Accepted Manuscript

Honest signalling in trust interactions: smiles rated as genuine induce trust and signal higher earnings opportunities

Samuele Centorrino, Elodie Djemai, Astrid Hopfensitz, Manfred Milinski, Paul Seabright

PII: S1090-5138(14)00100-7
DOI: doi: 10.1016/j.evolhumbehav.2014.08.001
Reference: ENS 5932

To appear in: *Evolution and Human Behavior*

Received date: 18 October 2013
Revised date: 23 July 2014
Accepted date: 5 August 2014

Please cite this article as: Centorrino, S., Djemai, E., Hopfensitz, A., Milinski, M. & Seabright, P., Honest signalling in trust interactions: smiles rated as genuine induce trust and signal higher earnings opportunities, *Evolution and Human Behavior* (2014), doi: 10.1016/j.evolhumbehav.2014.08.001

Honest signalling in trust interactions: smiles rated as genuine induce trust and signal higher earnings opportunities

Samuele Centorrino (Stony Brook University), Elodie Djemai (PSL, Université Paris Dauphine, LEDa, IRD, UMR DIAL), 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: August 2014

Abstract: We test the hypothesis that smiles perceived as honest serve as a signal that has evolved to induce cooperation in situations requiring mutual trust. Potential trustees (84 participants from Toulouse, France) made two video clips averaging around 15 seconds for viewing by potential senders before the latter decided whether to 'send' or 'keep' a lower stake (4 euro) or higher stake (8 euro). Senders (198 participants from Lyon, France) made trust decisions with respect to the recorded clips. If money was sent to the trustee, stakes were tripled and trustees could decide to keep all, two thirds or one half of the tripled stakes. Clips were further rated concerning the genuineness of the displayed smiles. We observe that smiles rated as more genuine strongly predict judgments about the trustworthiness of trustees, and willingness to send them money. We observe a relation between costs and benefits: smiles from trustees playing for higher stakes are rated as significantly more genuine. Finally, we show that those rated as smiling genuinely return more money on average to senders. An increase of one standard deviation in rating of smile genuineness is associated with an unconditional expected gain of about one dollar and thirty cents to senders in the two trials of the experiment. Potential gains for senders could be significantly increased from taking smiles rated as genuine into account.

Keywords: honest signaling, smiling, experiment, trust game, video

* Corresponding author. E-mail address: Paul.Seabright@tse-fr.eu

1. Introduction

Smiling is a form of behavior that is found in all human societies and plays a central part in human communication (Darwin 1872; Ekman 1982; Niedenthal et al. 2010). There is scientific consensus that viewers perceive smiles as varying in their degree of "genuineness" or "convincingness". Since the work of Duchenne (1862) and Darwin (1872) many researchers have attempted to identify objective measures of honest smiles, concluding that genuine smiles are characterized by use of the *orbicularis oculi* (the muscle surrounding the eyes) in combination with the *zygomatic major* (raising the corners of the mouth); symmetry is also an important characteristic. More recent research focuses on the importance of temporal dynamics such as smile onset, apex, and offset durations (Krumhuber et al. 2007). Smiles perceived as genuine 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. Smiles also induce both conscious and unconscious mimicry (Niedenthal et al. 2010). 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?

A large literature exists in affective sciences concerning the identification of honest smiles (see for example Ekman and Friesen 1982, Ekman and Rosenberg, 2005) through coding of facial activity. Naive untrained observers correlate in their evaluations with categorizations by experts, but are far from reaching the same accuracy as experts. Since the effect of a signal depends on the perception of this signal by the receiver, we will in this paper focus on smiles that are subjectively 'perceived' as being genuine. We test the hypothesis that smiles perceived as genuine are an honest signal of cooperation opportunities for situations requiring mutual trust. We observe trust and trustworthiness behavior in a two person trust game where senders observed short video clips of trustees before taking their decisions. These video clips were further rated by participating senders in the study along a number of dimensions, among which the genuineness of the trustee's smile. We hypothesize that:

H1: Senders will be more willing to trust those trustees who are able to produce smiles rated on average as genuine.

H2: Higher stake trust games will provoke more genuine smiles (as rated by all senders).

H3: On average, trusting those who produce smiles rated on average as more genuine, will lead to higher earnings for senders.

H1 is important in explaining why human beings should have evolved the habit of communicating in this way; there would be no point unless it succeeded in influencing the behavior of others. H2 is important in distinguishing the idea that smiles rated as genuine are an honest signal of some cooperation opportunity from two alternative views: first, that it is a form of costless communication that solves pure coordination problems (like "cheap talk"), and second, that it is not communication at all but merely an outward sign of an inner emotional state. H3 is important to explain why human beings should also have evolved the tendency to be influenced by the smiles of others. The evolutionary process by which smiling developed surely involved a good deal of repeated interaction between individuals who knew each other well, but like many other human adaptations, smiling may have been used opportunistically for interactions with strangers as these became more frequent in later history.

There exists some corroborating evidence for H1 and H3 in the literature. Schug 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 H3. However, there is no test of any association between their emission of Duchenne smiles and their gestures of cooperation, and the sample is small (20 participants). Mehu et al. (2007a) suggest that human smiles are more prevalent in situations that involve sharing or exploitation of resources. By filming sixty pairs of friends during a neutral and a sharing decision they observe that significantly more Duchenne smiles are produced during sharing situations. Thus situations requiring sharing elicit smiles and laughter (Mehu and Dunbar 2008).

Whether trustworthy partners can be detected from still pictures is controversial and might depend on the moment when the picture was taken (Yamagishi et al. 2003; Verplaetse et al. 2007) and whether trust evaluations are explicit or based on actions (de Neys et al. 2013). Efferson and Vogt (2013) report that viewing still pictures of men’s faces does not lead to improved accuracy in predictions of trustworthiness. Dynamic pictures might in this respect be better (Brown et al., 2003).

However Vogt, Efferson and Fehr (2012) used short video clips of subjects in a variety of interactional settings that were not explicitly directed at a partner; other experimental subjects were not able to use these clips to infer trustworthiness.

H1 is the only one of the three hypotheses that has been be tested directly, and has received significant support (Scharlemann et al. 2001; Johnston et al. 2010). Scharlemann et al. (2001) use still pictures and observe that participants trust more when seeing a smiling image of their partner. Johnston et al. (2010) use video clips and observe more trust in response to enjoyment smiles. In contrast to our study they test cooperation in a prisoners' dilemma (where non-cooperation is a dominant strategy, unlike in the trust game, where non-cooperation is the unique subgame perfect equilibrium strategy but is not a dominant strategy). They do so on the basis of comparison of only two clips, and cannot control for other differences between clips. Mehu et al. (2007b) assess what characteristics are associated with honest smiles by rating fifty faces across ten attributes. It turns out that Duchenne smiles play a significant role in the assessment of generosity and extraversion.

The phylogeny of smiling further suggests that it leads observers to behave less aggressively. The "horizontal silent-bared teeth" display (involving strong horizontal, as well as vertical, lip retraction; teeth and gums are exposed, but the mouth itself is closed) in non-human primates can be regarded as an analogue of our human smiling. It is assumed to have an appeasing or re-assuring function; its sender is usually the inferior partner; it may also be a signal in a process of negotiation between two individuals (Preuschoft and Van Hoof 1997). This suggests that in humans, smiling could serve as a kind of mimicry of submission, used by dominant partners to assure others that they will not abuse the opportunities for betrayal of the trust of others.

More generally faces seem to be consistently rated concerning their trustworthiness, which is mirrored by actions. Van 't Wout and Sanfey (2008) observe that judgments of facial trustworthiness are related to sending money in a trust game. Trustworthiness ratings are also a significant predictor of how much money these players received in one-shot trust game, a finding replicated for repeated trust games (Chang et al. 2010).

Even though our investigation of perceived honesty of smiles as an honest signaling device is novel, 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; Ketelaar 2006) and the information conveyed by emotional display (Parkinson, 2005). Smiles are an expression of experienced happiness and might be used as a coordination device (Manzini et al. 2009), but might also be an important component in social exchange (Owren and Bacharowski 2001).

Signaling has been extensively studied both in economics since Veblen (1899) and Spence (1974), and independently in biology since Zahavi (1975). Signals have been defined as “an act or structure that alters the behavior of another organism, which evolved because of that effect, and which is effective because the receiver’s response has also evolved” (Maynard Smith and Harper, 2003). A costly signal (or handicap) further imposes a cost on its bearer (a pecuniary or non-pecuniary effort cost in economics, a fitness cost in biology) by which reliability is ensured. Specifically it indicates the presence of some advantageous hidden trait because the signal is more difficult to send for those individuals who do not possess the trait than for those who do (Grafen 1990). For an overview of different definitions, specifically concerning the type and size of costs and the type of information conveyed, see Maynard-Smith and Harper (2003). We conjecture that the hidden trait associated with smiles perceived as honest could be an intrinsic characteristic of the smiler (such as her degree of altruism or tendency to display reciprocity as in Gintis et al. 2003), a medium-term state (such as good mood) or a characteristic of the situation in which the smiler finds herself (such as the size of the pie she is proposing to share). It could also be a combination of any of these.

To test our hypotheses we observe non-verbal behavior in an economic experiment involving trust. In a trust game first movers (called "senders") each decide whether to send a sum of money to a second player, called a trustee. If they do so the sum is tripled, and the trustee may divide this sum between himself and the sender. In our experiment, trustees made short video clips to be shown to senders before the latter took their decision. Participants knew that this was their only mean to convince their partner to trust them.

To detect whether an interaction partner can be trusted we can normally 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). Indeed it has been observed that players in trust games are willing to spend money relying on visual information of their partner (Eckel and Petrie

2008). The kind of visual information used is however not clear, but smiles perceived as honest could play a crucial role in that respect (Cohn and Smith 2004).

Altruism and cheater detection in social dilemmas have 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 no longer 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. In order to detect trustworthy partners reliably in one-shot interactions, it is therefore necessary to base decisions on verbal or non-verbal signals sent by the partner.

2. Experimental Methods and Data Description

2.1. Methods

We use a simplified version of the original trust game, proposing senders a binary choice of trust or no trust and trustees three different return options (see Hopfensitz and Reuben 2009). Sessions for trustees were conducted first, to allow them to record their video messages. Their actions were obtained at the same time using the strategy method (in other words, they reported what they would return to senders in case the sender decided to trust them). Decisions for senders and trustees were incentivized and earnings were determined according to their partners' decisions. Therefore trustees were not rewarded until after senders had made their respective trust decisions.

Video messages were produced by eighty-four volunteers aged between 18 and 35 years recruited from the general population in Toulouse, France. We told trustees they would face two different but unknown partners who might be persuaded to send them a sum that would be tripled if they did so. For the first partner, trustees recorded their video message before being informed about the precise payoffs and thus before taking their decision, while for the second they recorded their message after taking their return decision. These two treatments were always presented in the same order, due to the impossibility of having participants first play a game where they are informed about payoffs and then a game where they do not know about the different payoff options, as the options are constant over the two games. The trustee had to choose between returning: nothing, the sender's original stake or 1.5 times the sender's original stake (thus half the total amount). Trustees were randomly split into two payoff treatments: senders' stakes in the Lower stake treatment were 4 euros (approx. 5 $) per game (thus trustees would have to decide about the split of 12 euros, approximately 15 $, in case of a trust decision), and senders' stakes in the Higher stake treatment were 8 euros (approx. 10$) per game (thus trustees would have to decide about the split of 24 euros, approximately 30 $, in case of a trust decision). Average earnings for trustees were 2.7 euros per game in the Lower and 6.9 euros per game in the Higher stake treatment (there were two games per experimental session).

Video clips were made on a professional TV platform located on the campus of the Toulouse School of Economics, and a practice clip helped participants become familiar with the environment (for detailed methods see the online appendix). A total of 168 video clips was obtained, two for each trustee. Verbal messages during clips were standardized by giving a predetermined sentence to trustees that had to be memorized. The sentence (in French) was: "Hello, I am [*name*], I am [*age*] years old. I have been living in Toulouse since [*year*]. I am [*occupation*]. I am very happy that you have accepted to participate in this game with me. I do hope that you will trust me and that you will play with me." To make the message natural for trustees it included their name, age and occupation, and trustees were reassured that the precise wording did not matter. Participants were not allowed to view their recorded clips, and only in exceptional cases (outside noise, complete black-out of text) allowed to re-record their clip. Video clips lasted around fifteen seconds on average, with the fastest at around ten and the slowest at around twenty seconds.

Senders' behavior and evaluation of video clips were obtained in a different experimental laboratory to minimize the risk that senders might recognize trustees. A total of 198 student participants were recruited at the University of Lyon; 84 senders participated in a first wave and 114 in a second. The difference between the two waves was that participants in the first wave were matched with trustees and their decisions determined trustees' payoffs, as trustees had previously been informed. Observations from the first wave therefore concern games where sender and trustee faced the same stake size. We implemented a second wave, in which new participants again made trust decisions with respect to the same video clips and were paid according to trustees' initial decision. In the second wave, senders were informed only that trustees might return nothing, their initial endowment or 1.5 times their endowment. Note that since trustees had by this time already been paid, decisions of senders in this second wave did not influence trustees' earnings. Senders in the second wave were endowed with initial stakes of 4 Euros (as in the Lower stake treatment), with 8 Euros (i.e. as in the Higher stake treatment), or with a new Extra-High treatment of 12 Euros.

In the first and second waves respectively, each sender viewed a total of 42 (respectively 28) clips in two series of 21 (resp. 14). Senders were presented twice with each series. When first seeing each clip senders were asked to decide whether to send money to the trustee. They were then again presented the same series of clips and asked to rank the trustee on an 8-point scale along a number of dimensions including how much they smiled, how genuine were their smiles, their attractiveness, their trustworthiness, their intelligence and their self-confidence (from: 1=not at all to 8=very much so). To determine payouts, senders were matched at random with two trustees (one from each series) and received payoffs based on the actual decisions of these partners. To ensure anonymity for trustees it was not revealed which of the clips viewed had been selected to determine senders' payoffs. From the first wave a total of 21 decisions and ratings concerning each of the 168 clips was obtained. From the second wave an average of 19 decisions and ratings for each clip were obtained. Clips in our dataset have an average of 40 ratings each, with a minimum of 38 and a maximum of 45. Figure 1 summarizes the timeline of choices made by senders and trustees.

Note that during the first wave senders and trustees faced the same stake size (i.e. initial stakes available to senders were either 4 or 8 euros). To disentangle the direct effect of stake size on senders (more money at stake for sender) from the indirect effects (more money as stake for trustee), half of a sender's video messages during the second wave came from trustees in the Higher stake treatment and half came from the Lower stake treatment.

This set-up enabled us to investigate separately the effect of stake size on trustee clip characteristics and on sender behavior. All results reported use the pooled data from the two waves of the experiment. A dummy variable distinguishing the two waves was never significant in any specification, indicating that the two waves were conducted under indistinguishable conditions.

[Figure 1 and Table 1 here]

2.2. Descriptive statistics

Characteristics of the clips are based on ratings by senders, who would be thereby motivated to observe each clip carefully and would be less likely to be influenced by irrelevant factors in their evaluations. For each clip the large number of ratings avoids idiosyncratic reactions of individual raters. In our analyses below we use the average rating by all senders to predict any individual sender's behavior in order to avoid possible reverse causality whereby senders might seek to "justify" their decisions to send money by rating clips accordingly. This averaging method also avoids biases arising from possible systematic differences in ratings by individual senders, some of whom may be systematically more "positive" than others. As a robustness check we standardized the individual ratings by the mean and the variance of the rater before computing the average for each trustee, and similar qualitative results (not reported here) were obtained.

Table 1 presents a correlation table between the six characteristics rated for each clip. We observe a strong correlation between clips rated as showing genuine smiles and ratings of trustworthiness, attractiveness and intelligence. The mean rating concerning the intensity of smiles was 4.85 (median 4.87; st. dev. 0.542) and of the genuineness of smiles was 4.95 (median 4.96; st. dev. 0.341).

To identify characteristics of clips receiving high ratings concerning smile genuineness, facial movements in clips were automatically analyzed with the commercial facial analysis tool FaceReader 5 (Bijlstra and Dotsch, 2011). We focus here on the action units related to genuine smiles (Ekman, 2005). Specifically the percentage of the clips duration showing activation of the *zygomaticus major* (AU12); and combinations with the *orbicularis oculi* (AU12+AU6). FaceReaders' classification of AU6 and AU12 has been validated on the Amsterdam Dynamic Facial Expression Set and reaches accuracy of 88 and 82 percent, respectively (den Uyl, 2013). Since raters could record their ratings at any time during the viewing of the clip, we consider the first two-second interval of each clip and the first four-second interval of each clip (i.e. the first 50 and 100 frames, respectively). In these intervals 57% and 59%, respectively, of clips show an activation of either AU6 or AU12. We observe a significant rank correlation between the combined activation of AU12 and AU6 and ratings of smile genuineness (Kendalls' tau: 2 seconds: 0.08, p=0.041; 4 seconds: 0.07, p=0.099) and a non-significant rank correlation with lip corner pulls (Kendalls' tau: AU12: 2 seconds: 0.075, p=0.117; 4 seconds: 0.066, p=0.178). Using a rank measure of correlation is required because of the substantial number of zeros in the activation unit data.

Characteristics of our participants for each group of players are summarized in Table 2. The first column reports the means and standard deviations for the senders. This subsample is exactly gender-balanced, the average age is 22 years and 92% are currently students. For a given sender, the percentage of decisions for which she decided to send money ranges from 0 to 100%. Of the 198 participants, 3 participants decided to send money to every partner, and 20 decided never to send money. On average, senders decided to cooperate with 38% of their partners. Column (2) concerns the trustees, who are 25 years old on average. The sample of trustees comes from a less homogeneous population than the senders since only 46% are students. The sample of trustees is fairly balanced with 55% of women.

[Table 2 here]

As described above, the trustees recorded two video clips and made a sharing decision for each clip. The distribution of choices in these two decisions is not statistically different. In the first decision, 55% of the participants decided to share equally the amount received (i.e. the stake multiplied by 3), 31% decided to send back the original stake to the sender and keep 2/3 of the total pie, and the remaining 14% decided to return nothing. In the second decision, the distribution suggests a slight shift from equal sharing to sending nothing (49% share equally; 32% return original stakes and 19% return nothing), but this is not statistically significant. Overall, trustees' decisions had the result that the unconditional expected gain to senders from sending money as opposed to keeping it was a little above zero (those in the Lower stake treatment could expect to make 4.6 Euros instead of 4; those in the Higher stake treatment could make 8.4 Euros instead of 8, while those in the Extra-High treatment could make 13.1 Euros instead of 12). This therefore provides an excellent environment in which to test how smile characteristics influence the conditional expected gain.

[Table 3 here]

We further observe that senders' average ratings were not significantly different whether they watched the first or the second clip.

Table 3 summarizes other observable characteristics for trustees in the Lower and Higher stake treatments. Only the probability of having a facial piercing differs significantly from one treatment group to another, suggesting that the treatment can be considered as random.

3. Results

Figure 2 shows a comparison of Lower and Higher stake treatments of trustees in terms of the average ratings by senders of the genuineness of their smiles, their average trustworthiness rating, the proportion of senders who decided to send money, and the proportion of trustees who chose to return a positive amount of money.

[Figure 2 here]

Figure 2 indicates that trustees under the Higher stake treatment are perceived as having more genuine smiles and as being more trustworthy, and are associated with a higher percentage of senders sending money, although a smaller percentage of trustees under the Higher stake treatment actually return any money to the senders. One could think that the positive correlation of the Higher stake treatment with perceived smile genuineness and with selfish behavior would imply that smiles perceived as genuine are positively correlated with selfish behavior. However, Figure 3 indicates that this is not so. The explanation is that those in the Higher stakes group who *succeeded* in making smiles perceived as genuine were not a random subset; they were a more unselfish group than those who did not succeed.

In Figure 3 clips are divided into those whose smiles were given average ratings above 5 (46% of the clips) and the rest. As already observed from the correlations, clips with smiles perceived as genuine were given higher ratings for trustworthiness, attractiveness and intelligence, and were associated with a higher willingness to send money, but were also associated with a slightly higher willingness to return at least some money to senders. The latter is not statistically significant, but the association is not negative as Figure 2 might have led us to expect. Similarly we observe no significant rank correlation between the decision by trustees and the concurrent activation of action units 06 and 12 during the clip (Kendall's tau: 0.054 p=0.191). So overall it appears that the Higher stake treatment created both a higher incentive to smile in a way perceived as genuine, and a higher incentive to be selfish instead of returning money to the sender, but that those in the High treatment who succeeded in creating smiles perceived as genuine were more likely to return the money than those who did not. In Figures 2 and 3, most of the mean comparisons are significant at or near 5% levels, except for the unselfishness comparison in Figure 3 which is insignificant; the majority (7 out of 9) are significant at well under 1%. Standard errors are clustered by trustee to take account of the correlation between the characteristics of the two clips made by each trustee.

[Figure 3 here]

We now turn to multivariate regression analysis. We consider our three hypotheses in turn. In all cases, in order to avoid possible "justification effects" in which users' ratings are influenced by the decisions to send money they have already taken,, we use as measures of smile quality, trustworthiness and attractiveness the average rating of each clip across all viewers, rather than the rating given by the individuals themselves; this requires, however, that standard errors be calculated clustering by clip.

Table 4 reports the tests of our first two hypotheses. Equation A tests hypothesis H2, that Higher stake trust games will provoke more smiles rated as genuine. The treatment effect is significant at under 2%: 0.12 points, which is about 36% of one standard deviation of the distribution of mean ratings by clip.

[Table 4 here]

Other notable features of Equation A are that smiles rated as genuine are associated with trustees who are rated as more intelligent, and also with older trustees. Trustees with beards are rated as having less genuine smiles, and women with part of their cleavage exposed are rated as having significantly more genuine smiles (it is not clear whether the causal mechanism is via the psychology of the smiler or of the viewer). There is neither a significant effect of gender for either the sender or the trustee, nor of perceived attractiveness of the trustee, and no effect of whether the clip is filmed before or after the decision to return the money has been made. These coefficients (like those of other controls) are not reported, though the full specification is available from the authors.

Equations B and C test hypothesis H1, that senders will be more willing to trust those trustees who are able to produce smiles rated as genuine. First, Equation B considers whether smiles rated as genuine are associated with judgments of greater trustworthiness. There is a massively significant correlation (t-ratio of over 8); a one-point deviation increase in genuineness rating of the smile is associated with slightly more than a half point increase in perceived trustworthiness. Perceived intelligence is also positively and very significantly correlated with perceived trustworthiness, and there is a significant effect of the Higher stake treatment independently of perceived smile genuineness, suggesting that trustees are putting effort into other dimensions of non-verbal communication as well.

Equation C examines whether smiles rated as genuine lead to an increased probability of sending money to the trustees. Once again there is a massively significant association: a one point increase in perceived smile genuineness leads to a 21% increase in the probability of sending the money, which is equivalent to a 7% increase per standard deviation of the rating of smile genuineness. Perceived intelligence is again a very important factor in the decision.

Table 5 reports two tests of hypothesis H3, that on average, trusting those that produce smiles rated as more genuine will lead to higher earnings for senders. We calculate for each sender the potential gain from sending money, which is defined for those who sent money to trustees as their actual gain, and is defined for those who did not send money as the amount of money they would have gained if they had sent it. In equation D we test the hypothesis that the potential gain is unconditionally associated with smiles rated as more genuine. We therefore regress the potential gain on the mean genuineness rating of the smiles, and on nothing else. Mean ratings of smile genuineness are significant at slightly under 7% and the magnitude of the effect is economically significant too (a gain of 1.5 Euros for each point on the scale). To put this in perspective, an increase of one standard deviation in rating of smile genuineness is associated with an unconditional expected gain of about 0.5 Euros per trial or 1 Euro for the whole experiment (equivalent to around one dollar and thirty cents).

The second variant of the hypothesis we test in Equation E is that, conditional on what the sender otherwise knows, she could significantly improve her potential gain by accounting for smiles rated as genuine. This requires regressing the potential gain on mean smile genuineness ratings plus other reported characteristics including the size of the stake (which the sender knows). Here the coefficient on mean smile rating is around half as large as in the unconditional regression, and the effect is significant at around 4%.

[Table 5 here]

Finally, we investigate whether smiles perceived as genuine are associated more closely with the amount of money the trustee has available to offer the sender or with the willingness of the trustee to behave unselfishly. We therefore regress a dummy variable for the Higher stakes treatment on the mean genuineness rating of the smile, and in a separate equation we regress on the same measure of smile genuineness a dummy variable indicating that the trustee takes an unselfish decision (i.e., returns a non-zero amount to the sender). The purpose is to see whether the rating of the genuineness of the smile is a reliable signal of the amount available to share, and of the character of the trustee. As reported in Table 6, smiles rated as genuine are positively related to both the size of the pie to be shared and the unselfish behavior by the trustee. Two caveats are in order, however. First, the coefficient on unselfish behavior by the trustee has a large standard error so we cannot be confident in its measurement, and thus it is not significantly different from zero at conventional levels (unlike the coefficient on the Higher stakes treatment). Moreover, as we saw in Figure 2, unselfish behavior itself appears to be influenced by the treatment, so we cannot be confident in treating it as a measure of the character of the trustee. So we should conclude that smiles rated as genuine are definitely informative about high cooperation opportunities, and *may* also signal the character of the smiler.

[Table 6 here]

4. Conclusions

We have tested a three component hypothesis that smiles perceived as genuine are an honest signal that has evolved to induce cooperation in situations requiring mutual trust. All three components are supported by the evidence. First, senders are more willing to trust those trustees that are able to produce smiles rated as genuine. Smiles rated as more genuine are strong predictors of judgments about the trustworthiness of trustees, and of the revealed willingness to send them money. Secondly, higher stake trust games provoke more smiles rated as genuine. This suggests either that they reflect a more positive affective state produced by the higher stake, or that they are produced when there are rewards to any additional effort required. Finally, we show that on average, trusting those who produce more smiles rated as genuine, will lead to higher earnings for senders. Trustees who were rated as smiling more genuinely return more money on average to senders. It is clearly informative of the amount the trustee has available to share with the sender; there is weaker evidence that it may also be a signal of the intrinsic trustworthiness of the trustee independently of the amount at stake.

There are various possible reasons why smiles perceived as genuine might have evolved to be honest signals and to be hard to fake. There might just be some intrinsic correlation between the ability to produce the signal and the possession of the trait. Alternatively, the signal might require conspicuous waste (e.g. cost in times of energy or cognitive effort) or conspicuous precision (e.g. coordination, well-timed and symmetric activation of eye and cheek muscles). The narrowing of the visual field implied in focusing attention on the person to whom the smile is directed might also be a form of signaling that the smiler is willing to forgo other objects of attention.

Indeed, smiling might be interpreted as a form of costly communication, a hypothesis, that was first suggested by Owren and Bacharowski (2001). It can be shown that the three hypotheses tested in this paper are in line with three components necessary for smiling to be a costly signal (Centorrino et al 2014). Concerning the behavior of the smiler and the target of the smile these components imply that smiling "genuinely" needs to be :

- causally effective in inducing the target to cooperate with the smiler, and
- a reliable signal of the likely benefits to the target of cooperating with the smiler.

Which is in line with our hypotheses H1 and H3.

Under the Owren and Bacharowski hypothesis, it is further necessary that producing smiles perceived as genuine is costly to the smiler, involving some degree of cognitive effort for the smiler. Anecdotal evidence (from job interviews, for example) suggest that individuals try to “make an effort” to smile convincingly when there is a good enough reason to do so. We suggest that the test of our hypothesis H2 is also indirect evidence for such a cost. Specifically larger benefits (from larger stakes) make it easier to create smiles perceived as genuine. However, our evidence cannot distinguish between the hypotheses that the larger benefits do this by making individuals more willing to undertake the necessary cognitive cost, or by simply putting the smiler in a more positive affective state.

The gains from each instance of smiling are not small (around one dollar and thirty cents in the whole experiment for each increase of one standard deviation in smile quality). In our experiment as doubtless in real life, smiling to engage others is an activity in which we engage many times a day. Our results suggest that the importance of smiling for building social trust may explain why we engage in this form of communication that might otherwise seem so pointless.

Acknowledgements: The authors are grateful to Lecio Costa-Simoes for excellent and very professional filming in Toulouse; to Marie-Claire Villeval and her team at GATE at the University of

Lyon, especially Sylvain Ferriol, for their hospitality, encouragement and technical support; and the Institute of Advanced Studies Toulouse (IAST) and the Max Planck Institute for Evolutionary Biology in Plön for financial support. Automated facial ratings were funded by ANR grant ANR-2010-JCJC-1803-01. We thank Ruxanda Berlinschi, Julia Bird, Claudia Capozza, Marie Lalanne, Maxime Marty, Andrea Matranga, Racha Ramadan, Karine Van Der Straeten and Stéphane Peysson for participation in pilot versions, and Jean-Louis Rullière, Alice Seabright and James Tremewan for valuable comments on preliminary results. Seminar audiences in Auckland, Hong Kong, Oxford, Santa Fe and at the HBES conference in Montpellier gave valuable feedback.

References:

Bijlstra, G. and R. Dotsch (2011). FaceReader 4, emotion classification performance on images from the Radboud Faces Database. Technical Report. Retrieved at: http://gijsbijlstra.nl/wp-content/uploads/2012/02/TechnicalReport_FR4_RaFD.pdf

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. (2014). A Model of Smiling as a Costly Signal of Cooperation Opportunities. Toulouse School of Economics, working paper.

Chang, L. J., Doll, B.B, van 't Wout, M., Frank, M.J, & Sanfey, A.G. (2010). Seeing is believing: Trustworthiness as a dynamic belief. *Cognitive Psychology*, 61, 87–105.

Cohn, J. F., & Schmidt, K.L. (2004). The timing of facial motion in posed and spontaneous smiles. *International Journal of Wavelets, Multiresolution and Information Processing*, 2, 1-12.

Cosmides, L., & 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, A. (1994). *Descartes' error: Emotion, reason and the human brain.* New York: Avon Books.

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

De Neys, W., Hopfensitz, A. and J.-F. Bonnefon (2013). Evidence for the modularity of trustworthiness detection in an economic game. *Journal of Experimental Psychology: General.* Vol. 142(1), 143-150.

Den Uyl, T. (2013). Validation Action Unit Module. *Technical report*. Noldus.

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

Eckel, C. C., & Petrie, R. (2008). Face Value. Experimental Economics Center Working Paper Series 2008-11, Experimental Economics Center, Andrew Young School of Policy Studies, Georgia State University.

Efferson, Charles and Sonja Vogt (2013): Viewing men's faces does not lead to accurate predictions of trustworthiness. Nature Scientific Reports 3: 1047, doi: 10.1038/srep01047

Ekman, P. (2005). What we have learned by measuring facial behavior. In *What the face reveals*. second edition, edited by Paul Ekman and Erika Rosenberg, Oxford University Press.

Ekman, P. (1982). *Emotion in the Human Face*. Cambridge: Cambridge University Press.

Ekman, P., & Friesen, W.V. (1982). Felt, false and miserable smiles. *Journal of Nonverbal Behavior*, 6, 238-252.

Elster, J. (1998). Emotions and economic theory. *Journal of Economic Literature*, 36, 47–74.

Fehr, E., & Fischbacher, U. (2005). Altruists with Green Beards. *Analyse & Kritik*, 27, 73–84.

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

Frijda, N., Manstead, A., & Fischer, A. (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, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in Humans. *Evolution and Human Behavior*, 24, 153-172.

Gintis, H., Smith, E.A., & Bowles, S. (2001). Costly Signaling and Cooperation. *Journal of Theoretical Biology*, 213, 103-119.

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

Hopfensitz, A. & Reuben, E. (2009). The importance of emotions for the effectiveness of social punishment. *The Economic Journal*, 119, 1534-1559.

Johnston, L., Miles, L. & Macrae, C.N. (2010). Why Are You Smiling at Me? Social functions of enjoyment and non-enjoyment smiles. *British Journal of Social Psychology*, 49, 107-127.

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

Ketelaar, T. (2006). The Role of Moral Sentiments in Economic Decision Making in *Social Psychology and Economics*, ed. David De Cremer, Marcel Zeelenberg and J. Keith Murnighan. Lawrence Erlbaum Associates, Publishers.

Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P.L., & Kappas A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behaviour. *Emotion*, 7, 730-735.

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

Lotem, A., Fishman, M.A., & Stone, L. (2003). From reciprocity to unconditional altruism through signalling benefits. *Proceedings of the Royal Society B*, 270, 199-205.

Manzini, P., Sadrieh, A. & Vriend, N.J. (2009). On Smiles, Winks and Handshakes as Coordination Devices. *The Economic Journal*, 119, 826–854.

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

Mehu, M., Grammer, K., & Dunbar, R. (2007a). Smiles when Sharing. *Evolution and Human Behavior*, 28, 415–422.

Mehu, M., Little, A.C., & Dunbar, R. (2007b). Duchenne smiles and the Perception of Generosity and Sociability in the face. *Journal of Evolutionary Psychology*, 5, 183–196.

Mehu, M., & Dunbar, R. (2008). Naturalistic Observation of Smiling and Laughter in Human Group Interactions. *Behaviour*, 145, 1747-1780.

Niedenthal, P., Mermillod, M., Maringer, M., & Hess, U. (2010). The Simulation of Smiles (SIMS) Model: Embodied Simulation and the Meaning of Facial Expression. *Behavioral and Brain Sciences*, 33, 417-433.

Owren, M. J., & Bachorowski, J. (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.

Parkinson, B. (2005). Do Facial Movements Express Emotions or Communicate Motives? *Personality and Social Psychology Review*, 9(4), 278-311.

Preuschoft, S., & J.A.R.A.M. van Hooff (1997). The social function of "smile" and "laughter"; variations across primate species and societies. pp 171-189 In: *Nonverbal Communication: Where Nature Meets Culture*. U. Segerstrale, P. Molnar, eds. Lawrence Erlbaum Ass, New Jersey,.

Scharlemann, J. PW, Eckel, C.C., Kacelnik, A., & Wilson, R.K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22, 617-640.

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, 31, 87-94.

Smith, E.A., & Bliege Bird, R.L. (2000). Turtle hunting and tombstone opening: public generosity as costly signalling. *Evolution and Human Behavior*, 21, 245-261.

Sommerfeld, R. D., Krambeck, H., & Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. *Proceedings of the Royal Society B*, 275, 2529-2536.

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

van ‘t Wout, M., & Sanfey, A. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108, 796–803.

Veblen, T. (1899). *The Theory of the Leisure Class*. Prometheus Books.

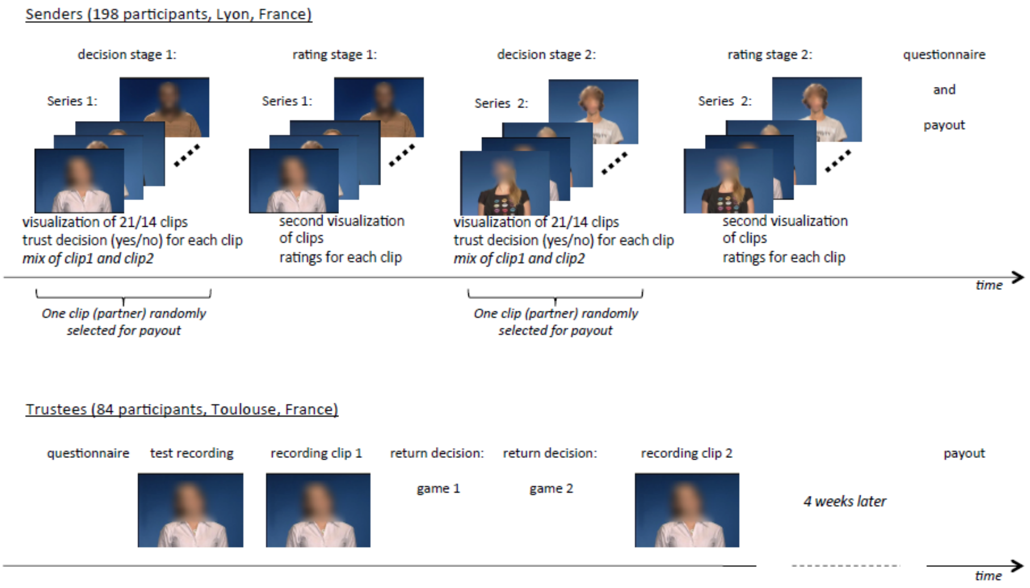
Verplaetse, J., Vanneste, S., & Braeckman, J. (2007). You can judge a book by its cover: the sequel. A kernel of truth in predictive cheating detection. *Evolution and Human Behavior,* 28, 260-271.

Vogt, S., C. Efferson & E. Fehr (2013). Can We See Inside? Predicting Strategic Behavior Given Limited Information. *Evolution and Human Behavior*. 34(4), 258-264.

Wylie, D.R., & Goodale, M.A. (1988). Left-sided oral asymmetries in spontaneous but not posed smiles. *Neuropsychologia*, 26, 823-832.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior,* 24, 290-301.

Figure 1: Timeline of experimental sessions for senders and trustees.



Note: faces are blurred in the Figure to preserve subject anonymity in publication; actual clips were not blurred

Figure 2: Differences in (a) ratings of trustees and (b) trust and trustworthiness by treatment (one tailed p-values for mean comparison clustered by trustee are reported)

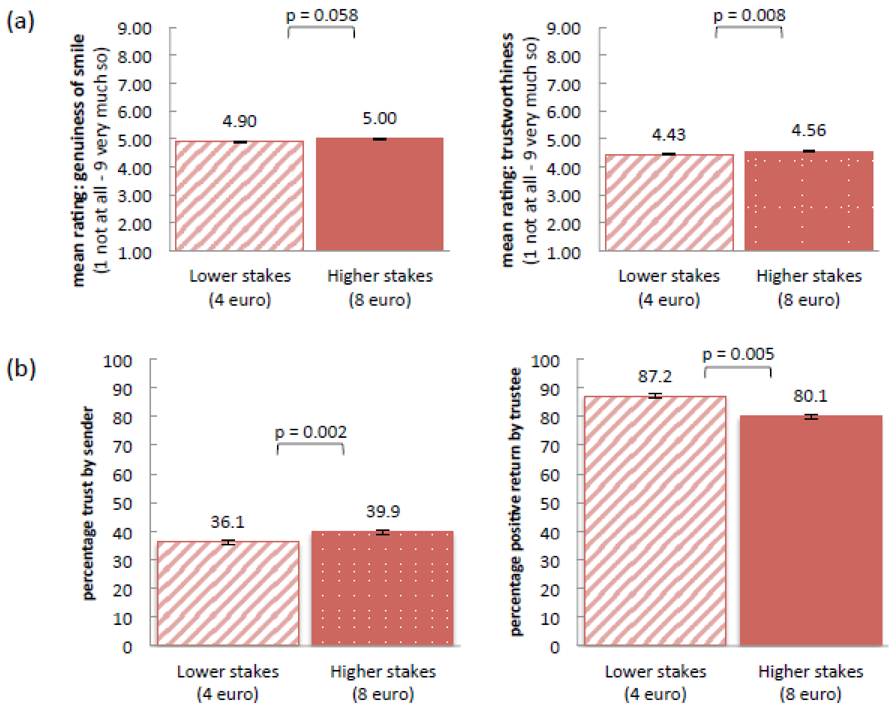


Figure 3: Differences in (a) ratings of trustees and (b) trust and trustworthiness by smile quality (one tailed p-values for mean comparison clustered by trustee are reported)

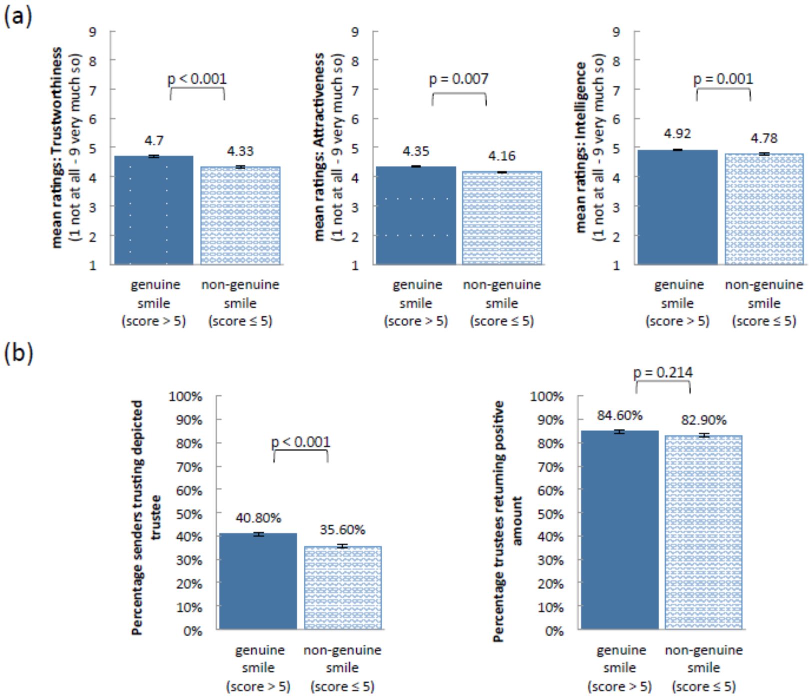


Table 1: Correlation between ratings of clips

	Genuineness	Smiles amount	Trustworthy	Attractive	Intelligent
Smiles amount	0.24***				
Trustworthy	0.64***	0.17***			
Attractive	0.25***	0.29***	0.28***		
Intelligent	0.25***	0.16***	0.37***	0.32***	
Self-confident	-0.01	0.36***	0	0.27***	0.25***

*Note: *** Correlation is significant at the 1% level.*

Table 2: Summary statistics of participants

	senders	**trustees**
Number of participants:	198	84
Percentage male	50 %	45 %
Percentage student	92 %	46 %
Mean age	21.58 (4.37)	24.86 (4.73)
Percentage of trust choices	38 %	

Note: standard deviations in parenthesis.

Table 3: General characteristics of trustees by treatment

	Lower Stakes treatment (N=42)	Higher Stakes treatment (N=42)	Difference
Age	24.7	25.0	-0.24
Men	45.2 %	45.2 %	0.00
Single (not in a relationship)	16.7 %	19.0 %	-2.40
African ethnicity	7.1 %	7.1 %	0.00
Facial (lips, eyebrows) piercing	9.5 %	0.0 %	9.6**
Wearing glasses	4.8 %	14.3 %	-9.50
Male participants with beard	9.5 %	14.3 %	-4.80
Female participants with cleavage exposed	11.9 %	7.1 %	4.80

*Note: ** mean difference is statistically different from 0 at 5% confidence level; differences between means are tested using the Wilcoxon rank sum statistic.*

Table 4: Tests of hypotheses

	Equation A	Equation B	Equation C
Dependent Variable:	**Mean genuineness rating of smile (scale 1-8)**	**Mean trustworthiness rating (scale 1-8)**	**Decision to send money (send=1)**
Independent variables (characteristics of trustee):			
Trustee in Higher stakes treatment	0.165*** (0.002)	0.079** (0.051)	-0.011 (0.800)
Mean genuineness rating of smile		0.542*** (0.000)	0.219*** (0.000)
Mean perceived amount of smiles	0.182*** (0.000)	0.013 (0.747)	0.035 (0.297)
Mean intelligence rating	0.190** (0.048)	0.310*** (0.000)	0.154** (0.012)
Age of trustee	0.012** (0.040)	0.007 (0.133)	0.005 (0.142)
Male participant with beard	-0.204** (0.020)	-0.006 (0.927)	-0.050 (0.404)
Female participant with cleavage exposed	0.187** (0.046)	-0.008 (0.929)	0.067 (0.188)

Note: p-values in parentheses, one-tailed values reported for High treatment and Smile quality, two-tailed values for other variables. Standard errors clustered by clip, *=significant at 10%; **=significant at 5%; ***=significant at 1%. Equations A, B and D are estimated by ordinary least squares, Equation C is a probit estimated by maximum likelihood. Other controls include gender of sender and trustee, perceived attractiveness, video sequence, sender's treatment, dummy variables for the trustee being of African ethnicity, having visible facial piercing and wearing glasses in equations A, in addition to perceived self-confidence, in Equation B. Equation C adds to the previous set of controls a dummy variable for senders self-reported unselfish behavior from the General Social Survey, income, age of sender and score on a simple intelligence test. The estimated coefficients for the other controls not reported, available from authors on request. Number of observations= 6,720. Variables Smile Quality, Intelligence and Trustworthiness are means by clip.

Table 4: Do ratings of smile genuineness predict potential gains from sending money?

	Equation D	Equation E
Dependent Variable:	**Unconditional potential gain (Euros) from sending money**	**Potential gain conditional on sender's knowledge**
Independent variables:		
Stake size of sender		1.034*** (0.000)
Mean genuineness rating of smile	1.514* (0.067)	2.515** (0.035)
Mean perceived presence of smiles		-0.301 (0.612)
Mean intelligence rating		0.581 (0.568)
Mean trustworthiness		-1.818 (0.105)
Mean attractiveness		-0.700 (0.357)
Mean self-confidence rating		1.475 (0.113)

Note: p-values in parentheses.

Table 5: Character or Opportunity? Determinants of smile genuineness ratings.

Dependent variable:	**Trustee is in Higher stakes treatment** (dummy variable)	**Unselfish behavior by trustee** (dummy variable)
Independent variable:		
Mean genuineness rating of smile	0.516** (0.038)	0.403 (0.126)

*Note: one-tailed p-values in parentheses, standard errors clustered by clip, **=significant at 5%. Probit equations estimated by maximum likelihood. Number of observations=6,720.*