Evolution and Human Behavior xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Evolution and Human Behavior

journal homepage: www.elsevier.com/locate/ens

Social observation increases deontological judgments in moral dilemmas

Minwoo Lee[a,1], Sunhae Sul[b], Hackjin Kim[a,*]

[a] Department of Psychology, Korea University, Seoul, Republic of Korea
[b] Department of Psychology, Pusan National University, Busan, Republic of Korea

ARTICLE INFO

Keywords:
Reputation concern
Moral dilemma
Social observation
Deontology
Warmth

ABSTRACT

A concern for positive reputation is one of the core motivations underlying various social behaviors in humans. The present study investigated how experimentally induced reputation concern modulates judgments in moral dilemmas. In a mixed-design experiment, participants were randomly assigned to the observed vs. the control group and responded to a series of trolley-type moral dilemmas either in the presence or absence of observers, respectively. While no significant baseline difference in personality traits and moral decision tendency were found across two groups of participants, our analyses revealed that social observation promoted deontological judgments especially for moral dilemmas involving direct bodily harm (i.e., personal moral dilemmas), yet with an overall decrease in decision confidence and significant prolongation of reaction time. Moreover, participants in the observed group, but not in the control group, showed the increased sensitivities towards warmth vs. competence traits words in the lexical decision task performed after the moral dilemma task. Our findings suggest that reputation concern, once triggered by the presence of potentially judgmental others, could activate a culturally dominant norm of warmth in various social contexts. This could, in turn, induce a series of goal-directed processes for self-presentation of warmth, leading to increased deontological judgments in moral dilemmas. The results of the present study provide insights into the reputational consequences of moral decisions that merit further exploration.

1. Introduction

1.1. The trolley problem: various determinants of judgments in moral dilemmas

The study of moral dilemmas has been one of the most fruitful venues for investigating the cognitive and motivational structure of morality. Markedly, a particular strand of research based on the "trolley problem (Thomson, 1976)" has garnered much interest from empirical minds across disciplines. In a typical trolley-type dilemma, people are asked to respond to a series of vignettes that pit a Kantian imperative against the consequential benefits expected from its violation (Greene, Sommerville, Nystrom, Darley, & Cohen, 2001) (e.g., "Is it appropriate to pull the lever and sacrifice a person to save five people from a runaway trolley?"). Coupled with recent advances in neuroscientific methodologies (e.g., Functional magnetic resonance imaging; fMRI), this now-famous portrayal of deontological and utilitarian schools of thought has become a widely-cited reference for the view that moral decision making is a dynamic process involving both affective and cognitive valuations in the brain (Greene et al., 2001; Hutcherson, Montaser-Kouhsari, Woodward, & Rangel, 2015).

The growing recognition of morality as a multifaceted construct stands in contrast to the conventional approaches that largely focused on the role of conscious reasoning in moral judgments (Haidt, 2008). While most rationality-centered models in philosophy and psychology postulate the universal norms or developmentally canalized moral rules as key anchors for moral judgments across individuals and social contexts (Piaget, 1948), a series of empirical investigations utilizing trolley-type dilemmas has revealed rather the opposite picture. For example, researchers have shown that people's characteristic decisions in moral dilemmas are not solely driven by the rigid implementation of moral principles. Instead, they are affected by seemingly-extraneous variables such as gender (Friesdorf, Conway, & Gawronski, 2015), personality (Bartels & Pizarro, 2011; Gleichgerrcht & Young, 2013), cultural background (Han, Glover, & Jeong, 2014), and genetic makeups of individuals (Bernhard et al., 2016). Evidence also suggests that even transient changes in mood (Valdesolo & DeSteno, 2006), cognitive state (Amit & Greene, 2012; Greene, Morelli, Lowenberg, Nystrom, & Cohen, 2008), and hormonal balance (Crockett, Clark, Hauser, & Robbins, 2010) could alter the responses to the identical set of moral vignettes.

* Corresponding author at: Department of Psychology, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 136-701, Republic of Korea.
E-mail address: hackjinkim@korea.ac.kr (H. Kim).
[1] Present address: Department of Anthropology, Emory University, 201 Dowman Drive, Atlanta, GA, 30322.

https://doi.org/10.1016/j.evolhumbehav.2018.06.004
Received 26 December 2017; Received in revised form 2 June 2018; Accepted 9 June 2018

These findings point to the malleable nature of moral cognition and call our attention to the factors that influence the way decision values are computed for deontological vs. utilitarian choices (Christensen & Gomila, 2012).

1.2. Reputation concern and judgments in moral dilemmas

One potentially important source of variability in moral judgment is the concern for reputation. People care a great deal about how their behaviors would be perceived by others (Milinski, 2016). Deeply rooted in our evolved social psychology, the sensitivity towards reputation is known to modulate a wide range of behaviors in ways conducive to one's positive image scoring (Fehr & Fischbacher, 2003; Nowak & Sigmund, 1998). That is, people can deviate from their internal decision criteria to obtain favorable evaluations from others in a given sociocultural context (Izuma, 2012). The best illustration of this process, often referred to as self-presentation or impression management (Izuma, 2012; Leary, 1983), comes from previous studies where participants displayed increased prosocial tendencies under conditions of low anonymity. For instance, the presence of third-party observers is known to facilitate charitable donations (Andreoni & Petrie, 2004), altruistic norm enforcement (Kurzban, DeScioli, & O'Brien, 2007), and cooperative economic decisions (Bateson, Nettle, & Roberts, 2006; Egas & Riedl, 2008).

Of critical relevance to the proposed link between reputation concern and decisions in moral dilemmas, a developing body of works indicates that deontological and utilitarian judgments could lead to opposing social impressions (Everett, Pizarro, & Crockett, 2016; Lee, Sul, & Kim, 2014; Rom, Weiss, & Conway, 2017; Sacco, Brown, Lustgraaf, & Hugenberg, 2017; Uhlmann, Zhu, & Tannenbaum, 2013). For example, people associate warmth- (e.g., friendly, sociable, and trustworthy etc.), and competence-related personality traits (e.g., rational, reasonable, and competent etc.) with others who make deontological and utilitarian choices in moral dilemmas, respectively (Lee et al., 2014; Rom et al., 2017). Moreover, participants in these studies largely favored deontologists over utilitarian individuals regardless of their own moral orientations (Lee et al., 2014). Several studies have so far replicated and extended this result in different contexts, with various prosocial attributes including trustworthiness, empathy and harm aversion preferentially ascribed to those who make deontological judgments (Everett et al., 2016; Sacco et al., 2017).

Considered together with the effects of reputation concern in promoting socially desirable behaviors that lead to positive impressions, these findings suggest an intriguing possibility that judgments in moral dilemmas may also be influenced by reputational consequences of deontological vs. utilitarian decisions. In other words, people's decisions whether or not to pull the lever in the trolley problem could involve consideration of how those behaviors would affect the impressions that others have of them. Although neither deontology nor utilitarianism is inherently more “desirable” than one another, it may still be the case that differential personality traits inferred from each type of response can mediate the relationship between reputation concern and judgments in moral dilemmas (Lee et al., 2014). For example, characteristically deontological behavior, due to its axiomatic rejection of harm, may elicit the perception of warmth that signals one's other-regarding preferences in social relationships (Fiske, Cuddy, & Glick, 2007). By contrast, a utilitarian choice can incur the perception of competence and self-profiting tendency since it has a direct bearing on one's motivation for effective goal-maximization (Wojciszke, 2005). Researchers have found that such a reverse inference is indeed central to deriving a set of interrelated personality traits from moral judgments (Everett et al., 2016; Lee et al., 2014; Rom et al., 2017).

Based on this evidence, we propose that people could adjust their idiosyncratic modes of decision making in moral dilemmas when their judgments could be subjected to social evaluation. The possible cross-talk between reputation concern and decision making in moral dilemmas is not trivial in that it may confound our understanding of moral cognition in humans and its underlying mechanisms. However, no study to date has directly examined the impacts of reputation concern on judgments in moral dilemmas. The present study thus aims to investigate how elevated concern for reputation modulates the known response patterns observed in moral dilemma task.

1.3. The present study: an overview

To this end, we devised three behavioral tasks. The first two tasks were variants of the well-established moral dilemma paradigm (Conway & Gawronski, 2013; Shenhav & Greene, 2014) each of which was designed to 1) measure participants' baseline moral judgment tendencies and 2) contrast them with subsequent decision bias induced by reputation concern, respectively. We specifically combined the second task with a social observation paradigm wherein participants' task performances are directly monitored by two independent observers (Izuma, Saito, & Sadato, 2010; Jung, Sul, Lee, & Kim, 2018). While a more popular approach based on artificial surveillance cues is often criticized for inconsistent findings and lack of ecological validity (Northover, Pedersen, Cohen, & Andrews, 2017), social observation involving real persons has been shown to reliably trigger audience effects among healthy adults in different experimental setups (Izuma, 2012; Jung et al., 2018). Our approach is also suited for modeling the effects of reputation concern on the real-world decision making which often has to be carried out in the absence of clear guidelines for normative behaviors. Finally, we used a lexical decision task (LDT) to get insight into whether and how social observation would affect participants' performance in the moral dilemma task.

1.4. Hypothesis and predictions

1.4.1. Decision bias in moral judgment

In principle, the specific direction of the decision bias caused by social observation may vary depending on how reputation is defined in a given evaluative context (Holoien & Fiske, 2013). Yet, a wealth of empirical evidence and theories in social psychology and evolutionary biology allow us to make a focused prediction about the possible impacts of social observation on moral judgments. For instance, numerous studies in person perception literature have revealed the fundamental asymmetry between warmth vs. competence in social interaction with the former being a major determinant of impression formation and interpersonal liking (Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012; Goodwin, Piazza, & Rozin, 2014; Wojciszke, Bazinska, & Jaworski, 1998).[2] It has also been suggested that such a predominance of warmth over competence in person perception might reflect the universal human adaptation for cooperation and successful partner choice within a group (Everett et al., 2016; Sacco et al., 2017). That is, warmth may be favored as a proxy for one's propensity to engage in mutually-beneficial social interaction and long-term relationship (Everett et al., 2016; Fiske et al., 2007). Note that traits such as stable cooperation, pair-bonding, and behavioral coordination among group members are thought to have played a crucial role in human evolution (Fehr, Fischbacher, & Gachter, 2002; Tomasello et al., 2012). Hence, it has been convincingly argued that humans have inherited psychological dispositions not only to favor those with agreeable social qualities, but also to signal in that manner (Chudek & Henrich, 2011; Gintis, Smith, & Bowles, 2001). Such an evolved bias for prosociality or warmth traits could also be potentiated by cultural institutions and social norms that fine-tune individuals' behaviors during development (Chudek & Henrich, 2011; Cushman, 2015; Fehr & Fischbacher, 2004).

[2] Recent progress in person perception literature offers a more fine-grained analysis of the relationship between personality traits (i.e., warmth, morality, and competence) and impression formation. Please also see: Goodwin et al. (2014).

It is thus fitting that people would generally be geared towards valuing warmth-related traits in a wide range of social contexts especially when no other explicit desirability cues are present (Brown & Sacco, 2017; Fiske et al., 2007).

Drawing from these lines of literature, we hypothesize that social observation would induce deontological decision bias in the context of moral dilemmas by making participants more tuned to the signaling aspects of moral judgments especially in relation to warmth-related traits (Everett et al., 2016; Lee et al., 2014; Rom et al., 2017). Several additional predictions follow this hypothesis. First, the deontological bias should only be present for moral vs. non-moral control dilemmas since responses to the latter would tell little about participants' social qualities that are often subjected to normative evaluation. Second, the effect would be more pronounced for personal moral dilemmas where stakes are high, or the conflict between deontological vs. utilitarian processes is salient so that participants' moral orientations would be revealed more clearly.

1.4.2. Reaction time

There exists more than one pathway through which reputation concern can influence moral judgments, which we think would differentially affect participants' RTs. While we remain open to multiple possibilities, one probable outcome is that reputation concern imposes a cognitive burden on participants' decision making processes. This is well-supported by evidence that impression management often involves active self-monitoring and behavioral adjustment based on perceived norms (Buckholtz & Marois, 2012; Vohs, Baumeister, & Ciarocco, 2005). Similarly, we predict that social observation can delay decision RTs in the context of moral dilemmas by making individuals consider reputational consequence of their responses and use the information to guide their actual decisions.

1.4.3. Lexical decision task

The main purpose of LDT is to obtain a separate measure showing that social observation affected participants' performance in the moral judgment task in a predicted way. Previous literature on goal-priming suggests that exposure to a certain physical/social environment or a stimulus imbued with a certain goal state (e.g., the scent of all-purpose cleaner) could facilitate the goal-fulfilling behaviors (e.g. cleaning) and increase sensitivities towards the related words presented in LDT (e.g., hygiene) (Holland, Hendriks, & Aarts, 2005). Here, we predict that participants would demonstrate an enhanced reactivity (e.g., faster RTs) towards a list of words that have close semantic affinities with reputation, or other related goal constructs (i.e. warmth and competence) if social observation in the present study renders reputation concern more salient and accessible in participants' knowledge network (Aarts & Dijksterhuis, 2000). To test this prediction in depth, we included multiple different word categories that could tap into more specific reputation-related goals (See Methods).

2. Methods

2.1. Participants

All participants were recruited from a larger pool of volunteers who signed up for a research project that investigates the moral cognition of healthy Korean adult population. To determine appropriate sample size, we conducted *a priory* power analysis using G-power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). Effect sizes ($\eta_p^2 = 0.120$ and $\eta_p^2 = 0.103$) were drawn from previous studies that used a similar social observation manipulation to test the impacts of reputation concern on altruism (Cage, Pellicano, Shah, & Bird, 2013) and prosocial decision making (Jung et al., 2018). Given the prediction that the impacts of social observation would vary across dilemma type, we focused on the estimated effect size for interaction effects. With an α of 0.05, the analysis yield a required sample size of 40 and 46 to provide 80% power for a within-between interaction effect in mixed ANOVA, respectively. To account for a possible data loss across task procedures, a total of 48 college students (Male: $N = 25$, Mean Age = 24.96 ± 3.37; Female: $N = 23$, Mean Age = 22.91 ± 1.64) were recruited as participants. Participants were screened for any prior exposure to the moral dilemma literature as well as for the history of taking advanced psychology courses. The present study was approved by the College Institutional Review Board. A written informed consent was obtained from each participant before the experiment. All participants were fully debriefed and received 7000 KRW (=6.50 USD) at the end of the experiment.

2.2. Task design and procedure

2.2.1. Pretest and pre-experiment questionnaire

Upon arrival, participants reported their current mood and the level of arousal on 7-point Likert scale (Mood: 1 = positive, 7 = negative; Arousal: 1 = low, 7 = high). Participants were then seated alone in a cubicle and completed a pretest with 20 moral dilemmas. Participants were led to believe that it was a practice prior to the main task. However, the goal of the pretest was to measure individual participant's baseline moral decision tendencies unaffected by experimental manipulation of reputation concern. For this purpose, we adopted a stimuli set from previous studies that applied the process dissociation analysis (PD) (Jacoby, 1991) to estimate the deontological (D) and utilitarian inclination (U) within individuals (Conway & Gawronski, 2013) (See Supplementary material S2-1 and S2-2 for details).

All tasks were implemented using the Cogent 2000 toolbox in MATLAB (version 8.4.0; Mathworks, Natick, MA). In each trial, a brief onset of fixation point was followed by three consecutive segments of a moral dilemma presented through a computer screen. All twenty PD-dilemma scenarios were shown to participants in a randomized order. Participants were instructed to read each segment at their own paces and to make binary decisions on the appropriateness of the behaviors described at the end of the scenarios (Inappropriate = Deontological judgment; Appropriate = Utilitarian judgment). Once responses were made, participants were further prompted to indicate the level of confidence for each decision on a 4-point Likert scale using a number pad. This format was taken from recent moral dilemma studies (Shenhav & Greene, 2014), yet with an important modification that extreme values correspond to the *strong confidence* for either deontological (i.e., 1: Very inappropriate) and utilitarian judgment (i.e., 4: Very appropriate), while the moderate values (i.e., 2 and 3) are to depict diminished decision certainty. In short, participants were first asked to make binary moral judgments and then indicated how they felt about their own decisions using a continuous scale. Every response was followed by a visual feedback (i.e., a red box) highlighting the choice made by participants. To dissociate people's responses with the motor execution, the assigned response keys for binary decisions and the numbering of the Likert scale were counterbalanced across individuals.

2.2.2. Main task

The main task was based on the moral dilemma paradigm originally developed by Greene and colleagues (Greene et al., 2001). After the pretest, participants respond to a series of 50 moral dilemmas: 20 personal-moral dilemmas, 15 impersonal-moral dilemmas and 15 non-moral control dilemmas (See Supplementary material S1-1). In order to avoid any demanding characteristics and memory effect due to repeated presentation of the identical stimuli, we excluded the scenarios that had been used for the pretest. The distinction between personal- vs. impersonal moral dilemmas was made based on the presence of direct, serious bodily harm towards a person(s) (Greene et al., 2001). The non-moral dilemmas included a set of scenarios that evokes value-neutral cognitive processes such as mathematical calculation and logical reasoning. All dilemma scenarios were translated into Korean and matched for the average word counts.

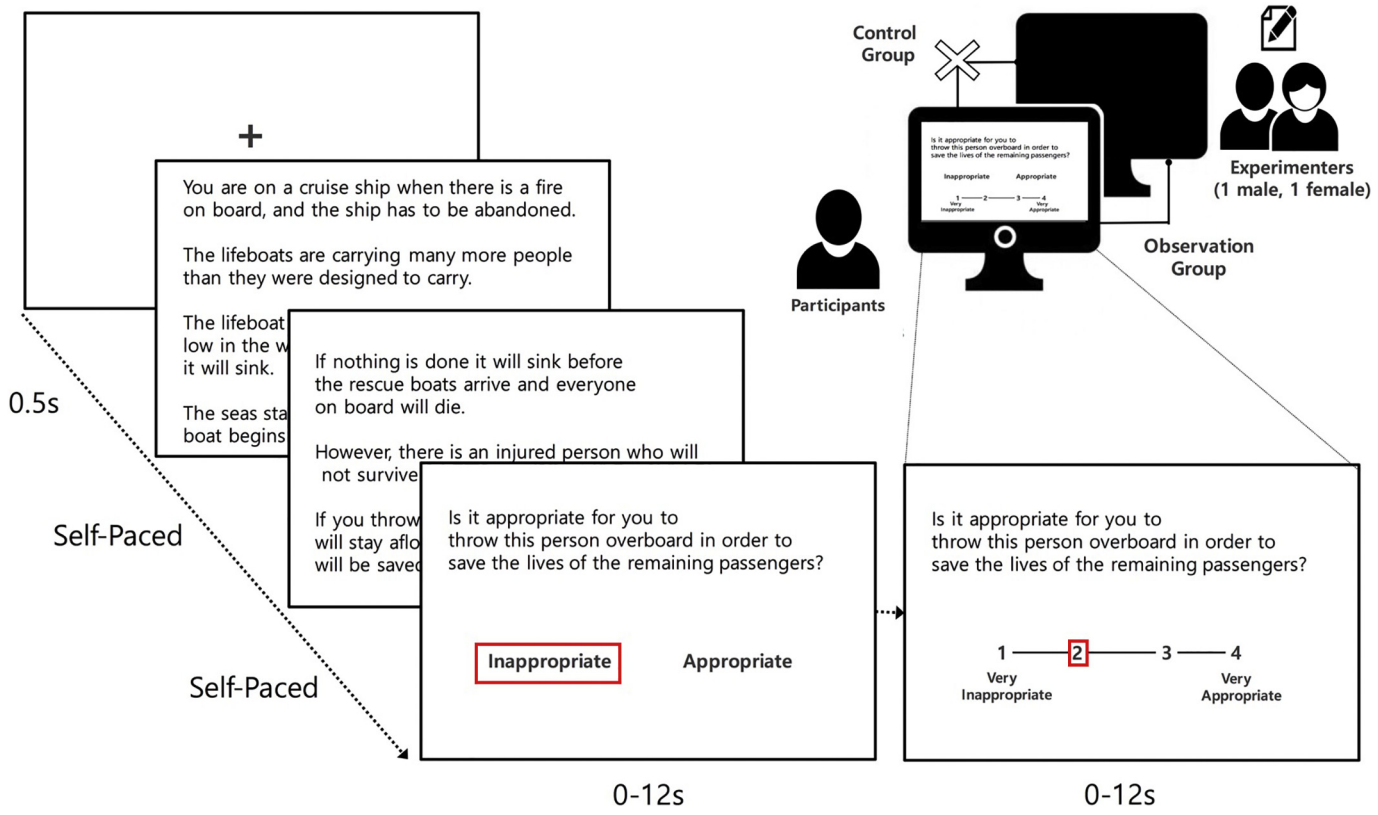

Fig. 1. Schematic representation of the moral dilemma task. Each moral dilemma scenario was presented in three sequential slides. First two slides described the background information. The third slide contained the dilemma question. Participants indicated *i*) their choice between characteristically deontological and utilitarian decision, and *ii*) how confident they were with their responses. Participants in OB group, but not in the CON group, believed that their task performances were being monitored and recorded by two experimenters.

A general process through which participants performed the main task was identical to that of the pretest. However, one critical feature was added to the procedure: a manipulation of reputation concern through social observation (Fig. 1). Before proceeding to the main task, all participants were randomly assigned to either the observed (OBS) or the control group (CON). Two groups were equivalent in size ($N = 24$), gender composition (OBS: M = 13, F = 11, CON: M = 12, F = 12) and mean age (OBS = 23.8 ± 3.8, CON = 24.2 ± 2.4). Participants in the OBS group were told that their performances will be monitored and manually recorded by two experimenters due to some technical glitch that may cause data loss (Izuma et al., 2010). To increase credibility of the manipulation, participants were introduced to an ostensibly-real experimental setup for social observation: they were brought to a separate cubicle wherein two experimenters (i.e., observers: 1 male and 1 female) were sitting with a monitoring device connected to the participants' own screen (e.g., a parallel display). Those in the CON group, on the other hand, were told that their responses will be spontaneously encrypted and remain anonymous after the experiment. Participants in the CON group were also led to come across the same pair of observers during the instruction. However, the monitoring device was removed and no additional information was provided regarding the presence of observers. Participants did not interact with the observers in either condition and returned to the original cubicle after the instruction. No actual social observation or manual recording was made throughout the experimental procedure (See Supplementary material S1-2 for the instruction for experimental manipulation).

2.2.3. Lexical decision task

After the main task, participants performed LDT. The task used the Go/no-go paradigm to minimize response error (Perea, Rosa, & Gomez, 2002). Participants were asked to make a button response as quickly and accurately as possible whenever a string of two Korean letters appearing in the visual display was a real word. A visual feedback was provided in case of wrong responses. According to the hypothesis that social observation would activate various reputation-related goal constructs, we used four word categories that may tap into different facets of reputation concern: Reputation (Repu), Non-Reputation (Control), Warmth-Competence, and Non-word. The word stimuli used for Repu and Control category were selected from the Sejong Corpus (Kim, 2006) with the matching degrees of frequency, familiarity, and valence but semantic-relatedness to reputation (See Supplementary material S2-3). Warmth-Competence category was comprised of six warmth- and six competence-trait words that had previously been shown to depict the impressions of deontologists and utilitarian individuals, respectively (Lee et al., 2014). Finally, Non-word category was comprised of pseudo-word adopted from an independent study that used LDT to assess linguistic processing of healthy Korean adults (Gweon, Kim, & Lee, 2006).

2.2.4. Post-experiment questionnaires

Judgments in moral dilemmas are known to be affected by individual differences in cognitive motivation and emotional sensitivity towards harm (Conway & Gawronski, 2013). To test the potential intergroup difference in psychological traits on the moral dilemma task, participants completed the self-report questionnaires measuring empathy (Interpersonal Reactivity Index; IRI) (Davis, 1983) and need for cognition (NfC) (Cacioppo & Petty, 1982). We also used the fear of negative evaluation (FNE) scale (Leary, 1983) to explore whether the individual difference in evaluation apprehension would correlate with the impacts of reputation concern in the OBS group. Finally, participants' mood and arousal were measured again for comparing the subjective experiences of physiological states before and after the task.

2.3. Statistical analysis

All statistical analyses were carried out using SPSS 23.0 (Statistical Package for the Social Sciences, IBM Corp., Armonk, NY, USA) with the type I error rate set to $\alpha = 0.05$ (two-tailed). Greenhous-Geisser correction was applied for violations of sphericity in analysis of variance (ANOVA) involving repeated measures. Bonferroni-corrected paired samples *t*-tests were conducted to analyze the simple effects for statistically significant interaction. When appropriate, Cohen's *d* and partial eta-squared (η_p^2) were reported as effect size measures.

2.3.1. Pretest

Based on PD analysis laid out in Conway and Gawronski (2013), the total number of deontological and utilitarian responses were algebraically combined at the individual level to estimate moral judgment parameters *D* and *U* (See Supplementary material S2-2 for details). Since *D* and *U* use different metrics (Conway & Gawronski, 2013), no direct comparison was made on the raw parameter values within each group. Rather, our analysis focused on whether participants in the OBS and the CON group significantly differed in terms of their baseline tendencies to make either deontological or utilitarian judgments. Accordingly, a 2 × 2 mixed ANOVA was performed on the standardized *D* and *U* parameters to test the interaction between the group (OBS vs. CON) and the parameter type (*D* vs. *U)*.

2.3.2. Main task

2.3.2.1. Moral judgments. The impact of social observation on judgments in moral dilemmas was assessed by a 2 × 3 mixed ANOVA conducted on the proportion of deontological responses made by each individual. The model included group (OBS vs. CON) as a between-subjects factor and the dilemma type (Personal vs. Impersonal vs. Non-moral) as a within-subjects factor. It should be pointed out that the conceptual distinction of "utilitarian vs. deontological" is not applicable to the judgments in the non-moral control dilemmas. For the sake of convenience, we henceforth use "deontological judgment" in the non-moral control dilemma only to refer to the trials where participants chose "*inappropriate*" during the main task.

2.3.2.2. Decision confidence. We analyzed individual participants' ratings on decision confidence by a 2 × 3 × 2 mixed ANOVA involving group (OBS vs. CON) as a between-subjects factor, and dilemma type (personal vs. impersonal vs. non-moral) and response type (appropriate vs. inappropriate) as within-subjects factors. All raw rating scores were converted into the values between 1 and 2 so that the higher numbers indicated stronger decision confidence for both types of moral judgments.

2.3.2.3. Reaction time. A 2 × 3 × 2 mixed ANOVA was applied to the individual participants' RTs, or the latencies between the presentation of the binary decision prompt and actual judgments. The model included group (OBS vs. CON) as a between-subjects factor while dilemma type (personal vs. impersonal vs. non-moral) and response type (appropriate vs. inappropriate) were entered as within-subjects factors.

2.3.3. Lexical decision task

In accordance with Holland et al. (2005), average RTs of the correct trials were analyzed in a 2 × 4 mixed ANOVA with the group (OBS vs. CON) as a between-subjects factor, and the word category (Reputation vs. Control vs. Warmth vs. Competence) as a within-subjects factor. For this analysis, the original WC category was divided into two subgroups, namely Warmth- and Competence category, and included into the model as two separate levels.

2.3.4. Post-experiment questionnaires

In order to compare baseline personality traits of participants in the OBS and the CON group, we ran a series of independent sample *t*-tests on IRI, NfC and FNE scores. The self-reported ratings on mood and arousal were also examined by a 2 (pre- vs. post) × 2 (group) mixed ANOVA to analyze the potential impact of task procedure (e.g., social observation) on participants' subjective experience of physiological states. A series of exploratory Pearson correlation analysis was carried out to test the relationship between the self-report measure of personality traits and the response data obtained from main task and LDT task.

3. Results

3.1. Pretest

No significant interaction effect involving group and parameter type was identified in the 2 (group) × 2 (moral parameter) mixed ANOVA on the standardized *D* and *U* parameters (all $p > .266$). This indicates that participants in the OBS (*D*: $M = -0.13 \pm 1.01$; *U*: $M = 0.08 \pm 0.98$) and the CON group (*D*: $M = 0.13 \pm 1.00$; *U*: $M = -0.08 \pm 1.03$) did not differ in their baseline inclinations to make either deontological or utilitarian responses in moral dilemmas. Inclusion of sex and motor-counterbalancing did not change the result in that no significant interaction involving group and moral parameter was identified (All $p > .321$). To further explore any intergroup difference, we also analyzed the average proportion of deontological judgment for the "incongruent" PD-dilemmas which are equivalent to personal moral dilemmas used in Greene et al. (2001) (Conway & Gawronski, 2013). Here, a two-sample *t*-test on the average proportions of deontological judgments in the incongruent PD-dilemmas replicated the initial results showing that participants in the OBS ($M = 37.08 \pm 15.17$) and the CON group ($M = 41.25 \pm 12.61$) were equivalent in their baseline moral decision tendencies ($p = .306$).

3.2. Main task

3.2.1. Moral judgments

The 2 (group) × 3 (dilemma type) mixed ANOVA on the proportion of deontological response first revealed a significant main effect of dilemma type, $F_{(1.742,\ 80.11)} = 75.059$, $p < .001$, $\eta_p^2 = 0.620$. Replicating the results of original study by Greene and colleagues (Greene et al., 2001), a series of post-hoc pairwise comparison found that personal moral dilemmas produced more deontological responses ($M = 58.13 \pm 17.30$) than both impersonal moral dilemmas ($M = 50.00 \pm 13.33$), $t_{(47)} = 3.479$, $p = .001$, Cohen's $d = 0.50$, and non-moral dilemmas ($M = 28.47 \pm 8.56$), $t_{(47)} = 9.529$, $p < .001$, Cohen's $d = 1.37$. Participants were also more likely to make deontological decisions in impersonal moral dilemmas vs. non-moral dilemmas, $t_{(47)} = 9.503$, $p < .001$, Cohen's $d = 1.37$. Most importantly, we found a significant interaction effect between the group and dilemma type, $F_{(1.742,\ 80.11)} = 4.779$, $p = .014$, $\eta_p^2 = 0.094$. Consistent with our hypothesis, this interaction was mainly driven by participants in the OBS group showing an increased tendency to make deontological judgments in personal moral dilemmas ($M = 63.96 \pm 19.84$) than participants in the CON group ($M = 52.29 \pm 12.16$), $t_{(38.126)} = 2.457$, $p = .018$, Cohen's $d = 0.70$ (Fig. 2). Finally, there was a marginally significant main effect of group, pointing to the overall increase in deontological responses among participants in the OBS, $F_{(1,\ 46)} = 3.757$, $p = .059$, $\eta_p^2 = 0.076$. Notably, these results were not substantially altered when participants' baseline moral orientation (i.e., *D*-parameter) was included in the model, with the same interaction effect between group and dilemma remaining highly significant, $F_{(2,\ 90)} = 6.772$, $p = .002$, $\eta_p^2 = 0.131$ (See Supplementary material S2-6 and S2-7). As in the results of the pretest, neither participants' sex nor motor-counterbalancing produced statistically significant effects (all $p > .134$). Therefore, we excluded those variables in the further analyses.

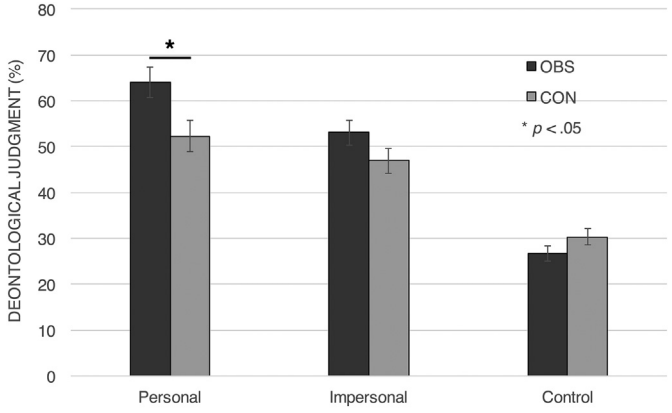

Fig. 2. The proportion of deontological judgments in the moral dilemma task. Participants in the OBS group showed an increased tendency to make deontological responses to personal moral dilemmas. Error bars indicate standard error mean (SEM).

3.2.2. Decision confidence

The 2 (group) × 3 (dilemma type) × 2 (response type) mixed ANOVA on the decision confidence yielded a significant main effect of dilemma type, $F_{(2,\ 88)} = 35.452$, $p < .001$, $\eta_p^2 = 0.446$. Participants generally indicated the lower decision confidence for the judgments in personal moral dilemmas ($M = 1.34 \pm 0.15$) vs. impersonal moral dilemmas ($M = 1.47 \pm 0.19$), $t_{(45)} = -3.736$, $p = .002$, Cohen's $d = -0.55$, personal moral dilemmas vs. non-moral dilemmas ($M = 1.63 \pm 0.18$), $t_{(45)} = -8.310$, $p < .001$, Cohen's $d = -1.23$, and impersonal moral dilemmas vs. non-moral dilemmas, $t_{(47)} = -5.33$, $p < .001$, Cohen's $d = -0.77$. Moreover, there was a significant main effect of response type, $F_{(1,\ 44)} = 8.912$, $p = .005$, $\eta_p^2 = 0.168$, reflecting that participants generally tended to give higher confidence rating for "appropriate" responses ($M = 1.55 \pm 0.18$) than "inappropriate" responses ($M = 1.42 \pm 0.21$), $t_{(45)} = 2.989$, $p < .005$, Cohen's $d = 0.46$. Yet, these main effects were qualified by a significant interaction between dilemma type and responses type, $F_{(1.868,\ 82.184)} = 12.098$, $p < .001$, $\eta_p^2 = 0.216$. A post-hoc analysis confirmed that this result was due to relatively higher confidence ratings for "appropriate" ($M = 1.77 \pm 0.19$) vs. "inappropriate" responses ($M = 1.48 \pm 0.37$) only in response to the non-moral dilemmas, $t_{(47)} = 4.345$, $p < .001$, Cohen's $d = 0.62$. Finally, there was a significant main effect of group, $F_{(1,\ 44)} = 4.793$, $p = .034$, $\eta_p^2 = 0.098$, showing that participants in the OBS group were characterized by a lower decision confidence ($M = 1.44 \pm 0.07$) than those in the CON group ($M = 1.51 \pm 0.14$) in general. No other statistically significant results were found (all $p > .07$).

3.2.3. Reaction time

For RTs in the main task, the 2 (group) × 3 (dilemma type) × 2 (response type) mixed ANOVA revealed a significant main effect of dilemma type, $F_{(1.713,\ 75.378)} = 49.089$, $p < .001$, $\eta_p^2 = 0.527$, with participants showing significantly longer RTs in the non-moral dilemmas ($M = 3366.49 \pm 1120.95$) than in personal moral dilemmas ($M = 2723.65 \pm 1102.20$), $t_{(47)} = 6.821$, $p = .006$, Cohen's $d = 0.99$, and in impersonal moral dilemmas ($M = 2506.61 \pm 1015.80$), $t_{(47)} = 8.638$, $p < .001$, Cohen's $d = 1.24$. The same analysis also found a significant interaction between dilemma type and response, $F_{(2,\ 88)} = 4.307$, $p = .016$, $\eta_p^2 = 0.089$. Closely aligning with previous evidence (Greene et al., 2001), this effect was produced by the longer RTs associated with utilitarian judgments ($M = 2916.14 \pm 1331.37$) vs. deontological judgments ($M = 2531.17 \pm 1040.66$) in personal moral dilemmas, $t_{(45)} = 2.829$, $p = .006$, Cohen's $d = 0.42$. Critically, a significant main effect of group was identified, reflecting a general prolongation of RTs among participants in the OBS group ($M = 3240.93 \pm 913.41$) vs. the CON group ($M = 2592.78 \pm 985.71$), $F_{(1,\ 44)} = 5.323$, $p = .026$, $\eta_p^2 = 0.108$. No other main effects or interactions turned out to be statistically significant (all $p > .246$).

3.2.4. The relationship between moral judgments and RTs

To explore a potential link between the impacts of social observation on moral judgments and RTs in the main task, the participants in each group were divided into two subgroups (i.e. Fast- vs. Slow-RT group) based on their overall RTs in moral dilemmas using a median split. The newly defined subgroup was then entered into a 2 (group) × 2 (RT subgroup) × 3 (dilemma type) mixed ANOVA on the average proportion of deontological decisions in the main task. The analysis replicated the significant interaction between group and dilemma type (See 3.2.1. Moral Judgments), $F_{(1.724,\ 75.860)} = 6.529$, $p = .004$, $\eta_p^2 = 0.129$, led by a selective increase in deontological responses to personal moral dilemmas among participants in the OBS group vs. CON group, $t_{(46)} = 2.75$, $p = .009$, Cohen's $d = 0.46$. Markedly, a significant three-way interaction among group, RT subgroup and dilemma type was found, $F_{(1.724,\ 75.860)} = 3.371$, $p = .046$, $\eta_p^2 = 0.071$. Post-hoc analyses showed that participants who were relatively faster (i.e., Fast-RT group), as opposed to those who were slower (i.e., Slow-RT group) in making moral judgments, tended to make deontological decisions more in both personal moral dilemmas (Fast-RT group: 73.33 ± 22.22; Slower-RT group: 58.33 ± 16.55), $t_{(22)} = 1.891$, $p = .031$, Cohen's $d = 0.91$, and impersonal moral dilemmas (Fast-RT group: 60.74 ± 13.92; Slower-RT group: 48.44 ± 8.90), $t_{(22)} = 2.653$, $p = .024$, Cohen's $d = 1.05$, but only if they had performed the task in OBS vs. CON (Fig. 3). Participants' responses in the non-moral dilemma, however, were unrelated to either social observation or RT subgroups.

3.3. Lexical decision task

The average hit rates were very high ($M = 97.3 \pm 2.9$) and did not differ across groups (OBS: 97.46 ± 2.4, CON: 97.14 ± 3.5) ($p = .705$). A 2 (group) × 4 (word category) mixed ANOVA on RT data from LDT revealed a significant main effect of word category, $F_{(3,\ 138)} = 4.002$, $p = .009$, $\eta_p^2 = 0.080$. A post-hoc pairwise comparison confirmed that the effect reflected the longer average RTs in Reputation ($M = 600.64 \pm 99.96$) vs. Warmth category ($M = 565.30 \pm 91.53$), $t_{(47)} = 3.186$, $p = .013$, Cohen's $d = 0.46$. Critically, a significant interaction between group and word category was found, $F_{(3,\ 138)} = 3.469$, $p = .025$, $\eta_p^2 = 0.065$, with participants in the OBS group ($M = 538.38 \pm 76.04$) showing faster RTs only in response to the words in the Warmth category than participants in the CON group

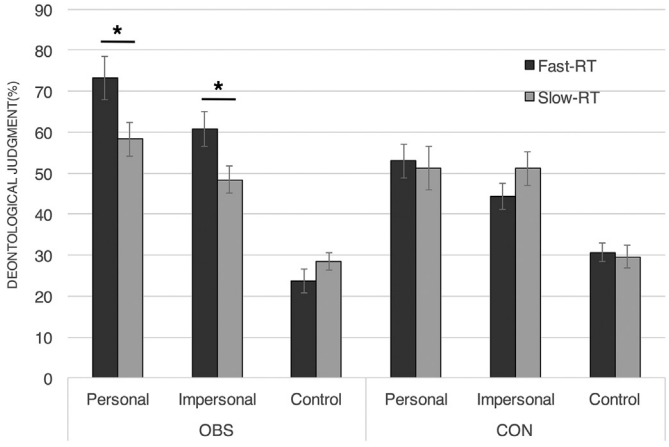

Fig. 3. Deontological judgments as a function of RTs in the OBS vs. CON group. Social observation promoted fast deontological responses to both personal and the impersonal moral dilemmas, but not to the non-moral dilemmas.

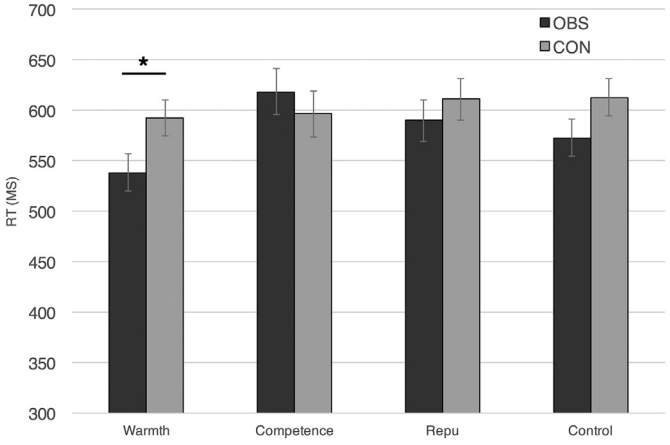

Fig. 4. LDT task performance of the OBS vs. CON group. RTs for the correct trials were averaged and compared between the OBS and CON group. Participants in the OBS group showed an enhanced reactivity towards the Warmth category than did those in the CON group.

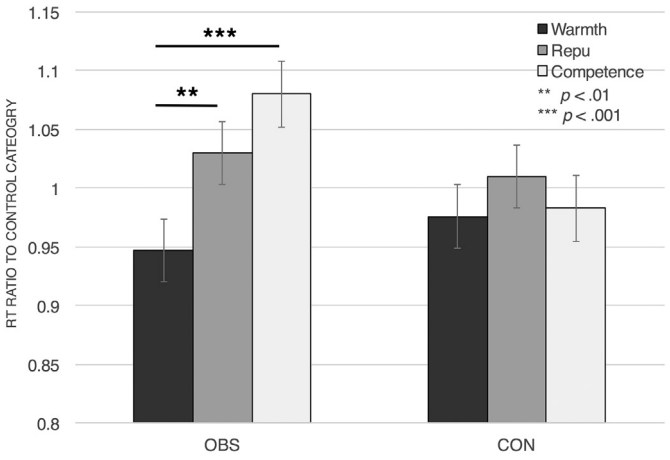

Fig. 5. RT ratios of each word category to Control category. Participants in the OBS group demonstrated response facilitation as well as inhibition for Warmth- and Competence category, respectively, with respect to Control category.

($M = 592.26 \pm 99.03$), $t_{(46)} = -2.116$, $p = .040$, Cohen's $d = 0.61$ (Fig. 4). A more pronounced intergroup difference was identified when the ratio of RTs for Warmth to Competence category (W/C) was compared, $t_{(46)} = -2.888$, $p = .006$, Cohen's $d = 0.83$. In other words, participants in the OBS group ($M = 0.88 \pm 0.11$) were relatively faster for warmth vs. competence trait words than participants in the CON group ($M = 1.01 \pm 0.19$). A follow-up analysis further revealed that comparatively small W/C in the OBS group reflects both increased reactivity towards Warmth category and decreased reactivity towards the Competence category. This is evidenced by a statistically significant difference between the ratio of RTs for Warmth ($M = 0.95 \pm 0.10$) and Competence category ($M = 1.08 \pm 0.10$) to RTs for Control category, $t_{(23)} = -4.828$, $p < .001$, Cohen's $d = -0.98$. No comparable results were identified from participants in the CON group (All $p = .454$) (Fig. 5).

3.4. Pre- and post-experiment questionnaire

Intergroup comparisons on the IRI, NfC, and FNE revealed no significant baseline difference in personality traits among participants in the OBS and the CON group (all $p > .140$). Intriguingly, however, a series of exploratory Pearson correlation analysis identified group-specific associations between personality traits and task performance. For example, FNE scores were inversely correlated with RTs for Warmth vs. Competence category in LDT (i.e., the smaller W/C) only in the OBS group, $r_{(24)} = -0.420$, $p = .041$ (The descriptive statistics of self-report measures as well as a full correlation table are reported in Supplementary material S2-4 and S2-5, respectively). Finally, the results of a 2 (pre-post) × 2 (group) mixed ANOVA on the emotional valence and arousal ratings showed that participants' self-reported physiological states were not significantly different across groups either before or after the experiments (all $p > .146$).

4. Discussion

Are decisions in moral dilemmas susceptible to our desire for positive reputation? The present study investigated how reputation concern modulates judgments in moral dilemmas using social observation paradigm. In addition to replicating the key results of previous studies, we found that participants in the OBS group made more deontological judgments in personal moral dilemmas than did participants in the CON group. Social observation also dwindled the overall decision confidence while promoting fast deontological judgments. Finally, participants in the OBS group showed enhanced sensitivity towards the warmth- vs. competence trait words in LDT (i.e. W/C)

4.1. Moral judgments

First and foremost, we found that social observation increased the proportion of deontological judgments in moral dilemmas. Our analyses showed that this effect was not driven by participants' baseline moral decision tendencies, psychological traits, and any reported changes in physiological states before and after the task. We also minimized demand characteristics by avoiding using the identical stimuli across the tasks, and, more importantly, by not providing participants with any explicit remarks on social evaluation as in some previous studies (Holoien & Fiske, 2013). It is still possible that difference between task environments of the pretest and the main task (i.e., social observation) might have led participants in the OBS group to alter their responses in some ways. However, our primary finding is not likely a product of implicit solicitation of decision shift since there seems no compelling reason to believe that changing task environments would selectively increase deontological judgments across participants rather consistently.

Rather, the overall direction of the audience effect observed in the present study is more resonant with the nascent body of evidence that deontological decisions in moral dilemmas evoke the perception of warmth-related positive traits such as trustworthiness and sociability (Everett et al., 2016; Lee et al., 2014; Rom et al., 2017; Sacco et al., 2017). When considered together with the robust link between self-presentational behaviors and reduced anonymity (Izuma, 2012), the increased deontological judgments in the OBS group could also reflect participants' attempts to manage their social impressions in the presence of observers (Izuma, 2012). Consistent with our interpretation is the result that the impacts of social observation were particularly evident in personal moral dilemmas where the conflict between affective vs. cognitive valuation is most salient (Greene et al., 2001). As noted earlier, people tend to infer others' primary modes of decisions in moral dilemmas and use the knowledge to subsequently form specific impressions of them (Lee et al., 2014; Rom et al., 2017). For example, the target individuals who make deontological judgments in emotionally arousing moral dilemma are considered higher on warmth-related traits via their perceived reliance on the affective vs. cognitive process (Lee et al., 2014; Rom et al., 2017). As such, personal moral dilemmas in the present study, with their higher relevance to affective valuation of “up and close harm” (Greene et al., 2001), might have been deemed critical to participants' reputation since they are the most effective stimuli by which participants could signal warmth or their propensities to avoid harming innocent others.

It is yet important to point out that whether or not certain

personality traits are considered desirable is contingent on the specific context of social evaluation. That is, competence-related traits could well lead to a good reputation in situations where one's technical mastery or rationality is valued over sociality or care for others (Cuddy, Glick, & Beninger, 2011). Evidence also indicates that people can strategically control their behaviors to appear more competent when the relevant attributes (e.g., "smart and intelligent") are presented as the evaluative criteria (Holoien & Fiske, 2013). Why, then, would participants in the present study have signaled warmth other than competence? As noted earlier, prosocial behaviors are not only appealing to our evolved psychological dispositions, but also widely internalized among people as a function of cultural norms (Fehr & Fischbacher, 2004; Rand et al., 2014). Upon encountering potentially judgmental observers, these cultural norms are known to be activated as default heuristics and guide our behaviors in a paucity of other directive cues (Herman, Roth, & Polivy, 2003; Rand, 2016). Note that reputation concern in the present study, too, was only indirectly triggered without any explicit indication of social evaluation. It is thus plausible that participants in the OBS group would have aligned their judgments with more a tenable cultural norm of warmth which, as a proxy for various prosocial traits, would help them earn a positive evaluation in a wider range of social interaction (Fiske et al., 2007).

One important caveat of this interpretation is that dominant cultural norms shared in a group or a society need not always be prosocial (Rand, 2016). As for the present study, however, one recent finding showed that deontological judgments in the trolley problem is considered stereotypically moral among Korean adults (Lee et al., 2014). This supports our view that social observation in the present study might have led participants to act in accordance with what is considered more normative and desirable in a given socio-cultural environment. Here, we emphasize again that our results do not establish an inherent causal connection between reputation concern and deontological judgments per se. Instead, they could be illuminating the role of social observation and reputation concern in activating a set of normative goals that direct participant's self-presentation strategy.

4.2. Decision confidence

Both norm-abiding behaviors and self-presentation often require situational goals (e.g., "obtaining positive social evaluation by signaling warmth,") to be incorporated into pre-existing behavioral repertoires within individuals (Buckholtz & Marois, 2012). This process can be cognitively taxing and uncertain since one must form the meta-cognitive representations of how others would think of oneself (Yeung & Summerfield, 2012). It could also take additional efforts to monitor one's own decisions and, if necessary, to override them for socially desirable responding (Buckholtz & Marois, 2012).

In the present study, participants' performances in the pretest were not different across both groups. Therefore, an overall decrease in decision confidence among participants in the OBS group is likely to reflect their increased needs for self-monitoring and internal conflict between socially desirable responses (i.e., deontological judgments) and alternative choices they would have made without evaluation apprehension. Supporting this possibility, previous studies have shown that the conflict between opposing intuitions as well as intentional false-responding could undermine decision confidence (De Neys, Cromheeke, & Osman, 2011) and elicit the neural activities in the brain regions responsible for response-monitoring and cognitive control (Ganis, Kosslyn, Stose, Thompson, & Yurgelun-Todd, 2003; Nunez, Casey, Egner, Hare, & Hirsch, 2005; Yeung & Summerfield, 2012). In addition, participants generally indicated lower decision confidence for both types of moral dilemmas vs. the non-moral dilemmas. This can be explained by the fact that non-moral dilemmas only require value-neutral cognitive processes (Greene et al., 2001). Many control dilemmas also have "correct" answers that can be identified with relative ease. The lack of such conflict-inducing elements could have made participants' responses to the non-moral dilemmas relatively immune to the mental processes associated with self-presentation and norm compliance, thereby yielding higher decision confidence than other dilemma scenarios that have moral relevance.

4.3. Reaction time

Our interpretation of the decision confidence is further supported by RT data. RTs have been considered an important measure of cognitive control and decision conflict in moral judgment literature (Greene, Nystrom, Engell, Darley, & Cohen, 2004; Rand, Greene, & Nowak, 2012). In the present study, we found the predicted effect of social observation that participants in the OBS group would be generally slower than those in the CON group when responding to the moral dilemmas. This result is not only consistent with the hindering influence of social observation on decision confidence, but also suggests that reputation concern might trigger additional mental processes associated with self-presentation and norm compliance.

Of course, longer RT does not necessarily imply colliding mental processes or response monitoring at work. For example, researchers have claimed that the mere presence of others could hamper peoples' cognitive performance by disrupting effective allocation of attentional resource (Huguet, Galvaing, Monteil, & Dumas, 1999; Sanders, Baron, & Moore, 1978). Previous findings in moral psychology also suggested that attentional process can affect both RTs and overall patterns of moral judgments without recruiting any motivational- or other high-order cognitive processes related to social norm compliance and self-presentation (Huebner, Dwyer, & Hauser, 2009). However, our results warrant more a nuanced approach, given the complex interaction between RTs, moral judgments, and group: the faster mean RTs (i.e., Fast-RT subgroup) were associated with the stronger deontological bias in moral vs. non-moral dilemmas, yet only in the OBS group. This outcome cannot be easily reconciled with a purely attentional account of social observation since resource depletion typically impairs cognitively demanding mental processes than intuitive or automatic processes (Engle, 2002). Empirical evidence also supports this, with cognitive load selectively disrupting RTs for utilitarian responses (Greene et al., 2008).

We suggest that the significant three-way interaction may be explained better by the proposed role of social observation in activating warmth-related cultural norms which, in turn, guide participants' self-presentation strategy. The perception of positive social traits is known to be partially mediated by how fast or slow target individuals make altruistic decisions (Critcher, Inbar, & Pizarro, 2013). Importantly, people are capable of exploiting this relationship to obtain reputational benefits. Jordan and colleagues, for instance, recently showed that participants choose to cooperate and help others much faster when their decision process was made observable in a series of economic games (Jordan, Hoffman, Nowak, & Rand, 2016). The authors concluded that such "uncalculating" prosociality is a product of participants' concern for positive social images (e.g., trustworthiness). Similarly, it is possible that increased concern for reputation in the OBS group, accompanied by concurrent activation of warmth-promoting social norm, rendered deontological judgments more accessible and intuitive as goal-congruent responses. This, in turn, might have contributed to the unique association between faster response and deontological judgments within the OBS group despite the general cognitive burden caused by social observation. This interpretation is also consistent with the growing body of evidence that a wide range of norm-congruent behaviors could gradually become intuitive as one receives continuous inputs from his or her cultural environment during development (Rand, 2016; Rand et al., 2012; Rand et al., 2014).

4.4. Lexical decision task

Finally, the results of the lexical decision task provide additional clues to how social observation might have affected judgments in moral

dilemmas. Researchers have long suggested that individuals' needs, concerns, and goals can increase the accessibility of the related mental constructs (Dijksterhuis & Aarts, 2010). These motivational states, once activated, are known to enhance processing of the goal-relevant target stimuli in LDT (Förster, Liberman, & Friedman, 2009). Intriguingly, previous studies have shown that this "goal-priming" effect can reverse if the target stimuli are linked to conflicting goals (i.e., goal shielding) (Shah, Friedman, & Kruglanski, 2002). In the present study, adding to the predicted response facilitation, we also found the evidence for goal shielding: participants in the OBS group not only reacted faster to the Warmth category than did participants in the CON group, but also showed slower responses towards the Competence category. Such a differential pattern of priming effects across LDT word categories again buttress the idea that the presence of observers did not modulate the task performance simply by hampering participants' attentional capacity in general. More specifically, these results imply that participants in the OBS group might have activated warmth-related goals while suppressing competence-related goals. This interpretation is also compatible with our exploratory analyses showing that the magnitude of goal-priming (i.e., Fast-W/C) in the OBS group correlated with individual participants' fear of negative evaluation (i.e. FNE), an avoidance motivation that predicts one's susceptibility to reputation-related goals (Schlenker & Weigold, 1990; Villarosa, Kison, Madson, & Zeigler-Hill, 2016). This result should not be over-interpreted given the small sample size and the multiple comparisons problem. However, the group-specific association between the FNE score and LDT performance result may still speak in favor of the motivational basis of the audience effect observed in the present study.

We, rather unexpectedly, did not find any significant relationship between participants' responses to the moral judgments and Repu category in LDT which had a direct semantic connection to "*pyungpan* (reputation)." One possible explanation is "multifinality effect (Kruglanski, Chernikova, Babush, Dugas, & Schumpe, 2015)." Researchers have suggested that the impact of goal priming is dependent upon the unique connections between goals (e.g., warmth-signaling) and a means (e.g., Warmth category in LDT) (Förster, Liberman, & Friedman, 2007). That is, if a single means can be related to multiple different goals, the association strength between the goals and the means decreases, or "diluted," which would, in turn, decrease the priming effect (Kruglanski et al., 2015). Similarly, "*Pyungpan* (reputation)" is an umbrella term which can be connected to various goal constructs including both warmth and competence. Due to their mutually-inhibiting dynamic (Judd, James-Hawkins, Yzerbyt, & Kashima, 2005), warmth and competence could dilute each other's effects when activated simultaneously (Kruglanski et al., 2015). We thus suspect that the average instrumental value of the Repu category in our LDT task would have been lower than that of the Warmth category due its linkage to competence, the opposing goal construct actively inhibited (i.e., goal shielding) among those in the OBS group during the moral judgment task.

5. Limitations and outstanding questions

The present study provided experimental evidence that social observation and its resultant activation of reputation concern can increase deontological judgments in moral dilemmas. We also suggested a potential role of a warmth-promoting cultural norm in incurring the decision bias. While most of our key predictions were confirmed, generalizability of the findings needs to be tested for multiple, larger study samples that may endorse distinct cultural norms. For instance, a developing line of studies recently showed that people may possess drastically different social heuristics depending on their respective socio-cultural environment. In effect, this can lead to opposing patterns of behaviors (e.g., Fast vs. slow cooperation) in the same experimental task in ways not solely explicable by a general preference for prosociality (Nishi, Christakis, & Rand, 2017). Similarly, it is possible that specific directions of audience effect in the context of moral dilemmas could also vary depending on the nature of norms people perceive from their own referent cultural groups.

It is also important to note that many behavioral data, RTs in particular, should be interpreted with caution since they often do not provide a full picture of specific cognitive mechanisms underlying a phenomenon of interest (Krajbich, Bartling, Hare, & Fehr, 2015). In the same vein, our discussion of the three-way interaction, for example, is post hoc and to a degree speculative. Further evidence would be necessary to reveal factors behind the complex relationship between RTs, dilemma type and actual moral judgments observed in the present study. A potentially useful approach would be to employ neuroimaging techniques such as fMRI that could offer insight into differential cognitive processes leading to the same behaviors.

Another important question not directly addressed in the present study is whether the observed deontological bias in the OBS group was truly volitional. As discussed earlier, numerous studies have suggested that social norm compliance and self-presentation are effortful processes and can be consciously orchestrated (Buckholtz & Marois, 2012; Gesiarz & Crockett, 2015; Vohs et al., 2005). However, researchers have also found that even a high-order cognitive activity could be guided by the subliminally activated goal-representations (Dijksterhuis & Aarts, 2010). Parsing out the unique contribution of conscious vs. unconscious processes to behaviors is difficult since they are often concurrent, and produce similar outcomes (Bargh, Schwader, Hailey, Dyer, & Boothby, 2012). In this regard, our data are not conclusive. The selectivity of the audience effects (e.g., moral vs. non-moral dilemma) suggests that at least part of the deontological decision bias can be attributed to participants' conscious engagement with the specific dilemma contents. Equally plausible, but not necessarily incompatible possibility is that the presence of observers activated warmth-related goals and a set of self-presentational behaviors in a form of subtle response-potentiation (Paulhus & Levitt, 1987). Additional studies are needed to gain more complete insight into the interaction between reputation concern and moral judgments.

6. Conclusion

Despite limitations, the present study adds to the growing recognition that judgments in moral dilemmas may have reputational consequences. We believe that this is an important research topic that merits further exploration for at least two reasons. First, many ethical conundrums in the real world are essentially social in that they require public disclosure of one's moral stance. This could influence how our moral beliefs translate to actions, which has often been missed out in laboratory experiments. The results of the present study specifically revealed that our day-to-day moral judgments are closely intertwined with the differential social outcomes of endorsing particular ethical positions. Second, the present study showed that our current knowledge of the cognitive processes underlying moral judgments can be enriched in the light of evolutionarily preserved self-presentational motivations. This would be particularly important for most fMRI research where participants' task performances are almost always monitored by experimenters. Our data showed that people may engage in desirable responding even without explicit social evaluation. Future studies should investigate these issues deeper to delineate the mechanistic contribution of reputation concern to behavioral profiles and neural substrates of moral judgments.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgement

This work was supported by the Ministry of Education of the

Republic of Korea and the National Research Foundation of Korea (NRF-2015S1A5B6036594). The authors would like to thank Doyo Choi, Hyunjoon Park, Gahyun Lim, and Seungchan Park for their assistance in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.evolhumbehav.2018.06.004.

References

Aarts, H., & Dijksterhuis, A. (2000). Habits as knowledge structures: Automaticity in goal-directed behavior. *Journal of Personality and Social Psychology, 78*(1), 53–63.

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science, 23*(8), 861–868. http://dx.doi.org/10.1177/0956797611434965.

Andreoni, J., & Petrie, R. J. (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics, 88*(7), 1605–1623.

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences, 16*(12), 593–605. http://dx.doi.org/10.1016/j.tics.2012.10.002.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154–161. http://dx.doi.org/10.1016/j.cognition.2011.05.010.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412–414. http://dx.doi.org/10.1098/rsbl.2006.0509.

Bernhard, R. M., Chaponis, J., Siburian, R., Gallagher, P., Ransohoff, K., Wikler, D., ... Greene, J. D. (2016). Variation in the oxytocin receptor gene (OXTR) is associated with differences in moral judgment. *Social Cognitive and Affective Neuroscience, 11*(12), 1872–1881. http://dx.doi.org/10.1093/scan/nsw103.

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *The British Journal of Social Psychology, 51*(1), 149–166. http://dx.doi.org/10.1111/j.2044-8309.2010.02011.x.

Brown, M., & Sacco, D. F. (2017). Is pulling the lever sexy? Deontology as a downstream cue to long-term mate quality. *Journal of Social and Personal Relationships*.

Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience, 15*(5), 655–661. http://dx.doi.org/10.1038/nn.3087.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116.

Cage, E., Pellicano, E., Shah, P., & Bird, G. (2013). Reputation management: Evidence for ability but reduced propensity in autism. *Autism Research, 6*(5), 433–442. http://dx.doi.org/10.1002/aur.1313.

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews, 36*(4), 1249–1264. http://dx.doi.org/10.1016/j.neubiorev.2012.02.008.

Chudek, M., & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences, 15*(5), 218–226. http://dx.doi.org/10.1016/j.tics.2011.03.003.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104*(2), 216–235. http://dx.doi.org/10.1037/a0031021.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science, 4*(3), 308–315.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America, 107*(40), 17433–17438. http://dx.doi.org/10.1073/pnas.1009396107.

Cuddy, A. J., Glick, P., & Beninger, A. (2011). The dynamics of warmth and competence judgments, and their outcomes in organizations. *Research in Organizational Behavior, 31*, 73–98.

Cushman, F. (2015). Punishment in humans: From intuitions to institutions. *Philosophy Compass, 10*(2), 117–133.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PloS One, 6*(1), e15954.

Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annual Review of Psychology, 61*, 467–490. http://dx.doi.org/10.1146/annurev.psych.093008.100445.

Egas, M., & Riedl, A. (2008). The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Biological Sciences, 275*(1637), 871–878. http://dx.doi.org/10.1098/rspb.2007.1558.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology. General, 145*(6), 772–787. http://dx.doi.org/10.1037/xge0000165.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791. http://dx.doi.org/10.1038/nature02043.

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185–190. http://dx.doi.org/10.1016/j.tics.2004.02.007.

Fehr, E., Fischbacher, U., & Gachter, S. (2002). Strong reciprocity, human cooperation, and the enforcement of social norms. *Human Nature, 13*(1), 1–25. http://dx.doi.org/10.1007/s12110-002-1012-7.

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences, 11*(2), 77–83. http://dx.doi.org/10.1016/j.tics.2006.11.005.

Förster, J., Liberman, N., & Friedman, R. S. (2007). Seven principles of goal activation: A systematic approach to distinguishing goal priming from priming of non-goal constructs. *Personality and Social Psychology Review, 11*(3), 211–233. http://dx.doi.org/10.1177/1088868307303029.

Förster, J., Liberman, N., & Friedman, R. S. (2009). *What do we prime? On distinguishing between semantic priming, procedural priming, and goal priming.* Oxford; New York: Oxford University Press.

Friesdorf, R., Conway, P., & Gawronski, B. (2015). Gender differences in responses to moral dilemmas: A process dissociation analysis. *Personality and Social Psychology Bulletin, 41*(5), 696–713. http://dx.doi.org/10.1177/0146167215575731.

Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., & Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: An fMRI investigation. *Cerebral Cortex, 13*(8), 830–836.

Gesiarz, F., & Crockett, M. J. (2015). Goal-directed, habitual and Pavlovian prosocial behavior. *Frontiers in Behavioral Neuroscience, 9*, 135. http://dx.doi.org/10.3389/fnbeh.2015.00135.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of Theoretical Biology, 213*(1), 103–119. http://dx.doi.org/10.1006/jtbi.2001.2406.

Gleichgerrcht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PLoS One, 8*(4), e60418. http://dx.doi.org/10.1371/journal.pone.0060418.

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology, 106*(1), 148–168. http://dx.doi.org/10.1037/a0034726.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144–1154. http://dx.doi.org/10.1016/j.cognition.2007.11.004.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389–400. http://dx.doi.org/10.1016/j.neuron.2004.09.027.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108. http://dx.doi.org/10.1126/science.1062872.

Gweon, H., Kim, S. L., & Lee, H. W. (2006). The relationship between word frequency and semantic priming effects in Hangul word recognition. *Korean Journal of Psychology: Experimental, 18*, 203–220.

Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*(1), 65–72. http://dx.doi.org/10.1111/j.1745-6916.2008.00063.x.

Han, H., Glover, G. H., & Jeong, C. (2014). Cultural influences on the neural correlate of moral decision making processes. *Behavioural Brain Research, 259*, 215–228. http://dx.doi.org/10.1016/j.bbr.2013.11.012.

Herman, C. P., Roth, D. A., & Polivy, J. (2003). Effects of the presence of others on food intake: A normative interpretation. *Psychological Bulletin, 129*(6), 873–886. http://dx.doi.org/10.1037/0033-2909.129.6.873.

Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit. Nonconscious effects of scent on cognition and behavior. *Psychological Science, 16*(9), 689–693. http://dx.doi.org/10.1111/j.1467-9280.2005.01597.x.

Holoien, D. S., & Fiske, S. T. (2013). Downplaying positive impressions: Compensation between warmth and competence in impression management. *Journal of Experimental Social Psychology, 49*(1), 33–41. http://dx.doi.org/10.1016/j.jesp.2012.09.001.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences, 13*(1), 1–6. http://dx.doi.org/10.1016/j.tics.2008.09.006.

Huguet, P., Galvaing, M. P., Monteil, J. M., & Dumas, F. (1999). Social presence effects in the Stroop task: Further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology, 77*(5), 1011–1025.

Hutcherson, C. A., Montaser-Kouhsari, L., Woodward, J., & Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas are encoded in separate areas and integrated in ventromedial prefrontal cortex. *The Journal of Neuroscience, 35*(36), 12593–12605. http://dx.doi.org/10.1523/JNEUROSCI.3402-14.2015.

Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience Research, 72*(4), 283–288. http://dx.doi.org/10.1016/j.neures.2012.01.003.

Izuma, K., Saito, D. N., & Sadato, N. (2010). The roles of the medial prefrontal cortex and striatum in reputation processing. *Social Neuroscience, 5*(2), 133–147. http://dx.doi.org/10.1080/17470910903202559.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*(5), 513–541.

Jordan, J. J., Hoffman, M., Nowak, M. A., & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences of the United States of America, 113*(31), 8658–8663. http://dx.doi.org/10.1073/pnas.1601280113.

Judd, C. M., James-Hawkins, L., Yzerbyt, V., & Kashima, Y. (2005). Fundamental dimensions of social judgment: Understanding the relations between judgments of competence and warmth. *Journal of Personality and Social Psychology, 89*(6), 899–913.

http://dx.doi.org/10.1037/0022-3514.89.6.899.
Jung, D., Sul, S., Lee, M., & Kim, H. (2018). Social observation increases functional segregation between MPFC subregions predicting prosocial consumer decisions. *Scientific Reports, 8*(1), 3368. http://dx.doi.org/10.1038/s41598-018-21449-z.
Kim, H. (2006). *Korean national corpus in the 21st century Sejong project paper presented at the the 13th NIJL International Symposium, Tokyo*.
Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications, 6*, 7455. http://dx.doi.org/10.1038/ncomms8455.
Kruglanski, A. W., Chernikova, M., Babush, M., Dugas, M., & Schumpe, B. (2015). The architecture of goal systems: Multifinality, equifinality, and counterfinality in means-end relations. *Advances in Motivation Science, 2*, 69–98.
Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Evolution and human behavior. *Audience Effects on Moralistic Punishment, 28*(2), 75–84.
Leary, M. R. (1983). Social anxiousness: The construct and its measurement. *Journal of Personality Assessment, 47*(1), 66–75. http://dx.doi.org/10.1207/s15327752jpa4701_8.
Lee, M., Sul, S., & Kim, H. (2014). The impact of moral decision style on impression formation. *Korean Journal of Social and Personality Psychology, 28*(2), 201–223.
Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 371*(1687), 20150100. http://dx.doi.org/10.1098/rstb.2015.0100.
Nishi, A., Christakis, N. A., & Rand, D. G. (2017). Cooperation, decision time, and culture: Online experiments with American and Indian participants. *PLoS One, 12*(2), e0171252. http://dx.doi.org/10.1371/journal.pone.0171252.
Northover, S. B., Pedersen, W. C., Cohen, A. B., & Andrews, P. W. (2017). Artificial surveillance cues do not increase generosity: Two meta-analyses. *Evolution and Human Behavior, 38*(1), 144–153.
Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573–577. http://dx.doi.org/10.1038/31225.
Nunez, J. M., Casey, B. J., Egner, T., Hare, T., & Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *NeuroImage, 25*(1), 267–277. http://dx.doi.org/10.1016/j.neuroimage.2004.10.041.
Paulhus, D. L., & Levitt, K. (1987). Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology, 52*(2), 245–259.
Perea, M., Rosa, E., & Gomez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition, 30*(1), 34–45.
Piaget, J. (1948). *The moral judgement of the child* (1st American edn). Glencoe, IL: Free Press.
Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science, 27*(9), 1192–1206. http://dx.doi.org/10.1177/0956797616654455.
Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature, 489*(7416), 427–430. http://dx.doi.org/10.1038/nature11467.
Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications, 5*, 3677. http://dx.doi.org/10.1038/ncomms4677.
Rom, S. C., Weiss, A., & Conway, P. (2017). Judging those who judge: Perceivers infer the roles of affect and cognition underpinning others' moral dilemma responses. *Journal of Experimental Social Psychology, 69*, 44–58.
Sacco, D. F., Brown, M., Lustgraaf, C. J., & Hugenberg, K. (2017). The adaptive utility of deontology: Deontological moral decision-making fosters perceptions of trust and likeability. *Evolutionary Psychological Science, 3*(2), 125–132.
Sanders, G. S., Baron, R. S., & Moore, D. L. (1978). Distraction and social comparison as mediators of social facilitation effects. *Journal of Experimental Social Psychology, 14*(3), 129–303.
Schlenker, B. R., & Weigold, M. F. (1990). Self-consciousness and self-presentation: Being autonomous versus appearing autonomous. *Journal of Personality and Social Psychology, 59*(4), 820.
Shah, J. Y., Friedman, R., & Kruglanski, A. W. (2002). Forgetting all else: On the antecedents and consequences of goal shielding. *Journal of Personality and Social Psychology, 83*(6), 1261–1280.
Shenhav, A., & Greene, J. D. (2014). Integrative moral judgment: Dissociating the roles of the amygdala and ventromedial prefrontal cortex. *The Journal of Neuroscience, 34*(13), 4741–4749. http://dx.doi.org/10.1523/JNEUROSCI.3390-13.2014.
Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59*(2), 204–217.
Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., Herrmann, E., Gilby, I. C., ... Melis, A. (2012). Two key steps in the evolution of human cooperation: The interdependence hypothesis. *Current Anthropology, 53*(6), 673–692.
Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition, 126*(2), 326–334. http://dx.doi.org/10.1016/j.cognition.2012.10.005.
Valdesolo, P., & Desteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*(6), 476–477. http://dx.doi.org/10.1111/j.1467-9280.2006.01731.x.
Villarosa, M., Kison, S., Madson, M., & Zeigler-Hill, V. (2016). Everyone else is doing it: Examining the role of peer influence on the relationship between social anxiety and alcohol use behaviours. *Addiction Research & Theory, 24*(2), 124–134.
Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-presentation depletes regulatory resources. *Journal of Personality and Social Psychology, 88*(4), 632–657. http://dx.doi.org/10.1037/0022-3514.88.4.632.
Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology, 16*(1), 155–188.
Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin, 24*(12), 1251–1263.
Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367*(1594), 1310–1321. http://dx.doi.org/10.1098/rstb.2011.0416.