A Model of Smiling as a Costly Signal of Cooperation Opportunities[1]

Samuele Centorrino (Toulouse School of Economics),
Elodie Djemai (University of Paris-Dauphine),
Astrid Hopfensitz (Toulouse School of Economics),
Manfred Milinski (Max Planck Institute of Evolutionary Biology, Plön),
Paul Seabright (Toulouse School of Economics and Institute for Advanced Study in Toulouse)[*]

This version: 16th October 2013

Abstract

We develop a theoretical model under which "genuine" or "convincing" smiling is a costly signal that has evolved to induce cooperation in situations requiring mutual trust. Prior to a trust interaction, individuals can send a signal to induce others to trust them; the signal takes the form of a smile that may be perceived as more or less convincing, and that can be made more convincing with the investment of great cognitive effort. Individuals differ in their degree of altruism and in their tendency to display reciprocity. The model generates three testable predictions. First, the perceived quality of Player B's smile is increasing in the size of the stake. Secondly, the amount sent by Player A is increasing in the perceived quality of the smile. Thirdly, the expected gain to player A from sending the stake to player B is increasing in the perceived convincingness of player B's smile.

Keywords: Smiling, costly signaling, experiment, trust game, video

JEL codes: D03, D85, D87, Z13

[1] **Acknowledgements:** The authors are grateful to the Max Planck Institute for Evolutionary Biology in Plön for financial support.

[*] To whom correspondence should be addressed: Paul.Seabright@tse-fr.eu

1. Introduction

The man who indulges us in this natural passion, who invites us into his heart, who, as it were, sets open the gates of his breast to us, seems to exercise a species of hospitality more delightful than any other. No man, who is in ordinary good temper, can fail of pleasing, if he has the courage to utter his real sentiments as he feels them, and because he feels them.

Adam Smith – The Theory of Moral Sentiments[2]

This paper develops a model of smiling as a form of signaling behavior whose purpose is to facilitate economic exchange in situations requiring mutual trust. Smiling is a form of behavior that is found in all human societies[3]. It appears to be more elaborate and more central to communication in humans than in any other species, and to play an important part in judgments of individuals about the character and general trustworthiness of others. Yet there is no scientific consensus as to why it has evolved to be like this, nor about what it is in smiling that makes it an appropriate basis for judgments of others.

There is consensus, however, about a number of the characteristics of smiling behavior. First, viewers perceive smiles as varying in their degree of "genuineness" or "convincingness". Since the work of Duchenne (1862) and Darwin (1872) in the 19th century it has been known that smiles perceived as genuine (known as enjoyment or "Duchenne" smiles) are characterized by use of the *orbicularis oculi* (which surrounds the eyes) in combination with the *zygomatic major* (which raises the corners of the mouth); symmetry is also an important characteristic of Duchenne smiles. More recent research focuses on the importance of temporal dynamics such

[2] Smith (2000), p.497.
[3] See Darwin (1872), Ekman (1982), Niedenthal et al. (2010).

as smile onset, apex, and offset durations for perceived genuineness[4]. Second, Duchenne smiles are not under straightforward voluntary control. Some individuals can make them more often and more easily than others, and all individuals find them easier to make when in certain affective states. Such states include a relaxed mood in general, and feeling well disposed to a communication partner in particular. Third, smiles induce mimicry, both in the sense that individuals viewing smiles by others have an increased tendency to smile themselves[5], and in the sense that individuals trying to make a good impression on others (as when posing for photographs) make an effort to smile well. Although individuals can smile when alone, smiling behavior seems to be a form of communication. But if so, what is it communicating, and why have we evolved a form of communication behavior that is under such imperfect conscious control?

In this paper we set out a model according to which smiling is a form of costly communication (costly in a sense we make precise below) that induces cooperation between individuals in situations requiring mutual trust. According to this view, the necessary costliness of smiling is precisely the reason why it is under such imperfect conscious control. In a nutshell, smiling is costly because otherwise it would be easy to fake, and would not reliably be associated with trustworthiness. This does not imply that smiling has evolved to be difficult to fake in order to act as a signal, but rather that natural selection has recruited as a signal a form of behavior that is already difficult to fake.

---

[4] See Krumhuber et al. (2007).
[5] See Niedenthal et al. (2010).

This hypothesis is not original to us[6], but to our knowledge it has not previously been formalized in a way that would make it capable of being subjected to a comprehensive experimental test. As we show below, the hypothesis involves three distinct component hypotheses, namely that smiling "genuinely" or "convincingly" is:

a) costly to the smiler in terms of effort,

b) causally effective in inducing the target of the smile to cooperate with the smiler, and

c) a reliable signal of the likely benefits to the target of cooperating with the smiler, both because it is correlated with the smiler’s intrinsic trustworthiness, and because it is correlated with the size of the benefits that the smiler has to share.

In Section 2 we briefly review the relevant literature. Section 3 sets out our model. Section 4 discusses empirical implications; we have tested these in a companion paper (Centorrino et al 2013). Section 5 concludes.

2. Literature review

Costly signaling has been extensively studied both in economics since the work of Spence (1974), and independently in biology since the work of Zahavi (1975). A signal is any observable trait that imposes a cost on its bearer (a pecuniary or non-pecuniary effort cost in economics, a fitness cost in biology) but which reliably indicates the presence of some advantageous hidden trait because the signal is

---

[6] See Owren & Bacharowski (2001).

more costly for those individuals that do not possess the trait than for those who do[7]. The benefit from signaling the hidden trait is that it attracts partners in mating or in some other mutually beneficial cooperative activity, and the benefit to the signaler of doing so must exceed the cost of the signal. So what is the hidden trait that is signaled by smiling? In economic exchange the hidden trait could be an intrinsic characteristic of the smiler (such as her degree of altruism or tendency to display reciprocity[8]), or a characteristic of the situation in which the smiler finds herself (such as the size of the pie she is proposing to share).

Apart from the paper of Owren & Bacharowski (2001) which suggested the hypothesis, there has not been to our knowledge a significant application of the costly signaling approach to understanding smiling (smiles have been considered as a coordination device, which is not at all the same thing; see Manzini et al., 2009). More broadly, however, a large number of studies in economics and psychology have in recent years investigated the importance of emotions in games. Inspired by results from affective sciences that emotions are not just some random noise but an essential part of the decision making mechanism (Damasio, 1994), theoretical and experimental work has investigated the effect of different emotions and other visceral factors on decision making (Elster, 1998; Loewenstein, 2000; Kahneman, 2003; Frijda et al., 2004). While the focus has been mostly on such negative social emotions as anger and guilt (Bosman and van Winden, 2002; Sanfey et al., 2003; de Quervain et al., 2004; Hopfensitz and Reuben, 2009), increasing attention has been given to the use of rewards and the experience of happiness (e.g. Kahneman et al., 1999; Frey, 2008; Frey and Neckermann, 2009).

---

[7] See Grafen (1990).
[8] See Gintis et al. (2003).

Altruism and cheater detection in social dilemmas has received considerable attention in economics and biology (Cosmides and Tooby, 1992; Gintis et. al., 2001). It is evident that signals that can be used to identify altruists might quickly be imitated by non-altruists and would thus not be reliable (Fehr and Fischbacher, 2005). One suggestion is that altruism as such can serve as a reliable signal of trustworthiness (Smith and Bliege Bird, 2000; Gintis et al., 2001; Lotem et al., 2003). However, in many situations, behavior of the interaction partner cannot be observed. To detect whether an interaction partner can be trusted we can either rely on third party information regarding the target individual's reputation (Sommerfeld et al., 2008) or use visual signals concerning the individual's character (Frank, 1988). Reputation requires a track record, which is not possible in one-shot interactions. In order to detect trustworthy partners with some degree of reliability in these circumstances, it is therefore necessary to base decisions on verbal or non-verbal signals sent by the partner. Just observing the partner may not be enough, however: Vogt, Efferson and Fehr (2013) report an experiment in why they use "thin slices" (short video clips) of subjects in a variety of interactional settings; other experimental subjects were not able to make use of these clips to infer trustworthiness. However, this is entirely compatible with the possibility that in a communication setting (where subjects were making clips for transmission to interaction partners) such clips might indeed convey relevant information.

Brown and Moore (2002) stress that honest signals with reliable emotional basis may be needed to guarantee positive intentions of a counterpart. This leads to the importance of 'emotional expressivity' i.e. the ability to accurately communicate your internal feeling state (Boone and Buck, 2003). To be reliable, these signals must be costly and therefore difficult to mimic. Smiles, and especially honest smiles, might be just that. Brown et al. (2003) were the first to observe that videos from self reported altruists are rated differently by neutral observers than videos of non-altruists. Further, an analysis of video recordings from altruists and non-altruists showed that self reported altruists showed more *orbicularis oculi* activity and more symmetric smiles (see also Oda et al.; 2009).

The model we develop hypothesizes that individuals may be motivated, to a greater or lesser degree, both by reciprocity and by altruism. There is a large literature addressing ways of incorporating social preferences in individual utility functions (see Sobel, 2005, for a survey). It is safe to say that there is no consensus as to the appropriate way of modelling such motivations, and it is emphatically not our intention to propose a general theory here. For instance, in many models of behavior in public goods games, individuals are considered to be motivated either by reciprocity or by altruism but not both (Fehr, Fischbacher & Gaechter, 2003); this is a useful device for focusing on the distinction between unconditional contributors and conditional contributors. Other papers (Hwang & Bowles, 2010; Brülhart & Usunier, 2004) hypothesize that individuals may have both motivations simultaneously to different degrees, and that is the approach we adopt here. This is a plausible and parsimonious way to capture the phenomenon, clearly present in many experimental studies including our own (Centorrino et al 2013), that individuals vary in their degree of trustworthiness. It is not just that some are trustworthy while others are not, but also that among individuals who are trustworthy, some are more generously or fully so than others. The combination of reciprocity and altruism in our model captures this difference, but we make no claim that it is the only modeling strategy that would do so.

3. A model of costly signaling prior to a trust interaction

3.1. Outline

In our model players engage in a trust interaction to which, for ease of exposition, we give the rather specific structure of an experimental trust game. They are able to engage in a signaling interaction before they do so. Thus, although our model represents a rather particular type of interaction between the players, the general conclusion of the analysis applies to a much more general class of economic situations, in which the parties interact without expectation of an extended relationship, one of the parties must make a commitment before the other, and the other will therefore try to signal trustworthiness in order to induce that commitment to be made.

There are two players, A and B. To avoid confusion we shall refer to A as "he" and to B as "she", though there is no intrinsic gender difference in the roles.

Player A receives a stake of value s and must decide whether or not to send it to player B (we consider s to be greater than or equal to 1, without loss of generality). If it is sent it is multiplied by three, and player B may choose to send some part of the new enlarged stake back to A. There is nothing special about the number three except that it is significantly greater than two, indicating that if the parties are willing to trust each other they can each gain significantly more than the original stake. The analysis below could be undertaken for any multiplicative factor greater than two without affecting the qualitative results, but we use the number three both to keep the exposition intuitive and because this corresponds to the factor that has typically been used in experimental settings, including that in our companion paper.

Player A's decision will be influenced by his beliefs about Player B along two dimensions – how much Player B cares about strong reciprocity, and how altruistic she is (we make these terms precise below). With respect to strong reciprocity, Player B may be one of two types $\theta \in (L,H)$; for simplicity we assume there are equal proportions of the two types in the population, though nothing of importance turns on this. H-types have stronger preferences for reciprocity than L-types (we can call these High Reciprocators and Low Reciprocators respectively). With respect to altruism, Player B has a component of her utility that is a stochastic function of the amount she sends back to A. Player B knows her own type at the start of the game, and notably when she makes a video clip in order to persuade player A to send her his stake.

If player A sends the stake, player B must decide to send back to player A a multiple m of the original stake. In principle that multiple could be chosen from a continuous interval, but to aid intuition we are interested in the choice between three types of reply, which we can call "selfish", "reciprocating" and "generous", and which we represent by $m \in (0,1,1.5)$. Since the stake has been multiplied by 3, this means player B has a choice between keeping all the stake (the selfish strategy), keeping two-thirds of it (the reciprocating strategy), and keeping half of it (the generous strategy). Note that the since the generous strategy involves the parties splitting the gains equally, it could be motivated by a desire for equality rather than altruism; the latter is the motivation we shall employ in our model. For our purposes nothing of importance turns on this point, though in other contexts it might matter which of these motivations was at work.

Prior to this interaction, Player B communicates with Player A, sending him a costly signal in the form of a smile. Then A forms a belief about B's type based on the signal. If A chooses not to send the stake the game ends, A keeps the stake and B receives a zero monetary payoff (and a total payoff that may include a cost of effort involved in sending the signal). If A chooses to send the stake then B finally chooses what multiple of the stake to return to A, and the game ends.

As is standard we solve the game backwards from the end, finding a perfect Bayesian equilibrium.

3.2. Player B's move

We model player B's motivation for returning a multiple of A's original stake using a random utility function. It is separable in money and in two types of social preference. The first social preference is for strong reciprocity, which we model as a fixed utility derived from sending back at least the original stake to player A, but not otherwise varying according to the amount sent. This utility, which differs between types, is given by α_θ, where $1 > \alpha_H > \alpha_L > 0.5$.

The second motivation is altruism, which is increasing in the amount sent back by B to A (it can be thought of as reflecting B's pleasure at knowing that she is increasing A's payoff). We model this as a utility that is a multiple β of the amount returned, plus a random error term ε. The coefficient β is itself random and may be greater or less than one (capturing the fact that, of players who return at least some money, some return only the original stake while others return a larger amount). Specifically, $\beta \in \{0.5,1.5\}$ with probability $(1 - p_\theta, p_\theta)$. We assume that $p_H > p_L$ to reflect the fact that individuals with a greater propensity for reciprocity are also likely to be more altruistic.

We therefore model player B's utility function as follows:

(1) $$U_B = 3s - ms + \alpha_\theta + \beta ms + \varepsilon \qquad \text{if } m>0$$

(2) $$U_B = 3s \qquad \text{if } m=0$$

where the error term ε has a zero mean, and is uniformly distributed between -0.5 and +0.5.

It is straightforward to see that if β=1.5, player B will always choose m=1.5, since his utility is always strictly increasing in m. Thus either type of player will choose m=1.5 with probability p_θ.

If β=0.5 on the other hand, player B's utility is strictly decreasing in *m* once *m* is positive. Thus B will either choose *m*=0 or *m*=1. The probability of choosing *m*=1 is therefore the probability that:

$$(3) \qquad \alpha_\theta > \frac{s}{2} - \varepsilon$$

If *s*=1, $\frac{s}{2} - \varepsilon$ is distributed uniformly on [0,1], so the probability that *m*=1 is just $(1 - p_\theta)\alpha_\theta$.

If *s*=2, $\frac{s}{2} - \varepsilon$ is distributed uniformly on [0.5,1.5], so the probability that *m*=1 is just $(1 - p_\theta)(\alpha_\theta - 0.5)$.

We can write this probability as a function of *s*, namely as $(1 - p_\theta)(\alpha_\theta + \frac{(1-s)}{2})$.

We therefore summarize in Table 1 the probabilities of choosing different values of *m* according to whether the player is of high or low type and whether the stakes are high or low, as follows:

[Table 1 here]

3.3. Player A's move

Player A will send the money if the expected value of doing so is greater than the sure value of keeping it.

We also model player A's decision using a random utility function. We ignore altruism on the part of player A[9] and consider his utility as given by his expected payoff plus an error term η which is uniformly distributed between –e and 0 (we can consider this as a way of allowing for risk aversion while keeping the advantages of linear utility: η=0 corresponds to risk neutrality, while η=-e is the highest risk aversion in the population).

Player A's decision then depends on γ, his subjective probability of facing a High Reciprocator type. He will send the money if the gain from receiving a net profit of half the original stake, multiplied by the probability that B chooses m=1.5, exceeds the loss of the whole original stake, multiplied by the probability that B chooses m=0. Formally, A sends the money iff:

$$(4) \qquad 0.5(\gamma p_H + (1-\gamma)p_L) + \eta \quad > \quad \gamma(1-p_H)\left(\frac{(1+s)}{2} - \alpha_H\right) \quad + \quad (1-\gamma)(1-p_L)\left(\frac{(1+s)}{2} - \alpha_L\right)$$

[9] One reason for doing so is that it is plausible that A players would be less likely to feel altruism towards those B players they believed were likely not to return them any money, and therefore the calculation how likely the B player is to return money precedes and predetermines the effect of altruism on player A's decision. This is a hard phenomenon to analyze, and we have chosen to ignore it to focus on the issues more central to this paper.

Notice that the right hand side of equation (4) is strictly increasing in s. This means that, for given γ, player A is less likely to send the money when the stakes are high than when they are low. Thus if we observe a higher probability of sending the money when the stakes are high, this must mean that A players have higher levels of γ.

Because of the uniform distribution of η, we can write the probability that an A player sends the money, given the value of γ, as q_γ, where

$$(5)\ q_\gamma = \frac{[0.5(\gamma p_H + (1-\gamma)p_L) - \gamma(1-p_H)((1+s)/2 - \alpha_H) - (1-\gamma)(1-p_L)((1+s)/2 - \alpha_L)\,]}{e}$$

Differentiating (5) with respect to γ yields:

$$(6)\quad \frac{\partial q_\gamma}{\partial \gamma} = \frac{[0.5(p_H - p_L) - (1-p_H)((1+s)/2 - \alpha_H) + (1-p_L)((1+s)/2 - \alpha_L)\,]}{e} > 0$$

Differentiating (6) with respect to s yields

$$(7)\qquad \frac{\partial^2 q_\gamma}{\partial \gamma\, \partial s} = \frac{(p_H - p_L)}{2e} > 0$$

which shows that a given increase in γ will result in a larger increase in q_γ when *s*=2 than when *s*=1. So higher stakes make the probability of sending the money more sensitive to player A's subjective probability that player B is the High Reciprocator type.

3.4. The signal

Now consider the sending of the signal. Player B invests effort e, which has an increasing convex cost $c_\theta(e)$, where $c_H(e) < c_L(e)$ for all positive values of e.

This effort produces a smile whose quality is related to the effort exerted via an increasing function $g(e, \tau)$, where τ is a random variable, and the probability distribution function $f(g|e)$ has the Monotone Likelihood Ratio Property.

We begin by assuming that this smile has a predictable positive effect on player A's subjective probability γ that player B is the High Reciprocator Type. Without such an effect neither player would have any incentive to exert any effort at all. This effect can be represented by the "smile function" $\gamma = \gamma(g)$, where $\gamma' > 0$. The function $\gamma(g)$ need not be concave but if not $c_\theta(e)$ must be sufficiently convex to yield a unique interior solution.

We next go on to show that if player B knows this, and if the quality of the smile responds to her effort, she has reason to invest effort in smiling in such a way that the smile will indeed be a positive signal not just of her effort but also of the probability that she is the High Reciprocator type. Thus A's tendency to display greater trust in individuals who have more convincing smiles is one that could be expected to evolve under natural selection since it would correspond to a real empirical regularity.

To see this, write $V_{s\theta}$ for the expected utility B will receive if player A sends the money and note that $V_{sH} \geq V_{sL}$.[10] Writing $e^*_{s\theta}$ for the optimal choice of effort by a player B who is playing for stake *s* and is of type *θ*, since $c_H(e) < c_L(e)$ it follows that

(8) $$e^*_{sH} > e^*_{sL}$$

It is also straightforward that $V_{2\theta} > V_{1\theta}$, and therefore that

(9) $$e^*_{2\theta} > e^*_{1\theta}$$

Any function $g(e, \tau)$ that has the Monotone Likelihood Ratio Property will imply that the conditional probability that Player B is the High Reciprocator Type is increasing in the value of $g(e, \tau)$. To see this note that Bayes' Law with a uniform prior implies that

[10] The reason why the expected utility for B players of type H is higher than the utility for those that are L is that they have more altruism payoff than L players do. They could choose to return the same amount as L players do and would get at least as much utility as L players from doing so. In fact they choose to return more (in expected terms) than L players do, so their expected utility must be higher.

(10) $$prob(\theta = H|g(e,\tau)) = \frac{1}{1 + f(g|e_{sL}^*)/f(g|e_{sH}^*)}$$

which is monotonically increasing in g by equation (8) and the Monotone Likelihood Ratio Property. This means that an increasing smile function $\gamma(g)$ is indeed consistent with natural selection and therefore we can predict, substituting the smile function into equation (6), that

(11) $$\frac{\partial q_\gamma}{\partial g} > 0$$

Finally, given that the convincingness of smiles is the result of effort in the way described in equation (10), we can calculate how the expected gain to A from sending money is related to smile quality. We write the expected gain to A from sending the money, conditional on smile quality as follows

(12) $$E\ (U_A|g,s)$$
$$= \left(pr(\theta = H|g)\right)\left[1.5s.p_H + s(1-p_H)(\alpha_H + \frac{(1-s)}{2})\right]$$
$$+\left(1 - pr(\theta = H|g)\right)\left[1.5s.p_L + s(1-p_L)(\alpha_L + \frac{(1-s)}{2})\right] - s$$

We can rewrite (12) as

(13) $$E\ (U_A|g,s)$$

$$= \big(pr(\theta = H|g)\big)\left[1.5s.(p_H - p_L) + s(\alpha_H - \alpha_L) - \frac{s(1-s)}{2}(p_H - p_L) + s(p_L\alpha_L - p_H\alpha_H)\right] + \left[1.5s.p_L + s(1-p_L)(\alpha_L + \frac{(1-s)}{2})\right]$$

and therefore we can write the derivative of $E\ (U_A|g,s)$ with respect to $pr(\theta = H|g)$ as

(14) $$\frac{\partial E\ (U_A|g,s)}{\partial pr(\theta = H|g)} = s[(s+0.5)(p_H - p_L) + (\alpha_H - \alpha_L) + (p_L\alpha_L - p_H\alpha_H)\ ]$$

which is strictly positive because

(15) $$[(s+0.5)(p_H - p_L) + (\alpha_H - \alpha_L) + (p_L\alpha_L - p_H\alpha_H)\ ] >$$
$$(s+0.5)(p_H - p_L) + \alpha_H(1-p_H) - \alpha_H(1-p_L) =$$
$$[(s+0.5)(p_H - p_L) - \alpha_H(p_H - p_L)\ ]$$

and the expression on the RHS is positive for any $s \geq 0.5$.

From this it follows, given (10), that

(16) $$\frac{\partial E\ (U_A|g,s)}{\partial g} > 0 \qquad \forall g,s$$

which is just the statement that the expected gain to player A from sending the stake to player B is increasing in the perceived convincingness of player B's signal.

3.5. Testable implications

Our hypothesis that smiling convincingly is a costly signal has the following testable implications:

H1: The perceived quality of Player B's smile is increasing in the size of the stake: this follows from inequality (9) given that g(.) is an increasing function;
H2: The amount sent by Player A is increasing in the perceived quality of the smile: this follows from inequality (11);
H3: The expected gain to player A from sending the stake to player B is increasing in the perceived quality of player B's smile: this follows from inequality (16).

H1 is necessary in order to distinguish this hypothesis from two alternative views: first, that smiling is a form of costless communication that solves pure coordination problems (like "cheap talk"), and secondly, that it is not communication at all but merely an outward sign of an inner emotional state (like blushing, say). H2 is necessary to explain why human beings should have evolved the habit of communicating in this costly way. H3 is necessary to explain why human beings should also have evolved the tendency to be influenced by the smiles of others.

In our companion paper (Centorrino et al 2013) we subject these three predictions to an experimental test that significantly supports all three. In addition there exists some corroborating evidence for H2 and H3 elsewhere in the literature.

Shug et al. (2010) demonstrate that individuals who display relatively cooperative tendencies as proposers in an ultimatum game are more emotionally expressive in the face of unfair treatment by others than those who do not, including in the tendency to emit Duchenne as opposed to non-Duchenne smiles, which is consistent with H2 though not directly implied by it. However, there is no test of any association between their emission of Duchenne smiles and their gestures of cooperation, and the sample is small (only 20 participants). H3 is the only one of the three to be tested directly, and has received significant support (Scharlemann et al. 2001; Johnston et al. 2010). Scharlemann et al. (2001) use still pictures, a methodology that captures only a small part of the complex interactions involved in a smile. Whether trustworthy partners can be detected from still pictures is controversial and might depend on the moment when the picture was taken[11]. Dynamic pictures might in this respect be better (Brown et al. 2003). Johnston et al. (2010) uses video clips but tests cooperation in a prisoners' dilemma (where non-cooperation is a dominant strategy, unlike in the trust game) on the basis of comparison of only two clips and cannot control for other differences between clips.

6. Conclusions

We have developed a model of smiling convincingly as a costly signal that has evolved to induce cooperation in situations requiring mutual trust. Individuals differ both in their willingness to engage in reciprocity and in their degree of altruism, and it is in their interest to signal this to others. In order to do so they must smile

[11] Yamagishi et al. (2003), Verplaetse et al. (2007). Efferson and Vogt (2013) report that viewing still pictures of men's faces does not lead to improved accuracy in predictions of trustworthiness.

convincingly, but to do so involves costly effort. The model generates three testable predictions. First, the perceived quality of Player B's smile is increasing in the size of the stake. Secondly, the amount sent by Player A is increasing in the perceived quality of the smile. Thirdly, the expected gain to player A from sending the stake to player B is increasing in the perceived convincingness of player B's smile. We test, and find support for, these three predictions in our companion paper.

References:

Boone, R. Thomas, and Ross Buck. 2003. "Emotional Expressivity and Trustworthiness: the role of nonverbal behavior in the evolution of cooperation." *Journal of Nonverbal Behavior*, 27(3): 163-182.

Bosman, Ronald, and Frans van Winden. 2002. "Emotional hazard in a power to take experiment." *The Economic Journal*, 112:147-169.

Brown, William Michel, and Chris Moore. 2002. "Smile Asymmetries and Reputation as Reliable Indicators of Likelihood to Cooperate: An Evolutionary Analysis." In *Advances in Psychology Research*, ed. S.P. Shobov, vol. 11, 59-78. New York: Nova Science Publishers.

Brown, William Michel, Boris Palameta, and Chris Moore. 2003. "Are there nonverbal cues to commitment? An exploratory study using the zero-acquaintance video presentation paradigm." *Evolutionary Psychology,* 1: 42-69.

Brülhart, Marcus, and Jean-Claude Usunier. 2004. "Verified Trust: Reciprocity, Altruism, and Noise in Trust games", Working paper 04.15, Université de Lausanne.

Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M. & Seabright, P. 2013. Honest Signaling in Trust Interactions: the power of a smile. Toulouse School of Economics, working paper.

Cosmides, Leda, and John Tooby, J. 1992. "Cognitive adaptations for social exchange." In *The adapted mind: Evolutionary psychology and the generation of culture,* ed. Jerome H. Barkow, Leda Cosmides, and John Tooby. New York: Oxford University Press.

Damasio, Antonio. 1994. *Descartes' error: Emotion, reason and the human brain.* New York: Avon Books.

Darwin, Charles R.. 1872. *The Expression of the Emotions in Man and Animals.* London: John Murray, London.

de Quervain, Dominique, Urs Fischbacher, Valerie Treyer, Melanie Schellhammer, Ulrich Schnyder, Alfred Buck, and Ernst Fehr. 2004. "The neural basis of altruistic punishment." *Science*, 305 (5688): 1254-1258.

Duchenne de Boulogne, Guillaume. 1862. *The Mechanism of Human Facial Expression.* Paris: Jules Renard.

Efferson, C. & Vogt. 2013. Viewing men's faces does not lead to accurate predictions of trustworthiness. Scientific Reports, 3, 1047 DOI: 10.1038/srep01047

Elster, Jon. 1998. "Emotions and economic theory." *Journal of Economic Literature*, 36 (1): 47–74.

Fehr, Ernst, and Fischbacher, Urs. 2005. "Altruists with Green Beards." *Analyse & Kritik*, 27/2005, 73–84.

Fehr, Ernst, Urs Fischbacher, and Simon Gächter, S. 2003. "Strong Reciprocity, Human Cooperation and the Enforcement of Social Norms." *Human Nature*, 13 (1):1-25.

Frank, Robert H. 1988. *Passions within Reason. The Strategic Role of the Emotions*. New York: Norton.

Frey, Bruno. 2008. *Happiness: A Revolution in Economics*. Cambridge, MA and London, UK: MIT Press.

Frey, Bruno, and Suzanne Neckermann. 2009. "Awards: A Disregarded Source of Motivation." In *Perspectives in Moral Science – Contributions from philosophy, economics, and politics in honour of Hartmut Kliemt*, ed. Michael Baurmann and Bernd Lahno, 177-182. Frankfurt: Frankfurt School Verlag.

Frijda, Nico, Antony Manstead, and Agneta Fischer. 2004. "Feelings and emotions: Where do we stand?" In *Feelings and Emotions: The Amsterdam Symposium*, ed. Antony Manstead, Nico Frijda, and Agneta Fischer. Cambridge: Cambridge University Press.

Gintis, Herbert, Samuel Bowles, Robert Boyd, and Ernst Fehr. 2003. "Explaining altruistic behavior in Humans." *Evolution and Human Behavior,* 24: 153-172.

Gintis, Herbert, Eric Alden Smith, and Samuel Bowles. 2001. "Costly Signaling and Cooperation". *Journal of Theoretical Biology*, 213(1): 103-119.

Grafen, Alan. 1990. "Biological signals as handicaps." *Journal of Theoretical Biology,* 144: 517-546.

Hopfensitz Astrid, Ernesto Reuben. 2009. "The importance of emotions for the effectiveness of social punishment." *The Economic Journal,* 119: 1534-1559.

Hwang, Sung-Ha, and Samuel Bowles. 2010. "Is Altruism Bad for Cooperation?" Working paper, University of Massachusetts.

Johnston, Lucy, Lynden Miles, and C. Neil Macrae. 2010. "Why Are You Smiling at Me? Social functions of enjoyment and non-enjoyment smiles." *British Journal of Social Psychology,* 49: 107-127.

Kahneman, Daniel, Ed Diener, and Norbert Schwarz (Eds.). 1999. *Well-being: The foundations of hedonic psychology*. New York: Russell Sage Foundation.

Kahneman, Daniel. 2003. "A psychological perspective on economics." *The American Economic Review*, 93 (2): 162–168.

Krumhuber, Eva, Anthony S.R. Manstead, Darren Cosker, Dave Marshall, Paul L. Rosin, and Arvid Kappas. 2007. "Facial dynamics as indicators of trustworthiness and cooperative behaviour." *Emotion,* 7: 730-735.

Loewenstein, George. 2000. "Emotions in economic theory and economic behaviour." *American Economic Review*, 90(2): 426–432.

Lotem, Arnon, Michael A. Fishman, and Lewi Stone. 2003. "From reciprocity to unconditional altruism through signalling benefits." *Proceedings of the Royal Society B,* 270: 199-205.

Niedenthal, Paula, Martial Mermillod, Marcus Maringer, and Ursula Hess. 2010. "The Simulation of Smiles (SIMS) Model: Embodied Simulation and the Meaning of Facial Expression." *Behavioral and Brain Sciences,* 33 (06): 417-433.

Oda, Ryo, Noriko Yamagata, Yuki Yabiku, and Akiko Matsumoto-Oda. 2009. "Altruism Can Be Assessed Correctly Based on Impression." *Human Nature,* 20 (3): 331-341.

Owren, Michael J, Jo-Anne Bachorowski. 2001. "The evolution of emotional expression: A "selfish-gene" account of smiling and laughter in early hominids and humans." In *Emotions: Current Issues and Future Directions*, ed. Tracy J. Mayne, and George A. Bonnano, 152-191. New York: Guilford Press.

Sanfey, Alan, James Rilling, Jessica Aronson, Leigh Nystrom, and Jonathan J. Cohen. 2003. "The neural basis of economic decision making in the ultimatum game." *Science*, 300 (5626):1755-1758.

Scharlemann, Jörn PW, Catherine C. Eckel, Alex Kacelnik, and Rick K. Wilson. 2001. "The value of a smile: Game theory with a human face." *Journal of Economic Psychology,* 22: 617-640.

Schug Joanna, David Matsumoto, Yutaka Horita, Toshio Yamagishi, and Kemberlee Bonnet. 2010. “Emotional expressivity as a signal of cooperation.” *Evolution and Human Behavior,* 31 (2): 87-94.

Smith, Adam. 2000. *The Theory of Moral Sentiments*, New York: Prometheus Books.

Smith, Eric Alden, and Rebecca L. Bliege Bird. 2000. “Turtle hunting and tombstone opening: public generosity as costly signalling.” *Evolution and Human Behavior*, 21(4): 245-261.

Sobel, Joel. 2005. “Interdependent Preferences and Reciprocity.” *Journal of Economic Literature*, Vol. XLIII (June), 392–436.

Sommerfeld, Ralf D., Hans-Jürgen Krambeck, and Manfred Milinski. 2008. “Multiple gossip statements and their effect on reputation and trustworthiness.” *Proceedings of the Royal Society B*, 275: 2529-2536.

Spence, Michael. 1974. *Market Signaling: Informational Transfer in Hiring and Related Screening Processes*. Cambridge: Harvard University Press.

Verplaetse, Jan, Sven Vanneste, and Johan Braeckman. 2007. “You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection.” *Evolution and Human Behavior,* 28: 260-271.

Vogt, S., C. Efferson and E. Fehr (2013). Can We See Inside? Predicting Strategic Behavior Given Limited Information. Evolution and Human Behavior. http://dx.doi.org/10.1016/j.evolhumbehav.2013.03.003

Yamagishi, Toshio, Shigehito Tanida, Rie Mashima, Eri Shimoma, and Satoshi Kanazawa. 2003. "You can judge a book by its cover: Evidence that cheaters may look different from cooperators." *Evolution and Human Behavior,* 24: 290-301.

Zahavi, Amotz. 1975. "Mate Selection – A Selection for a Handicap." *Journal of Theoretical Biology*, 53 (1): 205-214.

Table 1: Probabilities that player B chooses various values of *m*

	***m*=0**	***m*=1**	***m*=1.5**
High Reciprocator type (θ=H)	$(1-p_H)(\frac{(1+s)}{2}-\alpha_H)$	$(1-p_H)(\alpha_H+\frac{(1-s)}{2})$	p_H
Low Reciprocator type (θ=L)	$(1-p_L)(\frac{(1+s)}{2}-\alpha_L)$	$(1-p_L)(\alpha_L+\frac{(1-s)}{2})$	p_L