The Strategic Display of Emotions

Chen, Daniel; Hopfensitz, Astrid; van Leeuwen, Boris; van de Ven, Jeroen

Publication date:
2019

Document Version
Early version, also known as pre-print

CentER

Discussion Paper

No. 2019-014

THE STRATEGIC DISPLAY OF EMOTIONS

By

Daniel Chen, Astrid Hopfensitz,
Boris van Leeuwen, Jeroen van de Ven

30 April 2019

ISSN 0924-7815
ISSN 2213-9532

TILBURG UNIVERSITY

The Strategic Display of Emotions*

Daniel Chen†
Astrid Hopfensitz‡
Boris van Leeuwen§
Jeroen van de Ven¶

April 2019

Abstract

The emotion that someone expresses has consequences for how that person is treated. We study whether people display emotions strategically. In two laboratory experiments, participants play task delegation games in which managers assign a task to one of two workers. When assigning the task, managers see pictures of the workers and we vary whether getting the task is desirable or not. We find that workers strategically adapt their emotional expressions to the incentives they face, and that it indeed pays off to do so. Yet, workers do not exploit the full potential of the strategic display of emotions.

*We thank numerous seminar and conference audiences, and especially Martin Dufwenberg, Uri Gneezy, Alex Imas, Theo Offerman, Giorgia Romagnoli and Sigrid Suetens for helpful suggestions and comments. We would like to acknowledge financial support from the ANR project SINT ANR-15-CE33-0005-01 as well as support from the Institute for Advanced Study in Toulouse (IAST).

†Toulouse School of Economics, University of Toulouse Capitole, daniel.chen@iast.fr

‡Toulouse School of Economics, University of Toulouse Capitole, astrid.hopfensitz@tse-fr.eu

§Department of Economics and CentER, Tilburg University. b.vanleeuwen@uvt.nl

¶Amsterdam School of Economics (University of Amsterdam) and Tinbergen Institute. j.vandeven@uva.nl

Emotions play an important role in strategic settings (Battigalli et al., 2015; Loewenstein, 2000; Frank, 1987, 1988; Hirshleifer, 1984).[1] People's choices are affected by their emotions, and others respond to this. Anger is a good illustration of this. Angry people are willing to destroy surplus (Bosman & Van Winden, 2002). In response, people that are known to have a predisposition to get angry enjoy a better treatment in bargaining settings. Even in one-shot interactions, in which predispositions are not revealed through reputation, emotions can convey intentions effectively (Frank, 1988). Emotions are accompanied by characteristic facial expressions that people are able to recognize and interpret (Frijda, 1986; Ekman, 1993). An accumulating body of empirical evidence supports these ideas. The display of anger during negations leads to more favorable offers (Reed et al., 2014; Andrade & Ho, 2009; van Kleef et al., 2004) and smiling fosters trust (Scharlemann et al., 2001; Centorrino et al., 2015).

Given the strategic value of expressing emotions, people have clear incentives to mimic expressions. While emotional expressions are not fully under a person's control (such as blushing), to some extent they are (Frijda, 1986; Ekman, 1993). This opens up the possibility that emotional expressions are used strategically. Surprisingly, there are virtually no studies that test the strategic use of expressed emotions. This is in sharp contrast to verbal communication, which has been intensively studied in economics. Yet it is clearly important for a good understanding of the role that emotions play. It speaks to the question whether it can be equilibrium behavior to pay attention to emotional expressions and what the level of people's sophistication is in this respect.

Our main contribution is to study whether people strategically adjust their expressed emotions in different economic games. In a laboratory experiment, participants play a task delegation game. In this game, a 'manager' assigns an investment task to one of two 'workers'. The main treatment variation is whether getting the task is desirable or not for the worker. This crucially changes the strategic nature

[1] Early accounts of the role of emotions in economics can be found in Smith (1759) and Bentham (1789), see also Ashraf et al. (2005). Our focus is on the role of emotions in social interactions. Emotions are also important in the domain of individual decision making. See for instance the work by Loomes and Sugden (1982) on regret and disappointment, or Gneezy et al. (2014) on the role of guilt in pro-social behavior.

of the game for workers. While they wish to be selected by the manager in one treatment, they do not want the task in the other. In either treatment, the manager benefits from a worker's investment in the task, while to the worker investing is always costly. It is therefore in the manager's interest to assign the task to the worker that is most likely to invest. Many tasks in organizational environments fit this setting. For instance, some tasks are perceived as relatively pleasurable (teach a course on game theory), while others are best avoided (administrative duties).

Our main interest is in whether the worker's emotional expression affects the likelihood of getting the task and whether workers exploit this. In the experiment, workers could take pictures of themselves on which they express emotions, and send these to the manager. We hypothesize that the expression of positive emotions, like happiness, is associated with higher trustworthiness. Likewise, we expect that negative emotions, and especially anger, are associated with lower trustworthiness. If so, workers can avoid getting the task by showing less positive and more negative emotions. We therefore predict that workers would show more negative emotions when the task is undesirable to get.

We present results from two experiments. In Experiment I, we instructed workers to take one picture on which they looked happy and one picture on which they looked angry. Workers could then choose which picture they wanted to show to the manager. Their choice was implemented probabilistically, such that in some rounds their nonpreferred picture was shown. The advantage of this setup is that it gives us a counterfactual, providing us with a clean causal identification of the effect of different pictures on the managers' decisions. Managers were not made aware of the fact that workers were instructed to express emotions. We did this so that the managers would not immediately question the sincerity of the expressed emotions. In Experiment II, workers were not given any instructions on how to look on the pictures. The advantage of this setup is that we can study the extent to which participants adjust their expressions spontaneously.

The main findings are that workers can indeed avoid getting the task by expressing negative emotions (anger), and they are more likely to show negative emotions when the task is not desirable. Specifically, managers are less likely to assign the task to workers that express more negative emotions. We hypothesized that managers associate negative emotions with a lower inclination to invest. Consistent with this, workers that look more negative are rated as less trustworthy by an independent panel of subjects. Anticipating this, workers in Experiment I are twice as likely to show anger when the task is undesirable compared to when the task is desirable. In Experiment II, workers also express more negative emotions when the task is desirable, as measured by automated facial analysis software. Thus, they spontaneously adapt their emotion expressions to fit the situation. However, they only exploit the potential benefits to a limited extent, and the difference in intensity of emotions between treatments is only 35 percent of that observed in Experiment I.

We also propose a model that can accommodate the findings from our experiments. In the model, we assume that senders ('workers') can send messages about their valence levels by varying their expressed valence. Senders have a cost of lying, i.e. expressing a valence level different from what they truly experience. We show that in equilibrium, senders will overreport their valence when the task is desirable and underreport their valence when the task is undesirable. If the cost of lying is sufficiently high, there will be partial separation of types. This implies that senders strategically adapt their messages to the situation they face, while receivers ('managers') respond to the messages they observe.

To our best knowledge, ours is the first paper to study the strategic display of emotions. It fits well in the literature on cheap talk (Crawford & Sobel, 1982; Farrell & Rabin, 1996), in the sense of sending signals about private information or intentions. In that literature, the focus is on verbal messages. Most closely related is the work by Andrade and Ho (2009), who show that people strategically over-report their level of anger in verbal messages. While verbal and nonverbal messages have many similarities, they are distinct in some key respects. Nonverbal cues can be harder to fake (Ekman, 1993; Centorrino et al., 2015). Furthermore, when the facial expression of a person does not match the contents of his or her verbal statement, people may rely more on the former. Reed et al. (2014) indeed document a much smaller impact of verbal messages if they are accompanied by a facial expression that does not match the statement. This suggests that nonverbal cues may be of first-order importance.

Another closely related paper is Gneezy and Imas (2014), who show that people strategically manipulate the emotions in others. Specifically, when people have the option to anger their opponent, they do so in environments in which this pays off to do. In a theoretical contribution, Winter et al. (2016) analyze equilibrium play when people can choose their own emotional states.

Our study also relates to a growing literature showing the importance of facial cues more generally. In many economic interactions, people highly value facial cues. For many jobs, face-to-face interviews are a key part of the recruitment procedure, and most people post pictures of themselves on their professional websites. Online marketplaces such as Airbnb and Uber make it possible for buyers and sellers to see each other on pictures before trading. Business people spend large amounts of time and money to meet with their business partners before signing a contract (Forbes, 2009). In trust games, many people are willing to pay to see a picture of the other (Eckel & Petrie, 2011) or to show their own picture to others (Heyes & List, 2016). Facial cues are somewhat predictive of behavior. Based on seeing pictures of unknown others, observers can predict behavior in strategic situations slightly above chance levels (Verplaetse et al., 2007; Bonnefon et al., 2013; Vogt et al., 2013; Tognetti et al., 2013; van Leeuwen et al., 2018).

1 Task delegation games and hypotheses

In the experiment we use a novel task delegation game, loosely inspired by the games used in Babcock et al. (2017). In this game, a 'manager' is paired with two 'workers' (neutral labels are used in the experiment). The manager sees pictures of the workers and has to allocate an investment task to one of them. The chosen worker becomes the 'designated' worker and is the only worker that has to make a decision. The designated worker can accept or refuse to invest. Investing is costly to the worker but beneficial to the manager. We implement two versions of the game, that differ only in the payoffs. We vary the payoffs such that in one version it is beneficial to the worker to get the task (treatment 'Desirable') while in the other version it is not (treatment 'Undesirable').

Figure 1 illustrates the two versions of the game when the manager (M) assigns the task to worker A (the case where the task is assigned to worker B is symmetric). In treatment 'Desirable', the designated worker (A) always earns more than the other worker (B), independent of his investment decision. The designated worker earns €2 if he invests, and €2.2 if he refuses to invest. The other worker always

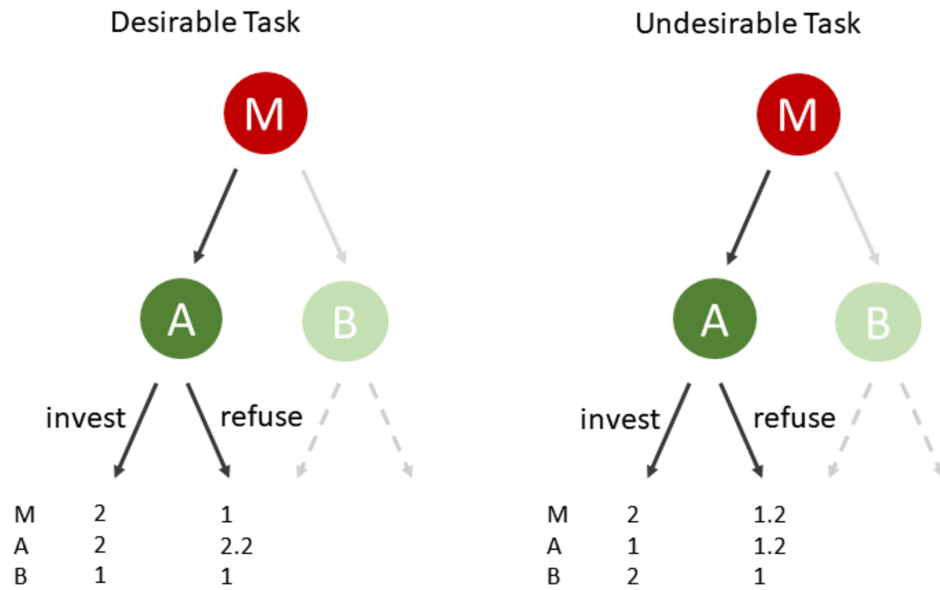


Figure 1: Payoffs to manager (M) and workers (A and B) when the task is assigned to worker A (i.e., worker A is selected to make a decision). The case in which the task is assigned to worker B is symmetric.

earns €1. The manager earns €2 if the designated worker invests, and €1 if the designated worker refuses to invest. In treatment 'Undesirable' (right part of Figure 1), a worker earns most if the task is assigned to the other worker and the other worker invests. In this version, if the designated worker (A) refuses to invest, the other worker (B) has to invest. The worker that ends up investing always earns €1. The worker not investing earns €2 if the designated worker invests and €1.2 if the designated worker refuses.

Note that in both versions of the task delegation game, the manager earns more money if the designated worker invests. This creates incentives to allocate the task to the worker that is most likely to invest, and therefore to select the worker that looks most trustworthy. If so, workers have incentives to appear trustworthy in Desirable in order to receive the task. In Undesirable, they may wish to appear untrustworthy to avoid getting the task.[2] We conjecture that negative emotions

[2]There is a caveat. If a worker is very pessimistic about the likelihood that the other worker will

(such as anger) are associated with lower trustworthiness, and positive emotions (such as happiness) are associated with higher trustworthiness.

Hypothesis 1. *In both treatments, managers are more likely to assign the task to workers that show more positive emotional expressions.*

Hypothesis 2. *Workers display more negative emotional expressions in treatment Undesirable compared to treatment Desirable.*

We test these hypotheses in two experiments. The two experiments differ in the way how workers can display emotions. The two experiments are complementary. In Experiment I we impose that workers either show a happy or an angry expression. While stylized, this allows us to cleanly identify the causal effect of emotional signals. Experiment II is more natural, allowing for more freedom in the way workers can display emotions. We will first discuss the design and main results of Experiment I before turning to Experiment II.

2 Experiment I

2.1 Experimental design and procedures

Prior to receiving any instructions about the game, participants in the role of workers were asked to take two pictures of themselves using a webcam. Participants were instructed to look natural on both photos and to display a happy expression on one picture and an angry expression on the other. Participants were allowed to retake pictures until they were satisfied with the result. The same two pictures were used throughout the experiment. All subjects - workers and managers - then received instructions on their screen (see Appendix D). Participants also received a hard-copy with a summary and were asked to respond to a series of control questions before continuing.

Participants played either the desirable or undesirable version of the task delegation game in a between-subject design. Each subject played a total of 12 rounds and was paid for every round. Subjects were rematched every round. Managers never saw the same worker more than once and workers never found out with which other

invest, he may be better off if getting the task himself and not investing.

worker they were matched. To achieve that managers never saw the same worker twice, managers were inactive in half of the rounds (receiving a fixed payment of €1).

At the beginning of every two rounds, workers could indicate which picture (the happy or the angry picture) they want to use for the next two rounds. In these next two rounds, the preferred picture was used in one round, while in the other round it was randomly determined (with equal probability) which picture was shown. This way, we have information about the worker's preferred choice, and at the same time the random variation creates a set of counterfactuals which allows us to establish the causal effect of emotional expressions on the likelihood of getting the task. Managers were not informed that workers were asked to express emotions or how pictures were selected and workers knew this. We did this so that the managers would not immediately question the sincerity of the expressed emotions.

At the end of each round, all subjects received feedback about their earnings and whether or not the (other) worker invested. Workers also learned which picture was shown. Earnings were between €12.00 and €23.40 (mean €15.60). Each session lasted around 75 minutes and ended with a survey in which we collected additional information. To preserve anonymity, worker and managers came from different cities in the Netherlands. Workers participated in Amsterdam (CREED lab) and managers in Tilburg (CentERlab). We recruited a total of 272 subjects, 136 in each role (51 percent female, mean age 22) for a total of 10 sessions.[3] Each session had 24, 28, or 32 participants, depending on the show-up. Sessions were gender-balanced, with a fraction of females that was always between 0.42 and 0.60.

Workers provided consent for the use of their pictures in the experiment. They were seated in closed, sound-proof cubicles. When taking pictures, participants were told to capture their entire face and to look into the camera. Pictures were taken with a high-quality webcam (Logitech HD 1080p). Ethical approval was granted by the IRB of the Faculty of Economics and Business at the University of Amsterdam.

[3]Prior to running this experiment, we conducted one pilot session in which we used different games (a trust game and ultimatum game). We decided to change the design to make the games more comparable and to create a situation in which looking angry has clearer potential benefits.

2.2 Measurements

Intensity of expressed emotions. We use facial recognition software (FaceReader 7.1, Bijlstra & Dotsch, 2011) to evaluate the intensity of expressed emotions. FaceReader classifies facial expressions based on the relative position of 538 grid-points on the face. It is based on an artificial neural network trained on over 10,000 images. For each emotion it gives a value between 0 and 1, reflecting the intensity of the expressed emotion on that picture. The software could successfully capture the face in all but one picture. Throughout the analysis, we use valence as a composite measure of the intensity of expressed emotions. Valence is defined as the difference between positive and negative emotions, and can vary from -1 (very negative) to +1 (very positive).[4]

Perceived trustworthiness and attractiveness. An independent group of raters evaluated pictures on trustworthiness and attractiveness on a 7-point scale. Each picture was evaluated by 8 raters (4 males, 4 females) on each dimension. Raters were recruited from the same subject pool as the managers. A total of 32 raters evaluated one picture of each of the 136 workers (either happy or angry, randomly selected). Each rater evaluated pictures on only one dimension and pictures were sorted by gender. Raters received a flat payment of €7. The interrater reliability is high (Cronbach's alpha: 0.809 and 0.814 for trustworthiness and attractiveness, respectively).

Strategic reasoning, emotional intelligence and sociodemographic data. We collected information on gender, the level of strategic reasoning, and emotional intelligence. The level of strategic reasoning is measured using an adapted version of the race game (Gneezy et al. (2010), and see also Dufwenberg et al. (2010)). This captures a subject's ability to perform backward induction. Emotional intelligence is measured in two ways. The first is performance on the 'Reading the Mind in the Eyes' test (Baron-Cohen et al., 2001), a standard test in psychology where subjects are asked to match one of four emotions to a pair of eyes. The second is the subject's ability to predict who would reject a low offer in an ultimatum game, based on facial

[4] Valence is calculated as the level of 'happy' minus the maximum level of any negative emotions (anger, sad, scared, and disgust). Subjects in our experiment express very little negative emotions besides anger, so that valence is essentially the level of 'happy' minus the level of 'anger'.

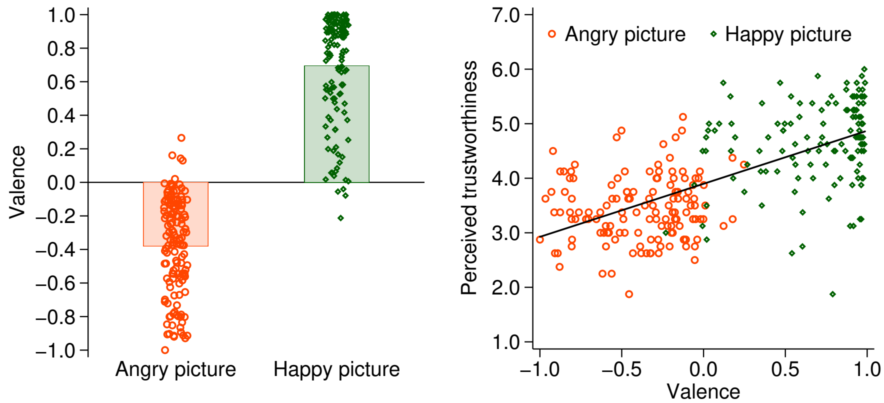


Figure 2: Valence and perceived trustworthiness on pictures. Valence is defined as the difference between positive and negative emotions. Bars indicate the mean valence by expression on the picture, circles/diamonds represent individual pictures.

expressions (see the task developed in van Leeuwen et al., 2018). We refer to this test as the 'angry button test'. The advantage of this test over the 'eyes test' is that it is more behavioral; subjects predict how people will behave. None of the measures were incentivized. For more details see Appendix A.

2.3 Results

2.3.1 Emotions and perceived trustworthiness

We confirm that expressed emotions are different between the happy and angry pictures. The left-hand panel of Figure 2 shows a substantial difference in valence between the two sets of pictures. As expected, mean valence is negative on the angry picture (-0.38) and positive on the happy picture (0.70). All but one subject display higher valence on the happy picture and the difference in valence between pictures is highly significant (signed-rank test, $p < 0.001$, $N = 135$).

We further observe a clear link between expressed emotions and trustworthiness ratings. The right-hand panel of Figure 2 shows that perceived trustworthiness is strongly positively associated with valence ($\rho = 0.654$, $p < 0.001$). A subject fixed-effects regression reveals that an increase of valence by one standard-deviation increases trustworthiness by 0.47 points ($p < 0.001$). Perceived trustworthiness is substantially higher on the happy picture (mean = 4.67) compared to the angry picture (mean = 3.43). Almost all subjects (125 out of 136, or 92 percent) look more trustworthy on their happy picture than on their angry picture, and this difference is highly significant (signed-rank test, $p < 0.001$, $N = 136$).

2.3.2 Behavior by managers

Our main research question is whether subjects strategically adapt their expression to the situation. This presumes that there is some benefit of choosing a certain expression.

We find clear evidence that the expression matters for the manager's decision. For both task types, managers are more likely to assign the task to a worker that looks happy. When two workers express different emotions, the one that looks happy is 17 percentage points more often chosen when the task is desirable (signed-rank test, $p = 0.019$, $N = 70$) and 29 percentage points more often when the task is undesirable (signed-rank test, $p < 0.001$, $N = 65$).

Table 1 shows the estimates of the likelihood that the task is assigned to a worker. The reported coefficients are in terms of odds ratios. In both task types, a happy expression significantly increases the likelihood of becoming the designated worker (columns 1 and 4). We find similar effects if we restrict worker-pairs to be of the same sex (columns 2 and 5). This implies that the effect is not driven by gender differences. As both valence and perceived trustworthiness are larger on happy pictures, it is interesting to see which factor is more important for managers, and if the intensity of valence or trustworthiness matters. Columns 3 and 6 show that only perceived trustworthiness is significant in the case of an undesirable task, supporting the idea that managers look for trustworthy-looking workers.

An important design feature is that in half of the cases it was randomly determined whether the happy or the angry picture was shown to the manager. This allows us to cleanly investigate the causal effect of emotional expressions on the likelihood being selected by the manager. To exploit this random variation, we compute for each worker the fraction of times they received the task when their happy or angry picture was *randomly* shown. We find that subjects in both treat-

Table 1: Assignment of tasks in Experiment I

	(1)	(2)	(3)	(4)	(5)	(6)
Task:		Undesirable			Desirable	
Sample:	All	Same sex	All	All	Same sex	All
Happy picture	1.818***	2.241***	1.663	1.405**	1.548**	1.351
	(0.270)	(0.534)	(0.520)	(0.207)	(0.306)	(0.399)
Valence			0.853			0.867
			(0.197)			(0.194)
Perceived trustworthiness			1.248*			1.142
			(0.148)			(0.139)
Observations	792	392	792	836	400	816
Wald test (p-value)	< 0.001	0.001	< 0.001	0.021	0.027	0.119

Notes: Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Column (6) has fewer observations because FaceReader did not capture every face. Robust s.e. in parentheses clustered at the manager level. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$.

ments are on average 13 percentage points more likely to be the designated player if the happy picture is randomly shown (signed-rank tests, $p = 0.024$, $N = 65$ for treatment Undesirable and $p = 0.018$, $N = 67$ for treatment Desirable).

In Appendix C, we show that the estimated effect of showing the happy picture survives several robustness checks. In particular, we find similar effects if we separately consider male-male and female-female pairs. Estimated coefficients are also unaffected by excluding managers that did not correctly answer all comprehension test questions on the first attempt, or restricting observations to photographs on which workers were rated to look 'natural' (see Table A.1).

2.3.3 Expression of emotions by workers

Our results indicate that there are clear benefits for workers of sending the happy expression when the task is desirable and the angry expression when the task is undesirable. We find that workers act in accordance, and are roughly twice as likely to send their angry picture when the task is undesirable compared to when the task is desirable (44 percent versus 21 percent, left panel of Figure 3), a difference that

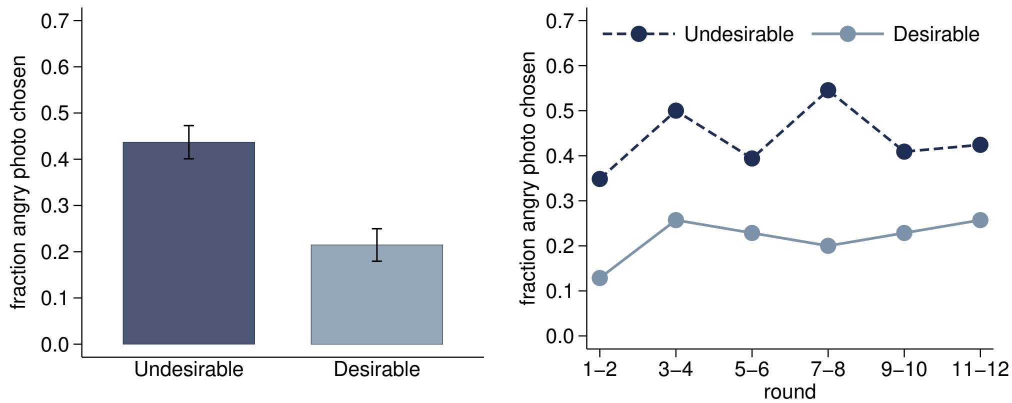


Figure 3: Fraction of workers choosing the angry picture. Left panel: fraction across all rounds (error bars indicate +/- 1 s.e.). Right panel: fraction over time (per decision made every two rounds).

is highly significant (ranksum test, $p < 0.001$, $N = 136$). The right panel of Figure 3 shows that this difference already exists in the first rounds and persists over the rounds.

Note that workers do not always show the hypothesized 'optimal' picture. However, this is not evidence per se that they are making the wrong choice. In the previous section we showed that workers are on average better off by sending their happy (angry) picture when the task is desirable (undesirable), but this may not always be the case. It may thus be that the workers who are sending the 'wrong' picture are, in fact, making the right choice. This, however, is not the case. To show this, we again exploit our design feature that it is sometimes randomly determined which of the worker's pictures is shown, but this time for the subset of cases in which workers sent the 'wrong' picture. For this subsample, we find again that sending a happy picture increases the likelihood of being the designated player. For cases in which workers want to show the 'wrong' picture, the happy picture increases the likelihood of being selected by 22 percentage points with the undesirable task (signed-rank test, $p = 0.009$, $N = 44$) and by 28 percentage points with the desirable task (signed-rank test, $p = 0.090$, $N = 16$).

2.3.4 Investment by workers

For completeness, we briefly report investment behavior by workers. Overall, in 35 percent of the cases workers invested when they were assigned the task. This investment rate is somewhat lower when the task is undesirable (28 percent versus 41 percent) but the difference is not significant (ranksum test, $p = 0.254$, $N = 136$).

Are managers right to select the workers that look happy? We only find a modest link between the expression on the picture and the worker's investment choice.[5] We note, however, that our experimental design is not well suited to answer this question. The sample of workers making an investment decision is a selective sample, chosen by managers. We therefore cannot establish any causal link between expressed emotions and investment behavior. This would require knowledge about the counterfactual, i.e., what the investment decision of the other worker would have been.

3 Experiment II

Subjects in Experiment I understand the strategic benefits of adapting their expressed emotion to the situation. Possibly, the request to express different emotions on the pictures made them aware of the potential benefits. In Experiment II, we study whether subjects spontaneously adapt their expressions.

The setup of Experiment II is very similar to Experiment I. The main difference is that this time we did not instruct them to express emotions on the pictures. Instead, in Experiment 2 subjects in the role of workers took a new picture at the start of each round. Emotions or expressions were never mentioned in the instructions. We predicted that the expressed emotions would be less extreme in this setup, making it harder to detect any differences. We therefore used a within-subject design to increase statistical power. Each subject played six rounds in each version of the game, and we varied the order of the games across sessions. Participants only received the instructions for the second part after completing the first part.

[5]In the Desirable task treatment, workers who prefer to show their happy picture invested in 45 percent of the cases. This is 27 percent for those who prefer to show their angry picture. In the Undesirable task treatment we find no difference (30 vs 27 percent investment).

We conducted 5 sessions, again with managers in Tilburg (CentERlab) and workers in Amsterdam (CREED lab). In total, 148 subjects participated in Experiment II, with 74 subjects in each role (54 percent female, mean age 22). Sessions lasted about 75 minutes, and subjects earned between €13.40 and €23.60 (mean €17.40). More details about the procedures can be found in Appendix B.

3.1 Measurements

Emotional expressions. Emotional expressions were again evaluated using FaceReader. FaceReader could successfully read emotions on 866 out of 888 pictures (97.5 percent).

Perceived trustworthiness. We followed similar procedures to obtain ratings of perceived trustworthiness. We recruited 24 raters (12 males, 12 females) at the CentERlab in Tilburg, who each rated one picture of each of the 74 workers. Half of the pictures that a rater saw came from treatment Desirable, and the other half from treatment Undesirable. Hence, each picture was rated twice, and each worker was judged by 24 raters.

Strategic reasoning, emotional intelligence and sociodemographic data. We used the same final survey collecting information on gender, the level of strategic reasoning, and emotional intelligence, as in Experiment I.

3.2 Results

3.2.1 Expression of emotions by workers

We again find evidence that subjects strategically express emotions. The mean valence when the task is desirable is 0.24, against a mean valence of 0.15 when the task is undesirable, a difference that is statistically highly significant (signed-rank test, $p < 0.001$, $N = 74$). The majority of subjects (73 percent) displays a higher valence in the desirable task. We also find that workers appear more trustworthy when the task is desirable. The average trustworthiness rating is 0.20 points higher in treatment Desirable (signed-rank test, $p = 0.012$, $N = 74$), and the majority of raters (83

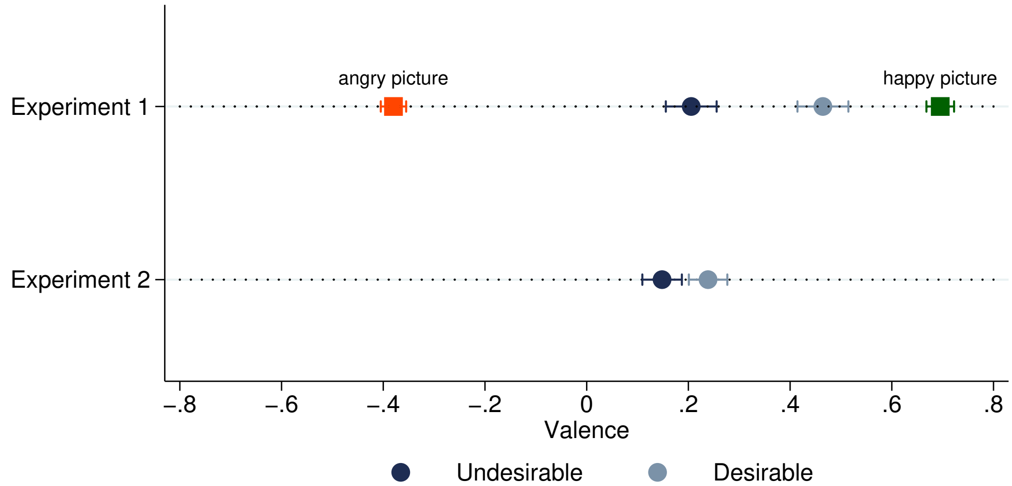


Figure 4: Mean valence on pictures taken and pictures sent. Squares indicate the mean valence on happy and angry pictures taken. Dots indicate the mean valence on pictures sent. Error bars indicate +/- 1 s.e.

percent) gives a higher average rating in treatment Desirable. In qualitative terms, this is consistent with the results from the previous experiment.[6]

In terms of magnitude, the difference between tasks is much more modest. Figure 4 shows the mean valence on the happy and angry pictures (squares, top line) from Experiment I. The difference of 1.08 gives an indication of what subjects can maximally achieve in terms of expressing different emotions. Workers do not achieve this maximal difference, because subjects do not always send the happy picture when the task is desirable or the angry picture when the task is undesirable. The observed difference in valence is 0.26 in Experiment I (dots, top line). This is about three times as high in Experiment II, where the difference in valence is 0.09 (dots, bottom line). Thus, the difference in valence across treatments in Experiment II is only 35 percent of the difference observed in Experiment I.

[6]An alternative explanation would be that subjects display higher valence with the desirable task because they 'enjoy' this task more, for example because payoffs are somewhat higher with this task. The data is not consistent with such an explanation. In Table A.2 in Appendix C, we show fixed effects regressions estimating the relation between valence and payoffs in the previous round. We find no significant effect of payoffs on valence with either task.

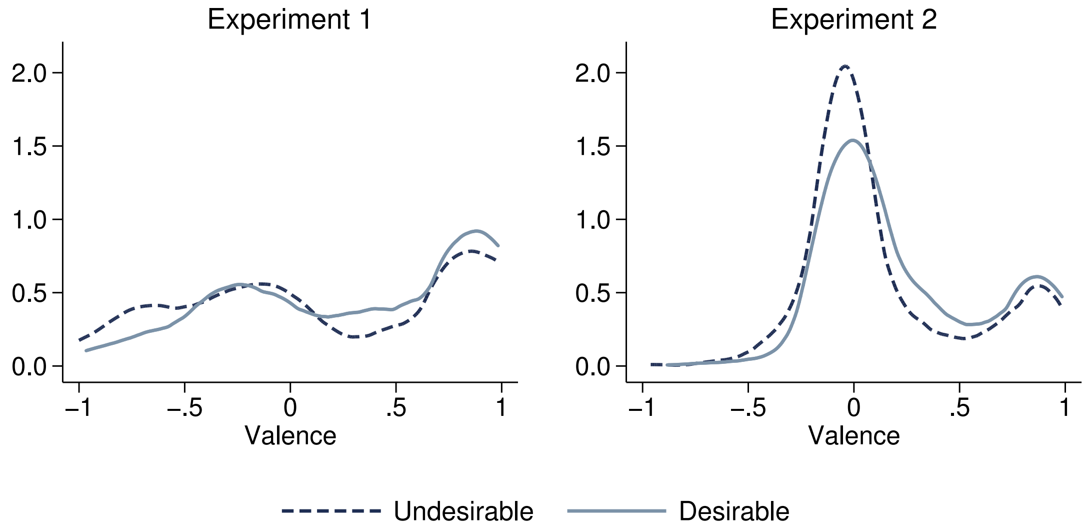


Figure 5: Kernel estimates of the distribution of valence on send pictures.

Kernel estimates of the distribution of valence on the pictures selected by subjects provide additional support (see Figure 5). In both experiments we see a shift in the distributions across the different tasks, illustrating that subjects do respond to the strategic situation they face. In Experiment I, valence has substantial mass almost everywhere on the scale. By contrast, in Experiment II it is much less dispersed and largely concentrated around 0, with little mass below zero.

3.2.2 Behavior of managers

While subjects in Experiment II vary their expressions to a lesser extent, the potential benefits appear to be of a similar magnitude. Table 2 shows the effect of valence on the likelihood of becoming the designated player. The odds ratios are above one for both tasks and both experiments. The coefficient is not significant for the undesirable task in Experiment II (column 2). This is probably due to a lack of power; few subjects express strong emotions in that treatment, and especially negative emotions are mostly absent. The odds ratio is, however, in the same ballpark as that in Experiment I (column 1). We also find a very similar correlation between valence and perceived trustworthiness: an increase of valence by one

Table 2: Assignment of tasks: Experiments I and II

	(1)	(2)	(3)	(4)
Task	Undesirable		Desirable	
Sample	Experiment I	Experiment II	Experiment I	Experiment II
Valence	1.447***	1.360	1.231*	1.898**
	(0.164)	(0.341)	(0.148)	(0.531)
Observations	792	432	816	412

Notes: Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Robust s.e.in parentheses clustered at the manager level. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

standard-deviation increases trustworthiness by 0.49 points ($p < 0.001$) in Experiment II, and this was 0.47 in Experiment I. We conclude that subjects spontaneously adapt their expressed emotions to the situation, but they only exploit the benefits to a modest extent.

4 Heterogeneity in the strategic display of emotions

We next examine if there is heterogeneity in how strategic people are in their emotional expressions. We collected information on gender, the level of strategic reasoning, and emotional intelligence, as described in subsection 2.2.

Our outcome variable concerns the level of strategic use of emotions. We evaluate this by the degree to which participants adjusted their expressions to the strategic context. For Experiment 1, we measure how strategic subjects are by the fraction of times that they send the picture that matches the situation; the happy picture when the task is desirable and the angry picture when the task is undesirable. For Experiment 2, we measure how strategic subjects are by the difference in displayed valence between treatment Desirable and treatment Undesirable.

Table 3 summarizes the results. In Experiment I, subjects who score high on the eyes test and angry button test are more strategic, while we do not find a significant impact of the level of strategic reasoning. In Experiment II, we find that people who score better on the angry button test adapt their valence more strongly to the strategic setting. Finally, only for Experiment I we observe a marginally significant

Table 3: Strategic display of emotions and individual characteristics

	(1)	(2)
Sample	Experiment 1	Experiment 2
Dep. var.	Matching picture	Increase in valence
Strategic reasoning score[a]	0.044	0.039
	(0.027)	(0.030)
Eyes test score[a]	0.105***	−0.027
	(0.031)	(0.021)
Angry button test score[a]	0.076***	0.049**
	(0.028)	(0.024)
Female	−0.094*	−0.036
	(0.056)	(0.054)
Constant	0.658***	0.102**
	(0.035)	(0.044)
Observations	136	74
R^2	0.173	0.084

Notes: OLS estimates. Column 1: Matching picture is the happy (angry) expression when the task is desirable (undesirable). Column 2: Increase in valence is mean valence when the task is desirable minus mean valence when the task is undesirable. [a]: standardized scores. See Appendix A for a description of the measures. Robust s.e. in parentheses. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

gender difference.

5 A model of expression of emotions

In our experiments, subjects in the role of workers adapt their expressed emotions strategically, though only partially, and those in the role of managers pay attention to the expressed emotions. In this section, we present a model that is consistent with this type of behavior.

We model the situation as a sender-receiver game with multiple senders. Each sender chooses an emotional expression (a 'message') to signal his inclination to invest (his 'type') to the receiver. The receiver then chooses which sender gets the task. The setup closely follows Kartik (2009), who studies a cost of lying in a cheap talk environment, thereby transforming the setting into a costly signalling game. We depart from his setup in three main ways. First, we introduce sender competition. Second, we add a comparative static, reflecting that the task can be desirable or undesirable. The case for an undesirable task resembles the ”Lure treatment” in Charness et al. (2018). Third, the payoff structure differs in different respects. In our setup the receiver always prefers the sender to invest, rather than having an interior optimum. Furthermore, our game falls in the class of ’monotonic signaling games’, as in Cho and Sobel (1990).[7]

There are two senders, $i = 1, 2$. Each sender has a type $t \in T = [0, 1]$. Types are independently drawn from a uniform distribution. The type reflects a sender's probability of investing when that sender is selected by the receiver. For simplicity, we assume that the probability of investing is fixed. The uncertainty may reflect that the sender's willingness to invest depends on idiosyncratic factors, such as his liking for a particular manager or co-worker.

After privately observing his type, each sender sends a message $m \in M = [0, 1]$.[8] In the experiment, the message is the expressed level of valence. The set M is taken to be the set of feasible expressions in terms of valence levels that subjects can achieve. We assume a natural language interpretation, where message m is interpreted by the receiver as ’I am of type m.’ After observing both messages, the receiver takes action $a_i \in A = \{0, 1\}$, where $a_i = 1$ means that the task is allocated to sender i and $a_i \neq a_j$ for $j \neq i$.

To keep things tractable, we assume specific functional forms for the payoff functions of senders and receivers. The payoff function of sender i is given by:

$$U_i^S(a_i, t, m) = \theta a_i - k(m - t)^2, \tag{1}$$

where θ reflects the value of getting the task (including the investment cost), which can be either positive ($\theta > 0$, reflecting the setting in which a task is desirable) or negative ($\theta < 0$, reflecting the setting in which a task is undesirable). The second part of the expression measures the cost of lying, i.e., showing an intensity level of valence that does not match the actual type, where we assume that the actual (expe-

[7]Loosely speaking, in a monotonic signaling game all types of senders rank the actions (including mixed strategies) by the receiver in the same way.

[8]Note that our message space is less rich than in Kartik (2009).

rienced) valence of a sender matches his type, and $k > 0$ is a parameter measuring the sender's aversion to lying.

The receiver's payoff function is given by:

$$U^R(a_i, t) = a_i t_i + (1 - a_i) t_j. \tag{2}$$

The receiver should thus select the sender with the highest type (in expectation). Note that her payoffs are not directly affected by the senders' messages.

In what follows we focus on Perfect Bayesian Equilibria (PBE) in which senders use pure strategies. Pure strategies of senders are functions from type to message. Let $\rho : T \to M$ be the map of a sender's type to the message sent. Strategies of the receiver are functions from the senders' messages to probabilities over actions $r(m_i, m_j) \in \Delta(\{0,1\})$. We assume that $r(m,m) = \frac{1}{2}$. $\mu(m_i)$ will denote the receiver's belief that sender i invests upon receiving message m_i. We further restrict the receiver's off-equilibrium path-beliefs to satisfy D1 (see Cho and Sobel (1990) or Fudenberg and Tirole (1991) for a definition).

We first establish that in any equilibrium the function $\rho(\cdot)$ must be monotonic and nondecreasing. All proofs are in Appendix E.

Lemma 1. *In any PBE, $\rho(t_2) \geq \rho(t_1)$ for any $t_2 > t_1$.*

That $\rho(\cdot)$ must be nondecreasing is intuitive: the benefits of getting the task do not depend on the type, and sending a higher message is less costly for higher types.

Despite the differences in setup, the equilibrium characterization closely matches that in Kartik (2009). It is easy to see that pooling equilibria exist for sufficiently small k, with all types sending the same message. In particular, such pooling equilibria exist if $\theta > 0$ and $2k < \theta$ (with all types pooling at $m = 1$) or $\theta < 0$ and $2k < -\theta$ (with all types pooling at $m = 0$). More interesting for our purposes are (partially) separating equilibria. On any open interval of types that separate, the map $\rho(\cdot)$ must satisfy:

$$\rho'(t) = \frac{\theta}{2k[\rho(t) - t]}. \tag{3}$$

When types on an interval separate, the receiver can deduce the sender's type from the message. This implies that the probability of being selected for sender i is given by $prob(t_j < t_i) = t_i$. Pretending to be a slightly higher type than their true type

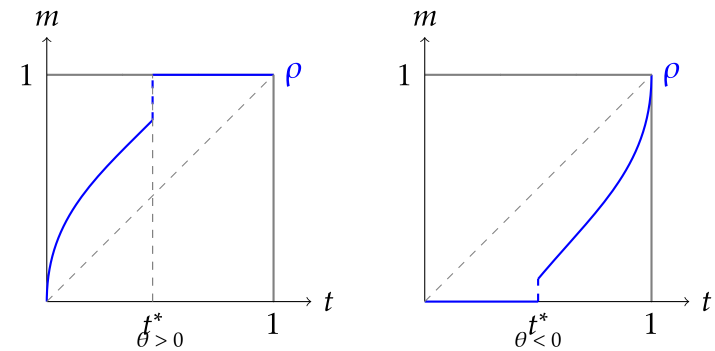


Figure 6: Equilibrium strategies for the case of a desirable task (left, $\theta > 0$) and an undesirable task (right, $\theta < 0$). The figure illustrates a partially separating strategy.

increases the payoffs from getting the task at a rate θ, while it increases the cost of lying by $2k(\rho(t) - t)\rho'(t)$ (evaluated at t). It is easy to verify that that in equilibrium for $\theta > 0$ (desirable task) this implies that $\rho^* > t$ and for $\theta < 0$ (undesirable task) that $\rho^* < t$. Compared to their actual valence, senders express a higher valence when the task is desirable and a lower valence when the task is undesirable. Charness et al. (2018) derive a similar result in a somewhat different tournament setting with two possible types of senders.

Proposition 1. *In any PBE satisfying D1, if $\rho^*(t)$ is part of an equilibrium profile, then (i) if $\theta > 0$, all types $t \in (0,1)$ overreport their valence, i.e., have $\rho^*(t) > t$, and (ii) if $\theta < 0$, all types $t \in (0,1)$ underreport their valence, i.e., have $\rho^*(t) < t$.*

With our assumptions on M, it is not possible that all types separate, and therefore some pooling will occur at the top (desirable task) or the bottom (undesirable task). Figure 6 shows a partially separating equilibrium in which there is a single pool (in Appendix E we show that there are no equilibria with multiple pools).

The analysis shows two basic facts. First, senders have an incentive to strategically 'overreport' their valence when the task is desirable and 'underreport' their valence when the task is undesirable. Second, there can be partial separation, so that receivers should pay attention to the expressed valence. With a slight modification, the model can also accommodate different investment rates across treatments. If senders' types are endogenous, and reflect senders' *intentions* to invest rather than necessarily reflecting actual investment decisions, then senders using strategy ρ^* are more inclined to invest when $\theta > 0$ compared to when $\theta < 0$. After overreporting their valence, investing will reduce lying costs. Of course, this assumes that senders are naive and do not anticipate their actual investment decision, otherwise ρ^* is no longer an equilibrium strategy.

6 Discussion and concluding remarks

In this paper we report evidence that people display emotions strategically and that it pays off to do so. In two experiments, managers are more likely to assign tasks to workers who express more positive emotions on a picture. Most importantly, in both experiments we find that the desirability of the task influences the workers' expressions. This result is observed in an environment where workers could choose between an angry and a happy picture (Experiment I), and in an environment where workers send photographs without any mentioning of emotions (Experiment II).

Interestingly, both experiments show evidence that people might not fully exploit the potential benefits of expressing emotions. In Experiment I, some workers would have been better off (in monetary terms) by sending a different picture. In Experiment II, the difference in intensity of expressed emotions between the Desirable and Undesirable task treatment is only 35 percent of the difference observed in Experiment I.

What can explain the reluctance to express certain emotions? One possibility is that expressing an insincere emotion comes at a cost. Just as talk may not be cheap, looks may not be cheap. These costs could come from lying or guilt aversion (Ellingsen & Johannesson, 2004; Gneezy, 2005; Charness & Dufwenberg, 2006) or simply a desire to look attractive. Faking emotions might also be cognitively demanding and thus not easy to achieve.

Given that workers strategically manipulate expressions, another question is why managers would respond to expressions. The work by Gneezy and Imas (2014) shows that people *are* aware of the role of emotions of others at some level. Possibly, this level of awareness is only of the first order and not of higher orders ("level-1" reasoning in the parlour of level-k reasoning, Stahl and Wilson (1995); Nagel (1995)). On the other hand, if workers experience costs of manipulating expressions, expressions may contain some informational value. This is similar to a cheap talk framework with lying costs, in which messages transmit information, like the model we propose in section 5. Indeed, previous work shows that facial cues can be informative of behavior. For instance, 'honest' smiles are associated with trustworthiness (Centorrino et al., 2015), and pupil dilation is related to deceptive behavior (Wang et al., 2010). There may also be costs on the side of the manager, for instance in scrutinizing whether an expressed emotion is sincere or not. In the presence of such costs or constraints, either on the side of the worker or the manager, displayed emotions could well be informative .

Our results demonstrate that emotions might be an important mode of communication. Subjects in our experiments use this form of communication. One interesting avenue for future research is to study how repeated interactions would affect our results. In such a setting, appearing angry might have short run benefits (you avoid receiving an undesirable task), but might have long run consequences as people might avoid future interactions with angry looking people (Elster, 1998).

References

Andrade, E. B., & Ho, T.-H. (2009). Gaming emotions in social interactions. *Journal of Consumer Research*, *36*(4), 539–552.

Ashraf, N., Camerer, C. F., & Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives*, *19*(3), 131–145.

Babcock, L., Recalde, M. P., Vesterlund, L., & Weingart, L. (2017). Gender differences in accepting and receiving requests for tasks with low promotability. *American Economic Review*, *107*(3), 714–47.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241–251.

Battigalli, P., Dufwenberg, M., & Smith, A. (2015). Frustration and anger in games. *CESifo Working Paper Series*, No. 5258.

Bentham, J. (1789). *The principles of morals and legislation.* Oxford.

Bijlstra, G., & Dotsch, R. (2011). Facereader 4 emotion classification performance on images from the radboud faces database. *Unpublished manuscript, Department of Social and Cultural Psychology, Radboud University Nijmegen, Nijmegen, The Netherlands*.

Bonnefon, J.-F., Hopfensitz, A., & De Neys, W. (2013). The modular nature of trustworthiness detection. *Journal of Experimental Psychology: General*, *142*(1), 143-150.

Bosman, R., & Van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal*, *112*(476), 147–169.

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

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579–1601.

Charness, G., Rustichini, A., & Van de Ven, J. (2018). Self-confidence and strategic behavior. *Experimental Economics*, *21*(1), 72–98.

Cho, I.-K., & Sobel, J. (1990). Strategic stability and uniqueness in signaling games. *Journal of Economic Theory*, *50*(2), 381–413.

Crawford, V. P., & Sobel, J. (1982). Strategic information transmission. *Econometrica*, *50*(6), 1431–1451.

Dufwenberg, M., Sundaram, R., & Butler, D. J. (2010). Epiphany in the game of 21. *Journal of Economic Behavior & Organization*, *75*(2), 132–143.

Eckel, C., & Petrie, R. (2011). Face value. *American Economic Review*, *101*(4), 1497-1513.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*, *48*(4), 384–392.

Ellingsen, T., & Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, *114*(495), 397–420.

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

Farrell, J., & Rabin, M. (1996). Cheap talk. *Journal of Economic Perspectives*, *10*(3), 103–118.

Forbes. (2009). Business meetings: The case for face-to-face. *Forbes Insight*.

Frank, R. H. (1987). If homo economicus could choose his own utility function, would he want one with a conscience? *American Economic Review*, *77*(4), 593–604.

Frank, R. H. (1988). *Passions within reason: the strategic role of the emotions.* WW Norton & Co.

Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. Cambridge, Massachusetts: Massachusetts Institute of Technology.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, *95*(1), 384–394.

Gneezy, U., & Imas, A. (2014). Materazzi effect and the strategic use of anger in competitive interactions. *Proceedings of the National Academy of Sciences*, *111*(4), 1334-1337.

Gneezy, U., Imas, A., & Madarász, K. (2014). Conscience accounting: Emotion dynamics and social behavior. *Management Science*, *60*(11), 2645–2658.

Gneezy, U., Rustichini, A., & Vostroknutov, A. (2010). Experience and insight in the race game. *Journal of Economic Behavior & Organization*, *75*(2), 144–155.

Heyes, A., & List, J. (2016). Supply and demand for discrimination: Strategic revelation of own characteristics in a trust game. *American Economic Review*, *106*(5), 319-323.

Hirshleifer, J. (1984). On the emotions as guarantors of threats and promises. *UCLA Dept. of Economics Working Paper*, No. 337.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, *76*(4), 1359–1395.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, *90*(2), 426–432.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal*, *92*(368), 805–824.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, *85*(5), 1313–1326.

Reed, L., DeScioli, P., & Pinker, S. (2014). The commitment function of angry facial expressions. *Psychological Science*, *25*(8), 1511-1517.

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

Smith, A. (1759). *The theory of moral sentiments*. D. D. Raphael and A. L. Macfie, eds.Liberty Fund: Indianapolis.

Stahl, D. O., & Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, *10*(1), 218–254.

Tognetti, A., Berticat, C., Raymond, M., & Faurie, C. (2013). Is cooperativeness readable in static facial features? an inter-cultural approach. *Evolution and Human Behavior*, *34*(6), 427–432.

van Kleef, G., De Dreu, C., & Manstead, A. (2004). The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology*.

van Leeuwen, B., Noussair, C. N., Offerman, T., Suetens, S., van Veelen, M., & van de Ven, J. (2018). Predictably angry—facial cues provide a credible signal of destructive behavior. *Management Science*, *64*(7), 3352-3364.

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*(4), 260-271.

Vogt, S., Efferson, C., & Fehr, E. (2013). Can we see inside? Predicting strategic behavior given limited information. *Evolution and Human Behavior*, *34*(4), 258–264.

Wang, J., Spezio, M., & Camerer, C. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, *100*, 984-1007.

Winter, E., Mendez-Naya, L., & Garcia-Jurado, I. (2016). Mental equilibrium and strategic emotions. *Management Science*, *63*(5), 1302–1317.

Appendix A Survey Measures

Level of Strategic Reasoning. The level of strategic reasoning is measured using an adapted version of the race game (see Gneezy et al. (2010)). In our version of the game, a number of chips is available. The subject and computer take turns, and can take 1 or more chips each turn. The player who takes the last chip wins. We implemented two versions. In the first game, they started with 15 chips and they could take 1,2, or 3 chips each turn. In the second game, they started with 17 chips and they could take 1,2,3, or 4 chips each turn. The subject always started first. We programmed the computer such that in every round, there was a winning strategy for the subject if she took the appropriate number of chips from that round onward. A subject's score is the number of rounds that she followed a winning strategy. Once the number of chips remaining is less than the maximum number of chips that a subject can take, the solution is trivial and those rounds are not scored. The mean score is 0.497 (median 0.5, standard deviation 0.183).

Level of Emotional Intelligence. We have two measures of emotional intelligence. The first is based on the 'Reading the Mind in the Eyes Test' (Baron-Cohen et al., 2001). This is a standard test in psychology, where pictures of eyes are shown together with a set of words describing emotions. The subject's task is to select the word that best fits the expression on the picture. There are 36 pictures in total and the average score is 25.7 (median 26, standard deviation 4.3).

Our second measure is based on van Leeuwen et al. (2018). We showed ten pictures of participants from a previous experiment who all received a low offer in an ultimatum game. Exactly five of those participants rejected the offer. We asked our subjects to predict which of the participants rejected the low offer. A subject's score is computed as the number of correctly identified rejecters and can ranges from 0 to 5. The average score is 3.0 (median 3.0, standard deviation 0.9).

Appendix B Details Procedures Experiment II

In the first session of Experiment II, some subjects used hand gestures on the pictures and some considerably slowed down the experiment by taking many selfies before submitting their picture. We therefore slightly modified the instructions in that treatment, and told participants that they were not allowed to use hand gestures, and limited the number of possible selfies to three. An experimenter inspected the submitted pictures. Subjects rarely violated the instructions. If they did, they were asked to retake the picture.

Appendix C Robustness

This section provides some robustness checks.

Table A.1 replicates the estimates in columns (1) and (4) of Table 1, showing that our results are robust across different subsamples. Column (1) is an exact replication of the entire sample. Column (2) reports the estimates for worker pairs of the same gender. Columns (3) and (4) split those results by gender. Column (5) shows the results for managers that correctly answered all test questions on their first attempt. Column (6) excludes pairs of workers where at least one of the workers displays deviant behavior. Whether or not a worker displayed deviant behavior is determined by our own coding. Three of the researchers independently rated pictures on deviant expressions. An expression is considered deviant if a participant used props or hand gestures, took abnormal expressions such as kissing or eye-crossing, or exaggerated the emotional expression making it look unnatural. A subject is classified as deviant if at least one of the researchers rated their expression as deviant. The researchers' codings showed strong agreement; about 98 percent of pictures are classified the same by all three researchers.

Table A.1: Assignment of task (Experiment I).

Sample	(1) All	(2) Same-gender All	(3) Same-gender Males	(4) Same-gender Females	(5) Understood	(6) Compliant
PANEL A: UNDESIRABLE TASK						
Happy expression	1.818***	2.241***	2.231**	2.250***	2.476***	2.000***
	(0.270)	(0.534)	(0.726)	(0.681)	(0.605)	(0.316)
Observations	792	392	172	220	348	722
PANEL B: DESIRABLE TASK						
Happy expression	1.405**	1.548**	1.423	1.750*	1.507**	1.410**
	(0.207)	(0.306)	(0.339)	(0.563)	(0.244)	(0.212)
Observations	840	400	236	164	756	830

Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Col. (2)-(4): workers have the same gender. Column (5): ”understood” means manager correctly answered all test questions on first try. Column (6): both workers complied with instructions to look natural.
Robust s.e. in parentheses clustered at the manager level. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

Table A.2: Effect of payoffs on emotional expressions (Experiment II).

	(1)	(2)
Task	Undesirable	Desirable
Dep. variable	Valence	Valence
Payoff in previous round	-0.015	0.032
	(0.036)	(0.025)
Constant	0.155***	0.203***
	(0.046)	(0.041)
Observations	364	357

Fixed effects linear regressions. Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D Instructions

What follows are the instructions for both types of players in the treatment with the desirable task in Experiment I. Instructions for the treatment with the undesirable task, as well as the instructions for Experiment II, are very similar and available upon request.

[Instructions for red players (managers).]

General information

Thank you for participating in this study. Please read the instructions carefully and make sure that your mobile phone is turned off. You can earn more money depending on your own choices and the choices of other participants.

In this experiment, there are two types of players: red and green. The green players are all students from another town (Amsterdam). The red players are students from Tilburg. You are one of the red players.

In every round, you are paired with two green players (green A and green B). Exactly one of the green players can make an investment.

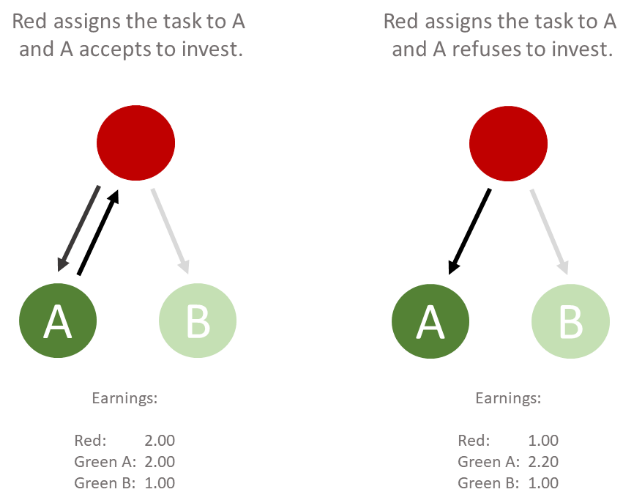
Your task is to assign the investment task to one of the green players. That green player becomes the ”designated” player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

The earnings are as follows:

You earn €2.00 if the designated green player accepts to invest, and €1.00 if the designated green player refuses to invest. The designated green player earns earns €2.00 if he or she accepts to invest, and €2.20 if he or she refuses to invest. The other green player always earns €1.00.

Suppose, for instance, that you assign the investment task to green A. If green A accepts to invest, green A and you earn €2.00 each, and green B earns €1.00. If green A refuses to invest, green A earns €2.20, and green B and you earn €1.00 each.

The picture below summarizes the possible earnings when you assign the investment task to green A. Of course, this is just an example; you could also assign the investment task to green B, and the earnings of A and B are then reversed.



Pictures and rounds

In total, there will be 12 rounds. You will be paid for every round.

When you assign the task to invest, you will see pictures of the green players.

In every round, you will learn whether the green player you assigned the task to accepted or refused to invest.

In half of the rounds, you will be inactive and do not have to make a decision. You will receive €1.00 for each round that you are inactive.

You will be paired with different players in each round. You will never be matched twice with the same green player throughout the entire experiment.

Summary

If you are active in a round, you will see the pictures of the green players you are paired with.

You assign the task to invest to one of the green players.

The designated green player can invest or refuse to invest. The other green player has no decision to make.

You earn €2.00 if the designated green player accepts to invest, and €1.00 if the designated green player refuses to invest. The designated green player earns earns €2.00 if he or she accepts to invest, and €2.20 if he or she refuses to invest. The other green player always earns €1.00.

There will be 12 rounds, you will be paid for all rounds.

You will be paired with different players in each round. You will never be matched twice with the same green player throughout the entire experiment.

In half of the rounds you will be inactive. If you are inactive you will receive €1.00

[Instructions for green players (workers).]

General information

In this experiment, there are two types of players: red and green. The red players are all students from another town (Tilburg). The green players are students from Amsterdam. You are one of the green players.

In every round, each red player is paired with two green players (green A and green B). Exactly one of the green players can make an investment.

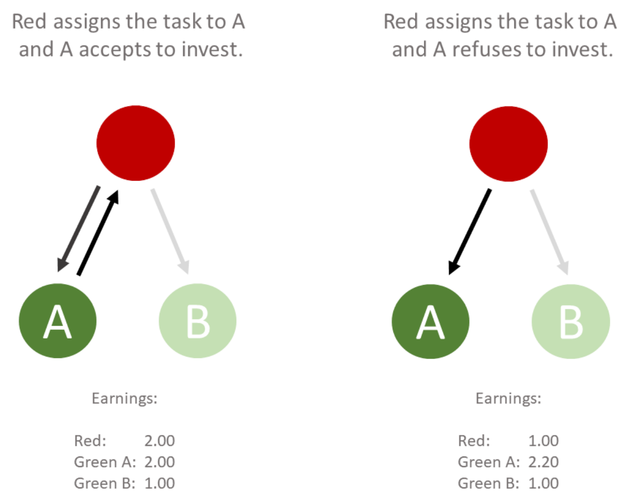
The task of the red player is to assign the investment task to one of the green players. That green player becomes the ”designated” player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

The earnings are as follows:

The designated green player earns earns €2.00 if he or she accepts to invest, and €2.20 if he or she refuses to invest. The red player earns €2.00 if the designated green player accepts to invest, and €1.00 if the designated green player refuses to invest. The other green player always earns €1.00.

Suppose, for instance, that the red player assigns the investment task to green A. If green A accepts to invest, green A and the red player earn €2.00 each, and green B earns €1.00. If green A refuses to invest, green A earns €2.20, and green B and the red player earn €1.00 each.

The picture below summarizes the possible earnings when the red player assigns the investment task to green A. Of course, this is just an example; the red player could also assign the investment task to green B, and the earnings of A and B are then reversed.



The red player will see a picture of you and the other green player when he or she assigns the investment task to one of you. The red player will either see your picture with the happy expression or the picture with the angry expression, and this partly depends on you.

In each round, one of the green players can choose which picture of him- or herself to show. For the other player, this is randomly determined (each picture is equally likely to be selected).

In total, there will be 12 rounds. Every two rounds, you first select which picture you want to show to the red players in the next two rounds. In these two rounds, in random order, your selected picture will be used in one round and a randomly selected picture will be used in the other round.

Feedback

In total, there will be 12 rounds. You will be paid for every round.

At the end of every round, you will learn which of your pictures the red player saw, and whether the red player assigned the investment task to you or to the other green player. If the other green player was chosen, you will also learn whether that player accepted or refused to invest.

Please also note the following:

The red player is not told that one of the green players could choose which picture to show. The red player simply sees the two pictures and then has to assign the task to invest to one of the green players. The red player is also not told that you were asked to express emotions on any of the pictures.

You and the other green player will never see each other's pictures, and you will never found out with whom you were paired in any round. You will be paired with different players in each round. You will never be matched twice with the same red player throughout the entire experiment.

Summary

One of the green players can choose which picture will be shown to the red player, for the other green player this is determined randomly.

The red player then assigns the task to invest to you or the other green player.

The designated green player can invest or refuse to invest. The other green player has no decision to make.

The designated green player earns earns €2.00 if he or she accepts to invest, and €2.20 if he or she refuses to invest. The red player earns €2.00 if the designated green player accepts to invest, and €1.00 if the designated green player refuses to invest. The other green player always earns €1.00.

There will be 12 rounds, you will be paid for all rounds.

You will be paired with different players in each round. You will never be matched twice with the same red player throughout the entire experiment.

Appendix E Proofs

Proof of Lemma 1. Fix an equilibrium strategy in which t_1 sends message m_1 and $t_2 > t_1$ sends message m_2. Let $r_1 = r(m_1)$ and $r_2 = r(m_2)$. The incentive compatibility constraints imply that:

$$\theta r_1 - k(m_1 - t_1)^2 \geq \theta r_2 - k(m_2 - t_1)^2, \tag{4}$$

and

$$\theta r_2 - k(m_2 - t_2)^2 \geq \theta r_1 - k(m_1 - t_2)^2. \tag{5}$$

The two constraints together imply that:

$$2(t_2 - t_1)(m_2 - m_1) \geq 0. \tag{6}$$

This implies that for any $t_2 > t_1$, $m_2 \geq m_1$.

Proof of Proposition 1. We first show that full separation is not possible. If $\theta > 0$, the solution to equation (3) is given by (see Kartik (2009)):

$$e^{-\frac{2k}{\theta}\rho(t)} = 1 - \frac{2k}{\theta}(\rho(t) - t). \tag{7}$$

Together with the boundary condition that $\rho(0) = 0$, this implies that the highest type that can separate is given by $t^h = 1 - \frac{\theta}{2k}(1 - e^{-\frac{2k}{\theta}}) < 1$. If $\theta < 0$, the solution to equation (3) is given by:

$$e^{-\frac{\theta}{2k}(\rho(t)-1)} = 1 - \frac{2k}{\theta}(\rho(t) - t). \tag{8}$$

Together with the boundary condition that $\rho(1) = 1$, this implies that the lowest type that can separate is given by $t^l = \frac{2k}{\theta}(1 - e^{\frac{2k}{\theta}}) > 0$.

We next show that there can be at most be a single pool. We do this for the case $\theta > 0$ (the case with $\theta < 0$ is symmetric). Consider an equilibrium that contains an interval (t_1, t_2) on which all types pool on some message $m_1 < 1$, and let $t_2 < 1$. We invoke the equilibrium refinement D1 to show that there is a profitable deviation for some type $\hat{t} = t_2 - \varepsilon$ for $\varepsilon \to 0$. Consider a deviation to message $\hat{m} = m_1 + \varepsilon$ and a receiver's response $\hat{r}$ that makes type $\hat{t}$ indifferent between sending message $\hat{m}$ and his equilibrium message m_1 (resulting in $r_1 = r(m_1)$):

$$r_1\theta - k(m_1 - \hat{t})^2 = \hat{r}\theta - k(\hat{m} - \hat{t})^2. \tag{9}$$

It is straightforward to show that for this action by the receiver, the difference in payoffs between sending $\hat{m}$ and sending m_1 for any type $\tilde{t} \in (t_1, \hat{t})$ is given by:

$$\Delta = -k(\hat{t} - \tilde{t})(\hat{m} - m_1) < 0. \tag{10}$$

Then, by monotonicity and D1, the receiver should not put any positive beliefs on the out-of-equilibrium action coming from type $\tilde{t}$. It is also straightforward to show that the same is true for any $\tilde{t} < t_1$. But then $\hat{r} > r(m_1)$ (strict, because there is a interval of types sending m_1). This means that there is some type $\hat{t}$ that can strictly increase the likelihood of getting the task by an arbitrarily small change in lying cost, a strictly profitable deviation.

The above implies that all pooling must occur at $m = 1$, implying $\rho(t) > t$ for all types $t < 1$ that send message $m = 1$. That $\rho(t) > t$ on an interval in which types separate is immediate from equation (3). Similarly, for $\theta < 0$, all pooling must occur at $m = 0$, implying $\rho(t) < t$ for all types $t > 0$ that send message $m = 0$. That $\rho(t) < t$ on an interval in which types separate is again immediate from equation (3).