The Strategic Display of Facial Expressions *

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

February 2024

Abstract

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

*We thank numerous seminar and conference audiences, and especially Martin Dufwenberg, Uri Gneezy, Alex Imas, Theo Offerman, Giorgia Romagnoli, Sigrid Suetens and Frans van Winden 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) and the LABEX CORTEX (ANR-11-LABX-0042) of Universite de Lyon, within the program Investissements d'Avenir (ANR-11-IDEX-007). A previous draft of this paper circulated under the title "The Strategic Display of Emotions".

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

[‡]EMLyon Business School, hopfensitz@em-lyon.com

[§]Department of Economics, Tilburg University. b.vanleeuwen@uvt.nl

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

1 Introduction

Facial expressions are rapidly perceived and evaluated (Ambady & Rosenthal, 1992; Ambady et al., 2000). They are used to judge personality, values and dispositions of interaction partners, and the most prominent facial expressions are linked to emotions (Frijda, 1986; Ekman, 1993). Emotions play an important role in strategic settings (Battigalli et al., 2015; Loewenstein, 2000; Frank, 1987, 1988; Hirshleifer, 1984).[1] They affect people's choices, 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). An accumulating body of empirical evidence supports these ideas. The display of anger during negotiations 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 those facial expressions. While facial 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 facial expressions are used strategically. Surprisingly, there are virtually no studies that test the strategic use of facial expressions. This is in sharp contrast to other types of communication, which has been intensively studied in economics. Yet it is clearly important for a good understanding of the role of nonverbal communication. It speaks to the question whether it can be equilibrium behavior to pay attention to facial expressions and what the level of people's sophistication is in this respect.

Our main hypothesis is that followers will adapt their facial expressions based on the strategic setting that they face. This is plausible if people infer intentions from facial expressions, which creates clear advantages of mimicking expressions (Frank, 1987, 1988). We present experimental evidence that supports our hypothesis. In

[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.

two laboratory experiments, we show that followers engage in (deliberate) strategic display of facial expressions, displaying more positive emotions when this is in their advantage and negative emotions otherwise.

In our experiments, participants play a task delegation game. In this game, a leader assigns an investment task to one of two followers. We varied whether getting the task is desirable or not for the follower (in monetary terms). This crucially changes the strategic nature of the game for followers. In either treatment, the leader benefits from a follower's investment in the task, while investing is always costly to the follower. It is therefore in the leader's interest to assign the task to the follower that is most likely to invest. In line with the empirical evidence that positive emotions are associated with higher perceived trustworthiness (Scharlemann et al., 2001; Oosterhof & Todorov, 2008; Centorrino et al., 2015; Sutherland et al., 2017), we find that leaders in our experiment tend to allocate tasks to followers who display positive emotions. Our main finding is that followers in both experiments, on their part, displayed more positive emotions when the task was desirable to get, compared to when the task was undesirable.

The two experiments we conducted are complementary. In the first experiment, we instructed followers to take one picture with a happy expression and one with an angry expression, and we let them decide which one they wanted to show to the leader. As we explain in the design section, this setup allows us to establish causal effects in a rigorous way. At the same time, it makes it salient to followers that there can be benefits of opting for a certain expression. In the second experiment, we gave no instructions to followers, making the benefits of adapting expressions less salient and creating a less stylized environment.

We also propose a model that can accommodate the findings from our experiments. In the model, we assume that senders ('followers') can send messages about their emotions by varying their facial expressions. Senders have a cost of lying or faking expressions, i.e. expressing an emotion different from what they truly experience. We show that in equilibrium, senders will overreport their emotional valence when the task is desirable and underreport their valence when the task is undesirable. If the cost of faking expressions 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 ('leaders') respond to the messages they observe.

To our best knowledge, ours is the first paper to study the strategic display of facial expressions, in the sense that people adapt their facial expression to fit the situation.[2] 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.[3] 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. Manzini et al. (2009) allow participants in an experiment to report that they are 'smiling', and find that such verbal smiles are used and recognized as signs of trustworthiness in coordination games. 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 like facial expressions, may be of first-order importance.

Some papers show that people strategically manipulate 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 (Gneezy & Imas, 2014). In contrast, our paper shows that people strategically manipulate their own facial expressions.

2 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 'leader' is paired with two 'followers' (neutral labels are used in the experiment). The leader sees pictures of the followers and has to allocate an investment task to one of them. The chosen

[2] Some of the existing literature uses the term strategic display to mean something different. They use strategic display to signify that emotional displays have strategic value. For instance, Kopelman, Rosette, and Thompson (2006) study if displaying different emotions has effects on outcomes, but they do not test if people make use of this fact.

[3] In a theoretical contribution, Winter et al. (2016) analyze equilibrium play when people can choose their own emotional states.

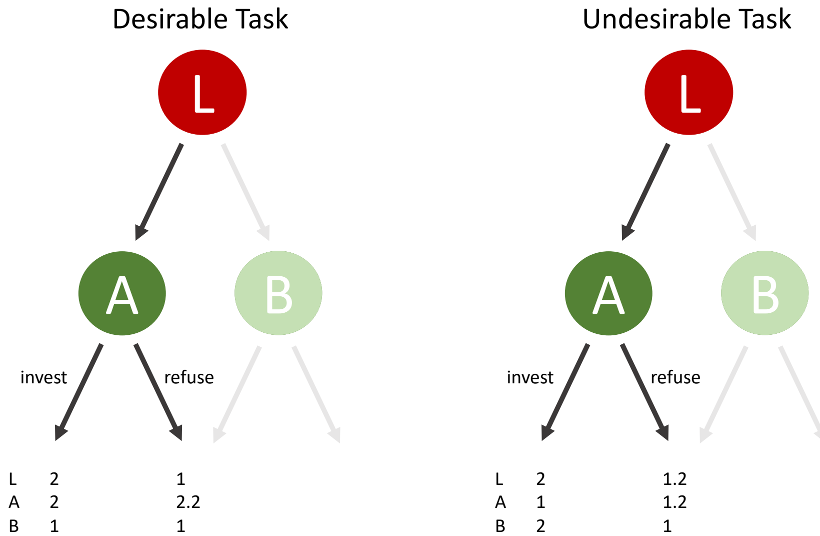


Figure 1: Payoffs to the leader (L) and followers (A and B) when the task is assigned to follower A (i.e., follower A is selected to make a decision). The case in which the task is assigned to follower B is symmetric.

follower becomes the 'designated' follower and is the only follower that has to make a decision. The designated follower can accept or refuse to invest. Investing is costly to the follower but beneficial to the leader. 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 follower 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 leader (L) assigns the task to follower A, and thus follower A becomes the designated follower (the case where the task is assigned to follower B is symmetric). In treatment 'Desirable', the designated follower (A) always earns more than the other follower (B), independent of his investment decision. The designated follower earns €2 if he invests, and €2.2 if he refuses to invest. The other follower always earns €1. The leader earns €2 if the designated follower invests, and €1 if the designated follower refuses to invest. In treatment 'Undesirable' (right part of Figure 1), a follower earns the most if the task is assigned to the other follower and the other follower invests. In this version, if the designated follower (A) invests, both the leader and the other follower (B) earn €2, while the designated follower earns €1. If the designated follower refuses to invest, the designated follower (A) and the leader both earn €1.2, while the other follower (B) earns €1. This payoff structure mimics a situation where the other follower has to complete the task if the designated follower refuses to invest.

The variation in payoffs between task delegation games creates the following incentives for leaders and followers. First, note that in both versions of the task delegation game, the leader earns more money if the designated follower invests, while investing is always costly for the designated follower. These payoffs create incentives for the leader to allocate the task to the follower that is most likely to invest, and therefore to select the follower that looks most trustworthy. If so, followers have incentives to appear trustworthy if they want to receive the task. In Desirable, followers always earn a higher monetary payoff when they are the designated follower, creating clear incentives for the followers to appear trustworthy. In Undesirable, followers may wish to appear untrustworthy to avoid getting the task. In terms of monetary outcomes, the first-best outcome for a follower is that the other follower gets the task and invests.[4] We conjecture that negative emotions (such as anger) are associated with lower trustworthiness, and positive emotions (such as happiness) are associated with higher trustworthiness.

Hypothesis 1. *In both treatments, leaders are more likely to assign the task to followers that show more positive facial expressions.*

Hypothesis 2. *Followers display more negative facial expressions in treatment Undesirable compared to treatment Desirable.*

We test these hypotheses in two experiments. The two experiments differ in the way how followers can display their facial expressions. The two experiments are complementary. In Experiment I we impose that followers either show a happy or an angry facial expression. While stylized, this allows us to cleanly identify the causal effect of facial expressions. Experiment II is more natural, allowing for more

[4] There is a caveat. If a follower is very pessimistic about the likelihood that the other follower will invest, he may be better off getting the task himself and not investing rather than refusing the task. A selfish player B would not invest. However, given the abundant literature documenting the existence of social preferences, we anticipated that many players would invest even if it is not in their own narrowly defined self-interest, and this is confirmed by our empirical results.

freedom in the way followers can display facial expressions. We will first discuss the design and main results of Experiment I before turning to Experiment II.

3 Experiment I

3.1 Experimental design and procedures

Prior to receiving any instructions about the game, participants in the role of followers 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 - followers and leaders - then received instructions on their screen (see Appendix B). 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. They kept the same role throughout the experiment. Each subject played a total of 12 rounds and subjects were rematched every round. Leaders never saw the same follower more than once and followers never found out with which other follower they were matched. For each session, we recruited as many leaders as followers. This way there were enough leaders to ensure that a leader never saw the same follower twice. Since followers were paired, this means half of the leaders were inactive in any round. Leaders received a fixed payment of €1 in each round in which they were not making a decision.

At the beginning of every two rounds, followers 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 follower'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 facial expressions on the likelihood of getting the task. Leaders were not informed that followers were asked to express emotions or how pictures were selected and followers knew this. We did this so that the leaders would not immediately question the sincerity of the expressed emotions.

At the end of each round, all subjects received feedback about their earnings. If they were not the designated player, they also learned whether or not the other follower invested. Followers also learned which one of their pictures was shown. Subjects were paid for all 12 rounds. They did not receive any show-up fee or additional payments beyond the payments for the 12 rounds of the task delegation game. 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, followers and leaders came from different cities in the Netherlands. Followers participated in Amsterdam (CREED lab) and leaders in Tilburg (CentERlab). Both leaders and followers were informed that they would interact with participants in another town (Amsterdam or Tilburg). From the CREED and CentERlab subject pools, we recruited a total of 272 subjects, 136 in each role (51 percent female, mean age 22) for a total of 10 sessions.[5] The sample size was mainly determined by our available budget and the number of subjects in our databases. For both subject pools, we used the online recruitment system to recruit participants. In the invitation, we explained to subjects in Amsterdam that their picture would be taken and shown to other participants from a different university, and they had to consent to this if they wanted to participate. Both at CREED and the CentERlab, the subject pool consists primarily of students, majoring in different fields of study. Each session had 24, 28, or 32 participants, depending on the show-up. Sessions were sex-balanced, with a fraction of female participants that was always between 0.42 and 0.60.

Followers 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

[5]Prior to running Experiments I and II, we conducted one pilot session in which we used different games (a trust game and an 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. None of the subjects who participated in the pilot took part in the experiments reported in this paper.

Amsterdam.

3.2 Measurements

Intensity of expressed emotions. We use facial recognition software Noldus FaceReader 7.1 to evaluate the intensity of expressed emotions (see Bijlstra and Dotsch (2011) for validation of the software). 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 in 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). Specifically, 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 few negative emotions besides anger so that valence is essentially the level of 'happy' minus the level of 'anger'.

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 leaders. A total of 32 raters evaluated one picture of each of the 136 followers (either happy or angry, randomly selected). Each rater evaluated pictures on only one dimension and pictures were sorted by sex. Raters received a flat payment of €7, the rating task lasted around 20 minutes. 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 sex, 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. In our version of the game, a number of chips are 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). The instructions can be found in Appendix B.3.1.

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. There are 36 pictures in total and the average score is 25.7 (median 26, standard deviation 4.3). The instructions can be found in Appendix B.3.3. The second measure of emotional intelligence is the subject's ability to predict who would reject a low offer in an ultimatum game, based on facial expressions (see the task developed in 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 range from 0 to 5. The average score is 3.0 (median 3.0, standard deviation 0.9). The instructions can be found in Appendix B.3.2. 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 of strategic reasoning or emotional intelligence were incentivized.

3.3 Results

3.3.1 Facial expressions and perceived trustworthiness

We confirm that facial expressions are different between happy and angry pictures. Panel A in Figure 2 shows a substantial difference in valence between the two sets of pictures. As expected, the mean valence is negative on the angry picture (-0.38) and positive on the happy picture (0.70). Panel C shows that all but one subject display higher valence on the happy picture and the difference in valence between pictures is highly significant (signed-rank test, $p < 0.001$, $N = 135$).[6]

We further observe a clear link between expressed emotions and trustworthiness ratings. Panel B in 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 in the happy picture (mean = 4.67) compared to the angry picture (mean = 3.43). Panel D in Figure 2 shows that almost all subjects (125 out of 136, or 92 percent) look more trustworthy in their happy picture than in their angry picture, and this difference is highly significant (signed-rank test, $p < 0.001$, $N = 136$).

3.3.2 Behavior by leaders

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

We find clear evidence that the expression matters for the leader's decision. For both task types, leaders are more likely to assign the task to a follower that looks happy. When two followers 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 follower. The reported coefficients are in terms of odds ratios. In both task types, a happy expression significantly increases the likelihood of becoming the designated

[6] For all non-parametric tests we report, we take a subject as the independent unit of observation.

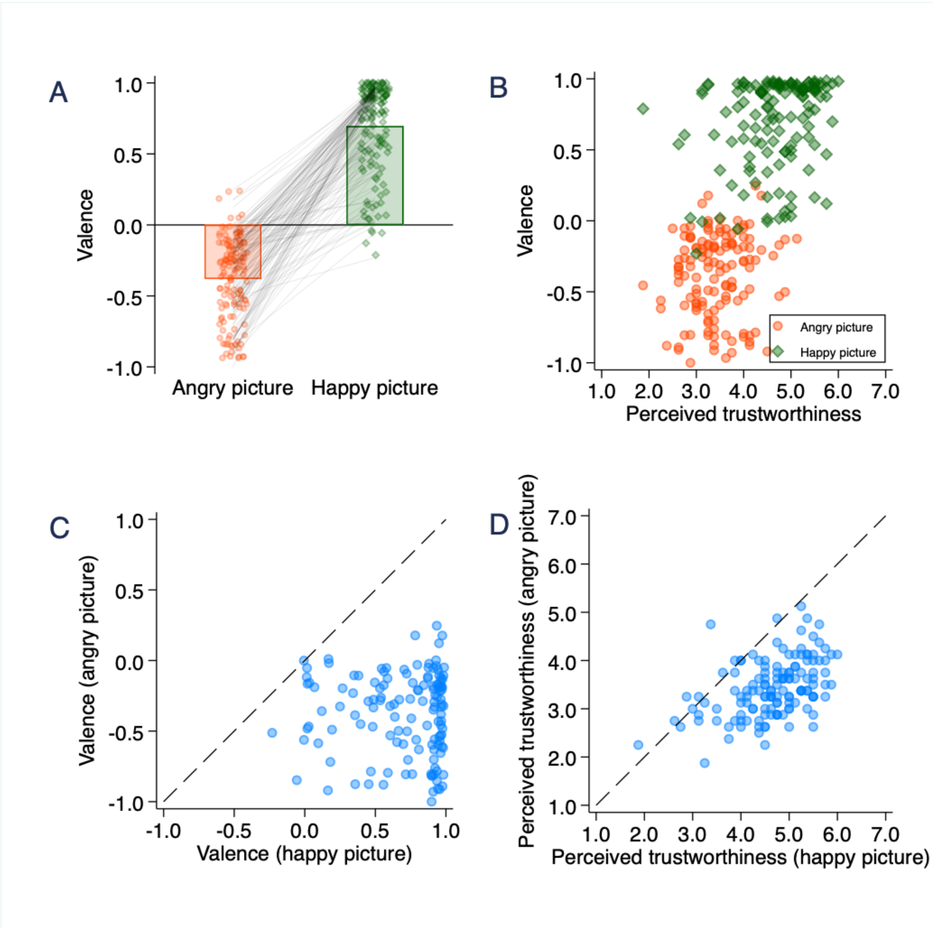


Figure 2: Valence and perceived trustworthiness on pictures. Valence is defined as the difference between positive and negative emotions. In panel A, bars indicate the mean valence by expression on the picture. Lines connect the pair of pictures of each subject. In all panels, circles/diamonds represent individual pictures.

Table 1: Assignment of tasks in Experiment I

	(1)	(2)	(3)	(4)	(5)	(6)
Task:	Undesirable			Desirable		
Happy picture	1.818***			1.405**		
	(0.270)			(0.207)		
Valence		1.447***			1.231*	
		(0.164)			(0.148)	
Perceived trustworthiness			1.399***			1.213**
			(0.127)			(0.108)
Observations	792	792	792	836	816	836

Notes: Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Valence is measured by FaceReader software, and is a score between -1 and 1. Perceived trustworthiness is the average rater score (between 1 and 7). Column (5) has fewer observations because FaceReader did not capture every face. Robust s.e. in parentheses clustered at the leader level. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$.

follower (columns 1 and 4). As both valence and perceived trustworthiness are larger in happy pictures, it is interesting to see if the intensity of valence or trustworthiness matters. Columns 2 and 5 show that followers displaying higher levels of valence are more likely to become the designated follower, while columns 3 and 6 support the idea that leaders look for trustworthy-looking followers.

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 leader. This allows us to cleanly investigate the causal effect of emotional expressions on the likelihood of being selected by the leader. To exploit this random variation, we compute for each follower the fraction of times they received the task when their happy or angry picture was *randomly* shown. We find that subjects in both treatments 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).[7]

[7] In Table A.2 we also show regression analyses based on only the random pictures. These analyses again indicate that leaders are more likely to select followers whose happy picture was randomly

In Table A.1 in Appendix A, we show that the estimated effect of showing the happy picture survives several robustness checks. In particular, we find similar effects if we restrict follower-pairs to be of the same sex.[8] Estimated coefficients are also unaffected by excluding leaders who did not correctly answer all comprehension test questions on the first attempt, or restricting observations to photographs on which followers were rated to look 'natural', or excluding the 20% of participants who display the strongest valence.

3.3.3 Facial expressions by followers

Our results indicate that there are clear benefits for followers of sending the happy expression when the task is desirable and the angry expression when the task is undesirable. We find that followers 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 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.[9]

Note that followers do not always show the hypothesized 'optimal' picture, i.e., the picture that yields the highest expected earnings. However, this is not evidence per se that they are making the wrong choice. The picture that maximizes expected earnings might not be the same for all followers. Conceivably, some followers are better off by sending their angry (happy) picture when the task is desirable (undesirable). This, however, is not the case. To show this, we again exploit our design feature that it is sometimes randomly determined which of the follower's pictures is shown, but this time for the subset of cases in which followers sent the 'wrong' pic-

shown, and that both randomly displayed valence and trustworthiness are associated with a higher likelihood of being selected. Note, however, that the statistical power is lower with these smaller samples, and the effects of valence and trustworthiness are not statistically significant in all specifications.

[8]We do not find evidence that tasks are more likely to be assigned male or female followers, see Table A.3 in Appendix A.

[9]The instructions to express emotions on pictures could possibly lead to an experimenter demand effect. In Experiment II (Section 4), we study the facial expressions of followers in a more natural setting where followers could freely take pictures and where we did not mention emotions at all. Experiment II shows that possible demand effects are not the only driver of the observed treatment effects.

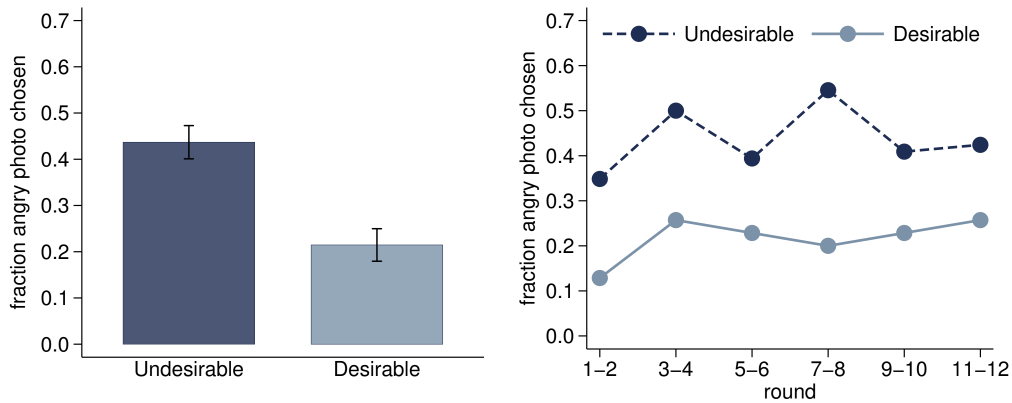


Figure 3: Fraction of followers choosing the angry picture. Left panel: fraction across all rounds (error bars indicate +/- 1 s.e.). Right panel: fraction over time (per decision made every two rounds).

ture. For this subsample, we find again that sending a happy picture increases the likelihood of being the designated player. For cases in which followers 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, $p = 0.090$, $N = 16$).

An interesting by-product of the probabilistic implementation of the shown pictures is that followers are 'forced' to experiment. While in real life they may not always experiment with their facial expressions, and therefore miss out on learning opportunities, in our experiment they get feedback about the counterfactual. As the right panel of Figure 3 indicates, we do not observe learning effects. Followers select the 'wrong' picture at roughly equal rates over the course over the course of the experiment.

3.3.4 Investment by followers

For completeness, we briefly report investment behavior by followers. Overall, in 35 percent of the cases followers 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, $p = 0.254$, $N = 136$). Possibly, kindness and reciprocity could play a role in these investment decisions. Getting the task is kind in Desirable, and followers may wish to reciprocate the leader's kindness by investing. Note, however, that the role of kindness is probably somewhat limited in our setup, because the leader is forced to select one of the followers.

Are leaders right to select the followers that look happy? We find a modest link between the expression on the picture and the follower's investment choice.[10] We note, however, that our experimental design is not well suited to answer this question. The sample of followers making an investment decision is a selective sample, chosen by leaders. 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 follower would have been.

3.3.5 Discussion

Subjects understand the strategic benefits of adapting their facial expressions to the situation. Requesting subjects to express emotions has the advantage that we can create counterfactuals. At the same time, this setup has some limitations. One potential concern is that the expressions in the pictures might be unnatural. We tried to minimize this risk by explicitly asking participants to keep the expression as natural as possible. Our results are also robust to excluding pictures that appeared unnatural to us, and to excluding pictures with the 20 percent most extreme values of valence (see Table A.1 in the Appendix). Another potential limitation is that by requesting participants to express different emotions on the pictures, we made them aware of the potential benefits. It could also affect the moral cost of trying to manipulate the leader if people feel less guilty about expressing different emotions if they are instructed to do so by others. Both factors could cause overuse of the different expressions. In Experiment II, we address these limitations. In Experiment

[10] In the Desirable task treatment, followers 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).

II, we study whether subjects spontaneously adapt their facial expressions, without instructing them to express any emotions.

4 Experiment II

The setup of Experiment II is very similar to Experiment I. The main difference is that, to address the above-mentioned potential experimenter demand concerns, this time we did not instruct them to express emotions in the pictures. Instead, in Experiment II subjects in the role of followers took a new picture at the start of each round, and leaders were informed about this. Emotions or facial 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.[11]

We conducted 5 sessions, again with leaders in Tilburg (CentERlab) and followers 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).

4.1 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 Cen-

[11]The instructions for Experiment II can be found in Appendix B. In the first session of Experiment II, some subjects used hand gestures in 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.

tERlab in Tilburg, who each rated one picture of each of the 74 followers. 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 follower was judged by 24 raters.

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

4.2 Results

4.2.1 Facial expressions by followers

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) display a higher valence in the desirable task. We also find that followers 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 percent) give a higher average rating in treatment Desirable. In qualitative terms, this is consistent with the results from the previous experiment.[12]

In terms of magnitude, the difference between tasks is much more modest. The top panel of 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. Followers 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

[12] 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.4 in Appendix A, we show fixed effects regressions estimating the relation between valence in the current round and payoffs in the previous round. We find no significant effect of payoffs on valence with either task.

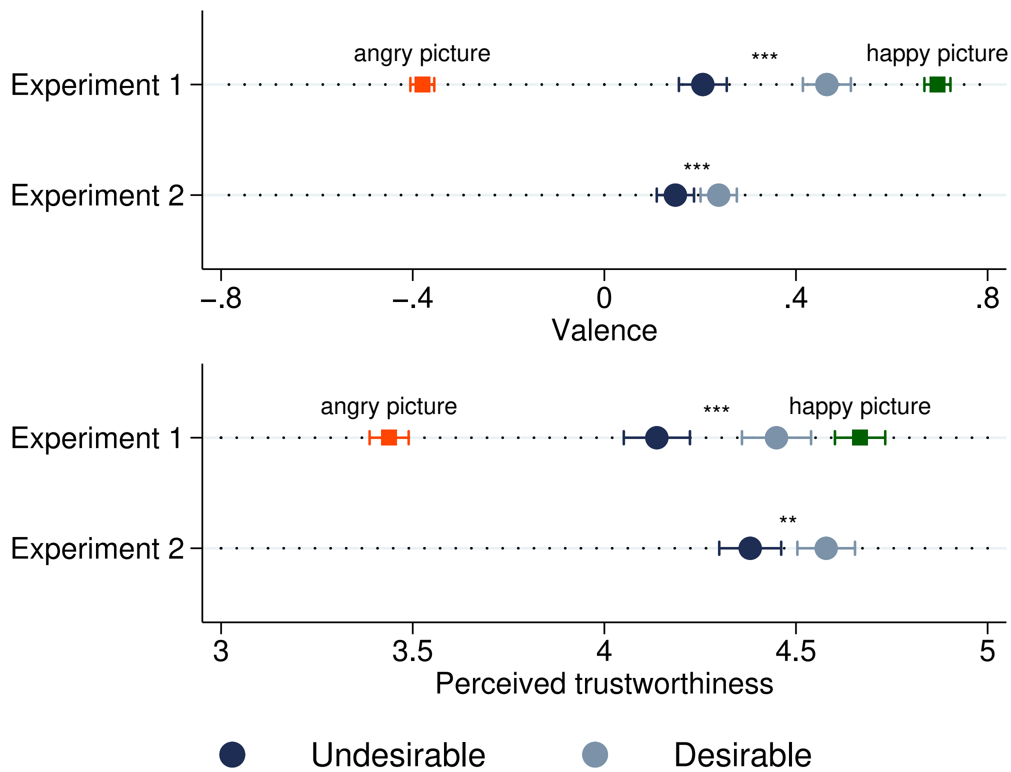


Figure 4: Mean valence and perceived trustworthiness on pictures taken and pictures sent. Squares indicate the mean valence (top panel) and perceived trustworthiness (bottom panel) on happy and angry pictures taken. Dots indicate the mean valence (top panel) and perceived trustworthiness (bottom panel) on pictures sent. Error bars indicate +/- 1 s.e. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$, relate to treatment differences.

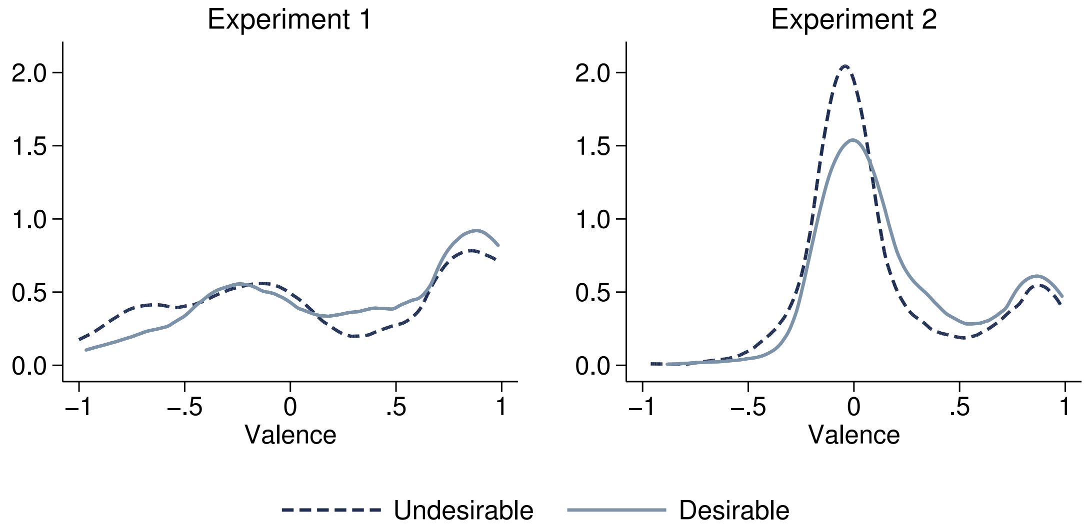


Figure 5: Kernel estimates of the distribution of valence on send pictures.

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. Similarly, the bottom panel of Figure 4 shows the differences in perceived trustworthiness in both experiments. The difference in perceived trustworthiness between treatments is again smaller in Experiment 2 than in Experiment 1. The treatment difference in perceived trustworthiness in Experiment 2 is only 63 percent of the difference in Experiment 1.

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.

4.2.2 Behavior of leaders

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

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 leader level. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

valence on the likelihood of becoming the designated player. Note that despite the weaker expressed emotions in Experiment II, leaders still react similarly to emotional valence: 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 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.

5 Heterogeneity in the strategic display of facial expressions

We next examine if there is heterogeneity in how strategic people are in their facial expressions. We collected information on sex, the level of strategic reasoning, and emotional intelligence, as described in subsection 3.2.

Our outcome variable concerns the level of strategic use of facial expressions. We evaluate this by the degree to which participants adjusted their expressions to

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 B.3.3 for a description of the measures. Robust s.e. in parentheses. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

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 in Experiment I we observe a marginally significant sex difference.

6 Discussion and concluding remarks

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

Our main interest is in whether the follower's facial expression affects the likelihood of getting the task and whether followers exploit this by adapting their facial expressions. It is worth emphasizing that those facial expressions need not necessarily be related to emotions. However, in one of our experiments, we explicitly ask followers to express emotions, and the software we use to read facial expressions also detects different expressed emotions. We hypothesize that followers use different facial expressions to induce (dis)trust by the leader. We indeed find that facial expressions affect trustworthiness perceptions, but it can equally well be that followers are trying to convey other signals, such as social preferences, or induce reciprocity by the leader. These interpretations are not essential to the questions we want to answer: Do facial expressions matter and do followers adapt their facial expressions? We also emphasize that our study is about *expressing* emotions, and not about actual emotions felt by followers.

In our design, the payoffs between the two treatments were not entirely symmetric. While this does not affect our main conclusion (that followers adapt their facial expressions depending on the strategic situation), it may affect our estimated effect sizes. In treatment "Desirable", followers are always better off if they get the task, while in "Undesirable" it depends on the decision of the other follower: followers who are pessimistic about the chance that the other followers will invest, may prefer to get the task even in the treatment "Undesirable". If that is the case, we might estimate a lower bound on what one may expect to find when it is never optimal to get an undesirable task. The incentives for leaders also differ between treatments, as leaders lose 1 euro in "Desirable" but only 0.8 euros in "Undesirable" if the follower does not invest. This difference is small, and we do not find that the leaders respond less to the emotional expression in the “undesirable” treatment (if anything, they respond stronger). Moreover, leaders who care strongly about equalizing payoffs, may even prefer that the follower does not invest in treatment ”Undesirable”. However, at least on average we do not find that leaders prefer followers who appear less trustworthy, and in general, the existing evidence does not support such a strong degree of inequality aversion (see e.g., Charness & Rabin, 2002; van Leeuwen & Alger, 2023).

Interestingly, both Experiment I and Experiment II show evidence that people might not fully exploit the potential benefits of displaying facial expressions. In Experiment I, some followers 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. This is especially striking in light of the fact they could only communicate through pictures.

What can explain the reluctance to display certain facial expressions? 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. Also, if people believe that others perceive them to be trustworthy, they are more likely to act trustworthy (Slepian & Ames, 2016). It could also be that people do not merely express different emotions, but actually feel them.[13] Faking emotions might also be cognitively demanding and thus not easy to achieve.

Given that followers strategically manipulate expressions, another question is why leaders 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)).

Even if leaders are aware of the fact that followers manipulate their expressions,

[13] Meshulam et al. (2012) incentivize recipients in a dictator game to feel angry upon receiving low offers, and find increased physiological arousal among those recipients.

they may still take them into account if they contain some useful information. Followers could all try to appear happier or angrier than they really are, but perhaps the followers who will actually invest still look happier than those who would not invest. Indeed, previous work shows that facial cues can be informative of behavior (Verplaetse et al., 2007; Bonnefon et al., 2013; Vogt et al., 2013; Tognetti et al., 2013; Slepian & Ames, 2016; van Leeuwen et al., 2018), though the predictive accuracy might be very weak (Rule et al., 2013; Foo et al., 2022). However, 'honest' smiles are associated with trustworthiness (Centorrino et al., 2015), and pupil dilation is related to deceptive behavior (Wang et al., 2010). Possibly, facial cues can be informative because people do not fully control their facial expressions, or perhaps lying aversion prevents them from doing so.

In Appendix C, we sketch a model based on the latter idea of lying costs. We show that expressions can have informational value in the equilibrium of a game in which each follower can send a 'message' to the leader, after which the leader decides to allocate a task to that follower or not. Those messages contain an indication of the follower's inclination to invest. We show that in the equilibrium of the game, followers will over-report their inclination to invest when the task is desirable, and under-report their inclination to invest when the task is undesirable, matching the behavior of participants in our two experiments. Although followers have incentives to misreport their inclination to invest, it can still be optimal for them not to exaggerate too much, as long as they experience some lying cost. As a consequence, leaders can distill some information from the messages. Intuitively, if lying costs are higher, followers will exaggerate their inclination to invest to a lesser extent. This could possibly also explain why subjects in Experiment I vary their expressions to a larger extent, compared to Experiment II. In Experiment I, we made the choice to express different emotions more salient, which may have lowered the moral cost of using the expressed emotions strategically.

Our results demonstrate that facial expressions might be an important mode of communication. This has especially high importance since people highly value facial cues. For many jobs, face-to-face interviews are still a key part of the recruitment procedure,[14] and most people post pictures of themselves on their profes-

[14]The recruiting platform *HireVue* analyzes facial expressions as part of the process (see this CNN article). In an experiment, Mobius and Rosenblat (2006) show that attractiveness has an impact on

sional websites. In trust games, many people and even children are willing to pay to see a picture of the other (Eckel & Petrie, 2011; Ewing, Caulfield, Read, & Rhodes, 2015) or to show their own picture to others (Heyes & List, 2016). Trustworthiness judgments of faces are fast and present even when faces are not consciously seen (Todorov et al., 2009; Freeman et al., 2014).

Of course, in terms of external validity, our experimental setup has some limitations. One interesting avenue for future research is to study how richer signals or repeated interactions would affect our results. Participants only had the possibility to signal through a photograph in our lab experiment. While this allowed us to cleanly identify the role of facial expressions, face-to-face interactions are much richer and might allow for more complex signals. It will be interesting to examine how richer signals affect the use and interpretation of facial expressions. It can facilitate getting the message across. At the same time, complex signals may distract or show inconsistencies, making it harder to interpret them. Furthermore, since not all expressions are under one's full control, people may start 'leaking' signals which can give away any insincerity of the expression[15]. Relatedly, it will be interesting to examine the coexistence of facial expressions and other (non)verbal expressions. Different ways to communicate may reinforce or work against each other. Finally, in our experiments, participants did not know each other and only interacted once. In the workplace, however, relationships are usually ongoing. While appearing angry or untrustworthy can have short-run benefits, there may also be long-run costs, as people might avoid future interactions with angry-looking individuals (Elster, 1998). Repeated encounters are also likely to raise leaders' awareness of any attempt to strategically regulate facial expressions. This could reduce the impact of displaying a certain expression on leaders' decisions, but does not necessarily take away the need to manipulate expressions for followers: if all other followers try to appear angry to avoid an undesirable task, it would still be in a follower's interest to do the same. Thus, our findings would still be relevant, and it would still be important for leaders to recognize this.

hiring decisions, and see e.g. Hamermesh and Biddle (1994) for empirical evidence.

[15]The literature shows only a modest ability to detect deception in other contexts. See for instance Bond Jr and DePaulo (2006) for a survey and Belot and Van de Ven (2017) and Serra-Garcia and Gneezy (2021) for some more recent evidence.

References

Ambady, N., Bernieri, F. J., & Richeson, J. A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. In *Advances in experimental social psychology* (Vol. 32, pp. 201–271). Elsevier.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, *111*(2), 256.

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.

Belot, M., & Van de Ven, J. (2017). How private is private information? the ability to spot deception in an economic game. *Experimental Economics*, *20*, 19–43.

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*.

Bond Jr, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

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., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

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.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science*, *18*(2), 327–334.

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

Foo, Y. Z., Sutherland, C. A. M., Burton, N. S., Nakagawa, S., & Rhodes, G. (2022). Accuracy in facial trustworthiness impressions: Kernel of truth or modern physiognomy? A meta-analysis. *Personality and Social Psychology Bulletin*, *48*(11), 1580–1596.

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.

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *Journal of Neuroscience*, *34*(32), 10573–10581.

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.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *The American Economic Review*, *84*(5), 1174–1194.

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.

Kopelman, S., Rosette, A. S., & Thompson, L. (2006). The three faces of eve: Strategic displays of positive, negative, and neutral emotions in negotiations. *Organiza-*

tional Behavior and Human Decision Processes, *99*(1), 81–101.

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.

Manzini, P., Sadrieh, A., & Vriend, N. J. (2009). On smiles, winks and handshakes as coordination devices. *The Economic Journal*, *119*(537), 826–854.

Meshulam, M., Winter, E., Ben-Shakhar, G., & Aharon, I. (2012). Rational emotions. *Social Neuroscience*, *7*(1), 11–17.

Mobius, M. M., & Rosenblat, T. S. (2006). Why beauty matters. *American Economic Review*, *96*(1), 222–235.

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

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, *105*(32), 11087–11092.

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

Rule, N. O., Krendl, A. C., Ivcevic, Z., & Ambady, N. (2013). Accuracy and consensus in judgments of trustworthiness from faces: behavioral and neural correlates. *Journal of Personality and Social Psychology*, *104*(3), 409.

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.

Serra-Garcia, M., & Gneezy, U. (2021). Mistakes, overconfidence, and the effect of sharing on detecting lies. *American Economic Review*, *111*(10), 3160–83.

Slepian, M. L., & Ames, D. R. (2016). Internalized impressions: The link between apparent facial trustworthiness and deceptive behavior is mediated by targets' expectations of how they will be judged. *Psychological Science*, *27*(2), 282–288.

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.

Sutherland, C. A., Young, A. W., & Rhodes, G. (2017). Facial first impressions from another angle: How social judgements are influenced by changeable and invariant facial properties. *British Journal of Psychology*, *108*(2), 397–415.

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social cognition*, *27*(6), 813–833.

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*, *86*(1), 57-76.

van Leeuwen, B., & Alger, I. (2023). Estimating social preferences and Kantian morality in strategic interactions. *TSE Working Paper*.

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 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 follower pairs of the same sex. Columns (3) and (4) split those results by sex. Column (5) shows the results for leaders that correctly answered all test questions on their first attempt. Column (6) excludes pairs of followers where at least one of the followers displays deviant behavior. Whether or not a follower 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. Column (7) excludes the 20% of pictures with the highest valence on happy pictures and the lowers valence on angry pictures.

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

Sample	(1) All	(2) Same-sex All	(3) Same-sex Males	(4) Same-sex Females	(5) Understood	(6) Compliant	(7) Not-extreme
PANEL A: UNDESIRABLE TASK							
Happy picture	1.818***	2.241***	2.231**	2.250***	2.476***	2.000***	1.872***
	(0.270)	(0.534)	(0.726)	(0.681)	(0.605)	(0.316)	(0.338)
Observations	792	392	172	220	348	722	556
PANEL B: DESIRABLE TASK							
Happy picture	1.405**	1.548**	1.423	1.750*	1.507**	1.410**	1.408*
	(0.207)	(0.306)	(0.339)	(0.563)	(0.244)	(0.212)	(0.266)
Observations	840	400	236	164	756	830	504

Notes: Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Columns (2)-(4): followers have the same sex. Column (5): "understood" means leader correctly answered all test questions on first try. Column (6): both followers complied with instructions to look natural. Column (7): excluding the 20% participants who display the strongest valence. Robust s.e. in parentheses clustered at the leader level.

Table A.2: Assignment of tasks in Experiment I (only using randomly shown pictures)

	(1)	(2)	(3)	(4)	(5)	(6)
Task:	Undesirable			Desirable		
Happy picture	0.135**			0.119**		
	(0.053)			(0.059)		
Valence		0.085*			0.066	
		(0.045)			(0.056)	
Perceived trustworthiness			0.072*			0.065
			(0.038)			(0.042)
Constant	0.413***	0.468***	0.188	0.444***	0.496***	0.241
	(0.035)	(0.025)	(0.156)	(0.037)	(0.024)	(0.174)
Observations	396	396	396	419	415	419

Notes: Linear regressions. Dep. var.: being the designated player. All regressions are based on data from followers whose picture (happy or angry) was randomly selected. Valence is measured by FaceReader software, and is a score between -1 and 1. Perceived trustworthiness is the average rater score (between 1 and 7). All regressions include fixed effects for each follower. Column (7) has fewer observations because FaceReader did not capture every face. Robust s.e. in parentheses clustered at the leader level. *** $p < 0.01$,** $p < 0.05$,* $p < 0.1$.

Table A.3: Sex differences in assignment of tasks: Experiments I and II

	(1)	(2)	(3)	(4)
Task	Undesirable		Desirable	
Sample	Experiment I	Experiment II	Experiment I	Experiment II
Female	0.835	1.314	1.202	1.107
	(0.117)	(0.253)	(0.181)	(0.224)
Observations	400	236	436	236

Notes: Conditional logit estimates (odds ratios). Dep. var.: being the designated player. Only follower pairs with one male and one female follower are included. Robust s.e. in parentheses clustered at the leader level. *** $p < 0.01$, ** $p < 0.05$,* $p < 0.1$.

Table A.4: 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 B Instructions

B.1 Experiment I

B.1.1 Instructions for red players (leaders)

General information [identical for both treatments]

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.

Investment decision [desirable task treatment]

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.

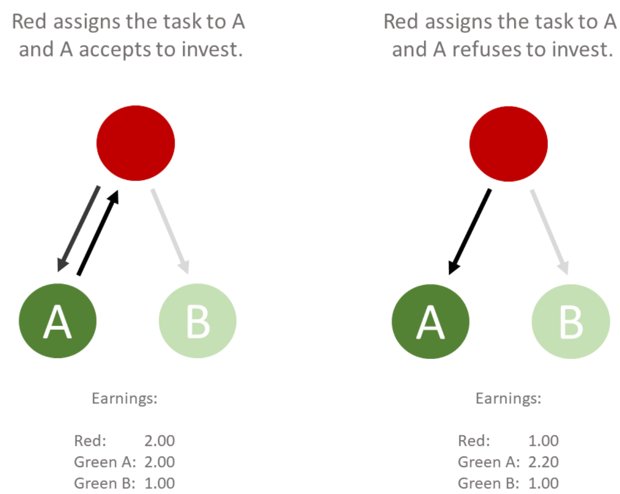
Your task is to assign the investment task to one of the green players. That green player becomes the ”designated” player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

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.



Investment decision [undesirable task treatment]

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

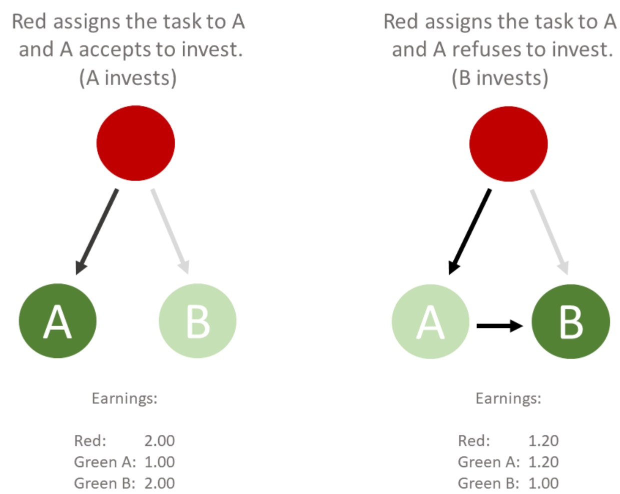
Your task is to assign the investment task to one of the green players. That green player becomes the ”designated” player. The designated green player can accept or refuse to invest. If the designated green player refuses, the other green player has to invest.

The earnings are as follows:

The green player that ends up making the investment always earns €1.00. The other green player and you earn €2.00 each if the designated green player accepts to invest, and €1.20 if the designated green player refuses to invest.

Suppose, for instance, that you assign the investment task to green A. If green A accepts to invest, he or she earns €1.00, and you and green B earn €2.00 each. If green A refuses to invest, then green B has to invest. Green A and you earn €1.20 each, and green B earns €1.00.

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 [identical for both treatments]

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 [desirable task treatment]

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

Summary [undesirable task treatment]

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. When the designated green player refuses, the other green player has to invest.

The (green) player that invests earns €1.00. The other (red and green) players earn €2.00 each if the designated player accepted to invest, and €1.20 each if the designated player refused to invest.

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

Quiz questions [identical for both treatments]

Please answer the following quiz questions. If you have any questions please raise your hand.

In each of the 12 rounds, you will be matched:

- ☐ with the same players
- ☐ with different players

Suppose the green player you assigned the task to refuses to invest. How much will you, the designated green player, and the other green player earn?

- You would earn: _____euro

- The designated green player would earn: _____ euro
- The other green player would earn: _____ euro

Suppose the green player you assigned the task to invests. How much will you, the designated green player, and the other green player earn?

- You would earn: _____euro
- The designated green player would earn: _____ euro
- The other green player would earn: _____ euro

B.1.2 Instructions for green players (followers)

Welcome [identical for both treatments]

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, you will be asked to take pictures of yourself. This is what we will do with your pictures:

- The pictures will be shown to other participants, but only to participants located elsewhere. Your pictures will not be shown to participants at the University of Amsterdam.
- The pictures will be analyzed with facial processing software.

You can choose not to participate in the study. You may withdraw at any time during the study. If you withdraw, no more information will be collected from you and the investigator will ask if the materials already collected in the study can be used. Results of this study may be used for teaching, research, publications, or presentations at scientific meetings. Your identity or picture will never be disclosed in any of these activities.

By continuing with the experiment you give consent to take part in this study.

- ☐ I have read and understood the information above and I want to continue with the experiment.
- ☐ I do not want to participate in the experiment.

Taking pictures [identical for both treatments]

In the next steps, you will take some pictures of yourself. To ensure that the pictures are useable, please pay attention to the following factors:

1. Look straight into the camera,
2. Capture your entire face on the picture,
3. Do not wear glasses or other items that cover your face.
4. If you have long hair, please hold them back with an elastic band so that they do not fall into your face.

We will ask you to express different emotions. When you express a certain emotion, try to do this as natural as possible. Try to express the emotion in a convincing way, without overdoing it.

Take a picture [identical for both treatments]

On the next picture we ask you to look angry

Try to look angry [identical for both treatments]

When you take a picture, there will be a 3-seconds countdown.

Picture result [identical for both treatments]

Below is the picture that you took. Please indicate whether you are satisfied with it or want to retake the picture

Take a picture [identical for both treatments]

On the next picture we ask you to look happy

Try to look happy [identical for both treatments]

When you take a picture, there will be a 3-seconds countdown.

General information [identical for both treatments]

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.

Investment decision [desirable task treatment]

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.

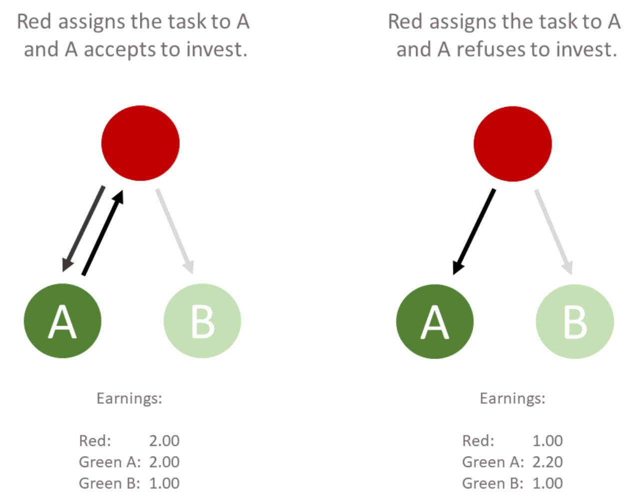
The task of the red player is to assign the investment task to one of the green players. That green player becomes the ”designated” player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

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.



Investment decision [undesirable task treatment]

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

The task of the red player is to assign the investment opportunity to one of the green players. That green player becomes the "designated" player. The designated green player can accept or refuse to invest. If the designated green player refuses, the other green player has to invest.

The earnings are as follows:

The green player that ends up making the investment always earns €1.00. The other green player and the red player earn €2.00 each if the designated green player accepts to invest, and €1.20 if the designated green player refuses to invest.

Suppose, for instance, that the red player assigns the investment task to green A. If green A accepts to invest, he or she earns €1.00, and the red player and green B earn €2.00 each. If green A refuses to invest, then green B has to invest. Green A and the red player earn €1.20 each, and green B earns €1.00.

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.

Pictures [identical for both treatments]

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 [identical for both treatments]

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 [desirable task treatment]

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.

Summary [undesirable task treatment]

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. When the designated green player refuses, the other green player has to invest.

The (green) player that invests earns €1.00. The other (red and green) players earn €2.00 each if the designated player accepts to invest, and €1.20 each if the designated player refuses to invest.

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.

Quiz questions [identical for both treatments]

Please answer the following quiz questions. If you have any questions please press the call button on the wall next to you.

In each of the 12 rounds, you will be matched:

☐ with the same players

☐ with different players

Suppose you are chosen to invest by the red player. You refuse to invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose you are chosen to invest by the red player. You invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose the other green player is chosen to invest by the red player. The other green player invests. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose the other green player is chosen to invest by the red player. The other green player refuses to invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

B.2 Experiment II

B.2.1 Instructions for red players (leaders)

General information [identical for both treatments]

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 money depending on your own choices and the choices of other participants.

Today's experiment consists of several independent parts. This is the first part. At the beginning of each part you will receive new instructions. Your decisions in one part do not influence the proceedings or earnings of any other part.

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.

Investment decision [desirable task treatment]

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.

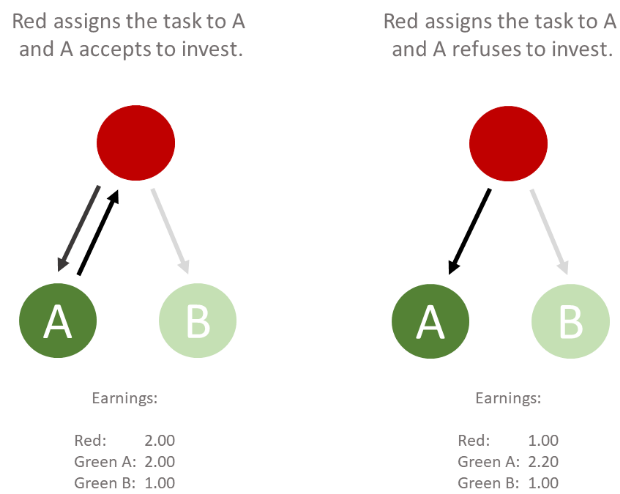
Your task is to assign the investment task to one of the green players. That green player becomes the "designated" player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

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.



Investment decision [undesirable task treatment]

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

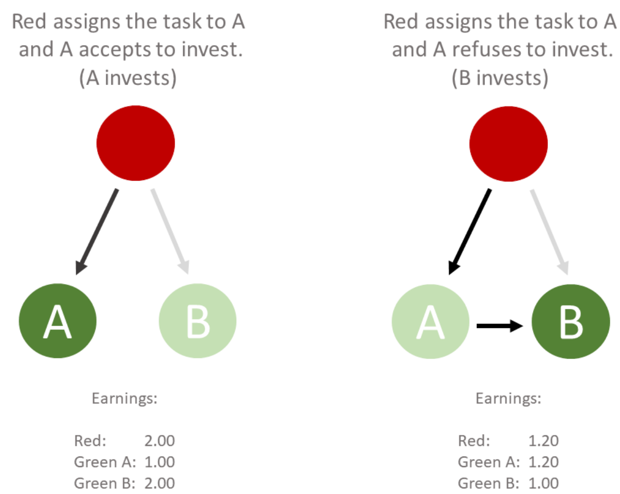
Your task is to assign the investment task to one of the green players. That green player becomes the "designated" player. The designated green player can accept or refuse to invest. If the designated green player refuses, the other green player has to invest.

The earnings are as follows:

The green player that ends up making the investment always earns €1.00. The other green player and you earn €2.00 each if the designated green player accepts to invest, and €1.20 if the designated green player refuses to invest.

Suppose, for instance, that you assign the investment task to green A. If green A accepts to invest, he or she earns €1.00, and you and green B earn €2.00 each. If green A refuses to invest, then green B has to invest. Green A and you earn €1.20 each, and green B earns €1.00.

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 [identical for both treatments]

In total, there will be 6 rounds in this part. You will be paid for every round.

When you assign the task to invest, you will see pictures of the green players. The green players make new pictures at the beginning of every round.

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 [desirable task treatment]

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 6 rounds in this part, 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

Summary [undesirable task treatment]

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. When the designated green player refuses, the other green player has to invest.

The (green) player that invests earns €1.00. The other (red and green) players earn €2.00 each if the designated player accepted to invest, and €1.20 each if the designated player refused to invest.

There will be 6 rounds in this part, 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

Quiz questions [identical for both treatments]

Please answer the following quiz questions. If you have any questions please raise your hand.

In each of the 6 rounds, you will be matched:

- □ with the same players
- □ with different players

Suppose the green player you assigned the task to refuses to invest. How much will you, the designated green player, and the other green player earn?

- You would earn: _____euro
- The designated green player would earn: _____ euro
- The other green player would earn: _____ euro

Suppose the green player you assigned the task to invests. How much will you, the designated green player, and the other green player earn?

- You would earn: _____euro
- The designated green player would earn: _____ euro
- The other green player would earn: _____ euro

B.2.2 Instructions for green players (followers)

Welcome [identical for both treatments]

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, you will be asked to take pictures of yourself. This is what we will do with your pictures:

- The pictures will be shown to other participants, but only to participants located elsewhere. Your pictures will not be shown to participants at the University of Amsterdam.
- The pictures will be analyzed with facial processing software.

You can choose not to participate in the study. You may withdraw at any time during the study. If you withdraw, no more information will be collected from you and the investigator will ask if the materials already collected in the study can be used. Results of this study may be used for teaching, research, publications, or presentations at scientific meetings. Your identity or picture will never be disclosed in any of these activities.

By continuing with the experiment you give consent to take part in this study.

- ◻ I have read and understood the information above and I want to continue with the experiment.
- ◻ I do not want to participate in the experiment.

General information [identical for both treatments]

Today's experiment consists of several independent parts. This is the first part. At the beginning of each part you will receive new instructions. Your decisions in one part do not influence the proceedings or earnings of any other part.

In this part, 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.

Investment decision [desirable task treatment]

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.

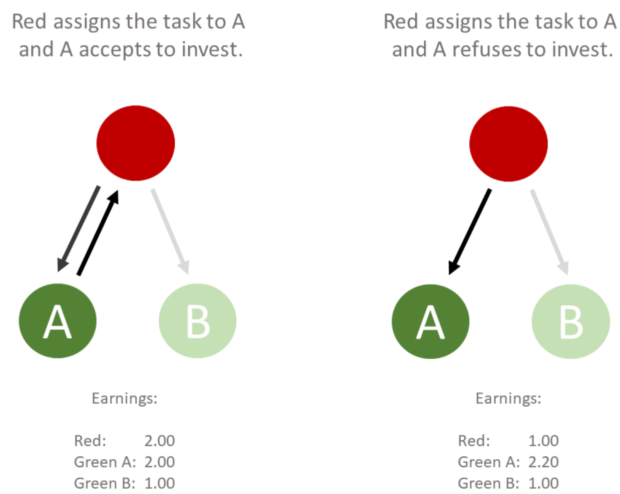
The task of the red player is to assign the investment task to one of the green players. That green player becomes the "designated" player. The designated green player can accept or refuse to invest. The other green player has no decision to make.

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.



Investment decision [undesirable task treatment]

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

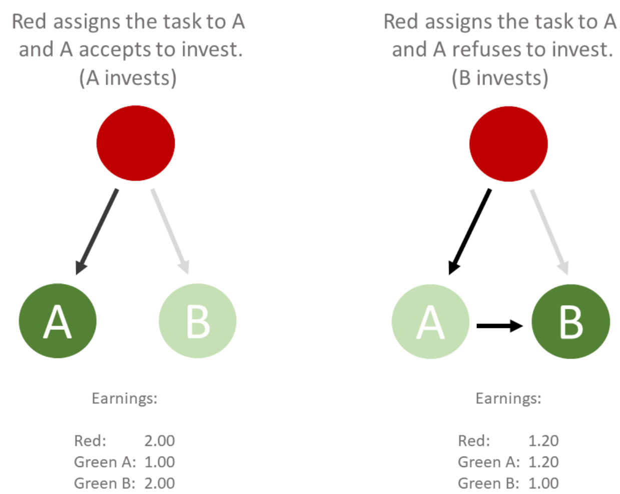
The task of the red player is to assign the investment opportunity to one of the green players. That green player becomes the "designated" player. The designated green player can accept or refuse to invest. If the designated green player refuses, the other green player has to invest.

The earnings are as follows:

The green player that ends up making the investment always earns €1.00. The other green player and the red player earn €2.00 each if the designated green player accepts to invest, and €1.20 if the designated green player refuses to invest.

Suppose, for instance, that the red player assigns the investment task to green A. If green A accepts to invest, he or she earns €1.00, and the red player and green B earn €2.00 each. If green A refuses to invest, then green B has to invest. Green A and the red player earn €1.20 each, and green B earns €1.00.

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.



Pictures [identical for both treatments]

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. At the beginning of every round, you will first take a new picture.

Feedback [identical for both treatments]

In total, there will be 6 rounds in this part. You will be paid for every round.

At the end of every round, you will learn 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:

- 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 [desirable task treatment]

The red player sees your picture and the picture of the other green player.

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 6 rounds in this part, 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.

Summary [undesirable task treatment]

The red player sees your picture and the picture of the other green player.

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. When the designated green player refuses, the other green player has to invest.

The (green) player that invests earns €1.00. The other (red and green) players earn €2.00 each if the designated player accepts to invest, and €1.20 each if the designated player refuses to invest.

There will be 6 rounds in this part, 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.

Quiz questions [identical for both treatments]

Please answer the following quiz questions. If you have any questions please press the call button on the wall next to you.

In each of the 6 rounds in this part, you will be matched:

- □ with the same players
- □ with different players

Suppose you are chosen to invest by the red player. You refuse to invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose you are chosen to invest by the red player. You invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose the other green player is chosen to invest by the red player. The other green player invests. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Suppose the other green player is chosen to invest by the red player. The other green player refuses to invest. How much will the red player, you, and the other green player earn?

- The red player would earn: _____ euro
- You would earn: _____euro
- The other green player would earn: _____ euro

Taking pictures [identical for both treatments, shown again at the beginning of every round]

In the next steps, you will take a picture of yourself. To ensure that the pictures are useable, please pay attention to the following factors:

1. Look straight into the camera,
2. Capture your entire face on the picture,
3. Do not wear glasses or other items that cover your face.
4. Do not show your hands on the picture.
5. If you have long hair, please hold them back with an elastic band so that they do not fall into your face.

Take a picture for the next round [identical for both treatments, shown again at the beginning of every round]

In total, you have 3 tries to make your picture. This is your first try.

When you take a picture, there will be a 3-seconds countdown.

Picture result [identical for both treatments, shown again at the beginning of every round]

Below is the picture that you took. Please indicate whether you are satisfied with it or want to retake the picture.

B.3 Survey measures

B.3.1 Strategic Reasoning

Instructions

You will now play the following game with the computer. Imagine there are 15 chips on the table. You and the computer take turns. Every time it is a player's turn, that player can remove 1, 2, or 3 chips from the table. The player who takes the final chip wins the game. You will be the player that starts.

Instructions

You will play another game with the computer. This time there are 17 chips on the table. You and the computer again take turns. Every time it is a player's turn, that player can remove 1, 2, 3, or 4 chips from the table. The player who takes the final chip wins the game. You will be the player that starts.

B.3.2 Angry button test

Decision

Below you see pictures of participants in another experiment. They played following game:

In the game, there were two types of players: 'proposers' and 'responders'. Each proposer could make an offer on how to divide €9. The proposer could choose between two options:

Option A gives €4 to the proposer and €5 to the responder,
Option B gives €7 to the proposer and €2 to the responder.

If the responder accepts the offer, the money was divided as proposed. If the responder rejects the offer, none of them earned money.

All the people on the pictures below were responders in this game. All of them were offered Option B (€7 for the proposer and €2 for the responder). 5 of the responders rejected the offer, the others accepted the offer. **Please select the 5 people who you think rejected the offer**.

B.3.3 Reading the mind in the eyes test

Instructions

For each set of eyes, choose and select which word best describes what the person in the picture is thinking or feeling. You may feel that more than one word is applicable but please choose just one word, the word which you consider to be most suitable. Before making your choice, make sure that you have read all 4 words. You should try to do the task as quickly as possible but you will not be timed. If you really don't know what a word means you can look it up by moving your mouse over the question mark. By doing so, you will also see the Dutch translation.

Below is an example. You can proceed by selecting a word and clicking on 'OK'.

Appendix C A model of expression of emotions

In our experiments, subjects in the role of followers adapt their expressed emotions strategically, though only partially, and those in the role of leaders pay attention to the expressed emotions. If followers adapt their expressions strategically to manipulate the leaders, one can expect that leaders will take this into account and ignore expressed emotions. If so, it is not clear whether there is still value for followers in adapting expressed emotions. Thus, it is not evident that the observed behavior can be equilibrium behavior. In this section, we present a model that is consistent with this type of behavior, showing that (under some conditions), such behavior can be equilibrium behavior. As in the experiment, the expressed emotions are not necessarily felt by followers, but merely used as a signal of intentions.

We model the situation as a sender-receiver game with multiple senders. In such a game, senders first send a message to the receiver, after which the receiver takes some action. The senders can be thought of as the followers in our experiment, and the receiver as the leader. Each sender chooses an emotional expression (a 'message') to signal his inclination to invest (his 'type') to the receiver. The receiver observes the expression and 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. 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).Loosely speaking, in a monotonic signaling game all types of senders rank the actions (including mixed strategies) by the receiver in the same way. Thus, either each type of sender prefers to be allocated to the task, or each type of sender prefers the other sender to be allocated to the task.

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 leader or co-follower.

After privately observing his type, each sender sends a message $m \in M = [0,1]$.[16] 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 or faking emotions, i.e., showing an intensity level of valence that does not match the actual type, where we assume that the actual (experienced) 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. In such equilibria, each player selects the strategy that maximizes his or her expected payoffs given his or her beliefs about the others. The beliefs must satisfy some rationality requirements.[17] 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

[16]Note that our message space is less rich than in Kartik (2009).
[17]For a precise definition, see for instance Fudenberg and Tirole (1991).

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 D.

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, in which different types can send different messages. 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 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.[18]

[18] In a somewhat different setting, Charness et al. (2018) derive a similar result. In their model, players can enter a competition. They can send messages about the strength of their type to their opponent. They show that in equilibrium, players will either overstate or understate their strength, depending on whether they want to discourage or encourage the other from competing with them.

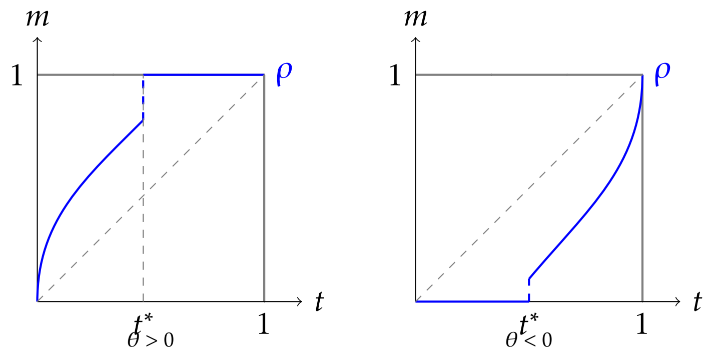


Figure A.1: Equilibrium strategies for the case of a desirable task (left, $\theta > 0$) and an undesirable task (right, $\theta < 0$). The figure illustrates a partially separating strategy.

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 A.1 shows a partially separating equilibrium in which there is a single pool (in Appendix D 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. As types are directly associated with the propensity to invest, observing a higher report indicates a higher investment rate. 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. Finally, we note that the degree of over- or underreporting depends negatively on the cost of lying: $|\rho^*(t) - t|$ (weakly) decreases in k.

Appendix D 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).