Abstract

The temptation to deceive others compares to a moral dilemma: it involves a conflict between the temptation to obtain some benefit and the desire to conform to personal and social moral norms or avoid aversive social consequences. Thus, people might feel different levels of emotional and moral conflict, depending on the target of the deception. Here, we explored, in a morally relevant setting, how social judgments based on two fundamental dimensions of human social cognition – the 'warmth' and 'competence' – impact on the decision to deceive others. Results revealed independent effects for warmth and competence. Specifically, while people are less inclined to deceive for self-gain those individuals they perceive as warm, they also tend to lie more to highly competent others. Furthermore, the perceived warmth and competence modulated the general tendency to reduce deceptive behaviour when there was a risk of disclosure, compared to when the lying was anonymous, highlighting the importance of these judgments in social evaluation processes. Together, our results demonstrate that the emotional costs and personal moral standards that inhibit engagement in deceptive behaviour are not stable but rather malleable according to the target and the consequences of the deception.

Keywords: Lie; Deception; Guilt; Stereotype Content Model; Bias; Moral

1. Introduction

Despite the fact that deception is a constant in daily life and may even be necessary for adequate social functioning in many situations, such behaviour is socially condemned (Mead et al., 2009; Mastroberardino, 2012). Typically, deceptive behaviour, such as lying, is regarded as dishonest and a betrayal of trust. The temptation to deceive others to obtain some kind of benefit involves a social dilemma, invoking a conflict between the temptation to obtain some benefit, versus the desire to conform to personal and social moral norms or avoid aversive social consequences (Mead et al., 2009; Mazar et al., 2008). Lying, is broadly defined as the voluntary decision to manipulate information in order to create a misbelief in others (Masip et al., 2004). It often entails some psychological cost to the liar, associated with the violation of moral codes (Mazar et al., 2008) and/or the risk of social penalties, such as reputational loss or punishment (Becker, 1968).

The social and moral aspects of intentional deception have recently been investigated using ecologically realistic experimental paradigms (e.g. Panasiti et al., 2011; Gneezy et al., 2005; Greene and Paxton, 2009; Baumgartner et al., 2009). However, little attention has yet been devoted to studying how social context impacts on spontaneous deception. One factor that should potentially influence the decision to engage in deceptive behaviour is the individual characteristics of the target of the deception. It is reasonable to expect that people might feel different levels of psychological or moral conflict when lying to different targets. Past research, based on self-report, shows that people tell more self-serving and fewer altruistic lies to strangers, compared to people that they care about (DePaulo and Kashy 1998; Ennis, Vrij, and Chance 2008). A similar pattern has been found in economic games, where people are more likely to deceive a stranger than a friend (Van Swol, Malhotra, and Braun 2012).

Interestingly, the use of deception during negotiations is decreased by the perceived trustworthiness of the other (Olekalns and Smith 2009) and increased (during social interaction) when the other is considered to be a liar (Tyler, Feldman, and Reichert 2006).

When navigating the complexities of social life, people rely on categorization processes to make inferences about others' personality traits and simplify the typically multifaceted information originating from other people (Tajfel, 1982; Van Bavel and Cunningham, 2011). Tellingly, automatic categorization processes impact numerous domains of social behaviour such as face processing (Golby et al., 2001; Cunningham et al., 2004), empathic resonance (Azevedo et al., 2013, 2014; Cikara et al., 2011), in-group/out-group categorization (Ponsi et al., 2015), learning of social fear (Olsson et al., 2005; Navarrete et al., 2012; Lindström et al., 2014) and social decision-making (Stanley, 2011). The Stereotype Content Model (SCM; Fiske et al., 2002) has proved particularly useful in understanding the social structure of intergroup processes and describing how inferences about others' personality traits shape our cognitive, affective and behavioural tendencies towards them. This Model has been validated within US society (where it was first developed) and also in many other countries, including Italy - where the present study took place (Durante et al., 2012). The SCM posits that stereotypes may be captured by two key universal dimensions of social cognition that shape interpersonal behaviour. The "warmth" dimension relates to the perception of others according to their perceived intent for good or ill and therefore comprises judgments of others as friendly, warm, trustworthy and sincere. The "competence" dimension is related to social status and the other person's perceived ability to pursue his/her intents - comprising traits such as efficacy, intelligence, confidence and skill. High warmth is associated with high morality and communality, while high competence is associated with power and high status. Although trait inferences along these two dimensions are usually assessed together, warmth

judgments are believed to take precedence over competence evaluations (Willis & Todorov, 2006) and to have a greater weight in impression formation. This is likely to be due to the fact that inferring another person's intentions, for ill or for good, is more important than judging the ability of that person to act on those intentions. The asymmetrical weight given to these types of trait inferences also predicts different patterns of behavioural tendencies. While the warmth dimension is associated with active behaviours (such as helping or, conversely, attacking), perceived competence predicts passive behaviours, such as associating or neglecting (Cuddy et al., 2008). Indeed, experimental research has shown that judgments of morality and trustworthiness predict approach-avoidance behaviours (Caccioppo et al., 1997) as well as trust (Delgado et al., 2005; Stanley et al., 2011) and also promote cooperative behaviour in negotiation contexts (Steinel and De Dreu, 2004).

In the present study, we used an ecological paradigm for assessing deception (Panasiti et al., 2011, 2014, 2016) in order to explore how social judgments along these two dimensions modulate the decision to lie to others. In this two-person online card game participants are always entirely free to decide whether to lie or tell the truth. Crucially, however, the decision to lie or tell the truth determines not only the participant's own payoff but also that of their opponent. Because participants have total control over their own and the opponent's payoff, their behaviour reflects only their intentional decisions to deceive, for self-gain or altruistic motives, and not any other competition-related strategy. The game is played under two conditions. In one condition the participant's decision is completely anonymous (thus, the participant is aware that the opponent will never know if his/her outcome was due to good/bad luck or to participant's decision to manipulate it). Conversely, in the other disclosed condition there is a risk that the opponent, may become aware of whether the participant lied or not. This paradigm allows us to simulate, in a laboratory setting, three key aspects of

deception behaviour that occur in daily life: 1) lying is intentional and instrumental, i.e.it reflects a personal decision motivated by the desire either (i) to obtain personal benefits ("self-gain" lie)or (ii) to lie "altruistically" i.e. to increase the opponent's' pay off at cost to the participant; 2) it implies a certain degree of moral commitment or conflict; 3) it involves some risk of social confrontation (Sip et al., 2008; Becker et al., 1968). Previous research using this paradigm has demonstrated that motivations and the neural and autonomic correlates (Panasiti et al., 2014; 2016) of the act of telling the truth or lying for self-profit are related to levels of commitment to moral standards, to manipulativeness and also to image management concerns (Panasiti et al., 2011).

Here, in an upgraded version of the original paradigm, participants were told that they were playing against four opponents, each of whose profiles fell into one of the four categories of the SCM (i.e., high warm-high competent (HW-HC); high warm-low competent (HW-LC); low warm-low competent (LW-LC); low warm-high competent (LW-HC)). This procedure allowed us to study how trait inferences impact on the tendency to deceive others. In particular, we were able to explore the relative contributions of others' perceived warmth and competence in modulating lie behaviour, in a morally relevant setting. We also investigated if and how concerns of social judgment, as indexed by different behaviour between the anonymous and disclosed conditions, have a different impact on decision-making according to the opponents' characteristics. Finally, a number of subjective ratings were collected to better understand the psychological processes underlying the decision to lie or to tell the truth.

Because moral traits are of primary importance in social evaluations (Cuddy et al., 2008; Wojciszke et al., 1998) and people tend to trust (Delgado et al., 2005; Stanley et al., 2011), cooperate with (Steinel and De Dreu, 2014) and identify themselves with (Farmer et al.,

2014) trustworthy others, we expected less lying for self-gain (i.e. egoistic lies) during interactions with opponents judged as highly warm. Conversely, competence is not directly associated with others' intentions but instead with their power and status. Thus, predictions about the influence of competence on spontaneous deception are not as obvious as for the warmth dimension. While research on cooperation suggests that people are less likely to cooperate with low competent others (e.g. Van Lange and Kuhlman, 1994), feelings of social responsibility towards powerless individuals (e.g. Handgraaf et al., 2008) may lead to less egoistic lies when dealing with low competent opponents. As for altruistic lies, these could be seen as acts of generosity, therefore we also predicted different modulation patterns for perceived warmth and competence. Warmth is tipically associated with active behavioural patterns, such as helping behaviour, whereas competence is linked to passive behavioural tendencies, such as passive facilitation and neglect (Cuddy et al., 2008). In view of this, we predicted increased altruistic lying to high warmth individuals but no modulation of lying as a function of competence. Finally, following past research (Panasiti et al., 2011,2014,2016), we expected that reputation concerns would lead to higher numbers of self-gain lies in the anonymous versus the disclosure condition. However, we had no specific hypothesis about how reputation may interact with perceived warmth and competence.

2. **Methods**

2.1 Participants

Sample size was determined using a power analysis based on previous study from our lab investigating the effects of the game outcome and reputation on deception (Panasiti et al., 2011; 2014). The effect size (partial eta square) of the interaction of interest ranged between

0.14 and 0.36 with an average of 0.25. We determined that a sample size of 36 participants would achieve 95% power to detect such effect size with an alpha of .05. A total of 36 healthy volunteers (19 males; mean age=23.7, s.d.=2.8) took part in the experiment. Careful recruitment procedures guaranteed that participants were naïve to research involving online interaction with other participants. Nevertheless, 4 participants were excluded from the analysis for reported suspicion about the procedure. Thus, the final dataset comprises 32 participants (17 males; mean age=23.8; s.d.=2.9). This sample size is similar to several other social decision-making studies (e.g. Panasiti et al., 2011; Tabibnia et al., 2008; Moretti and di Pellegrino, 2010; Halali et al., 2013; Leggenhager et al., 2013). Participants gave written informed consent and the study was approved by the local ethics committee.

2.2 The Temptation to Lie Card Game

The experimental task consisted of a two-card card game where the ace of hearts wins and the ace of spades loses. The participant (P) and the opponent (OP) had different roles. On each trial, the opponent's (OP's) task was to guess which one of two concealed cards was the ace of hearts, i.e. the card that gave him a win and the participant (P) a loss. However, OP was not able to see the true outcome of his/her choice, i.e. he could not know if the card s/he had chosen was the ace of hearts or the ace of spades. It was P's task to communicate the outcome to OP. We refer to trials in which OP's choice should have caused P to win as "Favorable" and to those in which OP's choice should cause P to lose as "Unfavorable". Crucially however, P had the opportunity to reverse the true outcome. That is, by lying P could chose to win when s/he had actually lost (a self-gain lie) or could lose when s/he had actually won (an altruistic lie). Participants were assured beforehand that they were always completely free to decide on each trial whether to lie or tell the truth.

The game was performed in two conditions. In the "No-Reputation" risk condition (No-Rep) P's decision to lie or tell the truth remained completely anonymous, i.e. OP could only be aware of what was communicated to him by P. In the "Reputation Risk" condition (Rep), however, participants were told that on two thirds of the trials, randomly chosen by the

computer and unknown to P, the OP would be able see both the true outcome and what P had communicated to him. Thus OP would be able to know if P had lied or told the truth.

Importantly, the game was played for real money so that a win by one player involved a corresponding loss to the other. Participants were guaranteed a minimum of 15 euro for taking part in the study but wins could increase this sum. To prevent participants from keeping track of their wins and losses, or from doing any other computation on wins/losses, they were told that each trial would be associated with an arbitrary and variable amount of reward and that this amount would be always unknown to both players. No maximum payoff was ever mentioned. Finally, they were told that they would meet their opponents at the very end of the experiment. This design allows simulation of the intentional, social and moral dimensions of lying behaviour in real life.

2.3.1 Experimental Procedures

Participants were comfortably seated in front of a computer monitor where stimuli were presented. The experimenter had no visual access to the screen or keyboard and had no way of knowing P's decisions. Each trial started with the presentation of a fixation cross for 2000 ms. Then two covered cards were shown on the screen until the OP had (allegedly) chosen one of them (between 1500ms and 2000ms). Immediately after this choice was made, the ace of spades and ace of hearts appeared uncovered on the screen. One of the cards would be bigger, reflecting which card OP had chosen. At this point, P was required to communicate the outcome of OP's choice to him, by pressing the left /right cursor key to indicate that the chosen card was the one on the left/right hand side of the screen, respectively. Thus, whenever P chose to tell the truth s/he pressed the key corresponding to side of the screen where the bigger card was presented. Conversely, when P chose to lie and thus to reverse the

true outcome, s/he pressed the key corresponding to the side of the screen where the smaller card was shown. P was instructed to press the chosen key within the 2500 ms time window in which the cards remained visible on the screen (see Figure 1). The left/right positions of the ace of hearts/spades and the bigger/smaller card were counterbalanced and randomized.

In order to study how social judgments modulate deception, participants played against 4 different supposed OPs (see below). They were told that they were playing online with participants who were sitting in different rooms of the same building. Stimuli presentation was organized in mini-blocks of opponent type, so that participants played 20 consecutive trials against a given OP. There were 12 experimental blocks in total, 3 for each OP, and an equal number of trials (15) for each condition (i.e. Rep/No-Rep and Favorable/Unfavorable), resulting in a total of 240 trials. Immediately before and after each block, participants were given forced-choice memory questions (see below) about the opponent they had played with in that block. Block order was randomized to prevent any systematic order effect. Given that the total number of players was uneven (one P and 4 OPs i.e. 5) and that players could not possibly play all others simultaneously, two two-minute waiting periods were included between blocks to simulate the times when the remaining participants were supposedly playingwith each other. Within blocks, Reputation (Rep trials; No-Rep trials) and Outcome (Favourable trials; Unfavourable trials) conditions were presented in a fully randomized fashion. On each trial, the supposed OPs name was visible at the top of the screen, together with an open or closed eye symbol which indicated the Rep or no-Rep condition, respectively (see Figure 1).

-------- Figure 1 ---------

218

219 *2.3.2 Profiles*

220 The profiles of the 4 OPs were created to fall at the extremes of the SCM's competence and
221 warmth dimensions, while remaining credible descriptions of potential research volunteers.
222 The (presumably) main stereotyping characteristic of each OP profile was based on the social
223 groups identified by the SCM literature (Fiske et al., 2002; Harris and Fiske, 2006; Cuddy et
224 al., 2008) and adapted to be relevant to the participants' population, which was university
225 students and young professionals. Thus, LC-LW was described as a low-class Eastern Europe
226 migrant; LC-HW was physically disabled; LW-HC was said to be an ambitious business
227 student; and HW-HC was a friendly music student. Additional personal information was
228 included, to reinforce the likelihood of participants correctly **evaluating OPs** along the
229 competence and warmth dimensions. An online pilot study was undertaken to validate these
230 opponent profiles (n=37; 11 males; mean age=27.9, s.d.=4.8). In this pilot each profile was
231 rated (0-10) on competence, i.e. how competent (e.g. intelligent, capable, efficacious), and
232 warmth, i.e. how warm (e.g. affectionate, friendly, trustworthy). Results confirmed that the
233 four opponent profiles differed significantly from each other, F-Greenhouse-Geisser (2.14,
234 77.03)=100.68, $p<0.001$, along the competence and warmth dimensions, as intended
235 ($ps<0.05$). It is worth noting that the HW-LC profile was judged as more competent than the
236 LW-LC ($p<0.001$), although both were deemed less competent than the HW-HC and LW-HC
237 ($ps<0.001$) (see Figure 2). Newman-Keuls correction was applied to all post-hoc
238 comparisons. The complete opponent profiles can be found in the Supplementary Materials.

239

240 -------- Figure 2 ---------

241

2.3.3 Cover story

An elaborate procedure was adopted to conceal the real purpose of the study. Participants were told that several experiments on different psychological phenomena were being carried out simultaneously for reasons of resource and time limitation. They were told that all these experiments had the card game in common and the fact that each participant had completed a profile with their personal information. They were also told that, because each study had different requirements, participants in each study would be assigned to different roles. Participants were told that in this particular experiment we were interested in studying how memory for personal and non-personal information was influenced by social interactions. After receiving the instructions about the card game, participants were required to answered some personal questions, on an online platform, that would constitute their own profile which would be shown to opponents. Shortly after this, they were informed that all the opponents' profiles had been completed and were finally available for the encoding phase. Participants were instructed to memorize as much information as possible about their 4 OPs in 10 minutes. During the card game, at the beginning and end of each block, multiple choice-type questions were asked about the corresponding OP's profile. Thus, for each question, participants were given an individual profile entry and were asked to identify which answer corresponded to their OP for that particular block. Immediately after the conclusion of the task (before the subjective ratings questioning, see below), an extensive memory test was administered. This whole procedure had the sole purpose of strengthening the cover story and thus no further consideration will be made in this regard.

To understand if participants were suspicious about our cover story, a verbal funnel debrief type of procedure was carried out. Thus, once participants had finished all their tasks and had been paid for their participation, the experimenter indicated that it was time to meet the other

OPs. However, before the OPs were actually introduced, the experimenter asked the following questions: “What do you think about this study and the game you just played?”, “Is there anything you want to ask or comment on before you meet the OPs?” and “Is there any OP in particular who you would or wouldn’t like to meet?”. Participants were excluded (n=4) where they spontaneously mentioned suspicion about the procedure and, in particular, about the existence of other players, e.g. said “Were there any other players?” or “I was wondering if the other players really were in other rooms.”.

2.3.4 Subjective Measures

To check how participants perceived each OP, at the end of the card game they completed a questionnaire about each OP. Specifically, they rated (on a scale ranging from 1 to 15) “How competent (e.g. intelligent, capable, efficacious) do you think he is?” (Competence); “How warm (e.g. affectionate, friendly, trustworthy) do you think he is?” (Warmth); “How much do you identify with him?” (Identification); and “How competitive do you think he his?” (Competitiveness). Participants were also asked to rate themselves along the warmth and competence scale (Self-categorization). To further explore the psychological processes underlying behavioural choices, we also asked participants to rate for each OP (on a scale ranging from one 1to 15): “How guilty do you feel about lying to him when he was not able to know your decision?” (Guilt NoRep); “How guilty do you feel about lying to him when he was able to know your decision?” (Guilt Rep); “How apprehensive do you feel about meeting him?” (Apprehension).

2.4.1 Data Analyses

Data analysis was performed with R, a free software programming language and software environment for statistical computing (R Development Core Team, 2013). We performed a multilevel mixed log-linear regression analysis with a logit link, a statistical method belonging to the family of linear mixed models (LMM or "mixed effects models"; Pinheiro & Bates, 2000; Garson, 2013), through the package lme4 ver. 0.999999-2 (Bates, 2014) (please see supplementary materials for standard analyses of variance, Table S1). We treated each decision of each participant as a separate observation, obtaining 240 observations per participant. We had one dependent variable (Lie/Truth) and four dummy predictors: Gain (0=other, 1=self); Reputation Risk (0=low risk; 1=high risk); Warmth (0=low; 1=high); Competence (0=low; 1=high). This approach allowed us to explore the independent influence of warmth and competence on lying. Moreover, participants were entered as random factors and fixed effects and interactions were also modelled as random slopes over participants (Barr, Levy, Scheepers, & Tily, 2013). This allowed us to deal with the issue of the non-independence of our dataset, i.e. the fact that we used multiple responses per subject. Multilevel mixed models are the preferred statistical method when the dependent variable is binary (Jaeger, 2008; Lindström et al., 2014) and are particularly useful in tackling multicolinearity problems in situations where observations (lie/truth behaviour) within a specific context level (participants) are highly correlated among themselves (across experimental conditions) (Baayen et al., 2008; Field et al., 2013). In other words, the model accounts for the fact that the probability of observing a lie/truth in a given experimental condition is largely dependent on the general tendency that each participant has to lie.

Reported main effects and interactions are based on model comparisons using the log-likelihood ratio statistics asymptotically approximated to a $\chi 2$ distribution. This allows the computation of a p-value that reaches statistical significance if the more complex model fits

the data better (Pinheiro & Bates, 2000). Post hoc pairwise comparisons were performed using least squares contrasts (lsc) and the Tukey correction. For each dummy regressor the low level (i.e. low warmth;low competence; No-rep risk; Other-gain was the reference category.

2.4.2. Subjective measures

To understand if participants perceived each OP as we had hypothesized, the subjective ratings of Warmth and Competence were entered in a 4 OPs (HW-HC; HW-LC; LW-HC; LW-LC) x 2 Dimension (Warmth; Competence) ANOVA[1]. The Newman-Keuls correction was applied to post-hoc comparisons. For the remaining subjective reports, dummy predictors were created for the variables Warmth (0=low; 1=high) and Competence (0=low; 1=high) and data was entered in different 2 Warmth (High Warm; Low Warm) x 2 Competence (High Competent; Low Competent) ANOVAs, each having reported values of Apprehension, Identification or Competitiveness as outcome variables. In the model that explored reported Guilt feelings, an additional dummy variable, namely Reputation (0=low risk; 1=high risk) was used to reflect feelings of Guilt under the anonymous or disclosed conditions.

3. **Results**

3.1 Behavioural results

The average numbers of lies to each OP, per condition, are depicted in Figure 3. The model that guaranteed the best interpolation with our data (see Tables S2-S3 in Supplementary Materials) was the following:

Lie/Truth=(1+Competence*Warmth*Gain*ReputationRisk|subject)+Gain+Warmth +Competence+ReputationRisk+Gain:ReputationRisk+Warmth:Gain+Competence:Gain+Warmth:ReputationRisk+Competence:ReputationRisk

The expression within parenthesis indicates the random effects defined in the model- i.e., the intercepts over participants and the slope of each main effect and interaction over participants. The expression outside the parentheses refers to the fixed effects. Both asterisks and colons between effects are used to indicate interactions. However, while asterisks are used to specify that the main effects and lower order interactions within that interaction are also included, colons are used to indicate that only the interaction is considered.

Results revealed that participants tended to lie more to obtain a gain than to cause the opponent to win ($\chi2$= 4.39, p<0.04). They were also more tempted to lie to low rather than high warmth OPs ($\chi2$=3.96, p<0.05). These main effects were qualified by the interactions Gain*Warmth interaction ($\chi2$=4.80, p<0.03) and Gain*Competence ($\chi2$=4.62, p<0.03). Post-hoc analyses revealed that participants tended to tell more self-gain lies (β=2.14, SE= 0.51, z.ratio=3.13, p<0.008) to low than high warmth OPs (Figure 4A), and more self-gain lies to high than low competent OPs (β =0.58, SE= 0.12, z.ratio= -2.59, p=0.04) (Figure 4B). Results also revealed that participants were more likely to lie during the No-Rep condition compared to the Rep condition ($\chi2$=377.06, p<0.001). The interaction Reputation*Gain was significant ($\chi2$=242.15, p<0.001) and explained by the fact that when reputation was not at risk participants told more self-gain lies (β=48.27, SE=28.87, z.ratio= 6.48, p<0.001) (Figure 4C). Interestingly, we found a significant interaction of Reputation*Warmth*Competence ($\chi2$=4.84, p=0.03), which was explained by the fact that the effect of reputation on lying was stronger for low warmth/ low competence OPs (β=14.25, SE=7.58, z.ratio=4.99, p<0.001) (Figure 4D). The remaining interactions and main effects were not significant (ps>0.05)

(please refer to Table S1 in Supplementary Materials for statistics using standard analyses of variance).

-------- Figure 3 ---------

--------Figure 4 --------

3.2 Subjective measures

A 4 OPs (HW-HC; HW-LC; LW-HC; LW-LC) x 2 Dimension (Warmth; Competence) ANOVA revealed that OPs were significantly differentiated by their perceived warmth and competence, $F(3,78)=58.71$, $p<0.001$. Both main effects of OPs $F(3,78)=10.90$, $p<0.001$, and Dimension $F(1,26)=7.38$, $p=0.012$ were significant. Post-hoc analyses revealed that, with the exception of the HW-LC that was rated as more competent than LW-LC ($p=0.004$) and equally competent as the HW-HC and LW-HC ($ps>0.52$)[2], all OPs were perceived as expected ($ps<0.05$) (see Figure 2).

A 2 (Rep; No-Rep) x 2 (High Warm; Low Warm) x 2 (High Competent; Low Competent) ANOVA on Guilt ratings revealed a main effect of Warmth ($F(1,31)=7.73$, $p=.009$), indicating that participants felt more guilt for lying to high warmth opponents; a main effect of Competence ($F(1,31)=4.73$, $p=.037$), indicating that they felt more guilt towards the low competence profiles; and a main effect of Reputation ($F(1,31)=4.56$, $p=.04$), indicating that participants felt more guilty when lying in the Reputation Risk Condition (Figure 5). No significant interaction was found. Ratings on Identification and Competitiveness and Apprehensiveness were entered into three separate 2 (High Warm; Low Warm) x 2 (High Competent; Low Competent) ANOVAs (Figure 5). For Identification we observed a

significant main effect of Warmth ($F(1,31)=9.82$, $p=0.003$). Neither the main effect of Competence ($F(1,31)=3.82$, $p=0.06$) nor the interaction ($F(1,31)=0.11$, $p=0.73$) were significant. These results suggest that Ps identified themselves more with warm Ops, regardless of their perceived competence. No main effect or interaction was significant in the Apprehensiveness model ($ps>0.72$).

-------- Figure 5---------

4. Discussion

The decision to lie inevitably entails a certain degree of conflict between the temptation to obtain some benefit and the desire to conform to moral norms or avoid aversive social consequences (Mead et al., 2009; Mazar et al., 2008). Such conflict is highly dependent on factors such as the context in which the lie is produced and the possible consequences of lying. Here, we used a paradigm of proven ecological validity in reproducing some of the moral and social pressures involved in voluntary deception behavior (Panasiti et al., 2011; 2014; 2016), in order to study how the decision to deceive someone depends on the perceived characteristics of the target. We show that in the same situational context, i.e. playing for monetary reward while having total control over both one's own and the opponent's payoff, people engage in different patterns of deceptive behaviour, depending on the perceived characteristics of the person they are interacting with. Specifically, we found that people are less likely to produce self-gain lies (which benefit themselves) when interacting with someone perceived as warm (here defined as affectionate, friendly and trustworthy) than

when playing with someone perceived as cold. Conversely, OPs who were perceived as competent (here defined as intelligent, capable, and efficacious) seem to increase the tendency to lie for egoistic purposes, as participants lied more to high than to low competent OPs. Notably however, interactions between these two dimensions were only observed when considering the risk of disclosure, suggesting that in most cases their independent effects are stronger than any interaction (Lee and Harris, 2014). In other words, lie behaviour in the different conditions can be better explained by independent evaluations along these two dimensions rather than by specific attitudes towards individual OPs. Finally, contrary to our hypothesis, neither warmth nor competence modulated the tendency to lie for altruistic purposes. Altogether, our results show that the decision to deceive depends not only on the evaluation of the possible benefits and costs that such behaviour brings to oneself but also on the costs and benefits that it carries to others (Gneezy, 2005). Furthermore, our study suggests that the personal moral standards that help to resist the temptation to deceive others are not stable but are rather malleable according to the context and the target of deception.

Previous research has shown that the decision to lie, and its neural and autonomic correlates, in this paradigm are strongly associated with both dispositional factors (e.g. personality traits like levels of manipulativeness or moral disengagement) as well as situational factors, such as the anonymity of the deceptive behaviour (Panasiti et al., 2011; 2014; 2016). Consistent with this finding, we argue that the observed modulation of behaviour according to an OP's characteristics is a reflection of moral evaluations rather than the product of competitive motives.

Even though they are playing for the same resources (i.e. the payoff), the game's rules clearly imply that no real competition is possible, as Ps always decide who is going to get the reward at the end of each trial. Because OPs are deprived of the power to decide the outcome, self-

gain lies are not likely to be the result of an open competition for the pay-off, nor of explicit defensive behaviour against possible attacks from the OP. Instead, self-gain are likely to reflect the triumph of temptation over fairness. Like in other socio-economic decision-making paradigms, such as the Dictator Game (Hoffman, McCabe, and Shachat 1994, Hibing et al., 2014) or the Ultimatum Game (Guth, Schmittberger, and Schwarze 1982; Lenggenhager et al. 2013; Mancini et al. 2011), in which the participant decides how to split the payoff, behaviour is guided by personal fairness and values of equity. Crucially however, in our paradigm, violations of fairness imply engagement in deceptive behaviour, i.e. choosing to provide a false statement. This makes the behaviour in our game different from the other economic games for at least two reasons: i) because by providing a false statement participants play unfairly but at same time protect their own reputation (the other players will think that they had bad luck); ii) because attempting to manipulate the beliefs of others (lying) is publicly condemned in our society (Nyberg, 1993). Importantly, we show here, for the first time in a controlled and morally relevant setting, how social factors, such as others' inferred traits, modulate lying behaviour.

As expected, participants made fewer egoistic lies to high than low warmth individuals, supporting the idea that moral information is of primary importance in guiding social judgments and interpersonal behaviour (Wojciszke et al., 1998; Brambrilla et al., 2011; Willis & Todorov, 2006; Delgado et al., 2005; Liuzza et al., 2014; Mancini et al., 2014; Panasiti et al., 2015). These results expand previous evidence that has demonstrated that people tend to trust (Delgado et al., 2005) and cooperate (Steinel and De Dreu, 2004) more with warm individuals, by showing that inferred warmth reduces the tendency to deceive others. These behavioural tendencies show how social evaluations, and in particular inferred traits of the

453 other's moral worth, impact on one's own codes of action, as well as the relative weight
454 given to dishonest behaviour.

455 We note that the uni-directional nature of our paradigm ensures that behaviour is not guided
456 by expectations of cooperation. Instead, cooperation-related evaluations may be relevant, to
457 the extent that they are a phylogenetically important factor in shaping social interactions and
458 attitudes towards others. According to several theoretical and evolutionary accounts, pro-
459 social and moral values are at the core of competition-cooperation behaviour (e.g. Van
460 Lange, 1999; Ohtsuki and Iwasa, 2004; Parks et al., 2013; Nowak, 2006). People with high
461 pro-social values not only are more likely to cooperate than pro-self individuals as they
462 expect others to cooperate as well, because they have greater feelings of social responsibility
463 and equality concerns (Van Lange & Kuhlman, 1994; De Cremer and Van Lange, 2001;
464 Steinel and De Dreu, 2004). Along similar lines, indirect reciprocity mechanisms posit that
465 helping others builds up a good reputation, which in turn makes pro-social people more liable
466 to be rewarded by others (Trivers, 1971; Wedekind and Milinski, 2000; DeSteno et al., 2010).
467 This explains why helping others or behaving fairly to them pays off, even when the others
468 cannot directly reciprocate. Crucially in the present paradigm, reciprocity computations are
469 unlikely as participants know that the roles will never be reversed and that it is impossible for
470 OPs to reward P's fairness/generosity or punish P's unfair behaviour, either directly or
471 indirectly. Our results thus suggest that such expectations about social conduct are embedded
472 in one's own codes of action and influence behaviour, even in situations where no one can act
473 upon them. Because high-warmth individuals are perceived to be pro-social and low warmth
474 others to be 'pro-self', the psychological cost of behaving dishonestly with trustworthy others
475 is higher than with cold individuals. Moreover, the fact that participants identified themselves
476 more with trustworthy opponents can also help to explain this behavioural pattern. When

coming from someone perceived to be similar to oneself (Mussweilera and Ockenfelsb, 2013), unfair behaviour is judged to be particularly unfair and increases the willingness to punish. Indeed, our participants reported increased guilt feelings for lying to high warm OPs, confirming that the moral costs associated with deception may be particularly high when interacting with these individuals.

An opposite pattern of results was found for the competence dimension. Participants lied less for egoistic reasons to OPs judged not to be competent, i.e. less capable and/or with lower social status. Even if in scenarios in which cooperation is possible people tend to cooperate less with non-competent others (Van Lange and Kuhlman, 1994), here participants were less inclined to engage in dishonest behaviour with these individuals. This is consistent with the reported heightened sense of social responsibility when interacting with individuals perceived to be particularly powerless (Handgraaf et al., 2008, Gino and Pierce, 2009). Because competence does not predict other's intentions nor their moral traits, it is unlikely that this behaviour is explained by trustworthiness evaluations or expectations of the OP's behavioural tendencies, e.g. "What would he do in my place?". Thus, it seems plausible that social responsibility, and possibly feelings of pity. may have encouraged participants to refrain from lying for self-gain towards low competent OPs. The reported increased feelings of guilt for lying to these individuals seem to support such interpretation. Interestingly however, while competence and warmth dimensions had opposite effects, they did not interact, showing that although they were assessed together they may have independent effects (Lee and Harris, 2014).

The present study also confirmed our previous results (Panasiti et al., 2011; 2014) by showing a reduction in lies for self-gain when participants' reputation was at risk. Importantly, in contrast with previous studies where the altruistic lies were not modulated by

the experimental manipulations (i.e., by the reputation risk or the physical presence of the OP in the room), here we found that these responses were shaped by reputation and social evaluations. That participants had information about the targets of their deception might have made the reputation manipulation stronger than in our previous studies. Interestingly, we found that the reputation factor interacted with perceived competence and warmth such that participants told more lies to low competence/low warmth OPs when their own behaviour was anonymous, compared to when there was a risk of disclosure. Typically, individuals with this profile, (i.e. perceived to be low in both warmth and competence dimensions) are judged particularly negatively and elicit feelings of contempt and disdain. It is therefore not surprising that they were associated with distinct behavioural patterns. The fact that reputational effects were particularly evident when interacting with these OPs suggests a strong dissociation between the personal and social costs of lying to them. Indeed, compared to the other OPs, participants exhibited equivalent behaviour when there was risk of social judgment but they lied more often to them when behaviour was anonymous and where therefore the participant was the only judge of his/her own actions. This suggests that honest/dishonest interactions with these individuals are less related to one's own personal moral standards and more to the desire to conform to social norms and reputation concerns.

Finally, previous research with this paradigm has shown that, even if expectations of direct reciprocity are not possible, participants will also lie to benefit the other person (Panasiti et al., 2011; 2014; 2016). These altruistic lies, although objectively lies, have remarkably different social and moral connotations, as dishonest behaviour is considered more acceptable when benefiting someone else (Gino et al., 2013; Eta and Gneezy, 2012). In fact because participants are sacrificing their own pay-off, this behaviour may be seen as an act of generosity. Our prediction was that perceiving others as warm would have encouraged active

helping behaviour, such as engaging in other-gain lies. However, warmth and competence did not modulate the tendency to lie altruistically. This indicates that information about the target influences egoistic and altruistic behaviour in different ways, depending on whether this choice implies deception or not. Specifically, other's inferred personality traits modulate the tendency to lie for self-gain but not altruistic lying, suggesting that it is more important for people to be honest than to be altruistic, if the latter this implies deception.

Together, our findings show that the decision to engage in deceptive behaviour involves complex evaluations, based on personal, social and situational factors. Because people need to act according to their own moral standards in order to preserve their positive self-concept as honest individuals (Mazar et al., 2008; Gino and Pierce, 2009), they evaluate their own behaviour in a malleable way. Even where the objective consequences of the action are identical, the psychological cost associated with lying varies according to the social context and interpersonal factors. All in all, the present study constitutes an important advance in the understanding of the moderators of spontaneous deception, by showing that biases in deceptive behaviour largely reflect internal motivations driven by subjective judgments of others' socio-moral worth. Whereas previous research has shown that levels of one's own moral commitment predict lying behaviour when interacting with a total stranger, we observed here that personal moral standards are malleable and may acquire a greater weight when the target of deception is a high warmth or low competence person.

Acknowledgements

We thank Vivien Ainley for proofreading andher helpful comments. Financial support: Fundação para a Ciência e a Tecnologia (FCT) and POPH/FSE (doctoral grant SFRH/BD/60517/2009) to RTA. Fondazione Santa Lucia, IRRCS and Sapienza University of Rome (MSP and SMA).

References

Azevedo, R.T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S.M. (2013). Their pain is not our pain: Brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping*, 34(12), 3168-81.

Azevedo, R.T, Macaluso, E., Viola, V., Sani, G. & Algioti, SM (2014). Weighing the stigma of weight: an fMRI study of neural reactivity to the pain of obese individuals. *NeuroImage*, 1, 91,109-19.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Bates, D., Maechler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version.

Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K. & Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron*, 64, 756–70.

Becker, G.S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76 (2), 169-217.

Brambilla, M., Rusconi, P., Sacchi, S. & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, 41(2), 135-143.

572 Cacioppo, J.T., Gardner, W.L. & Berntson, G.G. (1997). Beyond bipolar conceptualizations
573 and measures: the case of attitudes and evaluative space. *Personality and Social Psychology*
574 *Review*, 1, 3–25.

575 Cikara, M. & Fiske, S.T. (2011). Bounded empathy: neural responses to outgroup targets'
576 (mis)fortunes. *Journal of Cognitive Neuroscience*, 23(12), 3791-803.

577 Cuddy, A.J.C., Fiske, S.T. & Glick, P. (2008). Warmth and Competence as Universal
578 Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map.
579 *Advances in Experimental Social Psychology*, 40, 61–149.

580 Cunningham, W.A., Johnson, M.K., Raye, C.L., Gatenby, J.C., Gore, J.C. & Banaji, M.R.
581 (2004). Separable neural components in the processing of black and white faces.
582 *Psychological Science*, 15, 806–813.

583 deCremer, D., & Van Lange, P. A. M. (2001). Why prosocials exhibit greater cooperation
584 than proselfs: The roles of responsibility and reciprocity. *European Journal of Personality*,
585 15, 5-18.

586 Delgado, M.R., Frank, R.H. &, Phelps, E.A. (2005). Perceptions of moral character modulate
587 the neural systems of reward during the trust game. *Nature Neuroscience,* 8(11):1611-8

588 DePaulo, B. M., & Kashy, D. a. (1998). Everyday lies in close and casual relationships.
589 *Journal of Personality and Social Psychology*, 74(1), 63–79.

590 DeSteno, D., Bartlett, M.Y., Baumann, J., Williams, L.A. & Dickens, L. (2010). Gratitude as
591 moral sentiment: emotion-guided cooperation in economic exchange. *Emotion*, 10(2), 289-
592 93.

593 Durante, F., Fiske, S.T., Kervyn, N. & al. (2012). Nations' income inequality predicts
594 ambivalence in stereotype content: How societies mind the gap. *British Journal of Social*
595 *Psychology*, 52(4),726-46 .

596 Ennis, E., Vrij, a., & Chance, C. (2008). Individual differences and lying in everyday life.
597 *Journal of Social and Personal Relationships*, 25(1), 105–118.

598 Farmer, H., McKay, R. & Tsakiris, M. (2014). Trust in me: trustworthy others are seen as
599 more physically similar to the self. *Psychological Science*, 25(1),290-2.

600 Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical
601 power analysis program for the social, behavioral, and biomedical sciences. *Behavior*
602 *Research Methods*, 39, 175-191.

603 Field, A. (2013). Discovering statistics using IBM SPSS statistics. London: Sage.

604 Fiske, S.T., Cuddy, A.J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype
605 content - competence and warmth respectively follow from perceived status and competition.
606 *Journal of Personality and Social Psychology*, 82(6),878-902.

607 Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Thousand Oaks, CA:
608 Sage.

609 Garson, G.D. (2013). Fundamentals of Hierarchical Linear and Multilevel Modeling. In:
610 Garson GD (ed) Hierarchical Linear Modeling. Guide and Applications. Sage Publications.

611 Gino, F., Schweitzer, M.E., Mead, N.L., & Ariely, D. (2011). Unable to resist temptation:
612 How self-control depletion promotes unethical behavior. *Organizational Behavior and*
613 *Human Decision Processes*, 115 (2), 191–203.

614 Gino F., & Pierce L. (2009). Dishonesty in the name of equity. *Psychological Science*, 20(9),
615 1153-60.

616 Gino F, Ayal S, & Ariely D. (2013). Self-Serving Altruism? The Lure of Unethical Actions
617 that Benefit Others. *Journal of Economic Behavior & Organization*, 1, 93.

618 Gneezy, U. (2005). Deception: The Role of Consequences. *American Economic Review*,
619 95(1), 384-94.

620 Greene, J.D. & Paxton, J.M. (2009). Patterns of neural activity associated with honest and
621 dishonest moral decisions. *Proceedings of the National Academy of Sciences of USA*, 106,
622 12506–12511.

623 Golby, A.J., Gabrieli, J.D., Chiao, J.Y. & Eberhardt, J.L. (2001). Differential fusiform
624 responses to same- and other-race faces. *Nature Neuroscience*, 4, 845–850.

625 Halali, E., Bereby-Meyer, Y., & Meiran, N. (2013). Between Self-Interest and Reciprocity:
626 The Social Bright Side of Self-Control Failure. *Journal of Experimental Psychology:*
627 *General*, 143(2),745-54.

628 Handgraaf, M. J. J., van Dijk, E., Vermunt, R. C., Wilke, H. A. M., & De Dreu, C. K. W.
629 (2008). Less power or powerless? Egocentric empathy gaps and the irony of having little
630 versus no power in social decision making. *Journal of Personality and Social Psychology*, 95,
631 1136–1149.

632 Harris, L.T. & Fiske, S.T. (2006). Dehumanizing the lowest of the low: neuro-imaging
633 responses to extreme outgroups. *Psychological Science*, 17, 847–853.

Hilbig, B.E., Thielmann, I., Hepp, J., Klein, S.A. & Zettlerd, I. (2015). From personality to altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of Research in Personality*, 55, 46–50.

Hoffman, E, K McCabe, and K Shachat. 1994. "Smith, V. Prefer- Ences, Property Rights and Anonymity in Bargaining Games." *Games and Economic Behavior*, (7): 346.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434-446.

Lee, V.K. & Harris, L.T. (2014). Sticking with the nice guy: trait warmth information impairs learning and modulates person perception brain network activity. *Cogn Affect Behav Neurosci*, 14(4):1420-37.

Lehmann, E. L. (1986). Testing statistical hypotheses. New York: John Wiley & Sons, Inc.

Lenggenhager, B., Azevedo, R.T., Mancini, A., & Aglioti S.M. (2013). Listening to your heart and feeling yourself: effects of exposure to interoceptive signals during the ultimatum game. *Experimental Brain Research*, 230(2), 233-4.

Lindström, B., Selbing, I., Molapour, T. & Olsson, A. (2015). Racial Bias Shapes Social Reinforcement Learning. *Psychological Science*, 25(3) 711–719.

Liuzza, M.T., Candidi, M., Sforza, A.L., & Aglioti, S.M. (2014). Harm avoiders suppress motor resonance to observed immoral actions. *Social Cognitive and Affective Neuroscience*, 10(1),72-7.

Mancini, A., Betti, V., Panasiti, M. S., Pavone, E. F., & Aglioti, S. M. (2011). Suffering makes you egoist: Acute pain increases acceptance rates and reduces fairness during a bilateral ultimatum game. *PLoS ONE*, *6*(10).

Mancini, A., Betti V., Panasiti, M.S., Pavone, E.F. & Aglioti, S.M. (2014). Perceiving monetary loss as due to inequity reduces behavioral and cortical responses to pain. *European Journal of Neuroscience*, 40(2),2378-88.

Masip, J., Garrido,E., Herrero, C. (2004) "Defining deception" *Anales de Psicología*, 20, 147–171

Mastroberardino, S. (2012). Psicologia della Menzogna. Carocci Editore. Roma.

Mazar, N., Amir, O. & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45,633–644.

Mead, N.L., Baumeister, R.F., Gino, F., Schweitzer, M.E. & Ariely, D. (2009). Too Tired to Tell the Truth: Self-Control Resource Depletion and Dishonesty. *Journal of Experimental Social Psychology*, 45, 594–597.

Miller, S.L.; Maner, J.K. & Becker, D. V. (2010). Self-protective biases in group categorization: Threat cues shape the psychological boundary between "us" and "them". *Journal of Personality and Social Psychology*, 99(1), 2010, 62-77.

Moretti, L., & di Pellegrino, G. (2010). Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion*, 10(2),169-80.

Mussweiler, T. & Ockenfels, A. (2013). Similarity increases altruistic punishment in humans. *Proc Natl Acad Sci U S A*. 110(48):19318-23.

676 Navarrete, C.D., McDonald, M. M., Asher, B.D., Kerr, N.L., Yokota, K., Olsson, A., &
677 Sidanius, J.(2012). Fear is readily associated with an out-group face in a minimal group
678 context. *Evolution and Human Behavior*, 33, 590–593

679 Nowak, M. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–3.

680 Ohtsuki, H. & Iwasa, Y. (2004). How should we define goodness?—reputation dynamics in
681 indirect reciprocity. *Journal of Theoretical Biology*, 231(1),107–120.

682 Olekalns, M., & Smith, P. L. (2009). Mutually dependent: Power, trust, affect and the use of
683 deception in negotiation. *Journal of Business Ethics*, 85(3), 347–365.

684 Olsson, A., Ebert, J.P., Banaji M.R., & Phelps E.A. (2005). The role of social groups in the
685 persistence of learned fear. *Science*. 309(5735), 785-787.

686 Panasiti, M.S., Pavone, E.F., Merla, A. & Aglioti, S.M. (2011). Situational and Dispositional
687 Determinants of Intentional Deceiving. *PloS one*, 6 (4), e19465.

688 Panasiti, M.S., Pavone, E.F., Mancini, A., Merla, A., Grisoni, L. & Aglioti, S.M. (2014). The
689 motor cost of telling lies: electrocortical signatures and personality foundations of
690 spontaneous deception. *Social Neuroscience*, 9 (6), 573-89.

691 .

692 Panasiti, M. S., Puzzo, I., and Chakrabarti, B. (2015). Autistic Traits Moderate the Impact of
693 Reward Learning on Social Behaviour. *Autism Res*., 9: 471–479.

694 Panasiti, M.S., Cardone, D., Pavone, E.F., Mancini, A., Merla, A. & Aglioti, S.M. (2016).
695 Thermal signatures of voluntary deception in ecological conditions. Scientific Reports. 6:
696 35174.

697 Parks, C.D., Joireman, J. & Van Lange, P.A.M. (2013). Cooperation, Trust, and Antagonism:
698 How Public Goods Are Promoted. *Psychological Science in the Public Interest*, 14(3), 119-
699 165.

700 Pinheiro, J.C. & Bates, D.M. (2000). Mixed-Effects Models in S and S-PLUS. New. Springer
701 New York.

702 Ponsi, G., Panasiti, M.S., Scandola, M. & Aglioti, S.M. (2015). Influence of warmth and
703 competence on the promotion of safe in-group selection: Stereotype content model and social
704 categorization of faces. *Quarterly Journal of Experimental Psychology (Hove)*, 27,1-16.

705 Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics.
706 Dordrecht, The Netherlands: Reidel.

707 Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461–
708 464.

709 Sip, K.E., Roepstorff, A., McGregor, W. & Frith, C.D. (2008). Detecting deception: The
710 scope and limits. *Trends in Cognitive Sciences*, 12,48–53.

711 Stanley, D.A., Sokol-Hessner, P., Mahzarin, R., Banaji, E. & Phelps, A. (2011). Implicit race
712 attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the*
713 *National Academy of Sciences of USA*, 108(19), 7710-7715.

714 Steinel, W. & De Dreu, C.K.W. (2004). Social motives and strategic misrepresentation in
715 social decision making. *Journal of Personality and Social Psychology*, 86(3), 419-434.

716 Tabibnia, G., Satpute A.B., & Lieberman, M.D. (2008). The sunny side of fairness:
717 preference for fairness activates reward circuitry (and disregarding unfairness activates self-
718 control circuitry*). Psychological Science*, 19(4), 339-47.

719 Tajfel, H. (1982). Social Psychology of Intergroup Relations. *Annual Review of Psychology*,
720 33, 1-39.

721 Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior:
722 Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*,
723 42(1), 69–77.

724 Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46,
725 35–57.

726 Van Bavel, J. J. & Cunningham, W.A. (2011). A social neuroscience approach to self and
727 social categorisation - A new look at an old issue. *European Review of Social Psychology*, 21,
728 237–284.

729 Van Dijk, E., & Vermunt, R. (2000). Strategy and fairness in social decision making:
730 Sometimes it pays to be powerless. *Journal of Experimental Social Psychology*, 36, 1–25.

731 Van Lange, P.A.M., & Kuhlman, D.M. (1994). Social value orientations and impressions of
732 partner's honesty and intelligence: A test of the might versus morality effect. *Journal of*
733 *Personality and Social Psychology*, 67(1), 126-141.

734 Van Lange, P.A.M. (1999). The pursuit of joint outcomes and equality in outcomes:An
735 integrative model of social value orientation. *Journal of Personality and Social Psychology*,
736 77, 337–349.

737 Van Swol, L. M., Malhotra, D., & Braun, M. T. (2012). Deception and Its Detection: Effects
738 of Monetary Incentives and Personal Relationship History. *Communication Research*, 39(2),
739 217–238.

740 Wedekind, C. & Milinski, M. (2000). Cooperation Through Image Scoring in Humans.
741 *Science*, 288, 850-852.

742 Willis, J. & Todorov, A. (2006). First impressions: making up you mind after a 100-ms
743 exposure to a face. *Psychological Science*, 17,592–598.

744 Wojciszke, B., Bazinska, R. & Jaworski, M. (1998). On the dominance of moral categories in
745 impression formation. *Personality and Social Psychology Bulletin*, 24, 1245–1257.

746

747

Footnotes

1 - Note that in this case our purpose was to evaluate each opponent profile in relation to the others, therefore standard analyses of variance are appropriate.

2 - In contrast with the pilot study, Ps of the experimental session rated the HW-LC OP relatively high in competence. We believe such a discrepancy might be accounted for by a compassion-related effect that was enhanced by the (online) interaction with this individual profile, compared to the arguably detached ratings provided in the pilot study. Because in the experimental session the task involved a morally relevant interaction with this individual profile, pity and compassion feelings might have been exacerbated. This might have led to increased moral-based evaluations and motivation to provide positive ratings. While we cannot be certain of the reasons that drove this unexpectedly high ratings of competence, the fact that this individual opponent profile possess an undesirable characteristic previously identified by the SCM as a HW-LC characteristic (Harris and Fiske, 2006), and that Ps reported increased guilt feelings for lying to warm OPs, provides convincing evidence that this profile was associated with the predicted stereotypes and elicited the predicted feelings.

Figure captions

Figure 1. Schematic representation of the experimental procedure. The timeline of the various phases of the trial is provided in the furthest right part of the figure. Note that in the trial depicted here the P would be playing against the OP named “Simone” in a No-Rep trial, as shown by the closed eye symbol.

769 *Figure 2.* Mean values (s.d.) of the categorization of each OP and self-categorization according to
770 perceived warmth (y-axis) and competence (x-axis). Data from both studies (experimental session
771 and pilot study) was normalized to the same scale (1-10) for display purposes.

772

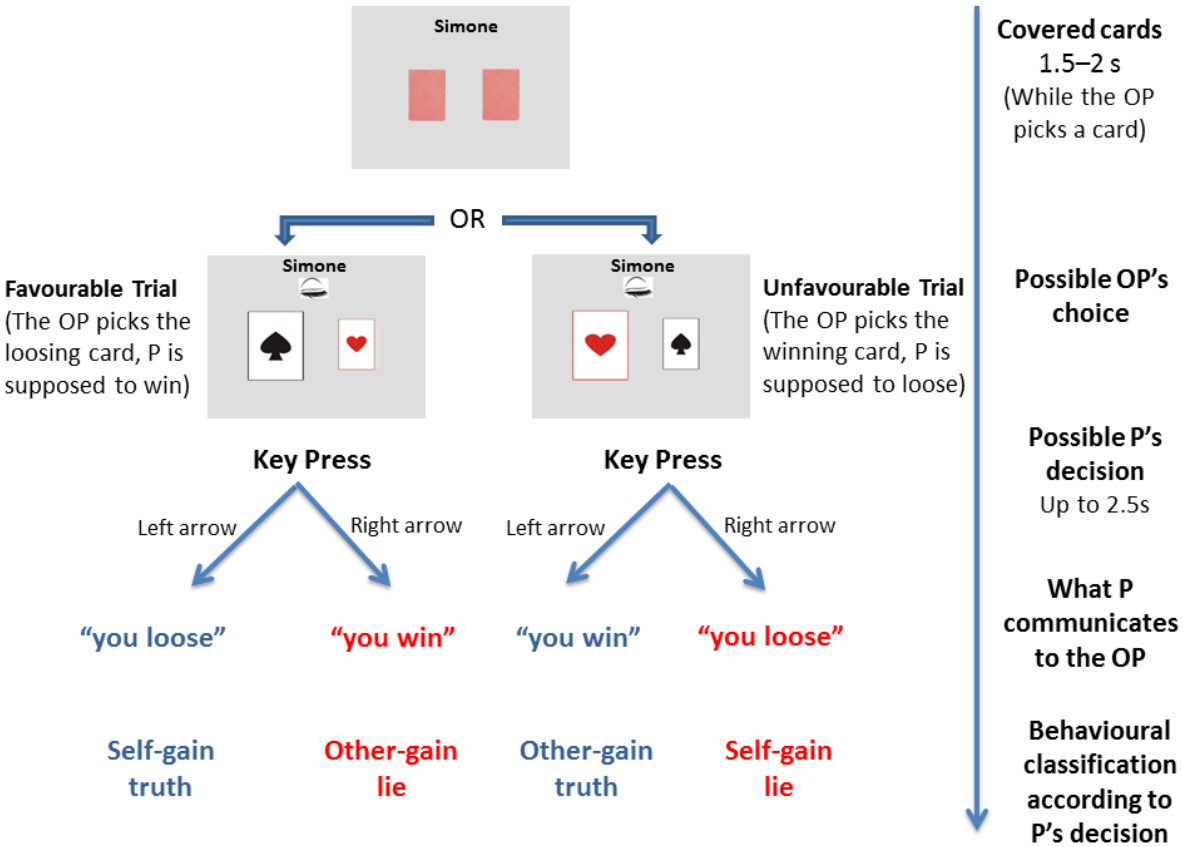
773 *Figure 3.* Average number of lies per condition to each opponent. Error bars indicate standard
774 errors.

*Figure **4***. Behavioural results in the online card game**.** A) Predicted probability for self-gain (blue) and other-gain/altruistic (red) lies when playing against high vs. low warmth opponents. B) Predicted probability for self-gain (blue) and other-gain (red) lies when playing against high vs. low competence opponent. C) Predicted probability for self-gain (blue) and other-gain (red) lies when playing during the reputation risk vs. the no-reputation risk condition. D) Predicted probability of lying to low competent (purple line) or high competent (green line) opponent when playing during the reputation risk vs. the no reputation risk condition**.** Shaded bands represent the 95% confidence intervals.

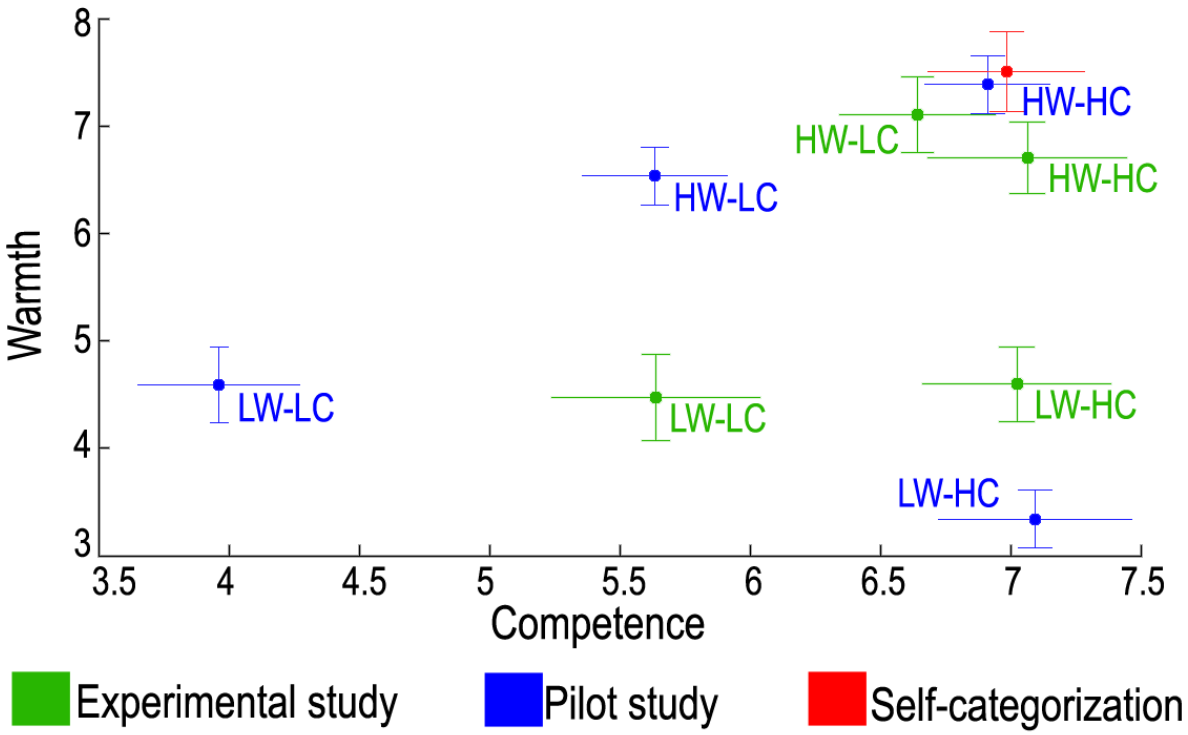
Figure 5. Subjective measures concerning the social categorization process. A) Higher identification is predicted for high warmth opponent ; B) Higher sense of guilt is predicted for high warmth and C) how competence opponent; Higher competitiveness is predicted by D) low warmth opponent and E) high competence opponent. Error bars indicate standard errors.

Figures

Figure 1

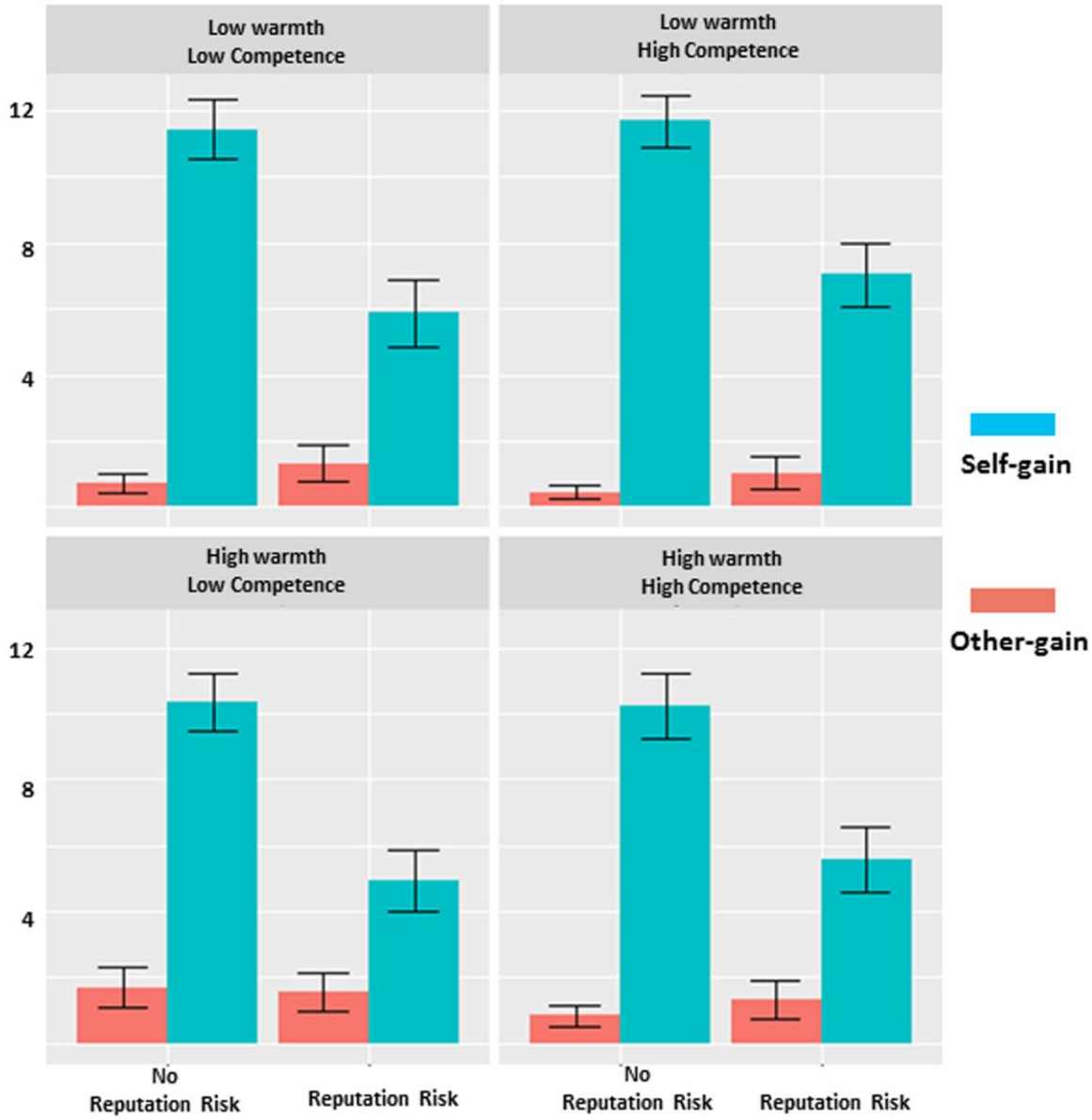


794 Figure 2



795

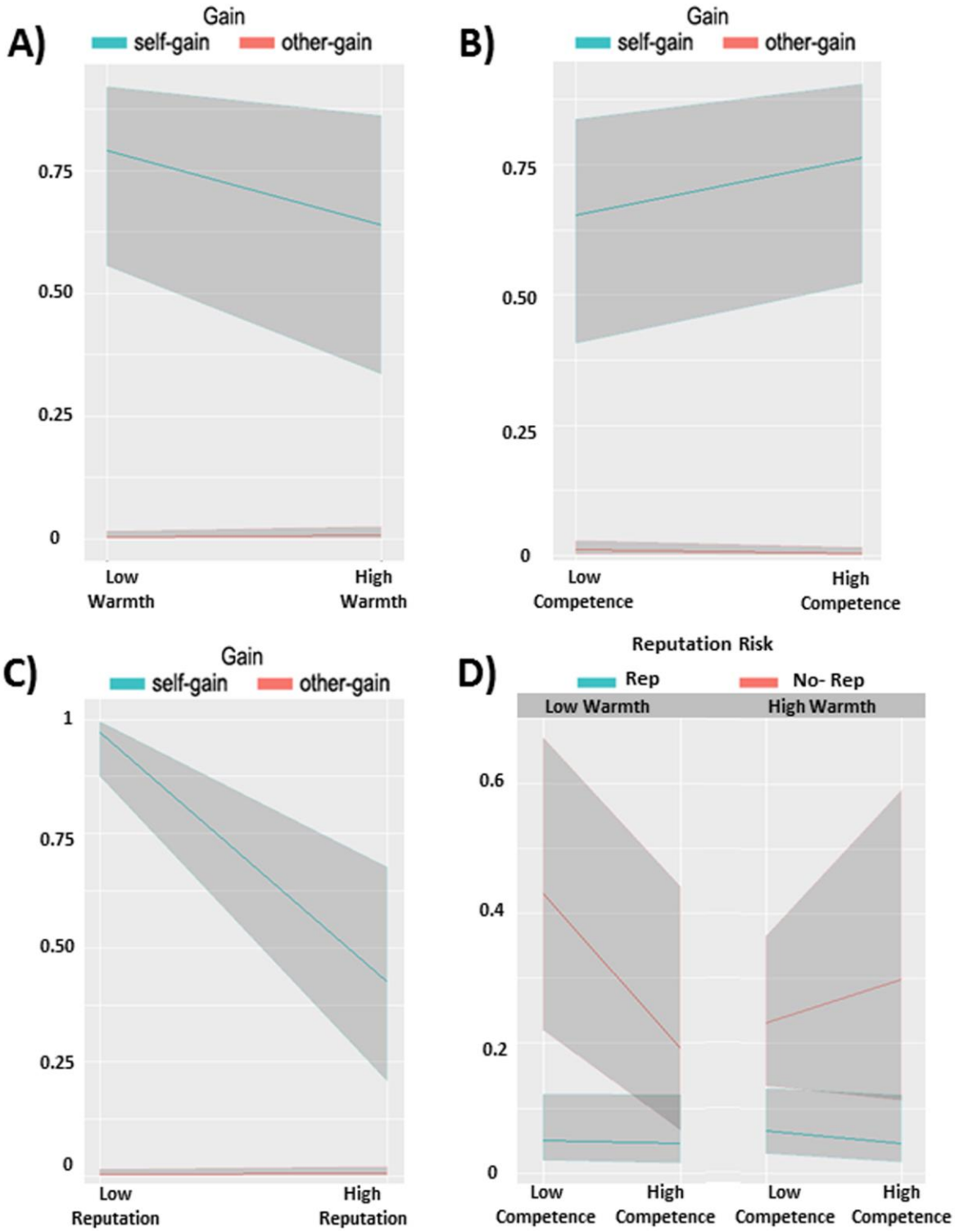
796 Figure3



797

798

799 Figure 4

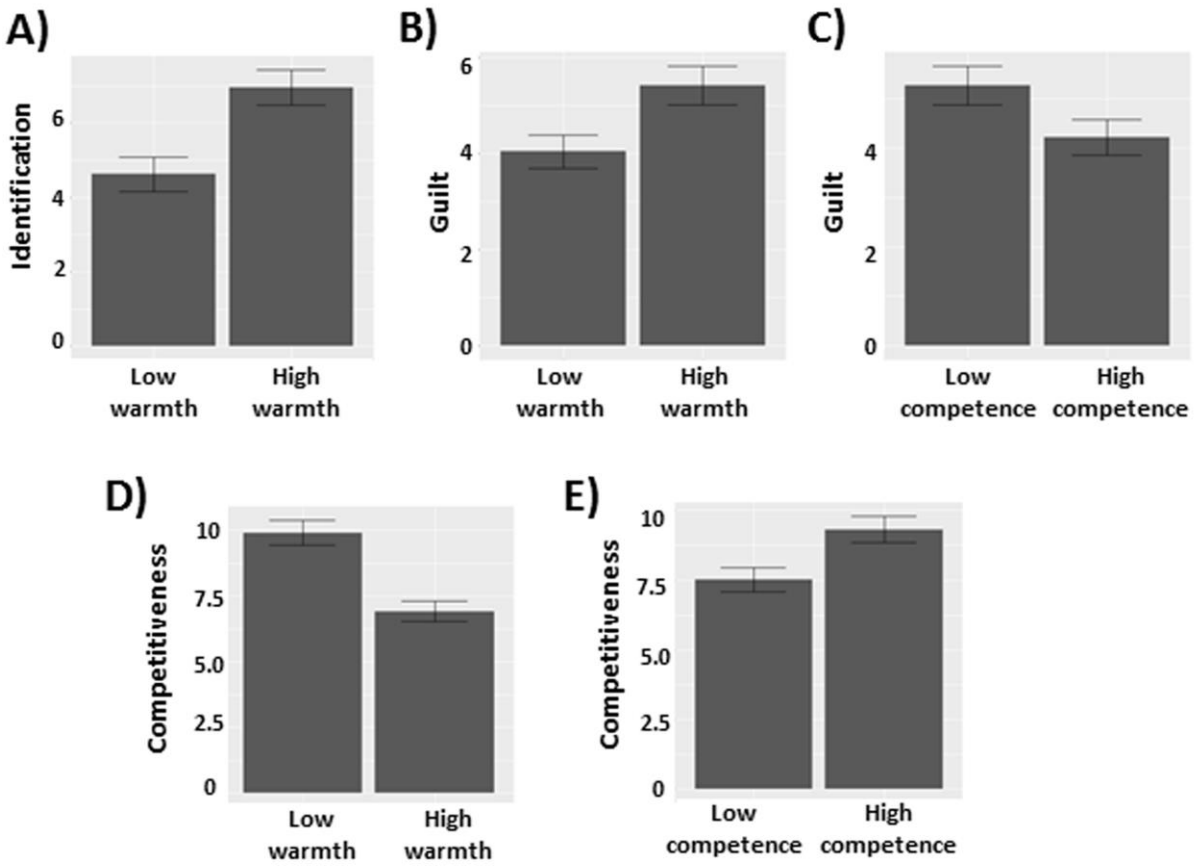


800

801

802

803 Figure 5



804