individual's visual field during production. However, even if one were to concede that smiles substantially reduce one's visual field (which seems dubious at best), there is little reason to suspect that squinting during a smile would carry a substantial cost for a dishonest smiler any more than squinting in reaction to a sunny day would carry a substantial cost for a person without sunglasses.

Smiles as Conventional Signals

Instead of searching for costs that maintain the reliability of smiles, it might be more profitable to conceptualize smiles as conventional signals, which can be thought of as signals for which the information conveyed is arbitrarily related to the signal (Guilford & Dawkins, 1995). West et al. (2011) speculated in passing that smiles might operate as signals between individuals whose interests are aligned, rather than signals of cooperative intent in the context of social exchange, where interests are not always aligned (Delton et al., 2012; Price, 2006; Price et al., 2002). Though it is possible for conventional signals to maintain their reliability in a signaling system between individuals with shared interests

(Searcy & Nowicki, 2005), the finding that smiles reliably indicate a cooperative disposition in the context of social exchange would suggest that smiles function as signals in a context in which interests diverge, and in which signalers therefore have an incentive to make dishonest displays.

There are, however, routes by which conventional signals can maintain their reliability even in contexts in which interactants' interests diverge: by imposing greater costs on dishonest signalers than on honest signalers (called "deterrence" costs; Scott-Phillips, 2008; and "receiver-dependent" costs; Searcy & Nowicki, 2005). There are two types of receiver-dependent costs that are of interest: (a) costs that accrue over repeated interactions because dishonest signalers are remembered and not trusted in successive interactions; and (b) costs resulting from social punishment because dishonest signalers are subject to aggression or other sanctions from receivers. Receiver-dependent costs, which low-quality signalers incur at higher rates than do high-quality signalers, can be viewed in contrast to handicaps, where high-quality signalers incur greater costs than do low-quality signalers (Scott-Phillips, 2008).

To clearly distinguish between handicaps and receiver-dependent costs, it might be useful to re-examine the peacock's tail (Grafen, 1990; Scott-Phillips, 2008; Zahavi, 1975). Peacocks of high quality (e.g., those in good condition, or with low parasite or mutation loads) can afford to pay the extra costs associated with the display and therefore invest in the display, whereas peacocks of low quality cannot afford to pay the extra costs and therefore do not invest in the display: For handicaps, signaling evolves through the reduced reproductive success of those who cannot pay the costs required to produce the signal. In contrast, with a conventional signal of cooperative intent, a non-cooperator

who is signaling as a cooperator might pay costs associated with punishment and/or having a bad reputation (which would need to outweigh any obtained benefits), whereas cooperators who signal as cooperators will, effectively, pay no cost for producing the display. In the case of conventional signals, signaling evolves through the reduced reproductive success of those who pay costs of dishonest signaling—that is, those who receive retaliatory aggression or other disincentives as a result of their dishonest signaling.

Several empirical studies with non-human animals have found that conventional signals can indeed be more costly for dishonest displayers than for honest displayers (Anderson, Searcy, Hughes, & Nowicki, 2012; Tibbetts & Izzo, 2010). Tibbetts and Izzo (2010), for instance, examined whether the facial patterns that display dominance in paper wasps (*polistes dominulus*) maintain their reliability as signals of dominance through deterrence costs—more specifically, social punishment. To do so, they (a) artificially enhanced subordinate wasps' facial patterns to make them appear more aggressive (the resulting individuals looked dominant, but behaved like subordinates); (b) artificially enhanced other subordinate wasps' aggression via hormone injection (the resulting individuals looked subordinate, but behaved like dominants); (c) artificially enhanced other subordinate wasps' features on both domains (the resulting individuals looked and behaved like dominants), and (d) recruited a group of naturally dominant controls. Each wasp then interacted with a single dominant conspecific. Tibbetts and Izzo found that wasps that looked dominant but behaved in a subordinate fashion received a greater number of attacks than did controls. In contrast, wasps that looked subordinate but behaved in a dominant fashion did not receive more attacks, but they also failed to

achieve dominance over the other wasp (i.e., the controls never submitted to the aggressive but subordinate-looking wasps). By showing that dishonest signalers of the conventional signal accrue more costs than honest signalers, these findings are consistent with the hypothesis that the reliability of conventional signals can be maintained via punishment from conspecifics. Experiments conducted on other species have similarly demonstrated that dishonest signalers of dishonest signals receive more aggression from their social partners than do honest signalers (e.g., Anderson et al., 2012).

The hypothesis that smiles have maintained their reliability as signals of cooperative intent over evolutionary time via punishment of dishonest smiles from receivers implies a number of testable predictions. First is the prediction that when faced with a cooperative dilemma, participants will trust smiling individuals more than non-smiling individuals—as has been found in previous studies (Centorrino et al., 2015; Krumhuber et al., 2007). A second, more unique prediction is that when participants interact with an untrustworthy partner, the participants will become angrier if that partner was smiling prior to the game than if the untrustworthy partner was not smiling prior to the game. A third prediction is that participants who are treated unfairly by smiling partners will punish their partners to a greater degree than will participants who are treated unfairly by non-smiling partners, much like the paper wasps that attacked subordinates who appeared dominant (Tibbetts & Izzo, 2010).

This experiment was designed to test these three predictions by having participants play a binary Trust Game with a confederate who is either smiling or not smiling, and who consequently behaves in either a trustworthy or untrustworthy fashion.

Chapter 2: Method

Participants

Participants were 385 students (Age: M=18.71, SD=3.07; 53.25% Female) from the University of Miami. Participants were randomly assigned to one of four conditions in a 2 (signal: smile vs. neutral) x 2 (treatment: fair vs. unfair) factorial design, blocking on sex (to ensure equal proportions of males to females in each condition). Participants received $9 as compensation in addition to course credit (if enrolled as part of a course requirement).

Procedure

After signing consent forms, all procedures were administered via computer using E-Prime 2.0 (v 2.0.10.242). Participants were told that they would be interacting with a second participant over networked computers, that the second participant may or may not have been in the same room as the participant, and that the two participants will be making a series of decision about what to with money. In reality, participants were interacting with a pre-programmed computer script, referred to here as a confederate. Participants were told that their decisions during the interaction would influence the amount of money they earned. Before the ostensible interaction occurred, participants were told to test the webcam, which was used to take pictures of participants and to display the pictures for the participants' approval. The picture test was also used to ensure that the participant knew pictures were actually being taken, thereby lending credibility to the notion that participants will actually be sending pictures to, and receiving pictures from, another real participant.

Trust Game

Participants played a modified binary Trust Game with their interaction partners, based on Berg, Dickhaut, and McCabe (1995). To reduce experimental demand that might result from telling participants they would be playing a game involving trust, the game was referred to as an "Economic Decision Making Game." In the Trust Game, participants and confederates were given a $3 endowment and were ostensibly randomly assigned to the position of either First Mover or Second Mover (referred to in the Trust Game literature as the "truster" and "trustee," respectively). In reality, participants were always assigned to the role of First Mover and were told that they could either send half (i.e., $1.50) or none (i.e., $0.00) of their endowment to the confederate. By using a binary trust game, rather than a trust game that enables continuous transfer options, I was able to control for one potential confound: Participants who invest more in smiling partners might become angrier and might exhibit more punishment behavior simply because more of their investments were lost, rather than because their partners smiled and did not behave in a trustworthy fashion. Participants were told that the amount of the transfer would be automatically tripled when it is deposited in the confederate's account. If the participant decided to send money to the confederate, participants were told that the confederate would be given the opportunity to return money from his/her account back to the participant. If the participant decided *not* to send money to the confederate, then participants were told that the confederate would not be given the opportunity to return any money.

Smiling Manipulation

Photographs were taken from still shots of videos from the MMI Face Database (Pantic, Valstar, Rademaker, & Maat, 2005). Participants saw one of two types of photographs before and during their money-transfer decision in the Trust Game: 1) A person with a smiling facial display or 2) a person with a neutral facial display. Males only saw a picture of one of three male faces, while females only saw a picture of one of three female faces. By using three different faces, I could be confident that it was the manipulation of the signal, and not the different faces, that would result in any effects. The same photograph was used throughout the experiment whenever any decisions regarding an interaction were made.

Fairness Manipulation

In the Trust Game, trusting behavior is typically measured by the amount of money the first decision-maker transfers to the second decision-maker, and trustworthiness is measured by the amount of money the second decision-maker returns to the first decision-maker (Berg et al., 1995). In my modified version of the game, the decision to trust was binary; therefore, trusting was measured by whether the participant decided to send $1.50 (which was multiplied by 3) to the confederate. In the fair (i.e., trustworthy) condition, confederates who were trusted (i.e., received the $4.50 transfer) returned $3, which left both players with $4.50 (a profit of $1.50 for both the participant and the confederate). In the unfair (i.e., untrustworthy) condition, confederates who were trusted returned $0.00, which left the participant with $1.50 and the confederate with $7.50 at the end of the trust game. Participants were told that the financial outcomes from this game were "banked," meaning that the resulting earnings were theirs to keep and that

they would be given that amount of money, plus whatever amount they earned in later interactions, at the end of the session.

Emotion Measurement

Following the Trust Game, participants completed a series of self-report items (administered by computer in a randomized order) to measure their feelings toward the confederate. Participants were asked to indicate "the extent to which you are feeling each of these emotional reactions toward the other player" by responding to a series of emotion words on a Likert-type scale (from *0* = "not at all" to *5* = "extremely"). These items included measurements of anger (i.e., angry, mad, outraged) and happiness (i.e., happy, satisfied, content).

Punishment Opportunity

When participants completed the self-report emotion assessment, they were instructed to play a second (and, unbeknownst to participants, final) economic decision-making game. In the second game, which was the modified Dictator Game used in Pedersen, Kurzban, and McCullough (2013), participants were told that they would be randomly assigned to the role of either the Decision-Maker or the Recipient, and that the confederate would take on the other role. In reality, the participant was always assigned to be the Decision Maker. At the start of this second game, each player received a $4 endowment. Decision Makers were allowed to make one of three choices: (1) Allocate any amount of his or her $4 (in $0.25 increments) to the Recipient; (2) deduct money from the Recipient (in $0.25 increments, without gaining the amount deducted for themselves) by paying one-fourth of the amount removed; or (3) do nothing with respect to the $4 in the Recipient's account, which would result in participants and confederates

both ending the second game with their $4 endowments intact. The amount of money (in $0.25 increments) removed from (interpreted as punishing), or transferred to (interpreted as rewarding), the Recipient served as the primary dependent measure.

Debriefing

After participants completed all of the tasks described above, they were probed for suspicion using an extensive procedure designed to assess the believability of the experiment (Aronson, Ellsworth, Carlsmith, & Gonzales, 1990). Following debriefing, participants were asked whether they had any additional questions regarding the study. Finally, before being dismissed, all participants were be paid a flat rate of $9, which was $0.50 greater than the maximum amount of money that could possibly have been earned in the experiment.

Data Management

Excluding Participants from Analyses

Responses during debriefing were coded for suspicion about the authenticity of the study. Analyses were conducted without including suspicious participants (n = 91). A breakdown of suspicion rates, trust rates, and final sample sizes for each analysis appears in Table 1.

Participants who were not initially trusting (i.e., those who did not send any money to the interaction partner during the trust game) were removed from analyses related to anger and punishment (n = 20): There is no possible way for the interaction partner to treat a non-trusting person fairly or unfairly, and as a consequence, it is impossible to interpret their levels of anger and retaliatory punishment. The analyses

predicting anger and punishment that were reported here only included participants who exhibited neither suspicion nor distrust (n = 272).

Aggregating Variables to Create Scales

Scores on the self-report measures of anger, comprising measures of how "angry," "mad," and "outraged" participants were (alpha = .94), and happiness, comprising measures of how "happy," "satisfied," and "content" participants were (alpha = .95), were aggregated into two separate outcome variables. Correlations among outcome variables appears in Table 2.

Variable Coding

Binary predictor and outcome variables were dummy coded. For the facial display predictor, smiling displays were coded as "1" and neutral displays were coded as "0". For the unfair treatment predictor, unfairness was coded as "1" and fairness was coded as "0". For the outcome of trust, people who trusted were coded as "1" and people who did not trust were coded as "0". Punishment and reward outcomes from the Dictator Game were coded in terms of the effect on the confederate's account. If the confederate lost money, the amount removed was coded as a negative number; if the confederate gained money, the amount gained was coded as a positive number; if nothing was done with respect to the confederate's account, that was coded as 0.

Chapter 3: Results

All analyses were conducted using R version 3.1.2 (R Core R Core Team, 2014). Multinomial logistic regression was conducted using the "nnet" package (Venables & Ripley, 2002) and all graphs and figures were created using the "ggplot2" package (Wickham, 2009). Effect sizes for Wilcoxon and Mann-Whitney U tests were computed using the methods from Fritz, Morris, and Richler (2012). Distributions for the three main outcome variables anger, happiness, and punishment, appear in Figures 1-3, respectively, for each condition.

Analyses

Fairness manipulation check. First, to test whether anger was greater for those who were treated unfairly (n=119) than those who were treated fairly (n=153), I conducted a Mann-Whitney U test comparing the fair and unfair groups. Results indicated that the two groups were different, $Z = 13.29$, $p < .001$, $\eta^2 = 0.65$, with those in the unfair group having higher levels of anger than those in the fair group, difference in location = 2.00, 95% CI [1.67, 2.00] [1]. Second, to test whether happiness was lower for those who were treated unfairly, I conducted a Mann-Whitney U test comparing the fair and unfair groups. Results indicated the two groups were different, $W = -13.52$, $p < .001$, $\eta^2 = 0.67$, with those in the unfair group having lower levels of happiness than those in the fair group, difference in location = -3.33, 95% CI [-3.33, -3.00].

Second, to test whether punishment increased for those who were treated unfairly, I conducted a Mann-Whitney U test to compare the two groups. Results indicated that the

---

[1] Due to rounding error, the upper bound of the confidence interval appears to match the estimate. This type of observation occurred on multiple occasions when requesting confidence intervals for Wilcoxon and Mann-Whitney U tests.

median was significantly lower for those who were treated unfairly, $Z = -6.57$, $p < .001$, $\eta^2 = .16$, difference in location = -0.5, 95% CI [-0.99, -0.00003]. To ensure that there was more punishment, rather than simply less giving, a single-sample Wilcoxon test against zero indicated that the median for people who were treated unfairly was negative and significantly different from zero, $Z = -2.20$, $p = .028$, $\eta^2 = .04$, (pseudo)median = -1.00, 95% CI [-2.75, -0.00002].

Trust. My first hypothesis was that people would be more likely to trust someone who is smiling than someone who is not smiling, which is a replication of previous studies (Centorrino et al., 2015; Krumhuber et al., 2007). I used logistic regression to test whether the presence of smiles predicted the binary option of trusting. Focusing only on subjects whose partners were not smiling (i.e., subjects whose partners took a score of zero on the "smiles" variable); participants were more than 11 times more likely to trust than to not trust, $OR_{intercept} = 11.33$, 95% CI [6.56, 21.59], $p < .001$. In this model, the smiles variable itself did not significantly affect the odds that participants would trust their partners, $OR_{smile} = 1.5$, 95% CI [.601, 3.94], $p = .391$, although this finding is unsurprising given that there was very little variance in trusting behavior.

Anger. My second hypothesis was that people would become angrier in response to someone who behaves unfairly while smiling, when compared to someone who behaves unfairly while not smiling. To test this prediction, I first conducted an omnibus test of a model predicting anger with the two treatment effects (unfairness and smiles) and their interaction using linear regression. The overall model was significant $F(3, 268) = 86.34$, $R^2_{adjusted} = 0.49$, $p < .001$ indicating that the three predictors successfully predicted variation in anger. There was a main effect for unfairness, $\beta = 2.23$, $SE = 0.18$,

$\eta^2_{partial} = .37$, $p < .001$, which was consistent with the manipulation checks, but there was neither a main effect for smiles, $\beta = 0.05$, $SE = 0.17$, $\eta^2_{partial} = .0004$, $p = .756$, nor an effect for the interaction of unfairness and smiles, $\beta = -0.47$, $SE = 0.25$, $\eta^2_{partial} = .013$, $p = .066$, which would indicate that the effect of unfairness did not differ across smile conditions.

The data were not normally distributed, so a single Mann-Whitney U analysis was conducted in addition to the regression analysis to test whether anger levels differed between those in the unfair smiles group (n = 57) and those in the unfair neutral group (n = 62). Results indicated that the two groups did not differ in self-reported anger, $Z = -1.44$, $p = .15$, $\eta^2 = .02$, difference in location = -0.33, 95% CI [-.99, .00003].

To further capture the emotional differences between treatment groups, the same set of analyses conducted for anger were also conducted for happiness. The overall model predicting happiness was significant, $F(3, 268) = 264.6$, $R^2_{adjusted} = .74$, $p < .001$. There was a significant main effect for unfairness, $\beta = -3.17$, $SE = .15$, $\eta^2_{partial} = .62$, $p < .001$, indicating that the presence of unfairness reduced levels of happiness. There were no significant effects for either smiles, $\beta = -.02$, $SE = .14$, $\eta^2_{partial} = .0001$, $p = .846$, or the smile by unfairness interaction, $\beta = .23$, $SE = .22$, $\eta^2_{partial} = .004$, $p = .298$. The distributions of these data were not normal, so a single Mann-Whitney U analysis was conducted in addition to the regression analysis to test whether happiness levels differed between those in the unfair neutral group and those in the unfair smile group. Results indicated that the two groups did not differ on levels of happiness, $Z = -1.50$, $p = .133$, $\eta^2 = .02$, difference in location = 0.00007, 95% CI [-.00004, .333], which would suggest that interacting with a smiler did not impact self-report levels of happiness.

Punishment. My third and final hypothesis was that people would punish unfair smilers more than fair smilers. To test this prediction, I first conducted an omnibus test of a model predicting punishment with the two treatment effects (unfairness and smiles) and their interaction using linear regression. The overall model was significant $F(3, 260) =$ 12.88, $R^2_{adjusted} = .12$, $p < .001$. Results for the intercept term indicated that for fair and neutral confederates, people gave a significant non-zero amount, $\beta_{intercept} = .77$, $\eta^2_{partial} = .08$, $p < .001$. There was also a negative main effect for unfairness, $\beta = -.94$, $\eta^2_{partial} = .06$, $p < .001$, but there was no main effect for smiles, $\beta = -.08$, $\eta^2_{partial} = .005$, $p = .711$, nor the smile by unfair interaction, $\beta = -.20$, $\eta^2_{partial} = .001$, $p = .547$.

Since the data on punishment were not normally distributed, I also conducted two different types of analyses. First, I grouped the punishment outcome into three separate categories for use in multinomial logistic regression: (1) "Punish"; (2) "Give"; and (3) "Nothing". For multinomial logistic regression, "Nothing" was used as a reference group, to which "Give" and "Punish" were compared. Results indicated that for fair and neutral confederates, people were just as likely to give as to do nothing, $OR = .97$, $p = .907$, but were much less likely to punish than do nothing, $OR = .03$, $p < .001$. Introducing unfairness had a negative impact on this relationship, only by reducing the likelihood of giving, $OR = .15$, $p < .001$, but not by significantly increasing the likelihood of punishing, $OR = 5.63$, $p = .113$. Further, there was no significant interaction between smiles and unfairness on the relative likelihood of either giving, $OR = 1.29$, $p = .714$, or punishing, $OR = 2.43$, $p = .561$.

As described above in the results for the manipulation checks, there was an overall effect of treatment on the punishment outcome, which appeared to be driven in

part by increases in punishment for the unfair treatment group. To directly test whether the unfair/smile group differed from the unfair/neutral group, I conducted a Mann-Whitney U analysis comparing the two groups. Results indicated that the two groups did not differ in their levels of punishment, $Z = 1.11$, $p = .268$, $\eta^2 = .01$. Though results indicated an overall effect of unfair treatment on punishment, when each group was separately tested against zero using a Wilcoxon test, neither the unfair/smiles group ($n = 54$), $Z = -1.81$, $\eta^2 = .06$, $p = .070$, (pseudo)median $= -1.62$, 95% CI [-3.00, 0.000007], nor the unfair/neutral group ($n=60$), $Z = -1.23$, $\eta^2 = .03$, $p = .218$, (pseudo)median $= -0.75$, 95% CI [-3.00, 1.62], differed significantly from zero.

Chapter 4: Discussion

Smiles and other facial displays have long been a central topic for scientists investigating emotions, communication, and other aspects of social psychology. However, no researchers to date have developed a paradigm to directly investigate *why* smiles might have maintained reliability over evolutionary time. That is, why has selection not favored those who exploit the benefits that come with smiling and, further, why has natural selection not favored those who ignored such displays after the signal could have lost its meaning? Here, I presented the first experiment designed to directly test hypotheses related to the evolved function of smiles and how they might have maintained reliability over evolutionary time. I hypothesized that smiles function in contexts of social exchange as signals of cooperative intent. Further, I hypothesized that smiles might have maintained their reliability in contexts of social exchange due to receiver-dependent costs or, more specifically, due to costs associated with punishment— if people smile and do not subsequently act cooperatively, they may be subjected to more costs from those who received, and acted on, the cooperative signal. From these hypotheses, I predicted that: (1) people who smiled would be trusted more often than people who did not smile; (2) people who trusted an unfair smiler would become angrier than people who trusted an unfair non-smiler; and (3) people who trusted an unfair smiler would punish the unfair person more than people who trusted an unfair non-smiler.

None of my three hypotheses were supported. First, people were not more likely to trust people who smiled than people who did not smile. Though the results were in the predicted direction, the effect was not significant. Second, people were neither angrier, nor less happy, after interacting with an unfair smiler than people who interacted with a

unfair non-smiler. Third, participants who were treated unfairly by a confederate who was smiling did not punish the confederate to a greater extent, nor were they more likely to engage in punishment, than those treated unfairly by a confederate who was not smiling.

Limitations

On the surface, it may be most surprising that I was unable to replicate previous research that showed a higher likelihood for people to engage in cooperative acts with a person who is smiling then with a person who is not smiling (Centorrino et al., 2015; Krumhuber et al., 2007). However, this expectation was perhaps overly optimistic given the way the study was set up with very low stakes so that people would decide to trust in the Trust Game regardless of condition. Though I still expected a statistically significant difference in trust between the smile and neutral conditions, it is likely that this study lacked enough power to detect an effect. In other words, I do not think that the results of this study alone should be enough to warrant dismissing the notion that smiles do function as signals of cooperative intent.

My central hypothesis for this thesis was that smiles maintained reliability due to receiver-dependent costs, specifically costs associated with punishment. One aspect of this hypothesis was that punishment would be positively related to anger. Though the data were consistent with this hypothesis, there were neither differences in anger nor in punishment when comparing the unfair/smile group to the unfair/neutral group. There are five main reasons for the lack of statistically significant results: (1) the hypothesis that smiles maintained reliability due to costs of punishment is incorrect; (2) the design of the study was not sensitive enough to detect a true effect; (3) the study design was not

appropriate for testing the hypothesis; (4) the population was not sensitive enough to the manipulations to produce an effect; and/or (5) the sample size was not large enough to detect what could be a very small effect.

The first, and simplest, explanation is that my hypothesis was incorrect. There are certainly alternative routes through which receiver-dependent costs can help to maintain the reliability of signals; that is, routes that do not require active punishment. As mentioned previously, signals can maintain reliability through receivers' memories for unreliable signalers (Maynard Smith & Harper, 2003; Searcy & Nowicki, 2005), which would lead to subsequent avoidance of those who deceptively signal cooperative intent with smiles. In this type of scenario, those who smile dishonestly may be able to exploit receivers who are willing to confer benefits in the short-term, but as receivers learn that the signal is no longer reliable from those specific signalers, the dishonest smilers would not be able to accrue benefits from the signal in the long term. This is similar to the notion that conventional signals could maintain reliability over evolutionary time due to the vigilance of receivers (Dezecache, Mercier, & Scott-Phillips, 2013); as receivers take note of who is signaling and how their signals match with their behaviors, receivers can use that information to avoid those whose signals are inconsistent with their behaviors.

Alternatively, though not mutually exclusive to memory and/or vigilance mechanisms, it is possible that I was analyzing the wrong type of punishment. A direct loss of tangible resources is a very explicit, obvious, and unambiguous type of punishment which can be easily interpreted, yet people may actually be more inclined to impose costs via the dishonest signalers' social resources. For example, people may be inclined to harm a dishonest signaler's reputation by sharing information regarding that

signaler's deceptive tendency, but may not be willing to impose costs by reducing any of the signaler's tangible resources. This form of receiver-dependent costs is dependent on memory, vigilance, and can still be thought of as a direct act of punishment (Barclay, 2013; Maynard Smith & Harper, 2003; Searcy & Nowicki, 2005).

If I am to consider that my original hypotheses were true, a second possibility for my results is that the design was not sensitive enough to detect an effect. One primary concern is how costly the punishment was in this study. The chosen rate of a 1-to-4 cost-to-punishment ratio is actually quite arbitrary and may actually be artificially higher than the ratio people would typically expect to pay for punishment in that situation. Further, since my goal in this study was not actually to demonstrate absolute levels of punishment, but rather *relative* levels of punishment (i.e., whether two groups differ in their willingness to engage in punishment), it is more important that there was sufficient variability in punishment to actually compare levels of punishment. Though results indicated that there was a significant level of punishment in the unfair treatment group, neither subgroup of the unfair treatment group (i.e., smiles vs neutral) actually had significant levels of punishment. Therefore, it could not be determined whether two groups differ in rates of punishment if neither group, independently, was willing to engage in punishment.

Still assuming that my original hypotheses were true, the design of the study, particularly with respect to the low stakes trust game, may have resulted in inadequate power to detect a true effect; those in the unfair treatment groups may not have perceived the loss as being high enough to warrant punitive action. This explanation may also be linked to the possibility that the sample in this study, undergraduate students at a private

university, was not sensitive to the manipulation. Not all students at private universities come from wealthy families but it is not unreasonable to suspect that a substantial proportion of students do. Therefore, the subjects in these experiments, on average, may not have felt as though they were at a significant loss due to the manipulation and, in turn, may have felt indifferent with regard to taking punitive action. On a related note, the effect sizes could actually be smaller than I had anticipated. In other words, the sample size may not have been large enough to detect true, but small, effects.

Future Directions

It is possible that, for smiles to maintain reliability as signals of cooperative intent, it is not necessary for people to have an enhanced punitive sentiment toward unfair smilers. Alternatively, a model that does not require any specialized psychology for increasing anger and punishment toward unfair smilers, relative to unfair non-smilers, may actually be a more accurate representation of how the signal maintained reliability. If it is true that (1) smiles do indeed increase the likelihood that receivers will cooperate and (2) unfair treatment causes people to become angry, then by extension, unfair smilers will be more likely to incur costs from punishment *just by virtue of engaging in unfair behavior*, which may be enough to maintain the reliability of the signal. In other words, punishment that functions to deter unfair behavior may, in and of itself, be enough to maintain the reliability of smiles as signals of cooperative intent. However, the data presented here cannot actually speak to this idea, as people were not more likely to cooperate with smilers than non-smilers. Therefore, it is imperative that smiles actually have an effect on perceivers' willingness to cooperate before any alternative models can be evaluated and compared. A first step in this direction would be to carefully examine

the discrepancies between the manipulations used in previous studies (e.g., Centorrino et al., 2015; Krumhuber et al., 2007) and in this study, as a way of further assessing why there was no effect for smiles in this sample.

Conclusion

This study was the first to directly assess the plausibility of one potential mechanism that may have maintained the reliability of smiles—the receiver-dependent costs associated with punishment for smiling in the absence of cooperative intent. Though results did not support my hypothesis, there were limitations in the design that preclude my ability to dismiss the hypothesis entirely. While my hypothesis may still lack support after further investigation, this paradigm still lends itself to future research as a template for directly assessing the plausibility of mechanisms that maintain the reliability of signaling systems in humans over evolutionary time.

References

Anderson, R. C., Searcy, W. A., Hughes, M., & Nowicki, S. (2012). The receiver-dependent cost of soft song: A signal of aggressive intent in songbirds. *Animal Behaviour, 83*(6), 1443-1448.

Aronson, E., Ellsworth, P. C., Carlsmith, J. M., & Gonzales, M. H. (1990). *Methods of research in social psychology*. New York, NY: McGraw-Hill.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*(4489), 1390-1396.

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution and Human Behavior, 34*(3), 164-175.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122-142.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior, 27*(3), 163-182.

Brown, W. M., & Moore, C. (2000). Is prospective altruist-detection an evolved solution to the adaptive problem of subtle cheating in cooperative ventures? Supportive evidence using the Wason selection task. *Evolution and Human Behavior, 21*(1), 25-37.

Brown, W. M., & Moore, C. (2002). Smile asymmetries and reputation as reliable indicators of likelihood to cooperate: An evolutionary analysis. In S. P. Shohov (Ed.), *Advances in psychology research* (pp. 59-78). New York, NY: Nova Science Publishers.

Brown, W. M., Palameta, B., & Moore, C. (2003). Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm. *Evolutionary Psychology, 1*, 42-69.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2010). Honest smiles as a costly signal in social exchange. *Behavioral and Brain Sciences, 33*(06), 439-439.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M., & Seabright, P. (2015). Honest signaling in trust interactions: Smiles rated as genuine induce trust and signal higher earning opportunities. *Evolution and Human Behavior, 36*(1), 8-16.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31*(3), 187-276.

Darwin, C. (1872). *The expression of emotions in animals and man*. London, UK: Murray.

Dawkins, R. (1999). *The extended phenotype: The long reach of the gene*. Oxford, UK: Oxford University Press.

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: Dissecting the architecture of a moral concept. *Journal of Personality and Social Psychology, 102*(6), 1252.

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior, 33*(6), 715-725.

Dezecache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *Journal of Pragmatics, 59*, 221-233.

Duchenne, G. B. A. (1990). *The mechanism of human facial expression* (R. A. Cuthbertson, Ed. and Trans.). Cambridge, UK: Cambridge University Press. (Original work published 1862).

Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry, 32*(1), 88-106.

Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural elements in facial displays of emotion. *Science, 164*(3875), 86-88.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. *Journal of Personality and Social Psychology, 60*(2), 229.

Fridlund, A. J. (1997). The new ethology of human facial expressions. *The Psychology of Facial Expression, 112*, 103-129.

Fridlund, A. J. (1997). The new ethology of human facial expressions. In J. A. Russel & J. M. F. Dols (Eds.), *The psychology of facial expression* (pp. 103-129). Cambridge, UK: Cambridge University Press.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2.

Fuentes, A., Wyczalkowski, M. A., & MacKinnon, K. C. (2010). Niche construction through cooperation: A nonlinear dynamics contribution to modeling facets of the evolutionary history in the genus Homo. *Current Anthropology, 51*(3), 435-444.

Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology, 144*(4), 517-546.

Gueguen, N., & De Gail, M. A. (2003). The effect of smiling on helping behavior: Smiling and good Samaritan behavior. *Communication Reports, 16*(2), 133-140.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour, 42*(1), 1-14.

Guilford, T., & Dawkins, M. S. (1995). What are conventional signals? *Animal Behaviour, 49*(6), 1689-1695.

Hess, U., Beaupré, M. G., & Cheung, N. (2002). Who to whom and why–cultural differences and similarities in the function of smiles. In M. H. Abel (Ed.), *An empirical reflection on the smile* (pp. 186-216). New York, NY: Edwin Mellen.

Hinde, R. A. (1985). Was 'The expression of the emotions'a misleading phrase? *Animal Behaviour, 33*(3), 985-992.

Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 201200155.

Jeffords, M., Sternburg, J., & Waldbauer, G. (1979). Batesian mimicry: Field demonstration of the survival value of pipevine swallowtail and monarch color patterns. *Evolution*, 275-286.

Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 380-402). Oxford, UK: Blackwell.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion, 7*(4), 730.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, UK: Oxford University Press.

Mehu, M., Grammer, K., & Dunbar, R. I. (2007). Smiles when sharing. *Evolution and Human Behavior, 28*(6), 415-422.

Mehu, M., Little, A. C., & Dunbar, R. I. (2007). Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology, 5*(1), 183-196.

Otta, E., Abrosio, F. F. E., & Hoshino, R. L. (1996). Reading a smiling face: Messages conveyed by various forms of smiling. *Perceptual and Motor Skills, 82*(3c), 1111-1121.

Pantic, M., Valstar, M., Rademaker, R., & Maat, L. (2005). *Web-based database for facial expression analysis.* Paper presented at the Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on.

Pedersen, E. J., Kurzban, R., & McCullough, M. E. (2013). Do humans really punish altruistically? A closer look. *Proceedings of the Royal Society B: Biological Sciences, 280*(1758).

Price, M. E. (2006). Monitoring, reputation, and 'greenbeard'reciprocity in a Shuar work team. *Journal of Organizational Behavior, 27*(2), 201-219.

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological device. *Evolution and Human Behavior, 23*(3), 203-231.

R Core Team. (2014). R: A language and environment for statistical computing. *R foundation for statistical computing, Vienna, Austria.* from http://www.R-project.org/

Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous Prisoner's Dilemma game. *Evolution and Human Behavior, 33*(3), 200-209.

Scott-Phillips, T. C. (2008). *On the correct application of animal signalling theory to human communication.* Paper presented at the Proceedings of the 7th International Conference on the Evolution of Language.

Scott-Phillips, T. C., Blythe, R. A., Gardner, A., & West, S. A. (2012). How do communication systems emerge? *Proceedings of the Royal Society B: Biological Sciences, 279*(1735), 1943-1949.

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science, 6*(1), 38-47.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton: Princeton University Press.

Tibbetts, E. A., & Izzo, A. (2010). Social punishment of dishonest signalers caused by mismatch between signal and behavior. *Current Biology, 20*(18), 1637-1640.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 35-57.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. New York, NY: Springer Science & Business Media.

West, S. A., El Mouden, C., & Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evolution and Human Behavior, 32*(4), 231-262.

West, S. A., Griffin, A. S., & Gardner, A. (2007a). Evolutionary explanations for cooperation. *Current Biology, 17*(16), R661-R672.

West, S. A., Griffin, A. S., & Gardner, A. (2007b). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology, 20*(2), 415-432.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer Science & Business Media.

Zahavi, A. (1975). Mate selection—A selection for a handicap. *Journal of Theoretical Biology, 53*(1), 205-214.

Table 1. Suspicion and trust rates for each condition

	Fair	Unfair
Smile	n=79/76	n=57/54
Suspicion	n=19	n=33
Trust	n=4	n=7
Neutral	n=74/74	n=62/60
Suspicion	n=14	n=25
Trust	n=7	n=8

Note: Numbers in the highlighted cells represent sample sizes for analyses predicting emotions and punishment, respectively. Numbers for Suspicion represent the total number of subjects excluded due to suspicion. Numbers for Trust represent the total number of subjects excluded due to not trusting in the Trust Game.

Table 2. Correlations among outcome variables

	Anger	Happiness
Happiness	-0.693	-
Punishment	0.384	-0.358

Note: For the zero-order correlations, punishment was reverse-coded so that higher levels of punishment (i.e., lower numbers) would be positively related to anger and negatively related to happiness.
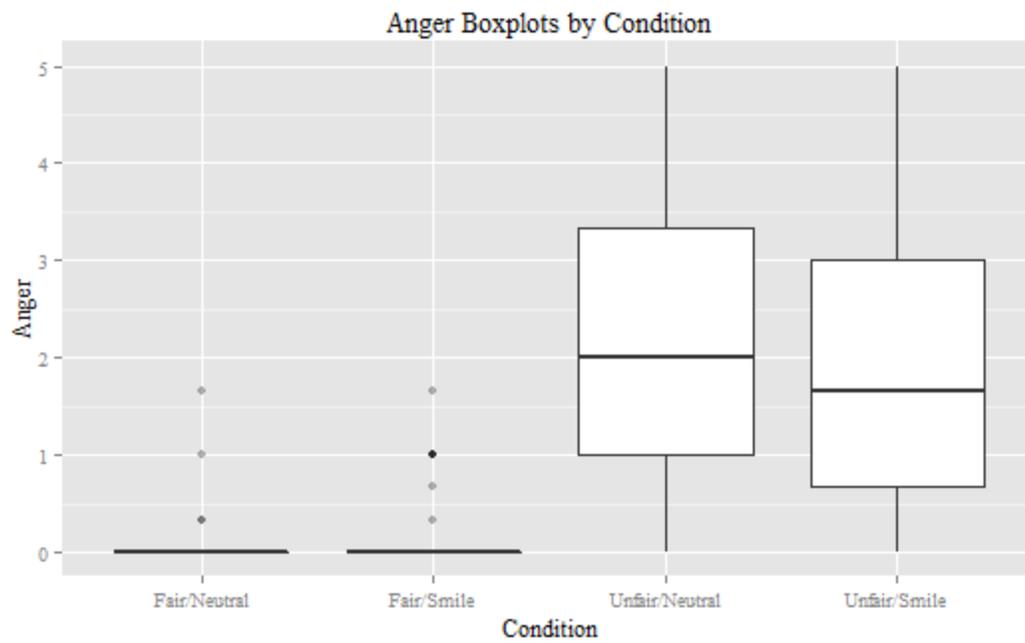
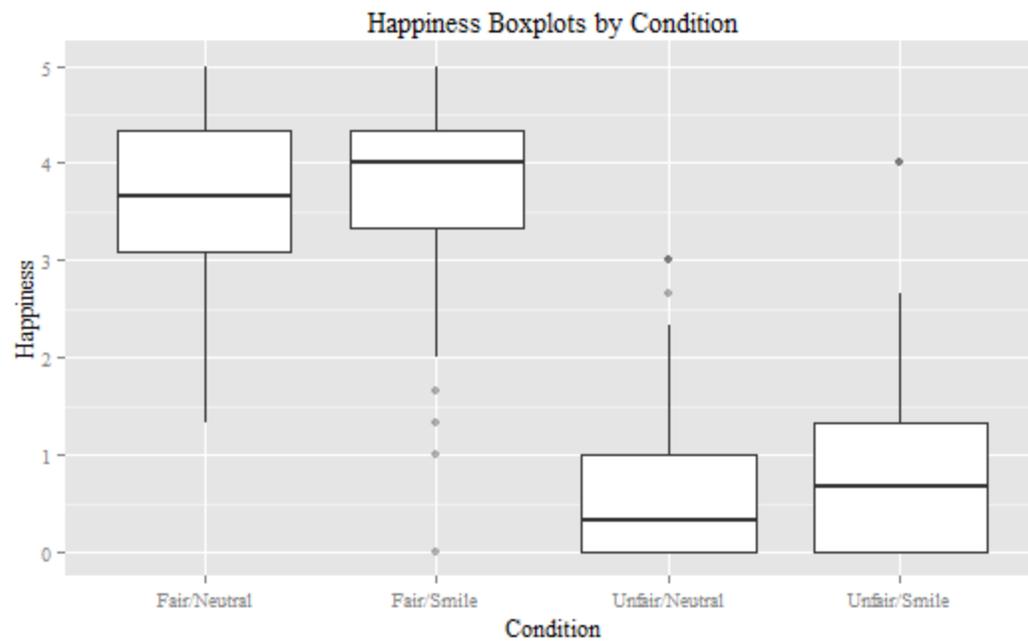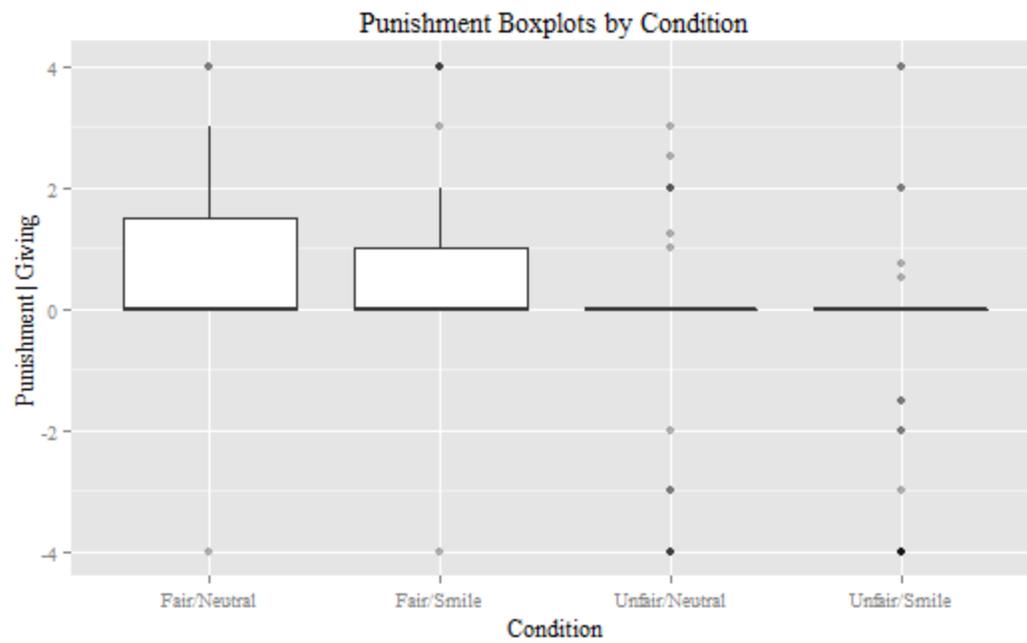*Figure 1.* Distributions of self-report anger levels for each condition.

Figure 2. Distributions of self-report happiness levels for each condition.

Figure 3. Distributions of behavior during the modified Dictator Game. Positive values

are interpreted as giving whereas negative values are interpreted as punishing.