

2020-04

If your device could smile: People trust happy-sounding artificial agents more

Torre, I

<http://hdl.handle.net/10026.1/16010>

10.1016/j.chb.2019.106215

Computers in Human Behavior

Elsevier BV

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

If your device could smile: People trust happy-sounding artificial agents more

Ilaria Torre ^{a,*}, Jeremy Goslin ^b, Laurence White ^c

^a *Electronic & Electrical Engineering, Trinity College Dublin, Ireland*

^b *School of Psychology, University of Plymouth, UK*

^c *Speech and Language Sciences, Newcastle University, UK*

A B S T R A C T

While it is clear that artificial agents that are able to express emotions increase trust in Human–Machine Interaction, most studies looking at this effect concentrated on the expression of emotions through the visual channel, e.g. facial expressions. However, emotions can be expressed in the vocal channel too, yet the relationship between trust and vocally expressive agents has not yet been investigated. We use a game theory paradigm to examine the influence of smiling in the voice on trusting behavior towards a virtual agent, who responds either trustworthily or untrustworthily in an investment game. We found that a smiling voice increases trust, and that this effect persists over time, despite the accumulation of clear evidence regarding the agent’s level of trustworthiness in a negotiated interaction. Smiling voices maintain this benefit even in the face of behavioral evidence of untrustworthiness.

1. Introduction

Emotional expressions are essential aspects of social interaction (Russell, Bachorowski, & Fernández-Dols, 2003). It is believed that one of their original functions was essentially social, to alert others in a group about a change in the environment (Ekman, 1999). Thus, emotional displays help inform others about what actions to take (Van Kleef, De Dreu, & Manstead, 2010). In accordance with this social function, there are some commonalities in the expression of emotion across all human societies (e.g. Elfenbein & Ambady, 2002). Nevertheless, there are cultural differences in the way emotions are expressed and experienced (e.g. Matsumoto, 1993-06). Importantly, it has been argued that emotional expressivity – i.e. the capacity for displaying emotions at varying intensities – facilitates social communication, and is therefore an adaptive evolutionary strategy (Boone & Buck, 2003): individuals who are able to display and detect emotions accurately might be at an advantage in social group interactions. Furthermore, when the visual channel is not present, people are able to display and detect emotional signals from the voice alone (e.g. Cowen, Laukka, Elfenbein, Liu, & Keltner, 2019).

As artificial agents become more present in our everyday lives, the question arises of whether the ability to detect and display emotions might serve a similar purpose for such agents. Specifically, whilst researchers in Human–Machine Interaction agree that it is important for a machine to be able to recognize and express emotions to increase the naturalness of the interaction (e.g. Breazeal, 2009), it is still not clear

what effect these displays will have on people’s behavior in the interaction, for example, in terms of trusting behavior. In particular, while studies have looked at the impact of artificial agents’ facial expressions, the effect of emotions expressed through the voice of artificial agents remains to be determined. As expressive Text-To-Speech technologies become less expensive and more realistic, it is important to understand the effect expressive synthetic voices will have on potential users.

We present an experiment in which participants played a trust game with a virtual agent that expressed emotion through its voice, in a manner congruent or incongruent with its behavior. Thus, the virtual agent’s voice was either smiling or neutral, and it behaved either trustworthily or untrustworthily. In this way, we examine how people form impressions of trustworthiness from a virtual agent’s voice, and how these impressions interact with actual experience of the agent’s behavior.

2. Background

Emotional expressivity is linked to attractiveness and other positive traits (Reis et al., 1990), including cooperation and trustworthiness (Lount & Robert, 2010; Schug, Matsumoto, Horita, Yamagishi, & Bonnet, 2010). Conversely, being able to detect trustworthy individuals might be a critical component in the evolution of cooperation (Boone & Buck, 2003; Russell et al., 2003). Insofar as emotional expression is linked to physiological arousal, emotional displays might be relative

'honest signals' of trustworthiness (Boone & Buck, 2003; Schug et al., 2010). However, while a significant amount of research has been conducted on trust as a motivation for rational behavior, not much has been dedicated to study the influence of emotional expression on trust (Lount & Robert, 2010).

Here, we focus on smiling: smiling is an emotional expression produced by moving muscles in the mouth and eye regions. In evolutionary terms, a relatively small body size can be a sign of non-threat; for this reason, Ohala (1980) hypothesized that retracting the lips, which produces smiling in the lower part of the face, might have become a sign of submission and non-threat, because the corresponding sound made with a raised pitch suggests a smaller body size (Ohala, 1980; Xu & Chuenwattanapranithi, 2007). This facial expression is also similar to the 'silent bared-teeth' display – which can still be observed in certain monkeys – which is a signal of submission (Preuschoft, 1992). Genuine or 'Duchenne' smiles (Ekman, Davidson, & Friesen, 1990) are mostly considered expressions of positive emotions such as happiness, joy, amusement or friendliness (e.g. Ekman & Friesen, 1982). However, smiles can also be false, derisive or sad (Ekman & Friesen, 1982).

Also, smiling can be heard from the voice even in the absence of a visual channel – a so called 'smiling voice'. Acoustically, smiling affects the vocal tract by shortening it (Shor, 1978), which contributes to raising the fundamental and formant frequencies (Fagel, 2009). Several perceptual studies have demonstrated that the presence of smiling is detectable in the speech signal alone (e.g. Drahota, Costall, & Reddy, 2008-04; El Haddad et al., 2017; Tartter & Braun, 1994), even when the speech stimuli are synthetic (El Haddad, Cakmak, Moinet, Dupont, & Dutoit, 2015). More details on the acoustics of smiling voice are given below, when describing the study stimuli (Section 3.4).

In terms of agent perception, displaying positive emotions, for example through smiling, is generally associated with positive trait attributions, including trustworthiness. For example, in a series of experiments with Chinese (Lau, 1982), Brazilian (Otta, Lira, Delevati, Cesar, & Pires, 1994) and American participants (Mueser, Grau, Sussman, & Rosen, 1984; Reis et al., 1990), it was found that pictures of smiling faces were rated higher in terms of attractiveness, friendliness and sincerity, among other traits, as compared to neutral faces. Looking specifically at trust, participants rated a smiling Embodied Conversational Agent as more trustworthy than a neutral one (Elkins & Derrick, 2013). Oosterhof and Todorov (2009) found a correlation between positive facial emotion displays (very happy and mildly happy) and trustworthiness perception, and between negative emotion displays (very angry and mildly angry) and untrustworthiness perception. They also artificially varied the trustworthiness of the faces, while keeping the emotional expressions constant, and this changed the emotion perception, suggesting that changes in structural features of faces affect the perception of emotions.

In actual decision-making tasks, participants trusted counterparts in a trust game who were represented as smiling photographs more than neutral ones (Scharlemann, Eckel, Kacelnik, & Wilson, 2001; Tortosa, Lupiáñez, & Ruz, 2013). Similarly, participants in Krumhuber et al. (2007) indicated that avatars displaying an authentic smile might be more cooperative in a trust game; subsequently, participants proceeded to invest more money with them in an actual game. An observation of participants' natural smiling behavior comes from Reed, Zeglen, and Schmidt (2012), who video-recorded pairs of participants getting acquainted before playing a one-shot trust game. Coding of participants' facial expressions before the game showed that people exhibiting smiles were more likely to cooperate and elicit cooperation in the game. However, as the authors point out, it is possible that participants used other information to infer future counterpart's behavior, such as their vocal characteristics. Indeed, it remains to be determined whether positive attributions to facial cues extend to the interpretation of smiling voice in artificial agents.

There are various accounts of why the display of a positive emotion – like smiling – leads to the attribution of another positive trait, such

as cooperativeness or trustworthiness. Lau (1982) argued that this could be due to a 'halo effect': smiling is good, so smiling individuals must also possess other positive traits. This has also been referred to as the 'what is smiling is good' stereotype (Reis et al., 1990). Additionally, since individuals who are particularly expressive might be less apt at disguising their emotions (DePaulo, Blank, Swaim, & Hairfield, 1992), it has been suggested that positive emotional expressions might be a sign of commitment to more cooperative, less deceptive behavior (Schug et al., 2010; Tsankova et al., 2015).

However, emotional display might not always function as a signal of trustworthiness. For example, the 'Emotions as Social Information' (EASI) model (Van Kleef et al., 2010) brings context into account. This model suggests that emotions are used to make sense of ambiguous situations, and that their effect depends on the situation in which the interaction takes place, specifically its cooperative or competitive nature. Thus, displaying a positive emotion, such as happiness, in a cooperative context will reinforce the parties' belief that everyone is gaining, and will elicit more cooperative behaviors. On the contrary, displaying a negative emotion, such as anger, in a cooperative context will hinder future cooperative behaviors, and so on. Supporting this, Antos, De Melo, Gratch, and Grosz (2011) found that, in a negotiation game, participants more often selected as partners those computer agents that displayed emotions which were congruent with their actions. These agents were also perceived as more trustworthy than agents whose emotional expression and action strategy did not match, even though their actual strategy was the same. Similarly, in de Melo, Gratch, and Carnevale (2015), people played a social dilemma game with virtual humans that always behaved in the same way, but that varied in emotional expressivity: one expressed a cooperative nature (happy facial expression when there was mutual cooperation, regretful facial expression when it defected) and one a competitive nature (happy facial expression when it defected, regretful facial expression when there was mutual cooperation). Perhaps unsurprisingly, participants behaved differently with these two virtual humans, generally cooperating more with the one that expressed cooperativeness.

These studies suggest that an emotional expression, such as smiling, might have a differential effect based on the context: smiling in a cooperative situation might elicit further cooperation, while smiling in a competitive situation might elicit further non-cooperation. However, some studies suggest that smiling consistently elicited cooperation, regardless of the counterpart's behavior. For example, Mieth, Bell, and Buchner (2016) found that smiling counterparts were punished less harshly (than neutral ones) when they defected in a trust game with an option for monetary punishment. Similarly, in LaFrance and Hecht (1995), people viewed photographs of two women, who were either smiling or not; they were told that the women in the photographs were on trial for having cheated in an exam, and were asked to rate how likely it was that they were guilty. The smiling models received significantly more benefit of the doubt and less punishment than non-smiling models, suggesting that smiling alleviates the consequences of a negative behavior. Also, smiling counterparts who were offering unfair shares in an Ultimatum Game evoked smaller Feedback-Related Negativity (FRN, an EEG component indicating that a feedback was negative and unexpected) activation than neutral ones (Mussel, Hewig, Allen, Coles, & Miltner, 2014). These studies suggest that untrustworthy individuals who appear to be smiling may actually alleviate the consequences of their behavior. Another possible explanation is that smiling, in accordance with the aforementioned 'halo effect', simply elicits positive emotions, which might reduce the negative response to a counterpart's untrustworthy behavior (Mieth et al., 2016).

Thus, given that smiling can be heard in the voice, and given that smiling elicits positive feelings and trusting behaviors, would an artificial agent with a smiling voice be trusted more than an agent with a neutral voice? And how would smiling interact with a congruent-positive speaker's behavior or with an incongruent-negative one? Thus far, the effect of smiling on trust towards artificial agents has mostly

been studied looking at facial expressions. Given that many artificial agents commonly in use do not have a body or a face – e.g. mobile phone assistants, navigator systems, vending machines... – the impact of emotions expressed in the voice of these agents should be examined as well.

The EASI model (Van Kleef et al., 2010) would predict that smiling might increase trustworthiness in a congruent context (such as a trustworthy behavior) and decrease it in an incongruent context (such as an untrustworthy behavior). Thus, a virtual agent who displays a positive emotion through smiling, while behaving in a trustworthy manner, might increase a feeling of mutual benefit and cooperation, while a smiling virtual agent who is consistently behaving in an untrustworthy manner might increase the feeling that the virtual agent is gaining at the participant's expense. On the other hand, a 'halo effect' would predict that the mere presence of smiling might alleviate the consequences of an untrustworthy behavior (LaFrance & Hecht, 1995; Mieth et al., 2016).

3. Method

3.1. The investment game

The experiment was designed as an iterated investment game with a pre-programmed artificial agent. Participants' interactions with the game were via a simple text-based interface, with vocal communications from the agent, which was not visually represented. In this paradigm, the trustworthiness of the agent is determined by how much money it returns to the participant, with trustworthy agents returning more money than the participant's original investment, and untrustworthy agents returning less (Berg, Dickhaut, & McCabe, 1995). This paradigm is particularly useful because it allows an assessment of implicit trusting behavior, indexed by monetary investments, rather than explicit subjective evaluation. In addition, iterative rounds of the game allow examination of the development of trust-related behavior over time. This provides a more naturalistic simulation of a real-life encounter, as our first impressions of a newly-met individual may be subject to modification according to experience of that individual's behavior (Asch, 1946). By playing multiple rounds of the game, we aim to establish not only how first impressions of the agent are modulated by its voice, but also how they are modulated by subsequent material demonstration of its trustworthiness. The voices of our artificial agents were recorded from four different British English speakers.

3.2. Participants

Participants were 110 native British English speakers (85 female, 24 male) aged 18-48 (median = 19, SD = 5.5). They were university students who received course credit for participation. One participant was excluded due to not being a native of the British Isles, and the data from another one were not recorded due to a technical error. Therefore, only the data of 108 participants were analyzed.

3.3. Stimuli

Recordings were obtained from 4 female native speakers of British English, all in their twenties. Two had Birmingham accents, and two spoke in Standard Southern British English (SSBE), labeled respectively B1, B2, R1, R2. Each speaker was recorded in a sound-proof recording booth reading two 20-sentence blocks each. Apart from the first sentence of each block, which served as an introduction (e.g. "Let's get started with the investment game"), all the other sentences were related to the conduct of the game (e.g. "The more money you invest, the more money the two of us will earn"). The sentences were balanced in terms of length (mean number of syllables per sentence = 16.6, SD = 1.08). After the speakers were recorded reading the sentences in their typical, 'neutral' voice, they were asked to re-read the sentences

while sounding 'amused'. We followed standard procedures to elicit genuine amusement in the speakers, and thus smiling voice, by showing them a selection of funny videos during the recordings (e.g. El Haddad et al., 2015). A researcher was in the same room as the speakers during the recordings, to ensure that spoken productions were consistent – apart from the required voice manipulation – across the two conditions. The speakers could be asked to repeat utterances to maintain this consistency. In particular, the researcher, a trained phonetician, asked the speakers to repeat the recordings in case smiling was not visible for the whole duration of the utterance, and in case laughter emerged at any point during the utterance.

3.4. Acoustic and perceptual correlates of smiling

As previously mentioned, smiling has acoustic repercussions on a voice. However, while the facial movements associated with smiles have been thoroughly studied, the same cannot be said for their audio representation, and to date speech scientists have not agreed on an acoustic definition of a smile (Arias et al., 2018). Still, there are some existing studies on the effect of smiling on speech acoustics, mostly investigating this effect in terms of fundamental and formant frequencies (e.g. Barthel & Quené, 2015; Fagel, 2009; Kohler, 2008; Lasarczyk & Trouvain, 2008).

Specifically, previous studies have shown that fundamental frequency (F0) consistently increases during smiling, in both acted and naturally-occurring conversations (Fagel, 2009; Lasarczyk & Trouvain, 2008; Tartter & Braun, 1994). This is partially due to the fact that activating the Zygomaticus Major muscle – the muscle responsible for pulling the lip corners upwards – has the additional effect of shortening the vocal tract, thus increasing the fundamental frequency (Ohala, 1980; Shor, 1978; Xu & Chuenwattanapranithi, 2007). While an increase in pitch seems to be an important factor in the perception of smiling voice (Fagel, 2009), it is not the only one, as it has been shown that smiling can be perceived in whispered (i.e., non-voiced) registers too (Tartter & Braun, 1994). Therefore, it has been suggested that smiling might be better characterized by looking at the utterance spectrum, rather than the prosody.

However, measurements of formant frequency changes reveal less consistent results. For example, some studies report an overall increase in the first three formants when the speech is smiled (El Haddad et al., 2017; Tartter, 1980; Tartter & Braun, 1994). Other studies report an increase only in F2 (Arias et al., 2018; Barthel & Quené, 2015; Erickson, Menezes, & Sakakibara, 2009; Podesva, Callier, Voigt, & Jurafsky, 2015), others only in F3 (Fagel, 2009; Torre, 2014), others in F2 and F3 (Robson & Beck, 1999), and others report no increase (Drahota et al., 2008-04; Lasarczyk & Trouvain, 2008). These differences might be due to the fact that speakers need to find a way to adjust their vocal cavity in a way to convey the smiling, while at the same time producing the intended phonemes (Fagel, 2009). Therefore, the effect of smiling on formant frequencies might be phoneme-dependent. In fact, some previous studies found a formant frequency difference only for certain sounds, e.g. only for rounded vowels (e.g. Arias et al., 2018; Barthel & Quené, 2015; Fagel, 2009). To further complicate matters, some studies found significant individual variation in this increase (e.g. El Haddad et al., 2017; Podesva et al., 2015). However, smiling affects the whole utterance, not just vowels.

Recently, Arias et al. (2018) showed that smiling increased the spectral centroid of each phoneme in their dataset. Spectral centroid is a measure of the spectrum, and it essentially shows how energy is distributed across different frequency bands. In a sound, it correlates with the perception of brightness of that sound (Schubert, Wolfe, & Tarnopolsky, 2004).

Based on the existing literature, we therefore decided to measure the following parameters in our smiled and neutral utterances: fundamental frequency (F0), as this seems to be one of the main discriminating

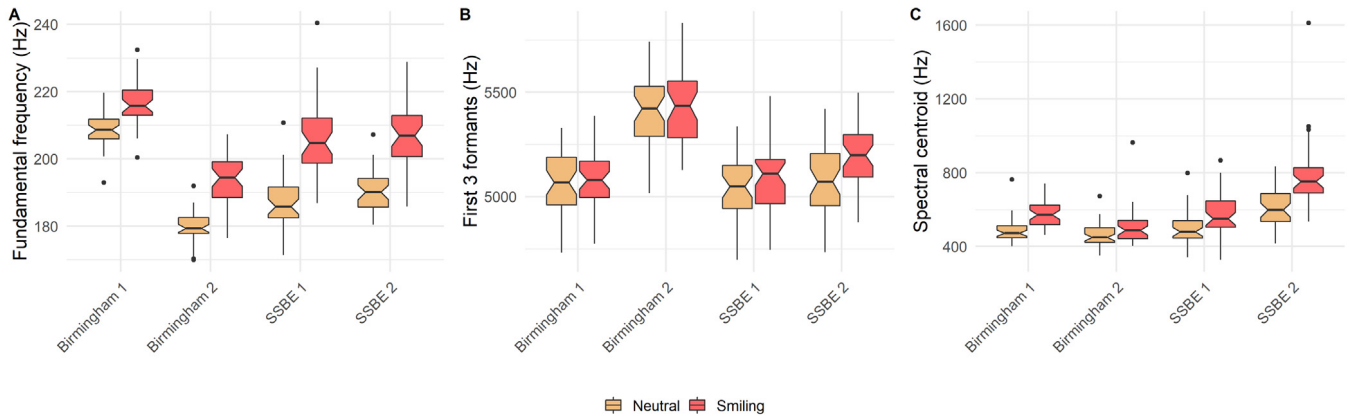


Fig. 1. Difference between smiling and neutral voice in three acoustics parameters, for each speaker: fundamental frequency (plot A), first three formants (plot B), spectral centroid (plot C).

factors in the perception of ‘smileyness’ in modal voice phonation (Drahota et al., 2008-04; Fagel, 2009); formant frequencies (F1, F2, F3), as these have been reported in the majority of the works on smiling voice, albeit with contrasting results; and spectral centroid, as a promising and theoretically-motivated measure of the quality of the sounds.

We measured these parameters on both the modal and smiled recordings of our four speakers, using Praat (Boersma & Weenink, 2017) and Parselmouth (Feinberg, 2019).

As shown in Fig. 1, average F0 was higher when the speakers were smiling while speaking, and this difference was statistically significant (mean smiling F0 = 206 Hz, mean neutral F0 = 191 Hz; $t(317.23) = 10.38, p < .001$). This confirms previous studies showing that raised F0 is a typical characteristic of smiling voice (e.g. Fagel, 2009; Lasarcy & Trouvain, 2008; Tartter & Braun, 1994). We also found a significant increase in the first three formants when the voice was smiled (mean smiling formants = 5198.63 Hz, mean neutral formants = 5149.36 Hz; $t(317.93) = -2.07, p = .039$). This is also consistent with some previous studies (Drahota et al., 2008-04; El Haddad et al., 2017; Tartter, 1980; Tartter & Braun, 1994). Finally, spectral centroid was also significantly higher in smiling voice than neutral voice (mean smiling centroid = 611.36 Hz, mean neutral centroid = 515.06 Hz; $t(268.02) = 6.42, p < .001$).

The presence of smiling voice in the speech recordings was further validated by ratings completed by participants in the post-game questionnaires. As detailed later (Section 4.2.2), people rated the smiling voices as happier than the neutral ones.

While there is no agreed-upon acoustic definition of smiling voice, as previously mentioned, we have taken several measures to ensure that the smiling voice in our sample is consistent with previous findings and theories. With the experimenter visually checking for the presence of smiles, the acoustic analysis, and participants’ perceptual evaluation, we are confident that our smiling voice sample was indeed ‘smiley’.

3.5. Procedure

Participants were told that the goal of the game was to earn as much money as possible, and that mutual cooperation with the virtual agent would lead to greater profit. They were informed that they could not verbally interact with the agent, but they would hear it speak an utterance at the beginning of each round. The participant started each of the 20 rounds of the game with a notional sum of £8, and s/he then decided whether to invest all, part, or none of it with the virtual agent. Whatever was invested, the agent received three times the invested amount. The turn then moved to the agent, which could return some of the investment to the participant. The agent was programmed with one of two behaviors: *generous* – returning between 120% to 240% of the investment; *mean* – returning between 0% and 120%. Here,

we chose a 120% range for both behavior conditions because we wanted our virtual agents to have a wide range of returns, so as to ensure that participants would not treat the agent as if it was making a dichotomous decision (cooperate/defect). For both conditions, the pattern of return percentages was randomly determined in advance for each of the 20 rounds, and the same *generous* or *mean* patterns were always used. For example, the *generous* agent always returned 150% of the investment in Round 1, 150% in Round 2, 180% in Round 3, 120% in Round 4, and so on. By keeping the returns constant in this way, we could control for the specific effect that different returns had on subsequent investments, and whether these would be different in the two agent voice conditions.

Each participant engaged in two games, with a different block of 20 utterances heard for each version of the game. Agent smiling was counterbalanced within participants, so that each participant played with one smiling agent and one neutral one; the behavior condition (*generous* or *mean*) was counterbalanced across participants in all possible combinations, so that some participants played two consecutive games with the same behavior condition, and some played one game with a *generous* virtual agent and one with a *mean* one. Each participant also played with two different accents, chosen at random, and the original speaker id was changed between participants (so that half of the participants heard speaker Birmingham 1, and half heard Birmingham 2). Finally, sentence block was also randomized across participants. See Fig. 2 for a visualization of the experimental design.

The game rounds proceeded as follows: Participants first heard the utterance from the virtual agent, then they indicated, by pressing a digit key, how much of £8 they wished to invest, in integers from 0 to 8. Then, participants saw a summary screen of the monetary transactions to and from the virtual agent during the round. This summary included the amount that the virtual agent had returned to them and also showed the total money that they had accumulated over all the rounds so far. Participants did not know that one game would consist of 20 rounds.

After completing each game, participants answered four questions about the virtual agents they played with on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). Two questions were about the sincerity and trustworthiness of the agent, and two were about the agent’s actions (see Table 2). These questions were taken from Rau, Li, and Li (2009), and were intended as a secondary measure, to verify that participants had explicitly understood the behavior of the virtual player and how the game worked. The inter-item correlation for these four questions was $r = .40$, which falls into the recommended range for scales with fewer than 10 items (Briggs & Cheek, 1986).

To confirm that the smiling and amusement could be perceived in the recorded voices, participants were played each utterance they had previously heard in the game, arranged in random order, and asked to rate them in terms of how happy they sounded on a 7-point Likert scale

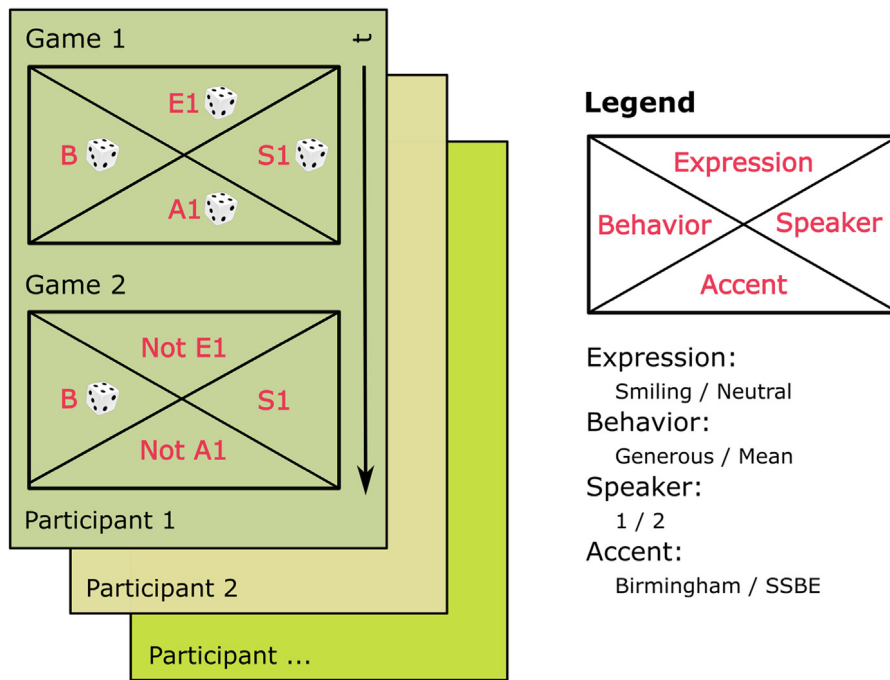


Fig. 2. Experiment design. The "Expression" condition is determined in the first game and follows a counterbalanced order. The "Behavior" condition also follows a counterbalanced order, but the value (*generous/mean*) can be repeated in two games in a row. The "Accent" condition is determined randomly for the first game, and the other value is used for the second game. The "Speaker" condition is determined randomly for the first game, and the same value is used in the second game.

Table 1

Number of participants assigned to each combination of experimental conditions for each game. Note that each participant played two games, so each participant was assigned two of these rows.

Accent	Smiling	Behavior	Number of participants
Birmingham	Neutral	Generous	28
Birmingham	Neutral	Mean	27
Birmingham	Smiling	Generous	27
Birmingham	Smiling	Mean	26
SSBE	Neutral	Generous	26
SSBE	Neutral	Mean	27
SSBE	Smiling	Generous	28
SSBE	Smiling	Mean	27

(1 = very unhappy, 7 = very happy). Here we used the word “happy”, rather than “amused”, as a widespread and easy to understand label. After rating the individual utterances, the participants then rated how much they liked the two voices on a 7-point Likert scale (1 = strongly dislike, 7 = strongly like), and had the option of stating the reason for their rating, in a free-text box. Finally, participants completed a short background questionnaire, where they were asked their age, gender, city of origin, and what accent they spoke. The total duration of the experiment was approximately 20 minutes. A summary of the number of participants assigned to each condition is shown in [Table 1](#).

4. Results

4.1. Investment game

To determine the effects of game behavior and voice on investments, a mixed-effects linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC; following standard practice, the models were compared with and without each of the key predictors to determine their effects. Investment was the dependent variable; behavior, game turn, and smiling were predictors; subject, speaker, and sentence block – i.e. the

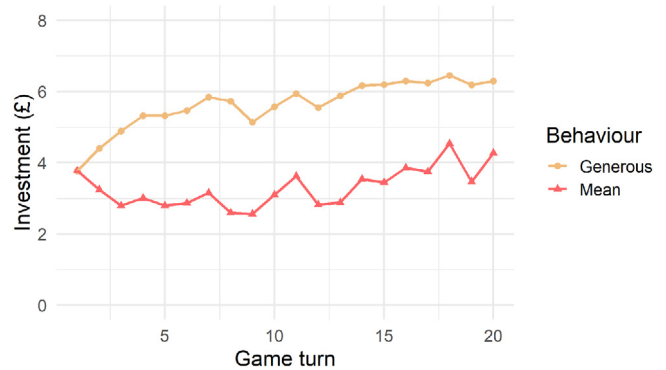


Fig. 3. Participants' average investments to the *generous* and *mean* virtual agents.

two different sets of 20 recorded sentences each – were random factors. All analyses were performed in R version 3.5.1.

There was a main effect of behavior, with an average investment of £5.63 to the *generous* virtual player and of £3.31 to the *mean* virtual player ($\chi^2(1) = 784.06, p < .001$). There was also a main effect of game turn ($\chi^2(1) = 278.12, p < .001$), with higher overall investments in the second half of the game, and a significant interaction between behavior and game turn ($\chi^2(1) = 27.50, p < .001$): investments increase in the *generous* condition and decrease initially, and then increase, in the *mean* condition as the game progresses ([Fig. 3](#)). These patterns are in line with previous observations of investment trajectories using this investment game paradigm (e.g. [Torre, Goslin, White, & Zanatto, 2018](#)).

There was also a main effect of smiling, with participants overall investing more in the smiling voice than the neutral voice (mean overall investments to smiling voice = £4.58, to neutral voice = £4.38; $\chi^2(1) = 13.60, p < .001$). There were no interactions between smiling and behavior ($\chi^2(1) = 0.36, p = .55$): as shown in [Fig. 4](#), people invest more money with the smiling agent in both behavior conditions.

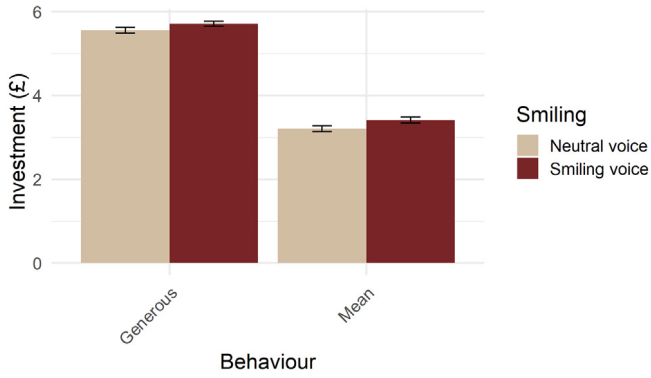


Fig. 4. Participants’ average investments to the two smiling and behavior conditions (error bars represent one standard error each way).

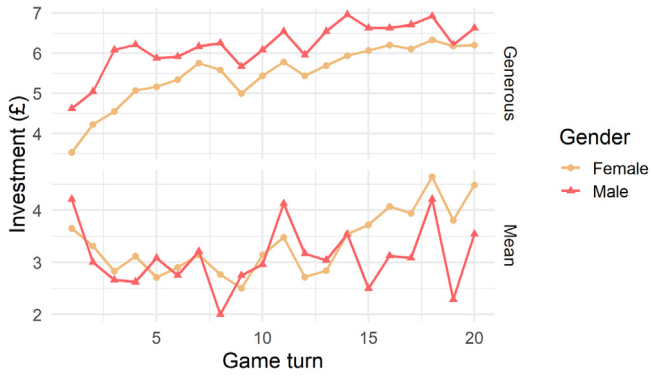


Fig. 5. Women’s and men’s average investments over time in the two behavior conditions.

There was also no interaction between smiling and game turn ($\chi^2(1) = 1.03, p = .31$), nor was there a significant three-way interaction ($\chi^2(2) = 2.34, p = .31$).

Looking at individual differences of participants, we added participants’ age and gender to the linear model. There were no main effects of either variable. The only significant effect was a three-way interaction between behavior, game turn and gender: $\chi^2(2) = 16.68, p < .001$. As shown in Fig. 5, men consistently invested more money than women in the *generous* condition (mean overall investment for men = £6.18, for women = £5.48), and less in the second half of the *mean* condition (mean investment for men in the second half of the game = £3.26, for women = £3.73).

4.2. Questionnaires

4.2.1. Trustworthiness evaluations

Summary statistics for the 4 questions about the perceived trustworthiness and behavior of the speaker are reported in Table 2. The aggregated answers to these questions were the dependent measure in a cumulative link mixed model with behavior and smiling as predictors and participant as random factor. There was a main effect of behavior ($\chi^2(1) = 122.39, p < .001$): participants overall gave higher ratings to the agents in the *generous* condition than in the *mean* condition (proportional odds ratio for *generous* behavior = 4.31, CI = [2.90, 6.41]). There was also a main effect of smiling, $\chi^2(1) = 17.00, p < .001$, with participants giving higher ratings to the smiling agents (proportional odds ratio for smiling agents = 1.44, CI = [0.99, 2.10]). There was no interaction between behavior and smiling ($\chi^2(1) = 0.88, p = .35$). We also added participants’ age and gender to the model, and found a main effect of participant age ($\chi^2(1) = 4.18, p = .04$), with trustworthiness ratings decreasing as age increases (proportional odds ratio = 0.97, CI = [0.94, 1.00]).

Table 2

Mean rating and standard deviation for the 4 trustworthiness questions, divided by agent’s behavior and smiling condition.

Question	Behavior	Smiling	Mean rating	SD
This agent wanted me to trust her.	Mean	Neutral	5.87	1.51
	Mean	Smiling	6.02	1.37
	Generous	Neutral	6.26	1.12
	Generous	Smiling	6.56	0.71
This agent was interested in talking with me.	Mean	Neutral	3.07	1.97
	Mean	Smiling	4.23	1.20
	Generous	Neutral	4.31	1.74
	Generous	Smiling	4.85	1.54
This agent was honest in communicating with me.	Mean	Neutral	2.17	1.32
	Mean	Smiling	2.25	1.42
	Generous	Neutral	4.26	1.74
	Generous	Smiling	5.15	1.39
This agent was sincere.	Mean	Neutral	2.30	1.22
	Mean	Smiling	2.55	1.46
	Generous	Neutral	4.44	1.37
	Generous	Smiling	5.16	1.44

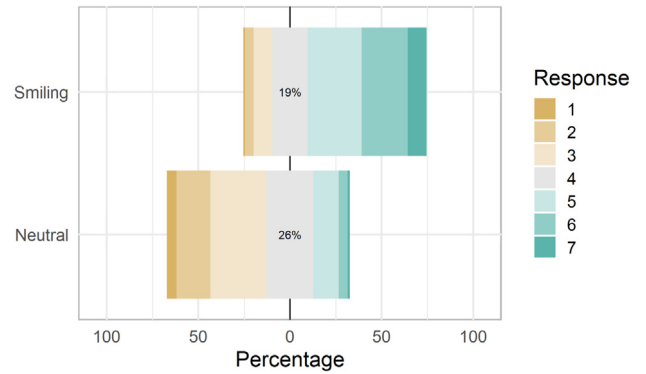


Fig. 6. Participants’ ratings of the smiling and neutral virtual agents in terms of happiness.

4.2.2. Happiness rating

To determine that smiling voice did indeed convey greater happiness, a cumulative link mixed model with happiness rating of game turn utterances as dependent variable, smiling as predictor, and participant id as random effect, was fitted to the data. There was a main effect of the smiling condition ($\chi^2(1) = 1345.64, p < .001$): as shown in Fig. 6, participants gave substantially higher ratings of happiness to the smiling voice than the neutral voice (proportional odds ratio for smiling = 8.76, CI = [7.45, 9.92]). We also added participants’ age and gender to the model, but there were no main effects, and no interactions.

4.2.3. Voice liking

A cumulative link model with liking rating as dependent variable, behavior and smiling as predictors, was fitted to the data. There was a main effect of behavior ($\chi^2(1) = 15.64, p < .001$), with people reporting liking the voices of agents that had a *generous* behavior more than agents that had a *mean* behavior (proportional odds ratio for *generous* = 2.46, CI = [1.24, 4.93]). There was also a main effect of smiling ($\chi^2(1) = 69.36, p < .001$), with people liking the smiling agents more than the neutral ones (proportional odds ratio for smiling agents = 8.66, CI = [4.15, 18.51]). There was no interaction between behavior and smiling ($\chi^2(1) = 0.10, p = .75$). We also added participants’ age and gender to the model, but there were no main effects, and no interactions.

5. Discussion

Using an investment game paradigm, we found that positive vocal emotional expression – smiling voice – increases participants’ implicit

trust attributions to virtual agents, compared with when agents speak with an emotionally neutral voice. As previously observed, the monetary returns of the agent also affected implicit trust, so that participants invested more money in the agent that was behaving *generously*. Critically, however, there was no interaction between behavior and vocal emotional expression: smiling voice enhanced trust regardless of the explicit behavioral cues that the virtual agent provided to its trustworthiness.

The effect of smiling voice in the game, supported by our questionnaire findings, adds to previous studies on emotional expression, showing that the display of a positive emotion increases trust and likeability, even in the vocal channel (Krumhuber et al., 2007; Penton-Voak, Pound, Little, & Perrett, 2006; Scharlemann et al., 2001).

Smiling was a consistent predictor of investments overall. That is to say, while participants' investments were primarily driven by the virtual player's generosity or meanness, they also overall invested more money in the smiling agents. This contrasts with the predictions of the EASI model (Van Kleef et al., 2010), according to which the display of a positive emotion in an incongruent context (such as the *mean* behavior condition) should elicit uncooperative behaviors. While Van Kleef et al. (2010) listed social dilemma tasks based on Prisoner's Dilemma among possible competitive situations, it is possible that participants in an iterated investment game view it as an essentially cooperative task. Specifically, while typical Prisoner's Dilemma tasks involve a dichotomous choice (cooperate/defect), in our experiment, even in the *mean* condition, the agent was still returning a (small) amount of money, which might have been seen as a partially cooperative signal by participants.

If participants are reluctant to give up on cooperation — as shown by the fact that investments increase in the second half of the game in the *mean* condition (Fig. 3) — they might be even more reluctant to give up on partners who seem to encourage them to cooperate, with their positive emotional expression. In Krumhuber et al. (2007), people explicitly and implicitly trusted smiling faces more than neutral faces, regardless of the sincerity of their smile, and genuine smiles were trusted more than fake smiles (Krumhuber et al., 2007). Similarly, Reed et al. (2012) found that people displaying either Duchenne or non-Duchenne smiles were more likely to cooperate in a one-shot investment game (Reed et al., 2012). Thus, displaying an emotion, even a feigned one, might be preferred to not displaying any emotion at all, hence the increased investments to the *mean* smiling agents.

Additionally, participants might have felt more positive emotions themselves upon hearing a smiling agent. In fact, emotional expressions can evoke affective reactions in observers (Geday, Gjedde, Boldsen, & Kupers, 2003), which may subsequently influence their behavior (Hatfield, Cacioppo, & Rapson, 1994), and this 'emotional contagion' might be transmitted through the auditory channel as well. If this is the case, participants might have trusted the smiling agents more because feeling a positive emotion themselves might have prompted them to behave in a cooperative manner (Mieth et al., 2016; Schug et al., 2010). These results show similarities with Tsankova et al. (2015), who found that people rated trustworthy faces and voices as happier (Tsankova et al., 2015). Although they addressed the issue from the opposite direction — "Are trustworthy stimuli perceived as happier?" rather than "Are happy stimuli perceived as trustworthy?" — taken together, the studies suggest a bidirectionality in the perception of trustworthiness and cues to positive emotion, congruent with a 'halo effect' of positive traits (Lau, 1982).

The smiling-voice effect suggests that, in the absence of visual information, the audio equivalent of a Duchenne smile might act as a relative 'honest signal' of cooperation. As mentioned before, Duchenne smiles are smiles describing genuine happiness or amusement (Ekman & Friesen, 1982). Traditionally, in the visual domain they can be distinguished from other types of smiles because they involve the contraction of the 'Orbicularis Oculi' muscle, which is a movement that is notoriously more difficult to fake (Ekman & Friesen, 1982; Schug

et al., 2010). Obviously, in the auditory channel it is not possible to detect a genuine smiling voice from this muscular movement. However, it is possible that a smiling voice which sounds happy might be the auditory equivalent of a Duchenne smile. As participants indicated that the smiling voices used in this study did sound happy, it is possible that the expression of happiness and amusement in the speech signal led listeners to believe that the agent could be trusted.

A limitation of this study is that no video recordings were taken during the audio recordings of the speakers used in this experiment. This means that, while every effort was made to ensure consistency in the smile production, it is possible that our speakers might have produced different kinds of smiles. As is well known in emotion theory, smiles can convey many different meanings, and several different facial expressions of smiles are known (e.g. Keltner, 1995; Rychlowska et al., 2017). However, much of the research on the effect of different types of smiles on person perception and decision making has concentrated on the difference between polite (non-Duchenne) and genuine (Duchenne) smiles (e.g. Chu et al., 2019; Krumhuber et al., 2007; Reed et al., 2012). Traditionally, these two are characterized by different muscle activation, with non-Duchenne smiles only activating the Zygomaticus Major muscle, and Duchenne smiles also activating the Orbicularis Oculi muscle (Frank, Ekman, & Friesen, 1993). However, recent studies have suggested that Orbicularis Oculi activation in Duchenne smiles might actually be a by-product of the Zygomaticus Major activation (Girard et al., 2019; Krumhuber & Manstead, 2009). Also, the acoustics of smiling are only affected by activation of the Zygomaticus Major muscle, which contributes to vocal tract shape, but not of Orbicularis Oculi. Following past research that Orbicularis Oculi activation is the only thing that distinguishes Duchenne from non-Duchenne smiles, we would still expect both smiles to sound the same, as the Zygomaticus Major activation would be the same.

Still, research on the acoustic characteristics of different types of smiles is lacking. Drahota et al. (2008-04) obtained three different smiling expressions — Duchenne smiles, non-Duchenne smiles, and suppressed smiles — as well as a neutral baseline, from English speakers, and asked participants to correctly identify these four expressions. Participants were only able to reliably distinguish Duchenne smiles from non-smiles, but the majority of the other smile types were classified as non-smiles. Furthermore, they only performed pairwise comparisons between a smile type and a non-smile, but they did not compare differences in identification between two different smile types. Even though they only had 11 participants, which warrants for a much-needed replication of this study, this finding suggests that people might only be able to acoustically discriminate between two categories, smile and non-smile.

Similar results were obtained in studies using different types of visual smiles in decision-making tasks. Previous work using cooperative games with Duchenne and non-Duchenne (facial) smiles have shown that people made the same decisions regardless of the type of smile (Krumhuber et al., 2007; Reed et al., 2012). This suggests that people might react according to a broad, dichotomous smile category (smile vs. non-smile), even though the smiles in the experiment stimuli were of different qualities. This corroborates previous findings in nonconscious mimicry, whereby facial EMG recordings were different when viewing a face with a Duchenne smile and a neutral expression, but not when viewing a face with a non-Duchenne smile and a neutral expression (Surakka & Hietanen, 1998). This contrasts with Chu et al. (2019), who found that participants cooperated more with a confederate expressing a non-Duchenne smile, than with a confederate expressing a Duchenne smile, following a breach of trust. However, in this study the confederate only showed the smiling expression after the cooperate/defect decision was made, whereas in Krumhuber et al. (2007), Reed et al. (2012), as well as in the current study, the smiling expression was displayed before the decision was made. As Chu et al. (2019) point out, this factor might have influenced the decisions and could explain the different behaviors. For example, participants might

interpret an emotional expression – such as a smile – after a decision as being an appraisal of that decision. People might put more cognitive effort into understanding this appraisal, as this is essential for shaping future interactions, hence the more accurate discrimination of different smile types. As [de Melo, Carnevale, Read, and Gratch \(2013\)](#), [de Melo et al. \(2015\)](#) suggest, a happy expression following the decision to cooperate conveys a different meaning than a happy expression following the decision to defect. This is also consistent with the EASI model ([Van Kleef et al., 2010](#)). On the other hand, a happy expression shown before the decision to cooperate/defect might rather convey some information about the emotional state of the person in question, and might be kept independent from that person’s actual behavior in the game. Also, counterparts’ smiles may lead people to anticipate positive social outcomes ([Kringelbach & Rolls, 2003](#)). Thus, it seems that the timing of emotional expression in relation to the behavior of interest drastically changes the interpretation of that, and future, behaviors.

It would be very interesting to replicate the current experiment with different smiling voices, shown before and after the action is taken in the game. Also, if a similar study were to be replicated, the actual facial expression of the speakers could be recorded in order to determine whether different facial expressions correspond to different auditory smiles, both in terms of objective measures (acoustics) and in terms of perception and behavior correlates in the game.

So far, we have compared our results with previous studies that used facial smiles. These comparisons are necessary, as at the time of writing there are virtually no studies that have employed trust games with expressive voices. However, emotional expressions are naturally multimodal, and it is possible that a certain emotion expressed only in the voice might elicit different behaviors than if it were expressed only in the face, or in a voice + face combination. In fact, previous research suggested that an ‘Emotional McGurk Effect’ might be at play ([Fagel, 2006](#); [Mower, Mataric, & Narayanan, 2009](#); [Pourtois, de Gelder, Bol, & Crommelinck, 2005](#)). Thus, our current results can only inform the design of voice-based artificial agents, but should not be extended to the design of embodied agents.

The results from questionnaires validate the behavioral measures obtained from the investment game. We found that people consistently gave higher ratings of trustworthiness and liking to the smiling agents, and to the agents that behaved *generously* in the game. Again, the lack of interactions between smiling and behavior suggests that the smiling voice mitigates negative reactions following an untrustworthy behavior.

We also found some evidence that individual differences among participants might play a role in trusting behavior, as shown by the 3-way interaction between behavior, game turn, and gender (Section 4.1). The effect of gender on trusting and trustworthiness has been widely studied using game theoretic paradigms, but so far there has been no definite conclusion on whether women trust more/are more trustworthy than men, or vice versa (e.g. [Bonein & Serra, 2009](#); [Chaudhuri, Paichayontvijit, & Shen, 2013](#); [Slonim & Guillen, 2010](#)). Our results support previous findings showing that we tend to trust people of the opposite gender more ([Slonim & Guillen, 2010](#)), as men in our experiment invested more money than women to the virtual agents, which had a female voice. They also support findings that men trust more than women in general ([Chaudhuri & Gangadharan, 2007](#)). However, these conclusions only hold insofar as the *generous* behavior condition is concerned, as in the *mean* condition men actually trusted the virtual agent less than women did. A similar behavior was previously observed in [Haselhuhn, Kennedy, Kray, Van Zant, and Schweitzer \(2015\)](#), who found that men showed less trust following a trust breach on the trustee’s part ([Haselhuhn et al., 2015](#)). Also, [Torre et al. \(2018\)](#) showed that people who formed a first impression of trustworthiness of a virtual agent punished it when the agent behaved in an untrustworthy manner, by investing less money than to an agent whose first impression was lower. Thus, a ‘congruency effect’ might be at play here: our male

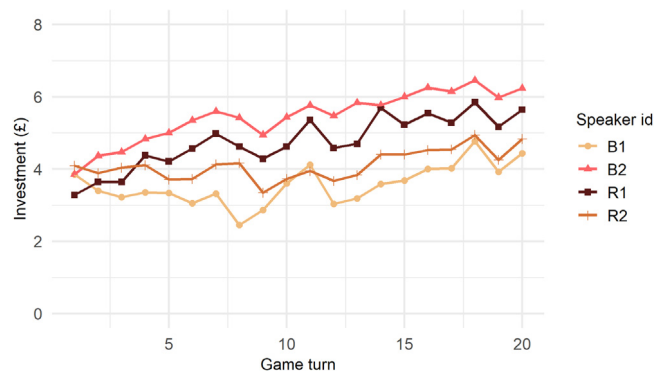


Fig. 7. Participants’ average investments to the four different voices.

participants might have formed a first impression of trustworthiness of the female agents ([Slonim & Guillen, 2010](#)); when this first impression was congruent with the observed behavior (in the *generous* condition), the agent received more monetary investments from the male participants. On the other hand, when the first impression was incongruent with the observed behavior (*mean* condition), it received less (cf. [Torre et al., 2018](#)). Participants’ age did not have an effect on the behavioral results from the investment game, but it did influence participants’ explicit ratings of the artificial agents’ trustworthiness, with older people indicating lower trust. This is consistent with the idea that younger people trust technology more, perhaps due to a higher degree of familiarity (e.g. [Czaja & Sharit, 1998](#); [Giuliani & Scopelliti, 2005](#); [Scopelliti, Giuliani, & Fornara, 2005](#)). However, we did not match participants’ age – or gender– systematically, so more research is needed on the role of individual differences on trust towards voice-based artificial agents.

Finally, speaker identity was varied randomly rather than wholly systematically in our experimental design, and so we included speaker identity as a random rather than fixed effect in our analyses. It is possible, indeed likely, that participants’ trust attributions were influenced by the virtual agents’ unique vocal profiles as well as their behavior and smiling status. In fact, [Fig. 7](#) shows that people invested more money with speaker B2, followed by speakers R1, R2, and B1 (mean overall investments = £5.46, £4.76, £4.11, £3.56, respectively). This is not unexpected: voices carry a wide variety of information about the speaker, such as gender, accent, age, emotional state, socio-economic background, etc., and all this information is implicitly used by listeners to form an initial impression of the speaker; a short exposure to someone’s voice is enough to determine if that someone can be trusted ([McAlear, Todorov, & Belin, 2014](#)). For example, in the free-text comments explaining the liking rating to each voice, one participant remarked that smiling speaker B2 “varied in tone and was much more interesting to listen to” and neutral speaker B2 was “calm and convincing”; on the other hand, smiling speaker R2 was “mellow and monotone” and neutral speaker R2 “sounded bored and insincere”. Smiling speaker B1 was “quite annoying” and the neutral version “did not seem trustworthy or reassuring”, “sounded too neutral” and even “too fake”. Thus, when designing a voice for an artificial agent, it is important to also keep in mind what effect its specific vocal imprint will have on the user (see also [McGinn & Torre, 2019](#)). Nevertheless, any potential between-speaker differences in the current experiment were nested within the effect of smiling voice, as all speakers were recorded in both smiling and neutral conditions.

6. Conclusion

We have shown that emotional expressivity in the vocal channel can influence trusting behaviors even in the absence of visual cues to speaker happiness. In particular, we found that people trusted virtual

agents that were smiling in their voice more than agents that had a neutral voice, an effect that was independent of how these agents actually behaved in an economic game. This has important implications for machine voice design. As many everyday artificial agents are disembodied, but communicate through vocal channels, it is important to design those voices appropriately. In particular, we have shown that a smiling voice might, in certain contexts, somewhat mitigate the effect of relatively negative behaviors on the part of a virtual agent.

CRedit authorship contribution statement

Iaria Torre: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing — original draft, Writing — review & editing. **Jeremy Goslin:** Conceptualization, Methodology, Writing — review & editing. **Laurence White:** Conceptualization, Methodology, Writing — review & editing.

Acknowledgments

This work was supported by CogNovo FP7-PEOPLE-2013-ITN-604764 (<https://CogNovo.eu>), a project funded by Marie Skłodowska-Curie Actions. The first author is funded from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713567 (<http://edge-research.eu/>), and from the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2016) and is co-funded by the European Regional Development Fund.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.chb.2019.106215>.

References

Antos, D., De Melo, C., Gratch, J., & Grosz, B. J. (2011). The influence of emotion expression on perceptions of trustworthiness in negotiation. In *Proceedings of the 25th AAAI conference on Artificial Intelligence*.

Arias, P., Soladie, C., Bouafif, O., Robel, A., Segurier, R., & Aucouturier, J.-J. (2018). Realistic transformation of facial and vocal smiles in real-time audiovisual streams. *IEEE Transactions on Affective Computing*.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41(3), 258–290. <http://dx.doi.org/10.1037/h0055756>.

Barthel, H., & Quené, H. (2015). Acoustic-phonetic properties of smiling revised: Measurements on a natural video corpus. In *Proceedings of the 18th international congress of phonetic sciences*.

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

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer (version 6.0.27)*. Retrieved from <http://www.praat.org>.

Bonein, A., & Serra, D. (2009). Gender pairing bias in trustworthiness. *The Journal of Socio-Economics*, 38(5), 779–789.

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

Breazeal, C. (2009). Role of expressive behaviour for robots that learn from people. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, 364(1535), 3527–3538.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54(1), 106–148.

Chaudhuri, A., & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, 959–985.

Chaudhuri, A., Paichayontvijit, T., & Shen, L. (2013). Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups. *Journal of Economic Psychology*, 34, 181–194.

Chu, V. C., Lucas, G. M., Lei, S., Mozgaj, S., Khooshabeh, P., & Gratch, J. (2019). Emotion regulation in the prisoner's dilemma: Effects of reappraisal on behavioral measures and cardiovascular measures of challenge and threat. *Frontiers in Human Neuroscience*, 13, 50.

Cowen, A. S., Laukka, P., Elflein, H. A., Liu, R., & Keltner, D. (2019). The primacy of categories in the recognition of 12 emotions in speech prosody across two cultures. *Nature Human Behaviour*, 3, 369–382.

Czaja, S. J., & Sharit, J. (1998). Age differences in attitudes toward computers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 53(5), P329–P340.

de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2013). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, 106(1), 73.

de Melo, C. M., Gratch, J., & Carnevale, P. J. (2015). Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, 6(2), 127–136.

DePaulo, B. M., Blank, A. L., Swaim, G. W., & Hairfield, J. G. (1992). Expressiveness and expressive control. *Personality and Social Psychology Bulletin*, 18(3), 276–285. <http://dx.doi.org/10.1177/0146167292183003>.

Drahota, A., Costall, A., & Reddy, V. (2008-04). The vocal communication of different kinds of smile. *Speech Communication*, 50(4), 278–287. <http://dx.doi.org/10.1016/j.specom.2007.10.001>.

Ekman, P. (1999). Basic emotions. In P. Dalglish, & M. Power (Eds.), (pp. 45–60). John Wiley & Sons, Ltd. <http://dx.doi.org/10.1002/0470013494.ch3>.

Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, 58(2), 342–353. <http://dx.doi.org/10.1037//0022-3514.58.2.342>.

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, 6(4), 238–252. <http://dx.doi.org/10.1007/bf00987191>.

El Haddad, K., Cakmak, H., Moinet, A., Dupont, S., & Dutoit, T. (2015). An HMM approach for synthesizing amused speech with a controllable intensity of smile approach for synthesizing amused speech with a controllable intensity of smile. In *Signal processing and information technology (ISSPIT), 2015 IEEE International Symposium on* (pp. 7–11). Abu Dhabi, UAE.

El Haddad, K., Torre, I., Gilmartin, E., Çakmak, H., Dupont, S., Dutoit, T., et al. (2017). Introducing amus: The amused speech database. In N. Camelin, Y. Estève, C. Martin-Vide (Eds.), *Proceedings of statistical language and speech processing conference* (pp. 229–240). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-68456-7_19.

Elflein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203–235. <http://dx.doi.org/10.1037/0033-2909.128.2.203>.

Elkins, A. C., & Derrick, D. C. (2013). The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, 22(5), 897–913.

Erickson, D., Menezes, C., & Sakakibara, K.-i. (2009). Are you laughing, smiling or crying? In *Proceedings of the annual summit and conference of the asia-pacific signal and information processing association* (pp. 529–537).

Fagel, S. (2006). Emotional McGurk effect. In *Proceedings of the international conference on speech prosody international conference on speech prosody (Vol. 1)*. Dresden, Germany.

Fagel, S. (2009). Effects of smiled speech on lips, larynx and acoustics. In *Proceedings of the auditory visual speech processing conference* (pp. 18–21). Norwich, UK.

Feinberg, D. (2019). Parselmouth praat scripts in python. <http://dx.doi.org/10.17605/OSF.IO/6DWR3>.

Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology*, 64(1), 83.

Geday, J., Gjedde, A., Boldsen, A.-S., & Kupers, R. (2003). Emotional valence modulates activity in the posterior fusiform gyrus and inferior medial prefrontal cortex in social perception. *Neuroimage*, 18(3), 675–684.

Girard, J. M., Shandar, G., Liu, Z., Cohn, J. F., Yin, L., & Morency, L.-P. (2019). Reconsidering the Duchenne smile: Indicator of positive emotion or artifact of smile intensity? In *Proceedings of the 8th international conference on affective computing and intelligent interaction*.

Giuliani, M. V., & Scopelliti, F. (2005). Elderly people at home: technological help in everyday activities. In *Proceedings of the IEEE international workshop on robot and human interactive communication* (pp. 365–370).

Haselhuhn, M. P., Kennedy, J. A., Kray, L. J., Van Zant, A. B., & Schweitzer, M. E. (2015). Gender differences in trust dynamics: Women trust more than men following a trust violation. *Journal of Experimental Social Psychology*, 56, 104–109.

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press.

Keltner, D. (1995). Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame. *Journal of Personality and Social Psychology*, 68(3), 441.

Kohler, K. J. (2008). 'Speech-smile', 'speech-laugh', 'laughter' and their sequencing in dialogic interaction. *Phonetica*, 65(1–2), 1–18.

Kringelbach, M. L., & Rolls, E. T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *Neuroimage*, 20(2), 1371–1383.

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New evidence on felt and false smiles. *Emotion*, 9(6), 807.

Krumhuber, E. G., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4), 730–735.

LaFrance, M., & Hecht, M. A. (1995). Why smiles generate leniency. *Personality and Social Psychology Bulletin*, 21(3), 207–214.

- Lasarcyk, E., & Trouvain, J. (2008). Spread lips+ raised larynx+ higher f0= smiled speech?-an articulatory synthesis approach. In *Proceedings of the 8th international seminar on speech production* (pp. 43–48). Strasbourg, France.
- Lau, S. (1982). The effect of smiling on person perception. *The Journal of Social Psychology*, *117*(1), 63–67. <http://dx.doi.org/10.1080/00224545.1982.9713408>.
- Lount, J., & Robert, B. (2010). The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology*, *98*(3), 420–433. <http://dx.doi.org/10.1037/a0017344>.
- Matsumoto, D. (1993-06). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an american sample. *Motivation and Emotion*, *17*(2), 107–123. <http://dx.doi.org/10.1007/bf00995188>.
- McAleer, P., Todorov, A., & Belin, P. (2014). How do you say Hello? Personality impressions from brief novel voices. *PLoS One*, *9*(3), e90779.
- McGinn, C., & Torre, I. (2019). Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robot. In *14th ACM/IEEE international conference on human-robot interaction* (pp. 211–221).
- Mieth, L., Bell, R., & Buchner, A. (2016). Facial likability and smiling enhance cooperation, but have no direct effect on moralistic punishment. *Experimental Psychology*.
- Mower, E., Mataric, M. J., & Narayanan, S. (2009). Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *IEEE Transactions on Multimedia*, *11*(5), 843–855.
- Mueser, K. T., Grau, B. W., Sussman, S., & Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, *46*(2), 469–478. <http://dx.doi.org/10.1037/0022-3514.46.2.469>.
- Mussel, P., Hewig, J., Allen, J. J., Coles, M. G., & Miltner, W. (2014). Smiling faces, sometimes they don't tell the truth: Facial expression in the ultimatum game impacts decision making and event-related potentials. *Psychophysiology*, *51*(4), 358–363.
- Ohala, J. J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America*, *68*, S33.
- Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128.
- Otta, E., Lira, B. B. P., Delevati, N. M., Cesar, O. P., & Pires, C. S. G. (1994). The effect of smiling and of head tilting on person perception. *Journal of Psychology*, *128*(3), 323–331. <http://dx.doi.org/10.1080/00223980.1994.9712736>.
- Penton-Voak, I. S., Pound, N., Little, A. C., & Perrett, D. I. (2006). Personality judgments from natural and composite facial images: More evidence for a “kernel of truth” in social perceptionkernel of truth in social perception. *Social Cognition*, *24*(5), 607–640. <http://dx.doi.org/10.1521/soco.2006.24.5.607>.
- Podesva, R., Callier, P., Voigt, R., & Jurafsky, D. (2015). The connection between smiling and GOAT fronting: Embodied affect in sociophonetic variation. In *Proceedings of the 18th international congress of phonetic sciences*.
- Pourtois, G., de Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex*, *41*(1), 49–59.
- Preuschoft, S. (1992). “Laughter” and “Smile” in Barbary Macaques (*Macaca Sylvanus*). *Ethology*, *91*(3), 220–236. <http://dx.doi.org/10.1111/j.1439-0310.1992.tb00864.x>.
- Rau, P. L. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595.
- Reed, L. I., Zeglen, K. N., & Schmidt, K. L. (2012). Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma games. *Evolution and Human Behaviour*, *33*(3), 200–209. <http://dx.doi.org/10.1016/j.evolhumbehav.2011.09.003>.
- Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., et al. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, *20*(3), 259–267.
- Robson, J., & Beck, J. M. (1999). Hearing smiles-Perceptual, acoustic and production aspects of labial spreading. In *Proceedings of the 14th international congress of phonetic sciencesinternational congress of phonetic sciences* (pp. 219–222).
- Russell, J. A., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*(1), 329–349. <http://dx.doi.org/10.1146/annurev.psych.54.101601.145102>.
- Rychlowska, M., Jack, R. E., Garrod, O. G. B., Schyns, P. G., Martin, J. D., & Niedenthal, P. M. (2017). Functional smiles: Tools for love, sympathy, and war. *Psychological Science*, *28*(9), 1259–1270.
- Scharlemann, J. P. W., 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*(5), 617–640.
- Schubert, E., Wolfe, J., & Tarnopolsky, A. (2004). Spectral centroid and timbre in complex, multiple instrumental textures. In *Proceedings of the international conference on music perception and cognition* (pp. 112–116).
- Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T., & Bonnet, K. (2010). Emotional expressivity as a signal of cooperation. *Evolution and Human Behaviour*, *31*(2), 87–94. <http://dx.doi.org/10.1016/j.evolhumbehav.2009.09.006>.
- Scopelliti, M., Giuliani, M. V., & Fornara, F. (2005). Robots in a domestic setting: a psychological approach. *Universal Access in the Information Society*, *4*(2), 146–155.
- Shor, R. E. (1978). The production and judgment of smile magnitude. *The Journal of General Psychology*, *98*(1), 79–96. <http://dx.doi.org/10.1080/00221309.1978.9920859>.
- Slonim, R., & Guillen, P. (2010). Gender selection discrimination: Evidence from a trust game. *Journal of Economic Behaviour and Organization*, *76*(2), 385–405.
- Surakka, V., & Hietanen, J. K. (1998). Facial and emotional reactions to duchenne and non-duchenne smiles. *International Journal of Psychophysiology*, *29*(1), 23–33.
- Tartter, V. C. (1980). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24–27. <http://dx.doi.org/10.3758/bf03199901>.
- Tartter, V. C., & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America*, *96*(4), 2101–2107.
- Torre, I. (2014). Production and perception of smiling voice. In *Proceedings of the first postgraduate and academic researchers in linguistics at york conference*.
- Torre, I., Goslin, J., White, L., & Zanatto, D. (2018). Trust in artificial voices: A “congruency effect” of first impressions and behavioural experience. In *Proceedings of APAS Science '18: Technology, Mind, and Society*.
- Tortosa, M. I., Lupiáñez, J., & Ruz, M. (2013). Race, emotion and trust: An ERP study. *Brain Research*, *1494*, 44–55.
- Tsankova, E., Krumhuber, E. G., Aubrey, A. J., Kappas, A., Möllering, G., Marshall, D., et al. (2015). The multi-modal nature of trustworthiness perception. In *Proceedings of the 1st joint conference on facial analysis, animation, and auditory-visual speech processing* (pp. 147–152).
- Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, *42*, 45–96.
- Xu, Y., & Chuenwattanapranithi, S. (2007). Perceiving anger and joy in speech through the size code. In *Proceedings of the 16th international congress of phonetic sciencesinternational congress of phonetic sciences* (pp. 2105–2108).