*Konrad Bocian **
*Bogdan Wojciszke ***

Unawareness of Self-interest Bias in Moral Judgments of Others' Behavior

Abstract: Previous studies (Bocian & Wojciszke, 2014) showed that self-interest biases moral perception of others' unethical actions. Moreover, affective changes in attitudinal responses towards the perpetrator of an immoral act drives the bias. In the present studies, we attempted to answer the question whether people are aware of the self-interest bias in their judgments of others' behavior. We conducted two experiments showing that moral judgments of verbally described and imagined actions were dominated by norms rather than self-interest (Study 1) and that people were not aware that self-interest distorted their moral judgment (Study 2). The unawareness of the self-interest bias among the participants was attributable to omission of their own emotional responses when forecasting their moral judgments. We discuss the importance of emotions presence in studies on moral judgments as well as contribution of the present research to the intuitionist approach to moral judgment.
Key words: self-interest; moral judgment; egotistic bias

Moral judgments play a paramount role in social life. Moral persons are lionized and held in high esteem while immoral villains are despised and rejected. When an issue becomes a really important problem, the social discourse on the issue quickly acquires moral undertones. Most actions, which are performed frequently or in the high intensity, tends to be seen as justified and morally good. Contemporary social cognition researchers distinguish between two fundamental dimensions of content – agency (competence) and communion (warmth) – and morality is a major component of communion, which is more important of the two dimensions. A great deal of research on social cognition shows that perceptions of individuals (Abele & Wojciszke, 2014) and social groups (Leach, Bilali & Pagliaro, 2013) is dominated by communal/moral over agentic contents. Information on others' morality is typically more sought for than information on other qualities; it is processed faster and exerts much greater influence on attitudes and evaluations than other sorts of information, even if balanced for their favorability (positivity-negativity) (Wojciszke, 2010). No wonder then, that for a long time moral judgment has been an important object of psychological theorizing and research.

There are two distinct traditions in theorizing about origins of moral judgments. According to an older, rationalistic tradition (having roots in the philosophy of Immanuel Kant) moral judgment involves thoughtful reasoning heavily embedded in language and requires time, effort and cognitive resources (e.g. Kohlberg, 1984; Thomson, 1990). According to a newer, intuitionist tradition (having roots in the philosophy of David Hume) moral judgment resembles perception rather than thinking and involves affective intuitions (gut feelings of right or wrong),which instantly and effortlessly appear in awareness and require little time or cognitive resources (Haidt, 2001).

A variety of findings reviewed by Haidt and Kesebir (2010) shows that moral judgments are typically dominated by automatic intuitions over thoughtful deliberations. For example, people make moral judgments instantly (within a quarter of a second) and developmentally before they acquire a full capacity to think in moral terms (i.e. in infancy - Hamlin, Wynn& Bloom, 2007; 2010). Those instant perceptions are hard to inhibit or change by conscious efforts and, therefore, frequently remain unchanged by deliberations. Such judgments are made without effort, intention and even when the judges

* Konrad Bocian, University of Social Sciences and Humanities, Sopot Campus, Polna 16/20, 81-745 Sopot, Poland; kbocian1@swps.edu.pl
** Bogdan Wojciszke, University of Social Sciences and Humanities, Sopot Campus, Polna 16/20, 81-745 Sopot, Poland; bwojciszke@swps.edu.pl.

cannot explain them. Changes in emotional states (such as unrelated surge of joy or disgust) influence moral judgments even when factual information about a judged act is held constant. Such affective influences are stronger in conditions where people lack proper information that is relevant in a particular situation (Van den Bos, 2003). Finally, impairments of brain areas related to emotions – such as the ventro-medial prefrontal cortex – impair moral judgments and healthy people showing low activation of these areas tend to ignore deontological considerations when solving moral dilemmas. Haidt and Kesebir (2010) call this domination of intuitions over thoughts "intuitive primacy but not dictatorship" (p. 801) and this primacy may be well understood in terms of a dual process model of moral judgment (Greene, 2007). According to this model, moral judgment is a product of two processes – an automatically arising intuition and a controlled deliberation which may change this intuition. The major difference between the two is that automatic processes operate constantly and independently of the conditions, whereas controlled ones operate only in generous conditions, when enough of time, motivation and cognitive resources are available. Using a process dissociation approach, Conway and Gawronski (2013) showed directly that introducing cognitive load selectively reduces operation of the controlled (but not automatic) processes, while inducing empathy for a victim increases operation of automatic processes with deliberative processes remaining intact.

Because moral judgments contain a deliberative component sometimes but comprise an automatic component always, they are highly susceptible to egocentric biases. This is because the automatic moral judgment is typically affective in nature and involves a "good-bad" response (although there are also exceptions to this rule – e.g. an embodied gesture of the hand-over-heart increases both moral judgment and moral behavior without affective mediation – Parzuchowski & Wojciszke, 2014). People automatically develop egocentric perceptions and evaluations in line with an unwritten rule "what is good or bad for me is generally good or bad" (Epley & Caruso, 2004). Continuing this line of reasoning we argue that people are prone to *the self-interest bias in moral judgments* of others. That is they perceive others' counter-normative behavior as immoral, but when the behavior in question benefits perceivers (and leads to their positive affect) they tend to evaluate this behavior in a more positive way up to seeing it as moral. "This self-interest bias arises because: (1) moral evaluations are automatically based on current affective states and (2) affective responses to stimuli are driven by their bearing on the perceivers' self-interest, such that the stimuli advancing self-interest lead to positive affect and the stimuli frustrating self-interest lead to negative affect" (Bocian & Wojciszke, 2014, p. 4).

Present Hypotheses and Studies

In line with this reasoning on self-interest bias, Bocian and Wojciszke (2014) showed in three studies that others' counter-normative behavior (breaking a rule or cheating for money) leads the perceiver to much more positive moral evaluations when he or she benefits materially from this behavior compared to a situation where only the perpetrator (the rule-breaker) benefits. The same studies showed that this change of judgment is entirely mediated by affective responses towards the perpetrator. A perpetrator whose behavior benefits observers becomes well-liked by them and this liking leads to heightened moral evaluations of his behavior (though changes in general mood do not serve as a mediator). Moreover, when observers are initially induced to dislike the perpetrator, the whole mechanism is being blocked, and the self-interest bias in moral judgment of the perpetrator does not emerge.

In the present studies, we want to continue this line of research attempting to answer the question whether people are aware of the self-interest bias in their judgments of others. Our previous studies (Bocian & Wojciszke, 2014) showed that the self-interest bias is driven by affective changes in attitudinal responses towards the perpetrator of an (immoral) act and it is well-known that such affective changes can emerge and operate without conscious awareness (Strack & Deutsch, 2004). Therefore, we predict that people can be unaware that their judgments are biased by their own interests, especially in the absence of de-biasing pressures, such as accountability (Lerner & Tetlock, 1999) or self-awareness (Wicklund, 1975).

To test this prediction we performed two studies. In first of them, we asked for moral evaluations of normative or counter-normative actions, which advanced or dwarfed interests of our participants (or other persons). Both actions and interests were verbally described, and participants were to imagine them. We reasoned that imagined interests may induce only weak affective responses (if any) making the moral judgment difficult to bias. Imagined norm-relevance, however, should still guide moral judgment because the latter is by definition about norms, whether imagined or observed. So, we predicted that norm maintenance behavior would lead to positive moral judgment, whereas norm-breaking would lead to negative judgment, self- or other-interests notwithstanding.

In second study, we arranged a situation in such a way that some participants benefited financially from rule-breaking by another person and other participants lost financially because the person kept to the rule. These two groups were called "experiencers" because they experienced the gain or loss. Additionally, we recruited two other groups of participants who only imagined one of these two situations described to them in detail. Because, these two groups were only to forecast their responses to an imaginary situation, we called them "forecasters". We predicted that moral judgments of experiencers would follow their interests and would be higher for those who experienced gain rather than loss. We also predicted that moral judgments of forecasters would not follow those interests. In effect, our hypotheses were that in the gain situation forecasters would yield lower moral evaluations than experiencers, whereas in the loss situation, forecasters would yield higher evaluations than experiencers.

Study 1

In this study we presented our participants with four short stories describing an allotment of some goods which either adhered to norms of justice or broke such norms and, at the same time, it either served or frustrated interests of a specific person. The participants were asked to imagine that they themselves were the persons or that other (unknown) individuals identified by name were the involved persons. After imaging each situation, the participants were asked to evaluate each allotment decision and to rate whether the goods were distributed in a morally appropriate or inappropriate way.

Method

Participants. Sixty-four university students participated (M_{age} = 22.06, *SD* = 4.70, 50 were women). They were randomly divided into the self-interest and other-interest groups of equal size.

Procedure and materials. Each participant received four vignettes describing morally relevant actions covering all combinations of a 2 (moral norm: maintained vs. broken) x 2 (personal interest: advanced vs. dwarfed) design. The four vignettes involved following actions: (a) promoting an employee whose qualifications were better or poorer than those of other candidates; (b) admitting for an attractive and free program a student whose test results were higher or lower than those of other candidates; (3) allotting a fringe benefit (flat for free) to an employee who qualified to a higher or lower degree than other employees; (d) allotting an attractive internship to a student who was more or less qualified than other candidates. Four versions of each story were generated covering all combinations of the norm by interest design and a Latin square was used to vary the specific vignette content in a systematic way. In the self-interest condition participants were asked to imagine that they themselves were the specific persons involved. In the other-interest condition the specific persons were given common first names (Adam, Tom etc.)

Participants read the stories in a random order and rated whether the goods in question were distributed in a morally appropriate or inappropriate way. The seven-point scale ranged from -3 (definitely inappropriate) to 0 (hard to say, somewhat appropriate, somewhat inappropriate) to 3 (definitely appropriate).

Results and Discussion

Moral judgments were subjected to an analysis of variance in a 2 (person involved: self vs. other) x 2 (moral norm: maintained vs. broken) x 2 (interest: advanced vs. dwarfed) design with repeated measures on the last two factors. This analysis revealed a strong main effect of norm $F(1,62) = 85.62$, $p < .001$, $\eta^2 = .58$, with judgments of norm maintaining behaviors being generally positive (*M* = 0.61, *SD* = 1.43) and much higher than judgments of norm-breaking, which were generally negative (*M* = -1.14, *SD* = 1.48). There was no main effect of the interest, $F(1,62) =$ 1.95, $p = .17$, ns, though an interaction between the norm and interest appeared significant, $F(1,62) = 13.88$, $p < .001$, $\eta^2 = .18$. As can be seen in Figure 1, the interaction meant that the effect of behavior normativity on judgments was more pronounced for actions dwarfing rather than serving personal interests. Specifically, norm breaking led to more negative judgments when the interest was dwarfed (*M* = -1.56, *SD* = 1.38) rather than advanced (*M* = -0.72, *SD* = 1.49) and this difference was significant, $t(63) = 3.47$, $p = .001$, *d* = .58. Similarly, judgments of norm maintenance tended to be more positive when the interest was dwarfed (*M* = 0.78, *SD* = 1.36) rather than advanced (*M* = 0.44, *SD* = 1.59) although this this difference failed to reach significance, $t(63) = 1.46$, $p = .15$.

Figure 1. Moral judgments of actions which break or maintain social norms while frustrating or advancing personal interests of involved individuals (Study 1).
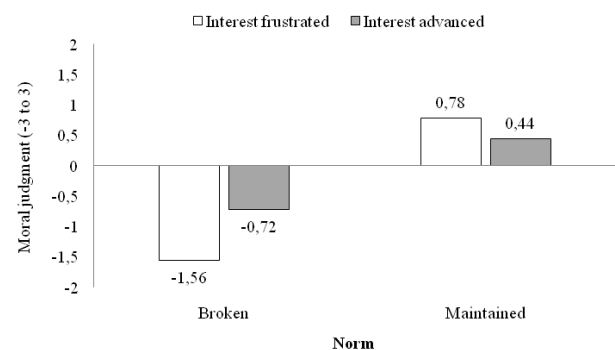
Clearly, this data shows that moral judgments of verbally described and imagined actions are dominated by norms. Analyzes with *t* tests for one sample showed that when a norm was broken, the action was condemned (i.e. the judgment was significantly below 0) both when the interest was dwarfed, $t(63) = 9.07$, $p < .001$, and when it was advanced, $t(63) = 3.87$, $p < .001$. But when the norm was maintained, the action was praised (i.e. the judgment was significantly above 0) both when the interest was frustrated, $t(63) = 4.59$, $p < .001$, and when it was advanced, $t(63) = 2.20$, $p = .032$. Interests played only a secondary role, moderating the effect of norms – the latter was more pronounced for actions minimizing rather than serving personal interests.

Importantly, the analysis revealed no triple interaction between norms, interests and whose interests were involved – the own (imagined) interest of the participants or interests of other persons described in the vignettes. In other words, the pattern depicted in Figure 1 kept both for self-interest and other-interest. That means that people do not realize the self-interest bias in moral judgment. Specifically, they do not know that when others' counter-normative behavior benefits their own interests moral judgments of this behavior become much more lenient.

Study 2

This study aimed at showing within the same design that people commit the self-interest bias (when judging others' norm-relevant behavior that is also relevant to the judges' interest) but do not realize committing the bias when they are asked only to imagine others' behavior. In the already mentioned series of studies, Bocian and Wojciszke (2014) found that participants judged counter-normative behavior in a more lenient way when they themselves gained from this behavior – compared to participants who did not gain personally from it. We borrowed and reanalyzed two conditions from Study 1 of Bocian and Wojciszke (2014). One was a condition where participants paid their fine for overdue books and another was a condition where participants had their fine waived by an arbitrary decision of a librarian. Because these two groups actually experienced the loss or gain, respectively, we call them here "experiencers". Additionally we recruited two new groups of "forecasters" for whom we described in detail the situation of experiencers, asked them to imagine themselves in this situation (of gain or loss) and to predict their responses by answering the same questions that were posed to experiencers.

We reasoned that if people realize committing the self-interest bias, the responses of forecasters would follow the responses of experiencers. That is, forecasters would predict their moral judgment and liking of the librarian to be higher when the librarian waived rather than imposed the fee (i.e. when she acted for the participants' interests, an effect which Bocian and Wojciszke found for the experiencers). However, if people do not realize the self-interest bias, their forecasted responses would diverge from the actual responses of experiencers and would not follow the interest-induced difference between waiving and imposing the fee. So, the main analyzes focused on comparing judgments of the experiencers and forecasters in two conditions of gain versus loss.

Method

Participants. Eighty-four undergraduate students participated (73 women; mean age = 24.66 years). All were approached when leaving their local library. Forty participants came from two conditions borrowed from Bocian and Wojciszke (2014, Study 1). At about the same time, we also recruited 44 other students who served as forecasters, that is they only imagined themselves in the condition of gain or loss (to which they were allotted on a random basis).

Procedure and materials. The study started when a student, who had to pay a fine for overdue books, showed up in the library. The librarian informed participants about the amount of fine and then based on a random selection either waived (M_{gain} = 15zł; SD = 12.53) or imposed the fine M_{loss} = 12.1zł; SD = 8.42); the difference between the two means was not significant, p = .396.

When the students left the library, they were followed by a female confederate for three minutes and then approached and asked to fill in a marketing department evaluation survey concerning two University departments: the library and the dean's office (with a random order of departments). Firstly participants evaluated the infrastructure of departments (e.g. chairs, lights, speed of internet) and then rated quality of the staff's work ("*imagine a person with whom you had the latest contact*"). Specifically, participants rated on 7–point scales (1 = *definitely not*, 7 = *definitely yes*) how much they agreed that, in their work, the employees were: *efficient, fast, sensible, accurate, capable and lazy* and those adjectives were fillers for the three items measuring moral evaluations: *honest, fair, and proper*. Finally, at the end, students rated on a similar scale (1 = *strongly negative*, 7 = *strongly positive*) their overall attitudes towards the library and the dean's office staff.

The forecasters were students who left the same library. They were approached by the same confederate and asked to fill in a questionnaire about imagination improvement technics. On the front page participants were informed that:

"*The survey, which you hold in your hands, concerns the ability to imagine hypothetical life situations. Read carefully the story described on the following page and as far as possible to imagine its parts: places, people, situation, etc. After you read the story, we will ask you some questions about it, so please visualize the story as best as you can. Note that there are not correct answers, because what matters to us most of all is what you think and feel making an imagination of characterized story*".

On the next page, participants read a story in the first person about the librarian whose behavior was exactly the same as for experiencers. Dependently on the experimental condition, half of the participants read that the librarian behaved in the counter-normative way waiving the fine, whereas the other half learnt that the librarian behaved in a normative way imposing the due fine. This one element (gain vs. loss) was the only difference between the two stories.

When participants read the whole story, we asked them to imagine that immediately after walking out the library another student approached them and kindly requested to take part in a brief survey concerning two University departments: the library and the dean's office (the same as for experiencers). The entire survey was on the next page with an instruction to "*Fill the survey in a way that you think you would have filled it after your fine for overdue books was imposed (waived) by the librarian*". After participants completed the survey the confederate thanked them for their participation and walked away in search of another person.

Moral ratings of the library and dean office employees' actions were averaged into a moral evaluation index (Cronbach's α = .86 and α = .93). The second dependent measure was an overall liking measured with a single item on a scale ranging from 1 to 7.

Results and Discussion

Moral evaluations were subjected to a 2 x 2 analysis of variance with the self-interest (gain vs. loss) and group (experiencers vs. forecasters) as between-participants factors. This analysis revealed a significant interaction between the two factors $F(1,84) = 17.63$, $p < .001$, $\eta^2 = .18$, as illustrated in Figure 2. As expected, the forecasters produced predictions of moral judgments which were far from reality (i.e. the judgments of experiencers). In the gain condition, the forecasters showed lower moral evaluations of the librarian ($M = 5.07$, $SD = 1.45$) than experiencers who actually gained from the librarian's counter-normative behavior ($M = 6.22$, $SD = .80$), $t(40) = 3.09$, $p = .004$, $d = .98$. Importantly, an opposite pattern emerged in the loss condition. This time the forecasters showed higher moral evaluations ($M = 6.11$, $SD = 0.80$) than those of experiencers ($M = 5.30$, $SD = 1.02$), $t(40) = 2.86$, $p = .007$, $d = .90$. Neither of the main effects was significant – for the self-interest $p = .807$, for the group $p = .472$. Of course, as reported by Bocian and Wojciszke (2014), the experiencers showed higher evaluations in the gain than loss condition.

Figure 2. Ratings of the librarian's morality as a function of self-interest and group conditions. Bars represent standard errors of the mean (Study 2).
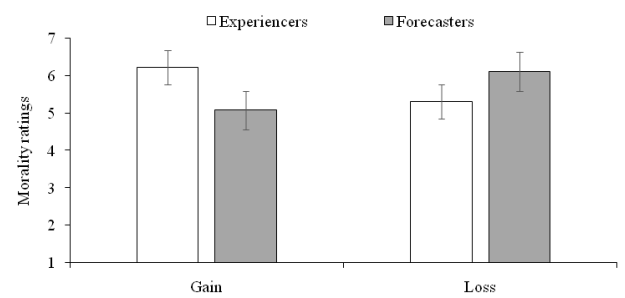
A similar analysis of variance was conducted on the overall attitude towards the librarian. The analysis revealed no main effect of the group, $p = .232$. The main effect of self-interest was significant $F(1,84) = 6.20$, $p = .015$, $\eta^2 = .07$, indicating higher liking in the gain condition ($M = 5.86$, $SD = .90$) than in the loss condition ($M = 5.33$, $SD = 1.39$). However, interpretation of this effect was constrained by the predicted group by interest interaction $F(1,84) = 19.93$; $p < .001$, $\eta^2 = .20$. As can be seen in Figure 3, in the gain condition the forecasters predicted that they would like the librarian to a lesser degree ($M = 5.50$, $SD = .86$) than she was actually liked by the experiencers ($M = 6.25$, $SD = .79$), $t(40) = 2.94$; $p = .005$, $d = .91$. However, in the loss condition the forecasters predicted to like the librarian more ($M = 5.95$, $SD = .95$), than the experiencers actually did ($M = 4.65$, $SD = 1.50$), $t(31,63) = 3.34$; $p = .002$, $d = 1.03$. This pattern of results shows again that students from our sample were poor forecasters of their judgments of others' behavior relevant for their own interests.

Figure 3. Liking of the librarian as a function of the self-interest and group conditions. Bars represent standard errors of the mean (Study 2).
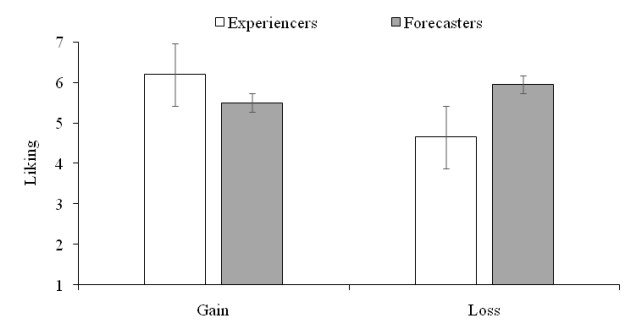
Study 2 confirms that people are lousy forecasters of own responses to imagined situations compared to responses of people who actually experience the same situations. Specifically, they do not realize that self-interest biases their moral judgments of norm-relevant behavior in such a way that the judgments become higher when the behavior in question advances their interests and become lower when the behavior dwarves their interests. Interestingly, our forecasters did not even realize that their liking responses would have been influenced by the simple self-interested logic – that people like well those who serve their interest but dislike those who frustrate those interests. Existing research suggests that the main reason for such inaccurate predictions is that when people try to forecast their future states or judgments they tend to forget about own emotions dismissing them because these emotions are absent during anticipation (though they are present in real experience, cf. Wilson, Gilbert & Centerbar, 2003).

General Discussion

The results of the present two studies suggest that people are aware that their moral judgments are guided by norms, though they are not aware that their moral judgments reflect also their own interests. Thus, they do not realize the self-interest bias their judgments of others' behavior. As already discussed, Bocian and Wojciszke (2014) evidenced this bias and showed that it is mediated by increases in liking of the person who immorally acts in a way which benefits the perceiver. Unawareness of the self-interest bias is consistent with the affective nature of moral intuitions – by definition people are always aware of their thoughtful deliberations (concerning norms), but they may be unaware of their affective responses and the ways the latter feed the conscious thoughts (cf. Strack & Deutsch, 2004).

This results also corresponds with intuitionist approach to moral judgments (Greene, 2007; Haidt, 2007) showing that emotional intuitions (liking vs. disliking) are underlying moral judgments as we have shown in previous studies (Bocian & Wojciszke, 2014). Thus, when true emotions are missing, as in the present studies, we are dealing with moral reasoning, which is a post-hoc structure (Haidt, 2001; Wheatley & Haidt, 2005), rather than a proper moral judgment.

Yet another issue emerging from the present studies concerns the validity of methods used to study moral judgments. From the venerable Heinz-type dilemmas studied by Kohlberg (1984) to classical trolley-type dilemmas studied to discern between deontological and utilitarian inclinations in moral decision making (Conway & Gawronski, 2013) psychologists require from their participants making moral judgments of imaginary behavior of others. Although this dilemma methodology provided many insights into moral judgments, the present results clearly suggest some limitations of this methodology. The validity of such methods seems to be constrained to the operation of rational contributions to moral judgments. Imaginary dilemmas are not useful to study automatic (affective and associative) operations because people are not aware of them and, therefore, cannot take them into account when forecasting their moral judgments. Studying judgments of real behaviors that concern the perceivers personally seems to be a more fruitful method to study these phenomena. However, this assumption needs to be investigated more precisely.

The current research brings evidence that moral judgment is guided by self-interest, but this mechanism is oblivious for people. Thus, it contributes to a growing body of research on intuitionist approach to moral judgments (Greene, 2007; Haidt, 2007), moral hypocrisy (Batson, Thompson, Seuferling, Whitney & Strongman, 1999; Valdesolo & DeSteno, 2007; 2008; cheating (Mazar, Amir & Ariely, 2008; Shu, Gino & Bazerman, 2011) and ethical actions of business leaders (Hoogervorst, De Cremer & van Dijke, 2010). Furthermore, as mentioned at the outset, Bocian and Wojciszke (2014) showed that perceiver's interests biased the moral recognition of other's immoral actions, and the present findings suggest that people are absolutely unaware that self-interest distorts their moral judgments. This underlines importance of emotions and personal concerns when it comes to research on moral opinions. It seems that only when we get real emotions involved, we are capable to capture the true nature of moral judgment since people make moral judgments instantly and really fast: within 200 to 250 milliseconds (Van Berkum, Holleman, Nieuwland, Otten & Murre, 2009).

We believe that self-interest bias, which is opaque for people committing the bias, can be reduced by accountability (Lerner& Tetlock, 1999), self-awareness (Wicklund, 1975; Batson, Kobrynowicz, Dinnerstein, Kampf, andWilson), primes of social control such as a pictureof human eyes (Bateson, Nettle& Roberts, 2006) and probably by raising people awareness about the self-interest bias and its underlying mechanism. These actions could help us with creating environments more conducive to moral behavior in everyday life.

References

Abele, A., & Wojciszke, B. (2014). Communal and Agentic Content in Social Cognition: A Dual Perspective Model. Article in press in *Advances in Experimental Social Psychology, Vol.50,* 195-255.

Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2,* 412–414.

Batson, C. D., Kobrynowicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72,* 1335–1348.

Batson, C. D., Thompson, E. R., Seuferling, G., Whitney, H., & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77,* 525-537.

Bocian, K., & Wojciszke, B. (2014). Self-Interest Bias in Moral Judgments of Others' Actions. *Personality and Social Psychology Bulletin*, *Vol. 40(7),* 898– 909.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology, 104,* 216-235.

Epley, N., & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research, 17,* 171-187.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual process theory of moral judgment explains. *Trends in Cognitive Science, 11,* 322-323.

Haidt, J., & Kesebir, S. (2010). Morality. W: S., Fiske,& D. Gilbert (red.) *Handbook of Social Psychology, 5th Edition.* (797 – 832) Hobeken, NJ: Wiley.

Haidt, J. (2007). The new synthesis in moral psychology. *Science, 316,* 998-1002.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

Hamlin, J. K., Wynn, K.,& Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557-559.

Hamlin, J.K., Wynn, K.,& Bloom, P. (2010). Three-month-old infants show a negativity bias in social evaluation. *Developmental Science, 13,* 923-929.

Hoogervorst, N., De Cremer, D. & Dijke, M. H. van (2010). Why Leaders Not Always Disapprove of Unethical Follower Behavior: It depends on the leader's self-interest and accountability. *Journal of Business Ethics, 95,* 29-41.

Kohlberg, L. (1984). *The psychology of moral development: Moral stages and the idea of justice.* San Francisco: Harper Row.

Leach, C. W., Bilali, R., & Pagliaro, S. (2013). Groups and morality. In J. Simpson & J. F. Dovidio (Eds.). *APA Handbook of Personality and Social Psychology, Vol. 2: Interpersonal Relationships and Group Processes.* Washington, DC: American Psychological Association.

Lerner, J. S., Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255-275.

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.

Parzuchowski, M., Wojciszke, B. (2014). Hand over Heart Primes Moral Judgments and Behavior. *Journal of Nonverbal Behavior, 38,* 145-165.

Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37,* 330-349.

Strack, F., Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220-247.

Thomson, J. J. (1990). *The realm of rights.* Cambridge MA: Harvard University Press.

Valdesolo, P. & DeSteno, D. (2007). Moral hypocrisy: Social groups and the flexibility of virtue. *Psychological Science, 18,* 689-690.

Valdesolo, P., & DeSteno, D. (2008). The Duality of Virtue: Deconstructing the Moral Hypocrite. *Journal of Experimental Social Psychology, 44,* 1334-1338.

Van Berkum, J. J. A., Holleman, B., Nieuwland, M. Otten, M., & Murre, J. (2009). Right or wrong? The brain's fast response to morally objectionable statements. *Psychological Science, 20,* 1092-1099.

Van den Bos, K. (2003). On the subjective quality of social justice: The role of affect as information in the psychology of justice judgments. *Journal of Personality and Social Psychology, 85,* 482-498.

Wheatley, T., & Haidt, J. (2005). Hypnotically induced disgust makes moral judgments more severe. *Psychological Science, 16,* 780-784.

Wicklund, R. A. (1975). Objective self-awareness. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 8, pp. 233-275). New York: Academic Press.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P.Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 35, pp.345–411). San Diego, CA: Academic Press.

Wilson, T. D., Gilbert, D. T., & Centerbar, D. B. (2003). Making sense: The causes of emotional evanescence. In I. Brocas & J. D. Carrillo (Eds.), *The psychology of economic decisions. Vol. 1: Rationality.* Oxford University Press.

Wojciszke, B. (2010). *Sprawczość i wspólnotowość. Podstawowe wymiary spostrzegania społecznego* [Agency and Communion as the Basic Dimensions of Social Cognition].Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

Acknowledgments