Honest signaling in trust interactions: the power of a smile

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 test the hypothesis that "genuine" or "convincing" smiling is a costly signal that has evolved to induce cooperation in situations requiring mutual trust. Potential trustees in a trust game made video clips for viewing by potential trusters before the latter decided whether to send them money. Ratings of the genuineness of smiles vary across clips; it is difficult to make convincing smiles to order. We argue that smiling convincingly is costly, because smiles from trustees playing for higher stakes are rated as significantly more convincing, so that rewards appear to induce effort. We show that it induces cooperation: smiles rated as more convincing strongly predict judgments about the trustworthiness of trustees, and willingness to send them money. Finally, we show that it is an honest signal: those smiling convincingly return more money on average to senders. Convincing smiles are to some extent a signal of the intrinsic character of trustees: less honest individuals find smiling convincingly more difficult. They are also informative about the greater amounts that trustees playing for higher stakes have available to share: it is harder to smile convincingly if you have less to offer.

Keywords: costly 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) it has been known that smiles perceived as genuine 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 for perceived genuineness (Krumhuber et al. 2007). Duchenne smiles are however 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?

In this paper we test the hypothesis that smiling is a form of costly communication (costly in a sense we make precise below) that has evolved to induce cooperation between individuals in situations requiring mutual trust. This hypothesis, was first suggested by Owren & Bacharowski (2001). In a companion paper

(Centorrino et al 2013) we develop a theoretical model of costly smiling and show that it implies three distinct component hypotheses about the behavior of the smiler and the target of the smile. These are that smiling "genuinely" or "convincingly" is:

a) costly to the smiler,

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

c) a reliable signal of the likely benefits to the target of cooperating with the smiler.

In that paper we translate these components into testable empirical predictions (that is, into relations between observable variables) in the context of a trust game played between a smiler and a target, with the target in the role of truster and the smiler in the role of trustee. These three testable predictions are:

H1: The quality of the smile as perceived by the target is increasing in the size of the stake.

H2: The amount sent by target is increasing in the perceived quality of the smile.

H3: The expected gain to the target from sending the stake to the smiler is increasing in the perceived convincingness of the smile.

H1 is important in distinguishing the costly signaling 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 important in explaining why human beings should have evolved the habit of communicating in this costly way. H3 is important to explain why human beings should also have evolved the tendency to be influenced by the smiles of others.

There exists some corroborating evidence for H2 and H3 in the literature, though in neither case have the components in question been rigorously tested. 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 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 (only 20 participants).

H2 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[1]. 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

[1] Yamagishi et al. (2003), Verplaetse et al. (2007).

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.

Costly signaling has been extensively studied both in economics since Spence (1974), and independently in biology since 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 *more* costly for those individuals that do not possess the trait than for those who do (Grafen 1990). We conjecture that the hidden trait of smiling 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), or a characteristic of the situation in which the smiler finds herself (such as the size of the pie she is proposing to share).

Emitting some kind of smile requires rather little effort, but when individuals are trying to impress others they often put a lot of effort into the task. There is evidence that smiling can itself alter felt emotional states (Strack et al. 1988). Trivers (2000) has hypothesized that self-deception may be a means of reducing the cognitive load required to deceive others successfully, a cognitive load that has been described in detail by Vrij et.al. (1996). Our hypothesis is that non-Duchenne smiles require negligible effort but Duchenne smiles require significant cognitive effort, to a degree that probably varies between individuals, unless the smiler is already in a state that makes him more likely to be cooperative towards the target of the smile.

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 choose to keep the money, to return the sender's original stake, or to return an additional 50% of the stake so as to share the total surplus equally.

We asked trustees to make short video clips to be shown to senders before the senders took their decision; 84 subjects produced a total of 168 clips. To our knowledge no earlier study observes reactions to dynamic pictures of trustees in a situation where video messages were obtained from participants knowing that this was their only means to convince their partner to cooperate[2].

Trustees were asked to present themselves in a simple common format to the senders, such as to convince the latter to send the money. Smiling was never mentioned. Clips were shown to senders before taking the decision about sending the money. Senders then rated the clips along a number of dimensions, one of which was the perceived genuineness of the smile.

Even though our investigation of smiles as a costly 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

[2] Vogt, Efferson and Fehr (2013) 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. The difference between contexts that do and do not permit the transmission of such information is an interesting subject for future research.

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). 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 & Bacharowski 2001).

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). Indeed it has been observed that players in trust games are willing to spend money on visual information of their partner (Eckel and Petrie 2008). Which visual information is used is however not clear, but honest smiles could play a crucial role in that respect (Cohn and Smith 2004). Because activation of the orbicularis oculi (one of the main markers of Duchenne smiles) is believed to be under emotional and involuntary control (Ekman and Friesen 1982), it might therefore be an informative signal.

Whether smiles are indeed perceived as a signal of trust has been subject of a number of recent experimental studies. Scharlemann et al. (2001) presented participants in a modified trust game with static pictures of sixty photographic models, which were portrayed smiling and with a neutral expression. The results show that smiling pictures are more often trusted than their non-smiling counterpart (68.3% versus 55%). Similarly Mehu et al. (2007b) assess which characteristics are associated with honest smiles by rating fifty faces across ten attributes. It turns out

that Duchenne smiles played a significant role in the assessment of generosity and extroversion. Van 't Wout and Sanfey (2008) observe that judgments of facial trustworthiness is related to sending money in a trust game. Trustworthiness ratings were further 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).

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. 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.

To test whether smiles are indeed related to actions in a game requires the observation of both behavior and facial expression. Mehu et al. (2007a) suggest that human smiles are more prevalent in situations which involve sharing or exploitations 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 smiles are also predictive of a specific sharing decision has so far not been studied and is subject of our work.

2. Experimental Methods and Data Description

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 by the strategy method (in other words, they reported what they would return to senders for each value of what the senders might send to them). Decisions for senders and trustees were incentivized and earnings were according to their partners' decisions. Therefore payout for trustees did not take place 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. The difference between the two games was that in the first, trustees recorded their video message before being informed about the precise payoffs and thus before taking their decision, while in the second game they recorded their message after taking their return decision[3]. 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

[3] 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.

split into two payoff treatments: stakes in the Low treatment were 4 euros, and stakes in the High treatment were 8 euros. Video clips were made on a professional TV platform, and a practice clip helped participants to become familiar with the environment. 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. To make the message natural for trustees they included in this sentence their name, age and occupation and were reassured that the precise wording did not matter. Video clips lasted around fifteen seconds on average, with the fastest at around ten and the slowest at around twenty seconds.

In order to measure the characteristics of the clips, we subjected them to rating by observers of the clips who were also first movers in the game, who would be thereby motivated to observe each clip carefully and would be less likely to be influenced by irrelevant factors in their evaluations. We subjected each clip to a large number of ratings (around 40 on average) to avoid idiosyncratic reactions of individual observers. In our analyses below we use the average rating by all observers to predict any individual sender's behaviour 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[4].

Senders' behavior and evaluation of video clips were obtained in two waves of experiments conducted in a different experimental laboratory to minimize the risk that

[4] Note that 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 are obtained. Results available upon request.

senders might recognize trustees. A total of 198 student participants were recruited at the University of Lyon; 84 senders participated in the first wave and 114 in the 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; on these data alone we would have been unable to control separately for the effect of stake size on the characteristics made by trustees and its direct effect on the behavior of senders. Having previously obtained trustees' consent to the re-use of their clips, we therefore implemented a second wave, in which participants made trust decisions and were paid according to trustees' initial decision[5]. However, the stake size for senders in the second wave varied: of the senders who saw each clip, some had Low stakes of 4 Euros, some had High stakes of 8 Euros, and some had a new Super-high treatment of 12 Euros. The actual stake size of their partner was varied: half of each sender's video messages came from participants in the High stake treatment and half came from the Low stake treatment. This 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.

In the first and second waves respectively, each sender viewed a total of 42 (respectively 28) clips in two series of 21 (resp. 14). For each clip the sender was asked to decide whether to send money to the trustee, and then to rank the trustee on

[5] In the second wave senders were informed only that trustees might return nothing, their initial endowment or 1.5 times their endowment.

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. Note that for each series, the two exercises were independent: first the sender made all his decisions, and second he watched all the clips again to rate each trustee. Senders were matched at random with one of the trustees from each series and received payoffs determined by the actual decision of this real partner. 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[6]. Figure 1 summarizes the choices made by the participants.

[Figure 1 and Table 1 here]

Table 1 summarizes the characteristics of our participants for each group of players. The first column reports the means and standard deviations for the senders. This subsample is exactly gender-balanced, the average age is 22 years old and 92% are currently students. For a given sender, the proportion of decisions for which she decided to send money ranges from 0 to 1. Of the 198 participants, 3 participants decided to send money to every partner, and 20 decided never to send money. On average, each sender decides to cooperate with 37% of their partners. Column (2) concerns the trustees, who are 25 years old on average. The sample of trustees comes

[6] Owing to some volunteers' not turning up for their session of the experiment there was a slight variation in numbers of ratings per clip: the clips in our dataset have an average of 40 ratings each, with a minimum of 38 and a maximum of 45.

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. In Table 2 Panels A and B investigate whether there was any difference in decisions and attitudes taken across both clips respectively. In the first decision, 55% of the participants decided to share equally the amount received (ie 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 (see Panel A). In the second decision, the distribution suggests a slight shift from equal sharing to sending 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 almost exactly zero. This therefore provides an excellent environment in which to test whether the conditional expected gain (conditional on smile characteristics) is positive, as indeed it is.

Panel B suggests that the ratings the trustors made on the trustees' video clips are not significantly different whether they watched the first or the second clip.

Panel C in Table 2 investigates whether there were any significant differences in characteristics between trustees in the Low and High treatments. Only the probability of having a 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 High and Low 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 High 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 High treatment actually return any money to the senders. One could think that the positive correlation of the High treatment with smile genuineness and with selfish behavior would imply that genuine smiles are positively correlated with selfish behavior. However, Figure 3 indicates that this is not so. The explanation is that those in the High group who *succeeded* in making genuine smiles were not a random subset of those in the High treatment; 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. 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. So overall it appears that the High 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 smiling genuinely 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 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 component hypotheses in turn. In all cases, in order to avoid possible "justification effects" in which users' ratings are influenced by the decisions they have already taken whether to send money to the trustees, 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 3 reports our tests of these three hypotheses. Equation A tests the hypothesis H1, that the quality of the smile as perceived by the target is increasing in the size of the stake. 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[7].

[7] Though this is not a large effect, our treatment was itself not very large. Although the difference between a 24 Euro maximum gain and a 48 Euro maximum gain may

[Table 3 here]

Other notable features of Equation A are that more convincing smiles are associated with trustees who are rated as more intelligent, and also with older trustees. Trustees with beards are rated as having less convincing smiles, and women with a significant décolleté are rated as having significantly more convincing smiles (it is not clear whether the causal mechanism is via the psychology of the smiler or of the viewer). There is no significant effect of gender of 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 sent. 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 H2, that the amount sent by target is increasing in the perceived quality of the smile. First, Equation B considers whether convincing smiles are associated with judgments of greater trustworthiness. There is a massively significant correlation (t-ratio of over 8); a one-point deviation increase in smile quality 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 High treatment independently of smile quality, suggesting that trustees are putting effort

---

seem important for an experimental session that lasted only around half an hour, the mean gain to trustees in the High treatment (which was of course reduced if senders did not send money, and if trustees returned part of the money) was just over 12 Euros, compared to a little over 7 Euros for those in the Low treatment. Although we considered trying larger treatments, it seemed likely that this would have induced much more selfish behavior by trustees, and that this might have been anticipated by senders.

into other dimensions of non-verbal communication as well. Equation C examines whether convincing smiles lead to an increased probability of sending money to the trustees. Once again there is a massively significant association: a one point increase in smile quality leads to a 21% increase in the probability of sending the money, which is equivalent to a 7% increase per standard deviation of smile quality. Perceived intelligence is again a very important factor in the decision.

Equation D tests hypothesis H3, that the expected gain to the target from sending the stake to the smiler is increasing in the perceived convincingness of the smile. We therefore regress gains from sending money on average smile quality, without other controls. The effect is small but positive (0.9 Euros per one point increase, or 0.3 Euros per standard deviation) and significant at under 1%[8].

[Table 4 here]

Finally, we investigate whether convincing smiles are associated more closely with the amount of money the trustee has available to offer the sender or with the intrinsic character of the trustee. We therefore regress a dummy variable for the High treatment on the mean smile quality, and in a separate equation we regress on the same measure of smile quality a dummy variable indicating that the trustee takes an unselfish decision. The purpose is to see whether the smile quality is a reliable informative of the amount available to share, and of the character of the trustee. As

[8] In the decision equation (though not in the others) the treatment effect for senders is large and highly significant, and follows an inverted-U shape: a High treatment makes senders more likely to send money, but a Superhigh treatment makes them much less likely to do so.

reported in Table 4, the smile quality is 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, which is not statistically significantly different from zero at conventional levels (unlike the coefficient on the High treatment). Secondly, 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 intrinsic character of the trustee. So we should conclude that smile quality is definitely informative about high cooperation opportunities, and *may* also signal the intrinsic character of the smiler.

4. Conclusions

We have tested three component hypotheses of the theory that smiling convincingly is a costly signal that has evolved to induce cooperation in situations requiring mutual trust. All three components are supported by the evidence. First, the quality of the smile as perceived by the target is increasing in the size of the stake, because smiles from trustees playing for higher stakes are rated as significantly more convincing, by over a third of one standard deviation. This strongly suggests that they are produced when there are rewards to the additional effort required. Secondly, the amount sent by target is increasing in the perceived quality of the smile: smiles rated as more convincing are strong predictors of judgments about the trustworthiness of trustees, and of the revealed willingness to send them money. Finally, we show that the expected gain to the target from sending the stake to the smiler is increasing in the perceived convincingness of the smile: those smiling convincingly return more money

on average to senders. It is clearly informative of the amount the trustee has available to share with the truster; there is weaker evidence that it may also be a signal of the intrinsic trustworthiness of the trustee independently of the amount at stake.

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 Max Planck Institute for Evolutionary Biology in Plön for financial support. 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. The usual disclaimer applies.

References:

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. (2013). 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.

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. (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*, 27/2005, 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.

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.

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.

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.

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.

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.

Strack, F., Martin, L., & Stepper, S. (1988). Inhibiting and Facilitating Conditions of the Human Smile: A Nonobtrusive Test of the Facial Feedback Hypothesis. *Journal of Personality and Social Psychology,* 54, 768-777.

Trivers, R. (2000). Elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences,* 907, 114-131.

van 't Wout, M., & Sanfey, A. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, 108, 796–803.

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 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

Vrij, A., Semin, G.R., & Bull, R. (1996). Insight into behavior displayed during deception. *Human Communication Research,* 22, 544-562.

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.

Table 1: Summary statistics

	1st movers (senders) (1)	2nd mover (trustees) (2)
Proportion Male	50 (0.50)	45 (0.50)
Mean Age	21.58 (4.37)	24.86 (4.73)
Proportion Student	92 (0.27)	46 (0.50)
Proportion Trusting	37 (0.23)	
Number of observations	198	84

Note: standard deviations in parenthesis.

Table 2: Treatments: decisions by 2nd movers, ratings of their clips and general characteristics

Panel A:	**First decision (N=84)**	**Second decision (N=84)**	**Difference (1)-(2)**
	(1)	**(2)**	**(3)**
Decision (in proportion)			
Equally share the triple stake	55	49	6
Send back the original stake	31	32	-1
Send back nothing	14	19	-5
Panel B:	**First clip (N=3342)**	**Second clip (N=3378)**	**Difference (1)-(2)**
	(1)	**(2)**	**(3)**
Genuine smile rating	4.93	4.96	-0.03
Trustworthiness rating	4.53	4.46	0.07
Attractiveness rating	4.25	4.24	0.01
Decision to be trusted (%)	38.00	38.00	0.00
Intelligence rating	4.85	4.84	0.01
Panel C:	**Low Stake treatment (N=42)**	**High Stake treatment (N=42)**	**Difference (1)-(2)**
	(1)	**(2)**	**(3)**
Men	45.2	45.2	0.00
Age	24.7	25.0	-0.24
Single (not in a relationship)	16.7	19.0	-2.40
Black	7.1	7.1	0.00
Beard	9.5	14.3	-4.80
Piercing	9.5	0.0	9.6**
Wearing a décolleté	11.9	7.1	4.80
Wearing glasses	4.8	14.3	-9.50

Note: ** mean difference is statistically different from 0 at 5% confidence level

Table 3: Tests of components of costly signaling hypothesis

	Equation A	**Equation B**	**Equation C**	**Equation D**
Dependent Variable:	**Smile Quality (scale 1-8)**	**Trustworthiness (scale 1-8)**	**Decision to send money (send=1)**	**Gain (Euros) from sending money**
Ind. variable (characteristics of trustee):				
High treatment	0.124** (0.017)	0.079** (0.026)	-0.011 (0.400)	
Smile quality		0.54*** (0.000)	0.219*** (0.000)	0.911*** (0.005)
Intelligence rating	0.213** (0.030)	0.310*** (0.000)	0.154** (0.012)	
Age of trustee	0.0149** (0.017)	0.007 (0.139)	0.005 (0.142)	
Beard	-0.262*** (0.003)	-0.006 (0.867)	-0.050 (0.404)	
Décolleté	0.229** (0.011)	-0.008 (0.821)	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 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 truster and trustee, perceived attractiveness, video sequence, truster's treatment, dummy variables for the trustee being black, having visible piercing and wearing glasses in equations A, in addition to perceived self-confidence, and smile frequency in Equation B. Equation C adds to the previous set of controls a dummy variable for A players' self-reported unselfish behavior from the General Social Survey, income, age of truster 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: Character or Opportunity? Determinants of smile quality

Dependent Variable:	**Trustee is in High treatment (dummy variable)**	**Unselfish behavior by trustee (dummy variable)**
Independent Variable:		
Mean smile quality rating	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.

Figure 1: Order of choices for second movers (trustee) and first movers (sender) in the experiment

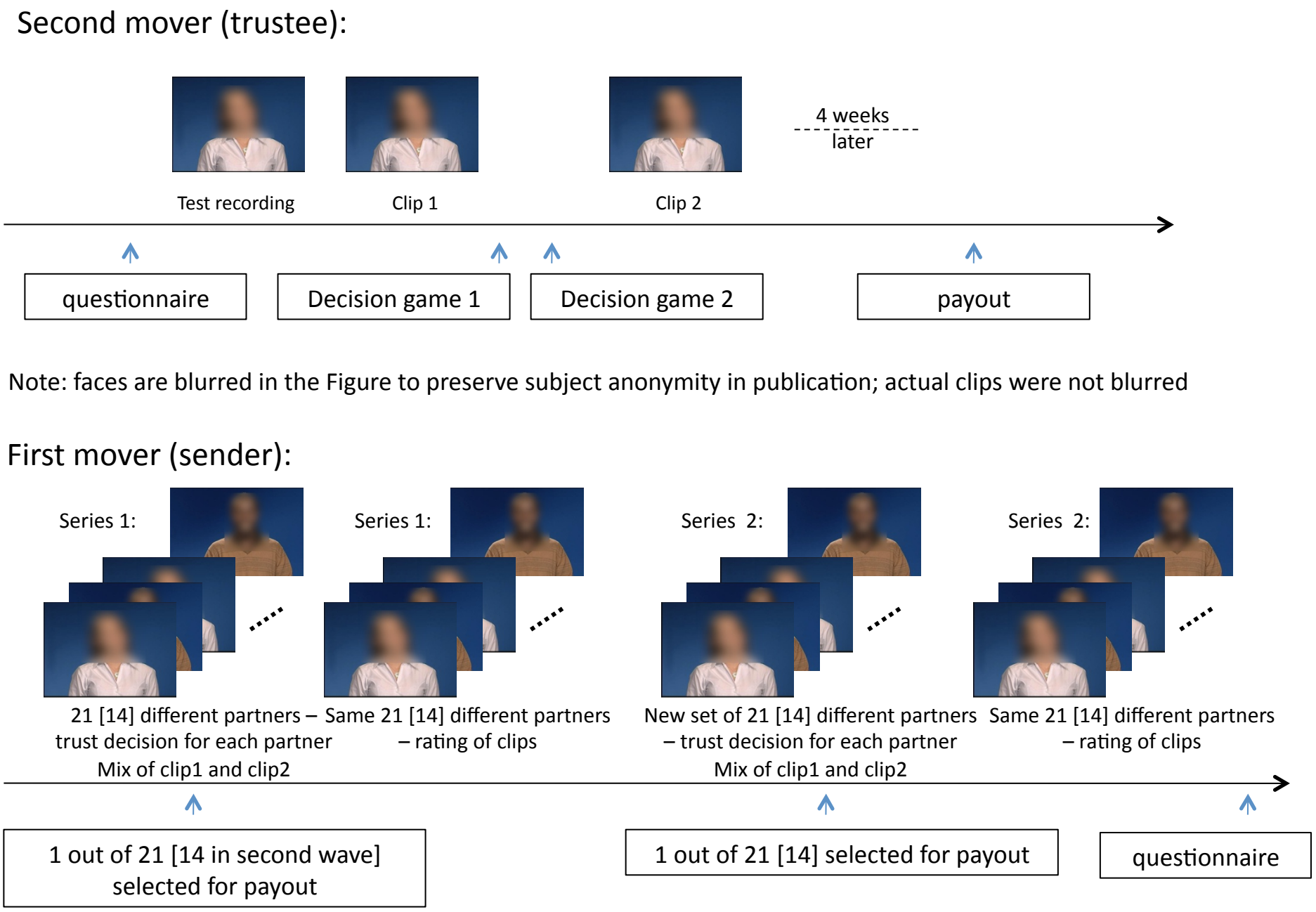


Figure 2: Differences in (a) ratings of trustees and (b) trust and trustworthiness by treatment (one tailed p-values for mean comparison clustered by trustee are reported)

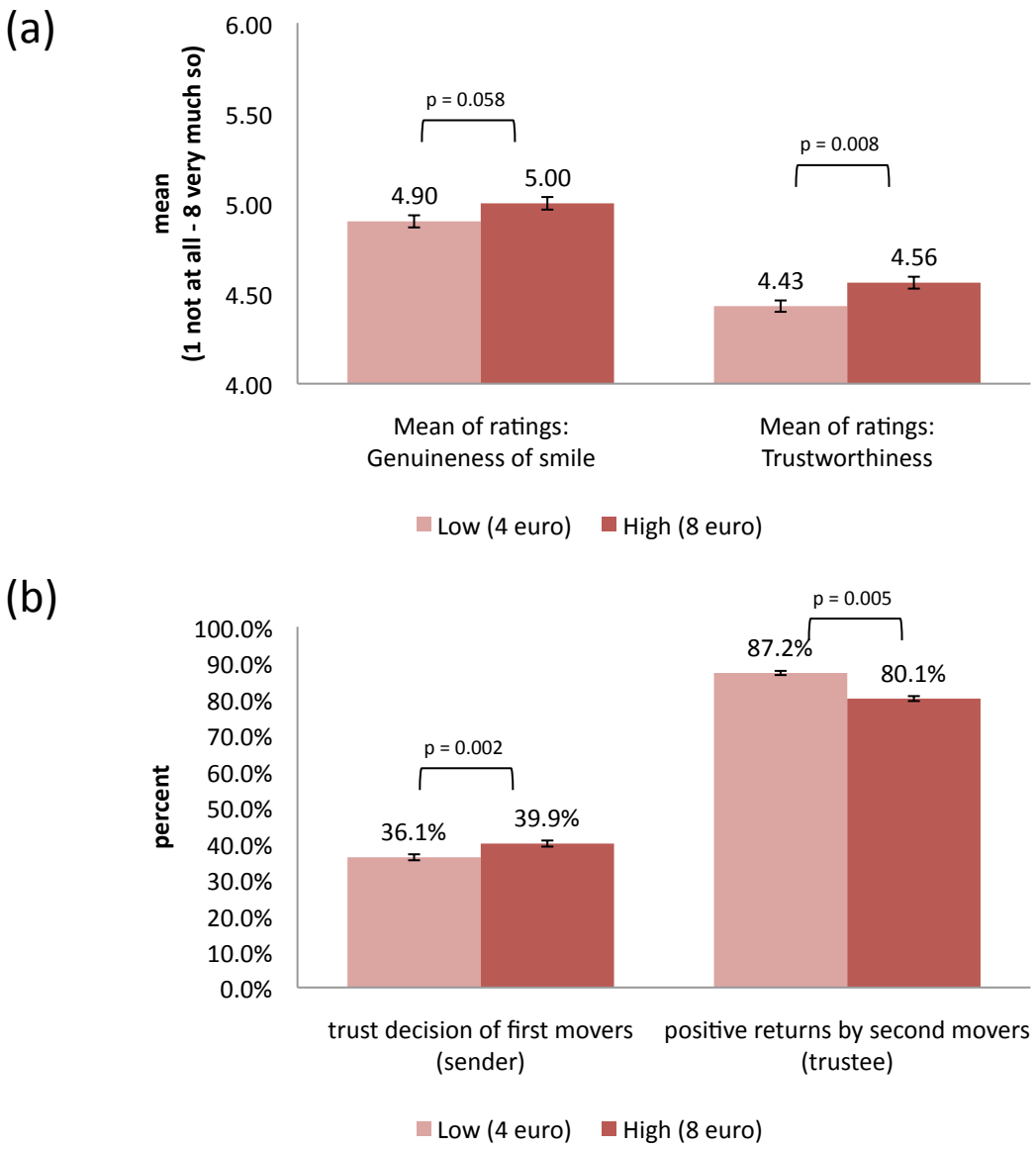
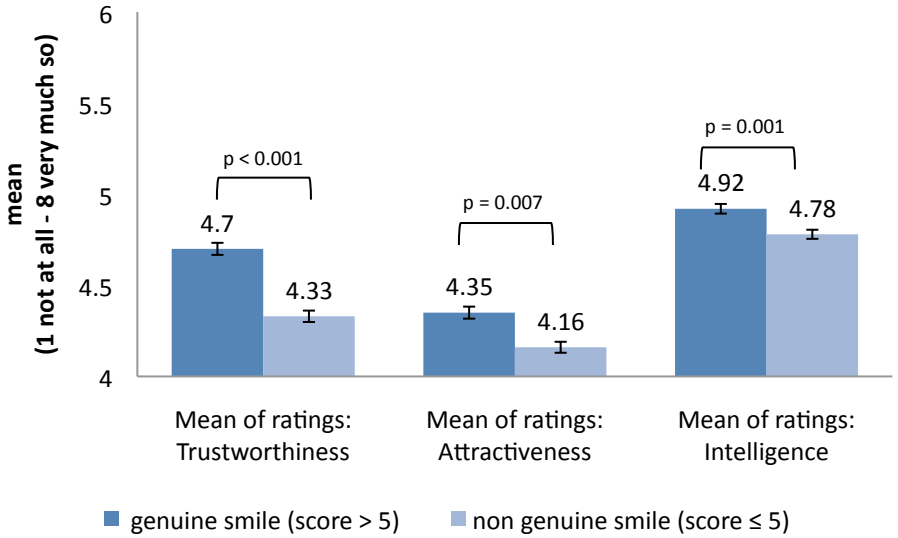


Figure 3: Differences in (a) ratings of trustees and (b) trust and trustworthiness by smile quality (one tailed p-values for mean comparison clustered by trustee are reported)

(a)



(b)