Journal of Personality and Social Psychology
2011, Vol. 101, No. 4, 791–811

0022-3514/11/$12.00 DOI: 10.1037/a0023556

On the Benign Qualities of Behavioral Disinhibition: Because of the Prosocial Nature of People, Behavioral Disinhibition Can Weaken Pleasure With Getting More Than You Deserve

Kees van den Bos
Utrecht University

Paul A. M. Van Lange
VU University Amsterdam

E. Allan Lind
Duke University

Leonie A. Venhoeven, Dagmar A. Beudeker, Florien M. Cramwinckel, Linda Smulders, and Jorien van der Laan
Utrecht University

This article focuses on social situations in which people are surprised about what is happening and inhibited about how to respond to the situation at hand. We study these situations by examining a classic topic in social psychology: how people respond to receiving better outcomes than are deserved. In these situations, the actions of an authority or a coworker push in the direction of accepting and enjoying the unfair outcome, whereas personal values for most people push in the direction of rejecting or being displeased with the outcome. This conflict may inhibit people's response to the advantageous but unfair outcomes. If people are indeed inhibited about how to respond to these kinds of outcomes, then lowering behavioral inhibition by reminding people of having acted in the past without inhibitions (in a manner that is unrelated to the outcomes participants subsequently receive) should affect reactions to the outcomes. Specifically, we hypothesize that because many people are prosocial and want to adhere to principles of fairness, reminders of behavioral disinhibition will lead to less pleasure with the unfairly obtained outcomes. The results of 8 experiments (conducted both inside and outside the psychology laboratory) revealed evidence for this benign disinhibition effect on various reactions to outcomes that are better than deserved. In further accordance with our line of reasoning, the effect is particularly pronounced among those who adhere to a prosocial orientation or who have adopted a prosocial mindset and is not observed among those with proself orientations or mindsets.

Keywords: unfairness, behavioral disinhibition, social value orientations, benign reactions

A recurring theme in social psychology is the notion that the social situation in which people find themselves can overwhelm their individual inclinations. This has been a core message of classic studies in our field, such as Asch's research on public conformity (e.g., Asch, 1951, 1955, 1956), Milgram's work on obedience to authority (e.g., Milgram, 1963, 1974), and Latané and Darley's research on bystander nonintervention (e.g., Darley & Latané, 1968; Latané & Darley, 1968, 1970). A central outcome of these studies has been that people ultimately comply with the situational pressures put on them. The present article focuses on a somewhat less noticed result and a rather different aspect of the social psychology of situations like those encountered in these classical studies. That is, we note that in these situations people are generally surprised, conflicted, and even flabbergasted (i.e., "extremely surprised"; Collins Cobuild Dictionary, 1992) by what is happening and often do not know how to respond to the situation at hand.

Consider, for example, the immediate reactions of the people who participated in the studies noted above. Asch (1956) reported

This article was published Online First May 16, 2011.

Kees van den Bos, Leonie A. Venhoeven, Dagmar A. Beudeker, Florien M. Cramwinckel, Linda Smulders, and Jorien van der Laan, Department of Social Psychology, Utrecht University, Utrecht, the Netherlands; Paul A. M. Van Lange, Department of Social and Organizational Psychology, VU University Amsterdam, Amsterdam, the Netherlands; E. Allan Lind, Fuqua School of Business, Duke University.

Dagmar A. Beudeker is now at the Department of Social and Organizational Psychology, Leiden University, Leiden, the Netherlands; Florien M. Cramwinckel is now at the Rotterdam School of Management, Erasmus University Rotterdam, Rotterdam, the Netherlands; and Jorien van der Laan is now at the Research Centre for Social Care, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands.

The research reported in this article was supported by VICI Innovational Research Grant 453.03.603 from the Netherlands Organization for Scientific Research awarded to Kees van den Bos. We thank Maurits Masselink, Mark Noort, and Martine van Dool for collecting the data reported in Footnote 2; Koen Bos, Myrna Hofstee, and Rutger Tjallema for their help with Study 3; and Marieke Boerkamp, Sylvia Van den Bosch, and Sanne van Huussen for their assistance with Study 7.

Correspondence concerning this article should be addressed to Kees van den Bos, Department of Social Psychology, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands. E-mail: k.vandenbos@uu.nl

that research participants in his studies could hardly believe what was happening and that they experienced difficulty in finding an appropriate response to the wrong answers given by the confederate participants (see also Jones & Gerard, 1967; Tuddenham & McBride, 1959). Milgram (1963) observed that his participants often showed signs of nervous tension and emotional distress and disturbance. In postexperimental interviews, Milgram's participants pointed out that this was because they were not sadistic types and were having trouble trying to figure out what was going on in the situation they had encountered (see also Reeder, Monroe, & Pryor, 2008). Latané and Nida (1981) noted that an important determinant of nonintervention when bystanders are around is behavioral inhibition: In the presence of a nonintervening audience, people may feel inhibited about intervening in the situation at hand (see also Van den Bos, Müller, & Van Bussel, 2009).

A largely unanswered question in all of these contexts is how people might overcome the pressure that is causing them such confusion and anxiety and act in accord with their own personal assessments of the situation. In the Asch, Milgram, and Latané and Darley studies, of course, the pressure of the social situation can be broken by the action of others to resist (see, e.g., Asch, 1951, 1955, 1956; Latané & Darley, 1968, 1970; Milgram, 1963, 1974), but this is just substituting one social force for another. We asked whether there was a mechanism that might allow individuals, without outside social support, to resolve things in terms of their own values.

In the current article we argue that feelings of surprise and inhibition often arise from a deep-seated conflict between social pressures and personal values. Because people are social beings (Aronson, 1999; see also De Waal, 1996), they want to act in concert with their fellows and the authorities present in the situation. Furthermore, because most people (but not all) adhere to prosocial values (e.g., Van Dijk, De Cremer, & Handgraaf, 2004; Van Prooijen, De Cremer, Van Beest, Stahl, & Van Lange, 2008), many want to do what is normatively appropriate and good and correct. In situations that pose a conflict between responding in accordance with what seems to be accepted by others versus reacting in terms of what seems the right thing to do, these two important determinants of social behavior are pushing in different directions. The net result is that people in such situations are flabbergasted, confused about what to do (Asch, 1956), emotionally distressed (Milgram, 1974), and inhibited regarding how to respond (Latané & Nida, 1981).

But suppose it were possible to "turn off" the social inhibitions that limit actions in line with personal preferences? Certainly most of us are able to throw off social pressure at times, and given this capacity for "disinhibition," there must be a mechanism for putting aside the social demands involved in the social confusion experiences we address here. We suggest that the activation of a disinhibition process—by recalling instances of relatively independent and unfettered action in the past (Studies 1–7) or even by simply attending to statements about not or only weakly caring about what others think of your actions (Study 8)—will allow people to overcome the pressures of conformity or social authority and follow their personal preferences to enact behaviors that fit with their personal values.

We study the possible implications of this analysis by examining reactions to a classic situation-based conflict of social and normative pressures, namely, people's reactions to receiving better outcomes than they deserve (see, e.g., Adams, 1963a, 1963b, 1965; Adams & Jacobsen, 1964; Adams & Rosenbaum, 1962; Blau, 1964; Jacques, 1961). We examined pleasure, displeasure, and related reactions to situations in which a person's own outcomes are inequitably better than the outcomes of comparable other people (Van den Bos, Peters, Bobocel, & Ybema, 2006). This situation contains the same essential elements as the studies by Asch (1956), Milgram (1974), and Latané and Darley (1968)—the actions of an authority (in our studies, the experimenter, a coworker who knows more than you, or your boss) push in the direction of accepting and enjoying the unfair outcome, whereas personal values for most people push in the direction of rejecting or at least discounting the outcome (Lind & Tyler, 1988). So, we ask how people react to getting more than they deserve.

Reactions to Getting More Than You Deserve

People are deeply affected when they get outcomes that are better than they deserve (Adams, 1963a, 1963b, 1965). Consider, for example, Adams's observations concerning workers at General Electric in the early 1960s, which stimulated him to begin his impressive research program on equity theory (see, e.g., Adams, 1963a, 1963b, 1965; Adams & Jacobsen, 1964; Adams & Rosenbaum, 1962). Adams noted that employees seek to maintain equity between the inputs that they bring to a job and the outcomes that they receive from it in comparison with the perceived inputs and outcomes of others (Adams, 1965). When individuals find themselves participating in inequitable relationships, they become distressed. As a result, the person who gets too much may feel guilt or shame (Adams, 1965) and often report "feelings of unease" about getting too much (Jacques, 1961; S. L. Peters, 2005).

Of course, those who get too little may feel angry or humiliated. It is noteworthy, though, that being underpaid is much easier to interpret than being overpaid (Adams, 1965; Jacques, 1961; S. L. Peters, 2005), and hence it is not very likely to lead to confusion or inhibition. After all, being underpaid often energizes immediate action or action tendencies (e.g., anger, reactance, protest). Yet being overpaid is an interesting situation in which immediate action is strongly tempered by inhibition—because guilt, shame, and feelings of unease do not directly imply action, and the conflict between advantageous outcomes and personal values triggers feelings of "what's going on here?" and behavioral inhibition.

Adams's (1965) research program studied many aspects of the social psychology of equity, but his true fascination with the topic focused on people's reactions to getting too much. Part of this fascination came from people's reactions regarding equity restoration. For example, the employees at General Electric who got too much tried to restore the inequity by working harder, thus raising the inputs they brought to the working relationship and making their situation more equitable (Adams, 1965). Reactions to overpayment are interesting not only in their own right but also because they carry built-in social–cognitive conflicts (Van den Bos et al., 2006), making it difficult for the overpaid person to know how to respond to the outcome. After all, when a person experiences advantageous inequity, there is a conflict between the tendency to accept a desirable outcome (and in doing so avoiding a challenge to the authority who has given the outcome) and doing what is fair. Furthermore, in advantageous inequity there is a source of positive affect and a source of negative affect: The positive source is the

egoism-based pleasure of receiving a relatively good outcome, whereas the negative source is the fairness-based feeling of being unfairly advantaged (Van den Bos, Lind, Vermunt, & Wilke, 1997; Van den Bos, Wilke, Lind, & Vermunt, 1998). Because of this mixed-motive quality of advantageous inequity arrangements (Jacques, 1961; S. L. Peters, Van den Bos, & Karremans, 2008), studying these situations may shed light on the relationship between people's egoism-oriented tendencies and their inclinations to do the right thing (Van den Bos et al., 2006).

In each study of the present article we confront people with situations in which they unexpectedly receive outcomes that are better than they deserve. Specifically, we study how people react to winning an iPod that they do not really deserve (Study 1); their responses to being promised €3 and then suddenly receiving €4 while another participant receives only €2 (Study 2); to being promised €2 and then receiving €3, with the result that a future participant will receive only €1 (Study 3); to receiving a bonus of €50 while a colleague receives a bonus of only €10 (Study 4); to winning a game of Trivial Pursuit among friends by unfair means (Studies 5, 6, and 8); and to obtaining an undeserved bonus of €2 (Study 7).

In these situations, we assess a number of conceptually relevant reactions people may have to these unfairly advantageous outcomes. In particular, we measure satisfaction with desirable but undeserved goods (Study 1), justice judgments and decision acceptance of outcomes that are better than deserved (Study 2), rejection of and protest against such unfair outcomes (Studies 3, 4, and 8), and feelings of pleasure (Study 5) and displeasure (Studies 6 and 7) with the unfair outcomes. Taken together, our eight studies (as well as our two pretests) enhance confidence in the robustness of the line of reasoning tested in all experiments. Importantly, what all experiments have in common is that people suddenly receive an outcome that is better than they deserved. How do people react to these kinds of outcome arrangements?

The line of reasoning that we propose begins with our assertion that when people are suddenly confronted with these advantageous but unfair outcomes, they will be surprised or even flabbergasted regarding what has happened and conflicted about how to respond. In fact, data from Pretest 1 (described in detail below) support this assertion and show that people (at least to some extent) are surprised, flabbergasted, and shaken when they receive the unfairly positive outcomes. Furthermore, we argue that the surprise and conflict elements are there in advantageous but unfair outcome situations, because the actions of an authority or a coworker push in the direction of accepting and enjoying the unfair outcome, whereas personal values for most people push in the direction of rejecting or discounting the outcome. In fact, consistent with our assertion that most people would personally be inclined to reject advantageous unfairness, data from Pretest 2 confirm that many people are prosocial beings (for an illustration with a sample representative of the adult population in the Netherlands, see Van Lange, Otten, De Bruin, & Joireman, 1997, Study 4).

The main experiments of our article examine an important and thus far untested implication of our line of reasoning. That is, we argue that if people are indeed flabbergasted and inhibited regarding how to respond to these kinds of outcome arrangements (as Pretest 1 suggests they are), then lowering behavioral inhibition by reminding people of having acted in the past without inhibitions (in a manner that is unrelated to the outcomes participants subsequently receive; see Van den Bos et al., 2009) should affect reactions to the advantageous but unfair outcomes. Moreover, we argue that because most people tend to adhere to prosocial value orientations (as Pretest 2 and Van Lange et al., 1997, suggest they do), then reminders of behavioral disinhibition should weaken pleasure with the advantageous but undeserved outcomes.

Consider the situation that we examine in Study 1. The aim of Study 1 was to demonstrate the disinhibition hypothesis in a social interaction context in which participants directly experienced and responded to the following dilemma: The participants worked together with another participant (in reality, a professional actor hired as a confederate) to arrive at correct answers on an intelligence test. The members of the best performing pair would each win an iPod. When the experimenter had left the laboratory, the actor suddenly pulled an illegally obtained note with the correct answers to the intelligence test out of his or her jacket and filled out the intelligence test using these answers, thus confronting the real participants with a dilemma regarding how to respond to now having a very good chance of obtaining an iPod but obtaining this desirable product in a patently unfair way.

Most people disapprove of unethical and unfair behaviors by others (Folger, 1984; Lind & Tyler, 1988; Tyler & Lind, 1992), but relatively few have the social fortitude needed to openly resist strong behavioral demands (e.g., Milgram, 1974), so we predicted that absent some intervention (as in the no-disinhibition conditions), participants would be inhibited from intervening against the other person's behavior and hence would indicate that they would be satisfied with the desirable product. However, when (as in the disinhibition conditions) participants had been reminded about having acted without inhibitions, they should find it easier to resist the other person's unfair behavior and hence to indicate that they would not be that satisfied with the product obtained by unfair means. Thus, following the line of thinking proposed here, we predicted that satisfaction with winning the iPod would be higher following no reminders of behavioral disinhibition than following reminders of disinhibited behavior. This is one example of our benign behavioral disinhibition hypothesis, predicting that following reminders of behavioral disinhibition people will be less pleased with advantageous unfair outcomes.

Behavioral Inhibition and Disinhibition

What do we mean when we talk about behavioral inhibition and disinhibition in this article? The concepts of inhibition and disinhibition have been used to refer to different processes in different research literatures (see, e.g., Amodio, Master, Yee, & Taylor, 2008; Carver, 2005). In the present article, we build our line of reasoning on theorizing on public inhibition, as defined by Latané and Nida (1981), and on work on the behavioral inhibition system (BIS), as developed by Carver and White (1994; see also Gray, 1972, 1990; Gray & McNaughton, 2000).

Latané and Nida (1981) noted that in public settings, such as bystander situations, the presence of others can constrain people from showing their personal inclinations. For example, in a bystander dilemma a person may want to engage in helping behavior but may be restrained from doing so because of the presence of others (bystanders) who are not helping. Similarly, we note that when responding to suddenly having a good chance of winning an iPod because of the unethical behavior of a coworker, or to getting

an outcome that is better than deserved because of an experimenter or a boss is unexpectedly giving you this outcome, people may feel inhibited to show to the coworker, experimenter, or authority that their personal inclination is to reject or be displeased with the advantageous but unfair outcome. Being in a state of behavioral disinhibition may then be helpful. After all, disinhibition—defined in this article as a state in which people do not or only weakly care about what others think of their actions (Van den Bos, 2010; Van den Bos, Müller, & Damen, in press; Van den Bos et al., 2009)—may make it easier for people to follow their own personal inclinations (which in the majority of people may be a prosocial orientation; see Pretest 2 and Van Lange et al., 1997).

Our ideas about behavioral inhibition and disinhibition are also grounded in the work of Carver and White (1994) on the BIS. Our Pretest 1 results show that when receiving an advantageous unfair outcome, people are not only flabbergasted and surprised but are also shaken and deeply affected by the situation. Carver and White argued that the BIS regulates people's responses to anxiety-related cues. This system inhibits behavior that can lead to negative or painful consequences. Certainly in social contexts, consequences such as rejection or disapproval by an authority or coworker can be negative or painful and may well inhibit people from showing displeasure with an advantageous but unfair outcome. If this is the case, then an important issue becomes how we can disengage people's inhibitory responses, with the result that they can respond more in line with their personal values when reacting to inequitably advantageous outcomes. In the studies we describe here, we attempt to produce this disengagement of the BIS by simply reminding people—using two different manipulations—that it is possible to behave without great concern for the reactions of others.[1]

The Current Research

In the current research we note that, by and large, disengagement of the BIS is seen as having negative social consequences. Whereas a very strong BIS is compatible with anxiety-related disorders (Fowles, 1993), a very weak BIS relates to primary psychopathy (Newman, MacCoon, Vaughn, & Sadeh, 2005). Low levels of BIS correspond to having no or very weak behavioral inhibitions. These low levels of BIS are usually called *behavioral disinhibition* and in the current article we use this label as well. Psychological research has shown that behavioral disinhibition may lead to antisocial acts (Lilienfeld, 1992) and psychopathological behaviors (Nigg, 2000). As a result, F. Peters et al. (2006) have referred to behavioral disinhibition as a source of unwanted acts.

Along the same lines, an important theme in moral and political philosophy has been that humans should refrain from disinhibited behavior. For example, Kant (1785/1959) proposed that if people thought more carefully about what is going on in the situation at hand before they started acting, this would lead people to do what is better for society at large. Thus, Kant was arguing that it would be conducive for the greater good if people acted with somewhat more inhibition than they normally tend to do.

In the present article we argue that although pernicious effects of behavioral disinhibition may often occur, at least some levels of behavioral disinhibition can have "benign effects" (Suler, 2004) on how people react to getting more than they deserve. That is, on the basis of the reasoning laid out in the previous sections, we hypothesized that behavioral disinhibition can weaken a person's pleasure with receiving advantageous but unfair outcomes. One of the more interesting aspects of this prediction is that it is counterintuitive and runs contrary to popular notions and some writing and research on the negative consequences of behavioral disinhibition (e.g., Kant, 1785/1959; Lilienfeld, 1992; Newman et al., 2005; Nigg, 2000; F. Peters et al., 2006). Because of this counterintuitive quality, we considered it important to show the robustness of our hypothesis concerning benign effects. For this reason, the aim of Studies 1–4 was to show evidence for the benign disinhibition hypothesis in the context of different outcome arrangements and on various reactions to these arrangements. In all four of these studies we reminded participants about times in which they acted without inhibitions (disinhibition conditions) or we reminded them about their normal actions on a regular day (no-disinhibition conditions). After this, in ostensibly unrelated parts of the studies, we confronted participants with advantageous but undeserved outcomes and observed their reactions to these outcomes. The four studies differ from each other in method and measures, but all four studies provide evidence that reminders of behavioral disinhibition can have benign effects on people's reactions to advantageous unfair outcomes.

We further argue that people are somewhat flabbergasted and inhibited regarding how to respond to advantageous but unfair outcomes because many people are social (e.g., Aronson, 1999; De Waal, 1996), indeed prosocial, beings (e.g., Van Dijk et al., 2004; Van Prooijen et al., 2008) and, as a result, care too much about what others think of their reactions (see Latané & Nida, 1981; Van den Bos et al., 2009). If this line of reasoning has merit, then lowering public inhibition by reminding people of having behaved in disinhibited ways (see Van den Bos et al., 2009) should weaken pleasure with unfairly obtained goods particularly for people who adhere to a prosocial orientation or who have adopted a prosocial mindset, but not among those with a proself orientation or mindset. Studies 5–8 show that the benign effects of reminders of behavioral disinhibition indeed are moderated in this way by social value orientations and associated mindsets.

We examine our disinhibition hypothesis in Studies 1–7 using a behavioral disinhibition manipulation that we developed and validated recently (see Van den Bos et al., 2009, in press). We also use a different manipulation of behavioral disinhibition in one of our studies (Study 8), and we note that individual differences in trait behavioral inhibition (as assessed by the Carver & White, 1994, scale) have been shown to have similar effects on reactions to moral dilemmas (Van den Bos et al., in press), as do video clips broadcast on public television in the Netherlands that affect state

[1] On the basis of the work by Gray (e.g., 1972, 1990; Gray & McNaughton, 2000) and others (e.g., Fowles, 1993), Carver and White (1994) assumed that two general and orthogonal systems orchestrate adaptive behavior. The first system is the behavioral inhibition system (BIS), and the second system is the behavioral activation system (BAS). This latter system controls appetitive motivation. The BIS and BAS represent different structures in the nervous system, and Carver and White and Gray (1987) assumed these systems to be orthogonal. In the present article we focus on the BIS. The disinhibition manipulation that we use in Studies 1–7 has been shown to influence state BIS and not state BAS (Van den Bos et al., 2009).

behavioral inhibition (as measured by a state version of the Carver & White, 1994, scale; Van den Bos & Griffioen, 2011).

Specifically, building on earlier manipulations successfully used in various domains of experimental social psychology (see, e.g., Greenberg, Solomon, & Pyszczynski, 1997; Loseman, Miedema, Van den Bos, & Vermunt, 2009; Van den Bos, Poortvliet, Maas, Miedema, & Van den Ham, 2005; Van Prooijen, Van den Bos, & Wilke, 2002), Studies 1–7 ask participants to complete three simple open-ended questions that remind them about their thoughts and feelings about having behaved without inhibitions. Participants were instructed as follows:

> The purpose of this questionnaire is to assess how people react to being disinhibited, that is, how people behave when they do not care about what others think of their reactions and what feelings they then experience. To this end, please complete the following three questions: Please briefly describe a situation out of your own life in which you acted without inhibitions. Please briefly describe how you behaved in the situation in which you acted without inhibitions. Please briefly describe the emotions that you experienced when you acted without inhibitions.

In the no-disinhibition condition participants received the following instruction:

> The purpose of this questionnaire is to assess how people experience a normal day in their lives, that is, how people usually behave on a regular day and what feelings they then experience. To this end, please complete the following three questions: Please briefly describe a situation out of your own life in which you acted in a normal way like you do on a regular day. Please briefly describe how you behave when you act in a normal way like you do on a regular day. Please briefly describe the emotions that you experience when you act in a normal way on a regular day.

Van den Bos et al. (2009) showed that reminding (vs. not reminding) participants in this way of having acted without behavioral inhibitions successfully lowers behavioral inhibition (assessed by a state version of the most popular and well-validated measure of BIS sensitivity by Carver & White, 1994). Furthermore, this manipulation does not trigger behavioral activation (no effects were found on state versions of the Carver & White, 1994, scales measuring behavioral activation), nor does it influence positive or negative affective states (no effects were found on the Positive and Negative subscales of the Positive and Negative Affect Schedule [PANAS] by Watson, Clark, & Tellegen, 1988). These findings indicate that our disinhibition manipulation is not some kind of action priming or an affect manipulation, but it is a manipulation that, as intended, does lower behavioral inhibition (Van den Bos et al., 2009). Furthermore, recent findings show that the manipulation does not affect self-monitoring or experienced accountability or self-awareness.[2]

Participants in the studies that used this manipulation indicated no suspicion of the procedures employed during the disinhibition manipulation, nor did they suspect a direct relationship between the manipulation and their subsequent reactions in other parts of the experiments they were taking part in (Van den Bos et al., 2009). Furthermore, Van den Bos et al. (in press) showed that the disinhibition manipulation yields comparable effects as do differences on Carver and White's (1994) measure of trait BIS. And in all the studies reported in the Van den Bos et al. (2009, in press) articles, gender did not interact with the effects of the disinhibition manipulation. Gender also did not affect the findings we report here. Thus, the reminders of behavioral disinhibition that we use in the majority of our studies here (Studies 1–7, but see Study 8 for a different manipulation) weaken behavioral inhibition, do not influence behavioral activation, do not influence affective states or self-monitoring or accountability, do not trigger strong experimenter demands, and are robust with respect to gender differences (Van den Bos et al., 2009, in press).

Pretest 1: Feelings of Surprise Following Advantageous Unfair Outcomes

Before conducting our eight main experiments, we checked in a pretest on 54 students at Utrecht University (29 men and 25 women)[3] whether a situation in which people suddenly receive an advantageous but undeserved outcome does indeed lead them to be somewhat surprised and flabbergasted, as our line of reasoning holds. Toward this end, we presented participants with the situation that in our eight main experiments may stimulate only mild surprise and trigger the least flabbergasted reactions. We reasoned that if we were to find effects in the least impactful situation present in our studies, we then would be on relatively safe ground in assuming that the effects are there in more impactful situations. Specifically, as in Studies 5, 6, and 8 of the present article, we asked some of our pretest participants ($n = 28$) to read and respond to the following situation:

> Each month you and your friends organize an evening of playing games. This month the night has nearly ended and you and your friends are playing Trivial Pursuit. One person who is playing the game with you has to answer one more question correctly to win the game. You happen to see that the person who is asking the questions is in fact posing a difficult question out of the wrong category to your opponent. Your opponent does not know the correct answer to this question. You end up winning the game.

In the control condition, the other participants ($n = 26$) read and responded to the following, nonsurprising version of the scenario:

[2] Using 42 students at Utrecht University, we assessed the effects of the disinhibition manipulation on situational self-monitoring, accountability, and self-awareness. Findings showed that the disinhibition manipulation did not significantly affect a state version of the Lennox and Wolfe (1984) Self-Monitoring Scale. This scale consists of two subscales (Ability to Modify Self-Presentation and Sensitivity to Experience Behaviors of Others), and neither subscale (αs $> .68$) yielded significant effects of the disinhibition manipulation, Fs$(1, 40) < 1.08$, ps $> .30$, η_p^2s $\leq .03$, nor were the two subscales combined affected by the disinhibition manipulation, $F(1, 40) = 1.69$, $p > .20$, $\eta_p^2 = .04$. Similarly, two questions that asked whether participants thought they were now accountable to others for their actions and whether they thought others would call them to account for their conduct ($\alpha = .76$) were not influenced significantly by the disinhibition manipulation, $F(1, 40) = 1.67$, $p > .20$, $\eta_p^2 = .04$. Furthermore, the Situational Self-Awareness Scale by Govern and Marsh (2001; $\alpha = .80$) was not affected by the disinhibition manipulation, $F(1, 40) = 0.00$, $p > .95$, $\eta_p^2 = .00$.

[3] In all studies of this article, gender was proportionally distributed among conditions. Furthermore, gender did not interact with the hypotheses under consideration and hence was dropped from the analyses.

> Each month you and your friends organize an evening of playing games. This month the night has nearly ended and you and your friends are playing Trivial Pursuit. One person who is playing the game with you has to answer one more question correctly to win the game. Your opponent does not know the correct answer to the question asked to him/her. You end up winning the game.

Participants were then asked to what extent they were flabbergasted by the situation, surprised by what happened, deeply affected by the situation, and shaken by what happened. All answers were given on 7-point scales (1 = *very weakly*, 7 = *very strongly*). Because the scales showed substantial correlations (rs $> .49$, ps $< .001$), participants' ratings were averaged to yield a reliable scale ($\alpha = .87$).

As expected, participants in the surprise condition were more flabbergasted by what happened ($M = 4.04$, $SD = 1.33$) than participants in the control, nonsurprise condition ($M = 2.94$, $SD = 1.09$), $F(1, 52) = 10.80$, $p < .01$, $\eta_p^2 = .17$. There was no main effect of gender or an interaction between gender and the surprise manipulation. Thus, even when responding to one of the least impactful situations studied in this article, receiving an advantageous outcome that was undeserved led participants to be more flabbergasted than receiving the same outcome in a nonsurprising way. This supports our contention that being surprised and flabbergasted are indeed part of the experience of the kinds of situations we are studying here.

Studies 1–4: The Benign Disinhibition Effect

The aim of our first four main experiments (Studies 1–4) was to show that reminders of behavioral disinhibition can weaken pleasure with various advantageous unfair outcomes and may affect other reactions to these outcome arrangements. In all four studies we reminded participants about having acted without inhibitions (disinhibition conditions) or how they act on normal days (no-disinhibition conditions). After this, in an unrelated part of the studies, participants were confronted with advantageous but undeserved outcomes and their reactions to these outcomes were measured.

Study 1

Method

Participants and design. Sixty-three students (eight men and 55 women) at Utrecht University were randomly assigned to either the disinhibition or no-disinhibition conditions. They were paid €4 or received course credit for their participation. Thirty-one participants took part in the disinhibition condition, and 32 participants took part in the no-disinhibition condition.

Procedure and materials. The experiment was presented to the participants as consisting of two unrelated parts. In the first part, participants were asked to complete the same three open-ended questions used in Van den Bos et al. (2009, in press) to induce behavioral disinhibition. In this way, participants in the disinhibition condition were reminded about having behaved without inhibitions, whereas participants in the no-disinhibition condition were reminded about having behaved in a normal way during a regular day.

All participants then answered the 20 items of the PANAS (Watson et al., 1988). The PANAS was included as a filler task and also to assess whether our disinhibition manipulation engendered positive and negative affect. The PANAS consists of two 10-item subscales (Watson et al., 1988), one measuring positive affect (PA) and one measuring negative affect (NA), and both subscales were averaged to form reliable scales (αs = .83 and .82, respectively).

After the PANAS, participants were informed that the first part of the study had ended and that the second part would now begin. In this part, participants worked together with another participant to complete as many correct answers as they could on an intelligence test. The other participant was one of six actors hired as a confederate. The particular actor with whom the participants interacted did not influence the results presented here, nor did gender of the actors have any effects. Participants were informed that they had 5 min to fill in as many correct answers as they could as a pair. The experimenter also told them that the persons in the best pair would each receive an iPod. The experimenter then left the lab, saying that she would be back shortly.

The actor then pulled a note with the correct answers to the questions of the intelligence test out of his or her pocket and showed it to the participant. The actor explained that his or her sister was a friend of one of the experimenters and that she had slipped him or her the answers. The actor then filled in the correct answers in order to win the iPod, thus creating for participants a dilemma about how to respond to now having a very good chance of obtaining an iPod but obtaining this desirable good by unethical or unfair means. When the experimenter came back, the participants were asked to fill in a questionnaire individually. The dependent variable was measured amidst filler questions and assessed to what extent participants agreed with the statement that if they won the iPod, they would be very pleased (1 = *strongly disagree*, 7 = *strongly agree*). After participants had been paid or had received course credit for their participation, they were debriefed thoroughly. During debriefing, as in all studies of this article, participants indicated no suspicion of the procedures employed, nor did they suspect a direct relationship between the disinhibition manipulation and their reactions to the other participant. They did not perceive the disinhibition manipulation to convey any demands to respond or behave in particular ways during the experiment.

Results

PANAS. Analyses of variance on the Positive and Negative subscales of the PANAS yielded no significant effects (Fs < 0.56, ps $> .46$, η_p^2s $\leq .01$; PA: $M = 4.56$, $SD = 0.76$; NA: $M = 2.11$, $SD = 0.76$). Furthermore, controlling for the PA and NA scores in analyses of covariance did not alter the effects of our disinhibition manipulation on our dependent variable. Thus, as in Van den Bos et al. (2009), affect cannot explain the effects of our disinhibition manipulation.

Product satisfaction. Not surprisingly, satisfaction with winning the iPod was high in the condition in which participants had not been reminded about disinhibited behavior ($M = 6.06$, $SD = 1.39$). Perhaps more interestingly, as predicted by our disinhibition hypothesis, satisfaction with winning the iPod was significantly less high when people had been reminded about

disinhibited behavior ($M = 4.97$, $SD = 2.23$), $F(1, 50.00) = 5.43$, $p < .03$, $\eta_p^2 = .08$.[4]

Study 2

Study 1 reveals first evidence for our hypothesis that reminders of behavioral disinhibition can weaken pleasure or satisfaction with unfairly obtained goods. The aim of Study 2 was to replicate the benign disinhibition effect in a different setting on different reactions. To this end, participants took part in an experiment for which participants were supposed to receive €3. However, at the end of the experiment they received €4 while a better performing other participant received only €2. We then assessed two important dependent variables in the literature on advantageous unfair outcomes (see, e.g., Adams, 1965), namely, the participants' outcome justice judgments and their acceptance of the payment decision. Both measures were assessed with multiple items. Following our benign disinhibition hypothesis, we predicted that reminders of behavioral disinhibition would lead participants to consider their outcome of €4 to be less just and would make them less likely to accept the outcome.

Method

Participants and design. Thirty-eight students (18 men and 20 women) at Utrecht University were randomly assigned to either the disinhibition or no-disinhibition conditions. They were informed that they would be paid €3 for their participation. Nineteen participants took part in the disinhibition condition, and the same number participated in the no-disinhibition condition.

Procedure and materials. The experiment was presented to the participants as consisting of two unrelated parts. In the first part, we again induced the disinhibition manipulation in the same way as in Study 1, this time excluding the PANAS measure so that we could affirm that the PANAS is not an essential component of this manipulation (as seems to be the case with respect to some other salience manipulations; see, e.g., Greenberg et al., 1997; Van den Bos et al., 2005). Excluding the PANAS has as additional advantage that it makes our stimulus materials shorter and makes it easier to apply the disinhibition manipulation in various contexts.

After the disinhibition manipulation, participants were informed that the first part of the study had ended and that the second part now would begin. In this part, participants worked together with another participant to answer rather difficult knowledge questions about recent news facts. The other participant was one of seven actors hired as a confederate. The particular actor with whom the participants interacted did not influence the results presented here, nor did gender of the actors have any effects. Participants were informed that the pair with the best performance on the knowledge questions would obtain free cinema tickets. After this, the experimenter left the lab, saying that she would be back when participants had finished the knowledge questions.

We had given the confederates the correct answers to the knowledge questions. As a result, when completing the knowledge questions, the participants found out that the confederate knew most answers to the questions posed. Thus, the performance of the pair was quite good, yielding a good chance of winning the cinema tickets, and the experimenter told them so when she had returned to the lab and had checked the answers given by the pair. The experimenter then paid the participant and the confederate for their participation: The experimenter gave the participant €4, and the confederate was given €2. After this, the experimenter asked both the participant and the confederate to fill in a short questionnaire.

Embedded in several filler questions were items measuring participants' justice judgments of the payment they received: Participants were asked whether they considered their payment to be just, fair, justified, and appropriate. All items were answered on 7-point scales (1 = *very weakly*, 7 = *very strongly*), and participants' answers to these questions were averaged to yield a reliable scale of outcome justice judgments ($\alpha = .86$). We assessed participants' acceptance of the experimenter's decision by asking participants to what extent they accepted the experimenter's decision, respected the decision, were intending to comply with the decision, and were intending to help the experimenter in implementing the decision. Answers were given on 7-point scales (1 = *very weakly*, 7 = *very strongly*) and averaged to form a reliable scale of decision acceptance ($\alpha = .88$). After participants had completed all questions, they were paid for their participation and thoroughly debriefed.

Results

Outcome justice judgments. When participants had not been reminded about disinhibited behavior, they judged their outcome to be somewhat unjust ($M = 3.07$, $SD = 1.23$). However, as hypothesized by our disinhibition prediction, participants judged their outcome to be more unjust when they had been reminded about disinhibited behavior ($M = 1.90$, $SD = 0.80$), $F(1, 36) = 12.10$, $p < .01$, $\eta_p^2 = .25$.

Decision acceptance. When participants had not been reminded about disinhibited behavior they were somewhat inclined to accept the experimenter's decision ($M = 4.34$, $SD = 1.57$). Perhaps more interestingly, in accordance with our disinhibition hypothesis, participants were less inclined to accept the experimenter's decision when they had been reminded about disinhibited behavior ($M = 3.21$, $SD = 1.26$), $F(1, 36) = 6.00$, $p < .02$, $\eta_p^2 = .14$.

These findings suggest that reminding people about disinhibited behavior leads them to judge their unequal outcome to be more unfair and more unjust and that these reminders may result in participants being less willing to accept this unfair outcome. More research is needed to explore in detail the relationship between justice judgments and acceptance ratings (see, e.g., Folger, Rosenfield, Grove, & Corkran, 1979; Van den Bos, 2005). This noted, in the next study we examine whether a nonstudent sample may show more actual protest behavior against an advantageous but undeserved outcome.

[4] In each study reported here we statistically tested for heterogeneity of variance and, when it was present, controlled for heterogeneity using the Welch-Satterthwaite approach (Satterthwaite, 1946; Welch, 1947). Controlling for heterogeneity was necessary when testing the dependent variables of Studies 1 and 7, explaining the degrees of freedom reported with these tests.

Study 3

Participants of Study 3 were people from various parts of the Netherlands, with different educational backgrounds, and from different age groups, who were interviewed in a shopping mall of the fourth largest city of the Netherlands, located in the middle of the country. They were promised €2 for their participation but received €3, as a consequence of which a future participant would receive only €1 instead of €2. We observed whether participants rejected their outcome that was the result of this advantageous but undeserved arrangement of outcomes.

Method

Participants and design. Fifty-nine adults (11 men, 47 women, one gender unknown) participated in the experiment and were randomly assigned to either the disinhibition or no-disinhibition conditions. They were promised €2 for their participation in a short psychological study. Participants were interviewed in a shopping mall (Hoog Catharijne) that is located near the central railway station and the city center of Utrecht and that attracts customers from different parts of the Netherlands. The mean age of the participants was 35.26 years ($SD = 16.31$), with the youngest participant being 18 years old and the oldest participant being 66 years old. Eight percent of the participants had completed a lower form of education, 20% had completed a middle-level form of education, 40% had completed a higher form of education, and 32% had completed a university degree. Twenty-nine participants took part in the disinhibition condition, and 31 participants took part in the no-disinhibition condition.

Procedure and materials. When participants agreed to participate, the experiment again was presented to the participants as consisting of two unrelated parts. In the first part, we again induced the disinhibition manipulation in the same way as in Study 1 (including the PANAS measure, which again yielded reliable scales of positive affect and negative affect; $\alpha s = .75$ and .88, respectively).

After this, participants were paid €3 for their participation and were told that a future participant hence would receive €1 instead of €2. Our dependent variable assessed whether participants rejected their unfair overpayment. After completing some filler questions, participants were thoroughly debriefed.

Results

PANAS. Analyses of variance on the Positive and Negative subscales of the PANAS yielded no significant effects (all $F s < 0.25$, $p s > .62$, $\eta_p^2 s \leq .01$; PA: $M = 3.24$, $SD = 0.44$; NA: $M = 1.60$, $SD = 0.66$). Furthermore, controlling for the PA and NA scores did not alter the effects of our disinhibition manipulation on our dependent variable. Thus, as in Study 1 and the Van den Bos et al. (2009) studies, affect cannot explain the effects of our disinhibition manipulation.

Rejection of unfair overpayment. In the condition in which participants were not reminded about disinhibited behavior, only three participants rejected the €3. However, when participants had been reminded about disinhibited behavior, 3 times as many participants rejected the unfair overpayment ($n = 9$). As predicted by our disinhibition hypothesis, this difference between the disinhibition and no-disinhibition conditions was statistically significant, $\chi^2(1, N = 59) = 4.03$, $p < .05$, $V = .26$, Nagelkerke $R^2 = .11$.

Specific for overpayment. To make sure that the effects we report here are specific for participants being overpaid in an unfair manner, we reran Study 3 in the same shopping mall with similar participants ($N = 61$; 20 men, 41 women; M age = 34.12 years, $SD = 16.86$, range 18–75; 10% lower education, 21% middle education, 39% higher education, and 30% university degree), but this time participants were promised and paid €2 and were told other participants would be paid €2 as well. The disinhibition manipulation had no effect on participants' rejection of their outcome, $\chi^2(1, N = 61) = 0.00$, $p = 1.00$, $V = .01$, Nagelkerke $R^2 = .00$, showing that the proportion of participants who rejected their outcome was the same in the disinhibition condition (16.12%) as in the no-disinhibition condition (16.67%). The disinhibition manipulation again did not influence the positive and negative affective states of participants (all $F s < 0.43$, $p s > .51$, $\eta_p^2 s \leq .01$). Thus, the effects we report here are specific for the rejection of unfair overpayment and are not found on reactions to equal payment.

Study 4

Study 3 reveals the benign disinhibition effect on real rejection of overpayment with real money. Furthermore, findings of the additional study just reported suggest that the disinhibition effect is specific for people's reactions to overpayment and are, as our line of reasoning suggests, not found on reactions to equal payment. To study this latter issue in somewhat more detail, and to rule out possible alternative explanations that might be put forward to our earlier high-impact studies, we ran a controlled experiment in which participants reacted to minimal stimulus materials. That is, we presented participants with scenarios that described situations in which participants' bonus would be either better than the bonus of a comparable colleague, equal to this colleague's bonus, or worse than this colleague's bonus. We assessed participants' intentions to reject this arrangement of outcomes. Because it is quite clear how to react to a fair, equal arrangement of outcomes as well as to an unfair arrangement in which one gets less than a comparable other person, we predicted that the benign disinhibition effect would show up most prominently (or only) in the condition in which people would have most difficulty finding how to respond, that is, the mixed-motive situation of being overpaid.

Method

Participants and design. One hundred and sixty-seven students (51 men and 116 women) at Utrecht University participated voluntarily in the experiment. Participants were randomly assigned to one of the six conditions of a 2 (behavioral disinhibition: disinhibition vs. no disinhibition) × 3 (outcome: overpayment, equal payment, underpayment) design. Between 26 and 29 students at the Uithof campus of Utrecht University were assigned to each cell of the experiment.

Procedure and materials. The experiment was presented to the participants as consisting of two unrelated parts. In the first part, we induced the disinhibition manipulation in the same way as in Studies 1 and 3 (including the PANAS, which again yielded reliable scales of positive affect and negative affect; $\alpha s = .79$ and .86, respectively).

After this, the first part of the experiment ended and the second part started. In this part, participants were asked to read and respond to a scenario. In this scenario, participants were asked to imagine that they had just completed a hard day of work at their job in a restaurant, together with another colleague. Their boss was very satisfied with the work they both had done and therefore decided to give them a bonus. The participant always obtained a bonus of €50. We varied whether participants were overpaid, equally paid, or underpaid by informing them in the scenario that the boss gave their colleague a bonus of €10, €50, or €100. Participants' intention to reject the bonus was assessed by asking participants to what extent they would want to refrain from accepting the bonus, to protest the bonus, to receive the bonus, and to accept the bonus. All items were answered on 7-point scales (1 = *very weakly*, 7 = *very strongly*), and after recoding the answers to the last two questions, participants' answers were averaged to form a reliable indicator of rejecting the bonus ($\alpha = .86$).

Results

PANAS. A 2 (disinhibition) × 3 (outcome) multivariate analysis of variance on the Positive and Negative subscales of the PANAS yielded no significant effects (Fs < 1, ps $> .37$, η_p^2s $\leq .01$; PA: $M = 2.93$, $SD = 0.61$; NA: $M = 1.65$, $SD = 0.61$). Furthermore, controlling for the PA and NA scores in analyses of covariance again did not alter the effects of our disinhibition manipulation on our dependent variables. Thus, as in Studies 1 and 3 and the Van den Bos et al. (2009) studies, affect cannot explain the effects of our disinhibition manipulation.

Rejection of bonus. A 2 (disinhibition) × 3 (outcome) analysis of variance on the scale that assessed to what extent participants indicated wanting to reject the bonus showed a main effect of the outcome manipulation, $F(2, 161) = 39.15$, $p < .001$, $\eta_p^2 = .33$. This effect indicated that participants wanted to reject the bonus more when they were overpaid than when their bonus equaled the bonus of their colleague, $F(1, 165) = 60.73$, $p < .001$, $\eta_p^2 = .27$. They also wanted to reject the bonus more when they were underpaid as opposed to equally paid, $F(1, 165) = 42.53$, $p < .001$, $\eta_p^2 = .20$. Reactions in the underpayment and overpayment conditions did not differ from each other ($F < 1$, $p > .42$, $\eta_p^2 = .00$). Importantly, the outcome main effect was qualified by a significant interaction effect between the disinhibition and outcome manipulations, $F(2, 161) = 3.22$, $p < .05$, $\eta_p^2 = .04$. Figure 1 shows this interaction effect together with the appropriate standard errors.

In correspondence with the results obtained in Studies 1–3, our findings revealed that when participants were paid more than their colleague, participants rejected their bonus more strongly when they had been reminded about disinhibited behavior ($M = 4.10$, $SD = 1.49$) than when they had not been reminded about disinhibited behavior ($M = 3.35$, $SD = 1.43$), $F(1, 161) = 5.24$, $p < .05$, $\eta_p^2 = .03$. The disinhibition manipulation had no such effect when participants' bonus equaled the bonus of their colleague (Ms = 1.70 and 2.09, SDs = 0.76 and 1.15), $F(1, 161) = 2.60$, $p > .10$, $\eta_p^2 < .02$, or when their bonus was less than the other's bonus (Ms = 3.54 and 3.42, SDs = 1.04 and 1.10), $F(1, 161) = 0.20$, $p > .66$, $\eta_p^2 = .00$. The main effect of the disinhibition manipulation was not significant in our 2 × 3 analysis of variance, $F(1, 161) = 0.79$, $p > .37$, $\eta_p^2 = .00$.

Pretest 2 and Studies 5–8: The Benign Disinhibition Effect Moderated by Social Value Orientations

Studies 1–4 reveal the benign effects of reminders of behavioral disinhibition on people's reactions to getting more than they deserve. That is, following reminders of having behaved without inhibitions, people are less satisfied with having good chances of winning an iPod by unfair means (Study 1), judge an undeserved outcome of €4 to be more unjust and are less willing to accept the €4 (Study 2), reject an undeserved outcome of €3 in a social interaction context (Study 3), and intend to reject an undeserved bonus in a scenario study (Study 4). Thus, more benign reactions to advantageous but unfair outcomes can be found on different, conceptually important reactions to a variety of interesting outcome arrangements (Studies 1–4). This attests to the robustness of the benign disinhibition effect. Furthermore, the effects are specific for being overpaid and are not found on people's reactions to

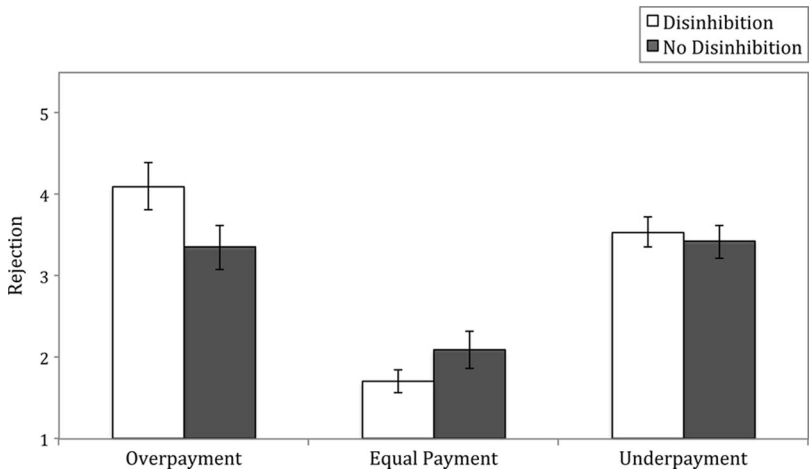

Figure 1. Rejection of outcome as a function of being reminded or not about disinhibited behavior and type of outcome received (Study 4). Error bars represent standard errors of the mean.

being paid equally or underpaid (Studies 3–4). This is in accordance with our suggestion that the benign disinhibition effect is most prominent when people are not sure how to respond to the situation at hand (see also Pretest 1), such as is the case when reacting to mixed-motive situations (Van den Bos et al., 2006; see also Adams, 1965).

In the remaining studies of this article, we examine our proposition that people are somewhat flabbergasted and inhibited regarding how to respond to advantageous but unfair outcomes (see Pretest 1) because many people are social (e.g., Aronson, 1999; De Waal, 1996), indeed prosocial, beings (e.g., Van Dijk et al., 2004; Van Prooijen et al., 2008) and, as a result, care too much about what others think of their reactions (see Latané & Nida, 1981; Van den Bos et al., 2009). If this proposition has merit, then reminders of behavioral disinhibition should especially weaken pleasure with advantageous unfair outcomes among those who adhere to prosocial values or who have adopted prosocial mindsets. In other words, the benign disinhibition effect should be prevalent especially among those with prosocial values or mindsets. The benign disinhibition effect should not be found among those with proself values or proself mindsets. The aim of Studies 5–8 was to examine whether the benign effects of reminders of behavioral disinhibition indeed are moderated by social value orientations (Study 5) and associated mindsets (Studies 6–8). However, before we discuss Studies 5–8, we think it is important to show that many of our participants indeed adhere to prosocial value orientations, as our line of reasoning argues. To this end we conducted Pretest 2.

Pretest 2

In Pretest 2 we tested our assumption that many of our participants (Utrecht University students) indeed are prosocially oriented. To this end, we assessed participants' social value orientation. Van Lange et al. (1997) and others (e.g., McClintock, 1978; Messick & McClintock, 1968) defined social value orientations as stable preferences for certain patterns of outcomes for oneself and others. Individual differences in social value orientation can be assessed by using a series of decomposed games (Messick & McClintock, 1968), which involve making choices among combinations of outcomes for oneself and for another person. These measures of social value orientation have generally good internal consistency and test–retest reliability (e.g., Kuhlman, Camac, & Cunha, 1986). An example of a decomposed game is the choice among three options: Option A, 480 points for self and 480 points for other (prosocial choice); Option B, 540 points for self and 280 points for other (individualistic choice); and Option C, 480 points for self and 80 points for other (competitive choice).

Van Lange et al. (1997) developed a nine-item decomposed game measure of social value orientation and classified participants as either prosocial, individualistic, or competitive if the majority of their choices were consistent with one of these three social value orientations. Using this measure, several studies have found that the largest group of participants tends to be prosocial, as opposed to individualistic or competitive, particularly in the Netherlands. For example, Van Lange et al. (1997, Study 4) observed that in a representative sample of the Dutch adult population (N = 1,728), 71% of the respondents could be identified as prosocials. Similarly, Van Prooijen et al. (2008) reported 61% of participants were prosocially oriented and Van Dijk et al. (2004) found between 55% and 63% were prosocially oriented. Van Lange (1999) concluded that it is common to find in student samples that more than 50% of the participants can be identified as prosocial. Prosocial participants assign a positive weight to the outcomes of others (i.e., other things being equal, seek to enhance others' welfare) and assign a positive weight to equality in outcomes (i.e., other things being equal, seek to minimize absolute differences in outcomes for self and others; Van Lange, 1999). Moreover, the prevalence of prosocials tends to be even more pronounced in the adult population in the Netherlands than in student samples in the psychology laboratory (Van Lange et al., 1997).

In our Pretest 2 we used the method developed by Van Lange et al. (1997) to assess the social value orientations of 100 students at Utrecht University (39 men and 61 women). Results showed that 76% of our participants could be categorized as prosocially oriented, 15% as holding an individualistic orientation, 1% as competitive, and 8% as not classifiable (i.e., not showing a consistent orientation response on six questions or more). Men and women did not hold significantly different value orientations, $F(1, 90) = 0.79$, $p > .37$, $\eta_p^2 = .01$. Thus, across different studies (e.g., Van Dijk et al., 2004; Van Lange et al., 1997; Van Prooijen et al., 2008) and among Utrecht University students (Pretest 2), the biggest group of participants tends to be prosocially oriented. Of course, this observation does not imply that there are no persons with a proself orientation, and we examine the implications of this notion in Studies 5–8 of this article.

Study 5

After having found evidence for our proposition that many of our participants may adhere to prosocial value orientations (Pretest 2), the aim of Study 5 was to show that these value orientations moderate the benign disinhibition effect. That is, evidence for our line of reasoning would be obtained if reminders of behavioral disinhibition weaken pleasure with unfairly obtained goods among those who adhere to a prosocial orientation. Individuals who adopt a proself value orientation should not show benign reactions following reminders of behavioral disinhibition. To test these predictions, we assessed people's social value orientations using the method of Van Lange et al. (1997). In an unrelated part of our study we then presented our participants with a scenario in which they won a game of Trivial Pursuit by unfair means (see Pretest 1), after which we measured a dependent variable central to our line of reasoning, participants' pleasure with winning the game. We predicted that participants with prosocial value orientations would be less pleased with winning the game following reminders of behavioral disinhibition than not following these reminders. Participants with proself value orientations were not expected to show this benign disinhibition effect.

Method

Participants and design. Two hundred and fifty-three students (41 men and 212 women) at Utrecht University completed the Social Value Orientation Measure by Van Lange et al. (1997), after which they were randomly assigned to either the disinhibition or no-disinhibition condition. One hundred and twenty-six students participated in the disinhibition condition, and 127 students participated in the no-disinhibition condition.

Procedure and materials. The study was presented to the participants as consisting of three unrelated parts. In the first part, we assessed participants' social value orientations (prosocial vs. proself) using the method described by Van Lange et al. (1997). As in many earlier studies (e.g., Kramer, McClintock, & Messick, 1986; McClintock & Liebrand, 1988; Van Lange & Kuhlman, 1994), we combined the individualistic and competitive participants to reflect those with proself orientations (42.0%) and contrasted this category in our analyses with those participants with prosocial orientations (58.0%).[5]

After participants had filled out the Social Value Orientation Measure, the first part of the study ended and the second part began. In the second part, we induced the disinhibition manipulation in the same way as in Study 2 (i.e., excluding the PANAS).

After this, the second part of the experiment ended and the third part started. In this part, participants were asked to read and respond to the scenario described in the experimental condition of Pretest 1 in which participants suddenly win a game of Trivial Pursuit by unfair means. We assessed participants' pleasure with winning the game in this unfair way by asking them to what extent they were pleased (1 = *very displeased*, 7 = *very pleased*), happy (1 = *very unhappy*, 7 = *very happy*), and satisfied (1 = *very dissatisfied*, 7 = *very satisfied*) with winning the game. Participants' answers to these questions were averaged to form a reliable indicator of pleasure with the outcome of the game, and this scale served as our dependent variable ($\alpha = .93$).

Results

Outcome pleasure. A 2 (social value orientation) × 2 (disinhibition) analysis of variance on the scale that assessed to what extent participants were pleased with winning the game in the unfair manner described in the scenario showed a significant interaction effect only, $F(1, 246) = 9.15$, $p < .01$, $\eta_p^2 = .04$. Figure 2 illustrates this effect. In correspondence with our predictions, the findings revealed that when participants were prosocially oriented, they were less pleased with the unfair winning of the game when they had been reminded about disinhibited behavior ($M = 3.00$, $SD = 1.17$) than when they had not been reminded about disinhibited behavior ($M = 3.45$, $SD = 1.22$), $F(1, 246) = 4.76$, $p < .04$, $\eta_p^2 = .02$. In contrast, the participants who were proself oriented did not show such a benign disinhibition effect. In fact, when participants were proself oriented, they were more pleased with winning the game when they had been reminded about disinhibited behavior ($M = 3.68$, $SD = 1.39$) than when they had not been reminded about disinhibited behavior ($M = 3.17$, $SD = 1.22$), $F(1, 246) = 4.49$, $p < .04$, $\eta_p^2 = .04$. We return to this issue in Studies 6–8.

As an aside, it can be mentioned that when participants had been reminded about disinhibited behavior, proselfs were more pleased with the outcome than prosocials, $F(1, 246) = 9.31$, $p < .01$, $\eta_p^2 = .04$. The effect of social value orientation was not statistically significant for participants who had not been reminded about disinhibited behavior, $F(1, 246) = 1.61$, $p > .20$, $\eta_p^2 = .01$.

Study 6

Our line of reasoning holds that reminding people of how they acted without inhibitions should lead them to care less about what others think of their reactions and hence to show their more genuine reactions to outcomes that are advantageous to them but are achieved by unfair means. We found in Study 5 that those participants who adhered to prosocial values were indeed less pleased with receiving their advantageous but unfair outcome following reminders of behavioral disinhibition than following reminders of how they normally act on regular days. This supports our line of reasoning.

Related to this, we also found that those participants who adhered to proself values did not show the benign disinhibition effect. This also supports our line of reasoning. In fact, we found that those who held proself values were more pleased with the advantageous outcomes following the reminders of behavioral disinhibition than following the reminders of how they normally react. One implication of these findings seems to be that reminders of behavioral disinhibition lead both prosocials and proselfs to react to advantageous unfair outcomes in ways that are more true to their genuine (prosocial or proself) selves. That is, following disinhibition reminders, those with prosocial orientations react to advantageous unfair outcomes in ways that are more true to their genuine (i.e., prosocial) selves, such that these prosocial participants are less pleased with receiving advantageous unfair outcomes. Furthermore, reminders of behavioral disinhibition lead those with proself orientations to show reactions that are more true to their genuine (i.e., proself) selves, such that these proself participants are more pleased with receiving advantageous unfair outcomes. However, before strong conclusions are drawn on the basis of the findings of Study 5, it is important to replicate them in other experiments, with greater experimental control on the value orientations variable in the studies. This was an important aim of Studies 6–8.

The goal of Study 6 was to test whether participants' reactions to the outcome arrangement that we examined in Study 5 (winning a game of Trivial Pursuit by unfair means) would be moderated by the prosocial versus proself mindsets participants adopted when reacting to the outcome arrangement. To this end we manipulated whether participants were told that they played the game with their friends just for fun (prosocial mindset condition) or that the real goal was to win the game (proself mindset condition). After this we assessed a dependent variable that is directly relevant to our line of reasoning, namely, the extent to which participants were displeased with winning the game. Following our line of reasoning and the findings we obtained in Study 5, we predicted that participants in the prosocial mindset conditions would be more displeased with winning the game following reminders of behavioral disinhibition than not following these reminders. Participants in the proself mindset conditions were not expected to show this benign disinhibition effect. In fact, we reasoned that if our proself mindset manipulation successfully led to an internalization of proself values and associated patterns of reactions (see, e.g., Van den Bos et al., 2010), then we should be able to replicate the findings we obtained in Study 5. That is, in the proself mindset conditions participants should be less displeased with winning the

[5] Although this time the percentage of prosocials was a bit lower than in Pretest 2 and Van Lange et al. (1997, Study 4), please note that also in Study 5 the largest group of participants adhered to prosocial value orientations.

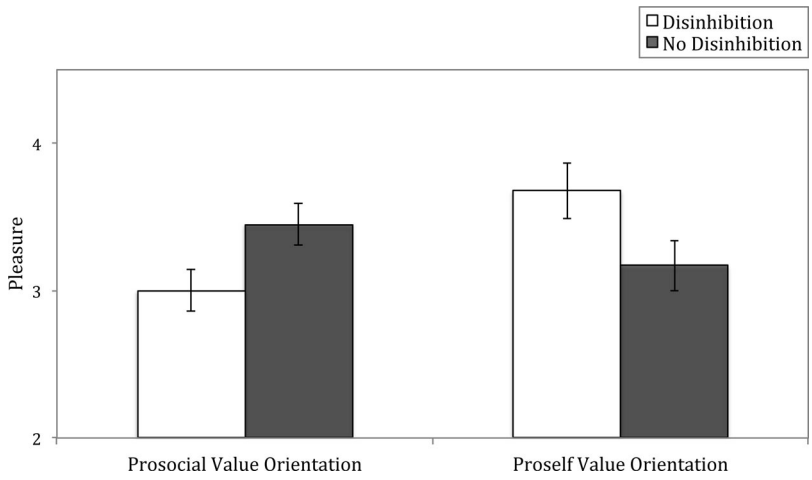

Figure 2. Pleasure with unfairly winning a game of Trivial Pursuit as a function of prosocial or proself value orientations and being reminded or not about disinhibited behavior (Study 5). Error bars represent standard errors of the mean.

game following reminders of behavioral disinhibition than not following these reminders. Study 6 examined whether our somewhat smallish manipulation of people's mindsets would be successful in achieving this aim.

Method

Participants and design. One hundred students (47 men and 53 women) at Utrecht University were randomly assigned to one of the conditions of the 2 (mindset: prosocial vs. proself) × 2 (behavioral disinhibition: disinhibition vs. no disinhibition) design. Between 23 and 27 students participated in each cell of this design.

Procedure and materials. The study was presented to the participants as consisting of two unrelated parts. In the first part, we induced the disinhibition manipulation in the same way as in Studies 1, 3, and 4 (including the PANAS, again yielding reliable scales of positive affect and negative affect; αs = .78 and .87, respectively).

After this, the first part of the experiment ended and the second part started. In this part, participants read and responded to the same scenario as in Study 5. The mindset manipulation involved how the scenario was introduced to the participants. In the prosocial mindset condition, we mentioned that the evenings playing games with friends were explicitly meant to be just for fun and that during these evenings the notion of playing fair games was highly valued. In the proself mindset condition, we told participants that everybody knew that during the evenings the real goal was to win as many games as possible and that the winner of the evening would receive a gift voucher of substantial value. After the mindset instructions, participants read the same scenario as in Study 5. The dependent variable was measured amidst filler questions and asked participants how angry (1 = *very weakly*, 7 = *very strongly*) and sad (1 = *very weakly*, 7 = *very strongly*) they would be with winning the game. Answers to these questions were averaged to yield a reliable scale of outcome displeasure (α = .79).

Results

PANAS. A 2 (mindset) × 2 (disinhibition) multivariate analysis of variance on the Positive and Negative subscales of the PANAS yielded no significant effects (Fs < 1.83, ps > .18, η_p^2s ≤ .02; PA: M = 2.85, SD = 0.60; NA: M = 1.64, SD = 0.63). Furthermore, controlling for the PA and NA scores in analyses of covariance again did not alter the effects of our disinhibition manipulation on our dependent variables. Thus, as in Studies 1, 3, and 4 and the Van den Bos et al. (2009) studies, affect cannot explain the effects of our disinhibition manipulation.

Outcome displeasure. A 2 (mindset) × 2 (disinhibition) analysis of variance on the scale that assessed the extent to which participants were displeased with winning the game in the unfair manner described in the scenario showed a significant interaction effect only, $F(1, 96) = 8.42$, $p < .01$, $\eta_p^2 = .08$. Figure 3 shows this effect. As predicted, we found that when participants were in a prosocial mindset they were more displeased with the unfair winning of the game when they had been reminded about disinhibited behavior (M = 1.83, SD = 1.14) than when they had not been reminded about disinhibited behavior (M = 1.33, SD = 0.68), $F(1, 96) = 4.12$, $p < .05$, $\eta_p^2 = .04$. In contrast, when participants were in a proself mindset they were less displeased with winning the game when they had been reminded about disinhibited behavior (M = 1.39, SD = 0.60) than when they had not been reminded about disinhibited behavior (M = 1.94, SD = 1.05), $F(1, 96) = 4.31$, $p < .05$, $\eta_p^2 = .04$.

In addition, it is worth noting that the effect of the mindset manipulation was only marginally significant for participants who had been reminded about disinhibited behavior, $F(1, 96) = 2.97$, $p < .09$, $\eta_p^2 = .03$, and was statistically significant when they had not been reminded about disinhibited behavior, $F(1, 96) = 5.69$, $p < .02$, $\eta_p^2 = .06$.

Specificity of emotions. To check whether anger and sadness yielded different responses in our experiment, we ran a mixed analysis of variance with mindset and disinhibition as between-subjects independent variables and type of dependent variable

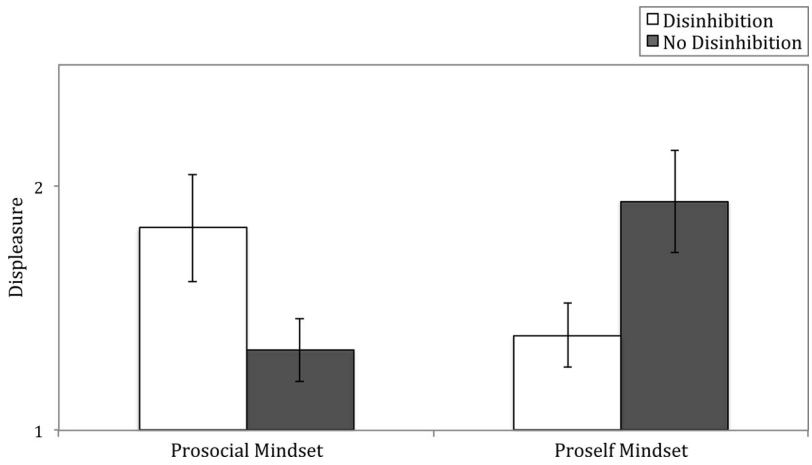

Figure 3. Displeasure with unfairly winning a game of Trivial Pursuit as a function of a prosocial or proself mindset and being reminded or not about disinhibited behavior (Study 6). Error bars represent standard errors of the mean.

(anger vs. sadness) as a within-subject variable. This yielded no significant main or interaction effects of type of dependent variable (all Fs < 1, ps $> .38$, η_p^2s $\leq .01$) and did not qualify the earlier reported Mindset $\times$ Disinhibition effect. Thus, type of emotion was not an important factor in Study 6. We return to this finding in Study 7.

Study 7

Replicating and extending the results of Study 5, the findings of Study 6 showed that participants who approached a game of Trivial Pursuit in a prosocial mindset (playing with friends just for fun) were more displeased with the unfair winning of the game following reminders of behavioral disinhibition than not following these reminders. This replicates the benign disinhibition effect. Those who reacted to the game in a proself mindset (playing to win) did not show the benign disinhibition effect. In contrast, in proself mindsets reminders of behavioral disinhibition led participants to be less displeased with the unfair winning of the game, compared with not following these reminders. This latter, more selfish effect of behavioral disinhibition replicates the findings of Study 5, suggesting that our manipulation of proself mindsets was successful in establishing the same reactions to reminders of disinhibition as internalized proself values. Thus, the results of Study 6 replicate those obtained in Study 5, this time with experimental control on the value orientations variable, enhancing confidence in the line of reasoning presented here.

Furthermore, these effects were robust for the types of negative emotions that we studied in Study 6. In addition, the experimental manipulation of proself mindsets in Study 6 led what had been a statistically nonsignificant trend in Study 5 to become statistically significant, such that in Study 6 we found that within the no-disinhibition condition, those in the proself condition were significantly less displeased with winning the unfair game than those in the prosocial condition. However, before strong conclusions are drawn on the basis of these findings, it is important to replicate them. This was one of the goals of Study 7.

Another goal of Study 7 was to examine the implications of an important assumption underlying our line of reasoning, stating that, all else being equal, the prosocial orientation is the more dominant orientation among our participants, including adult participants (Van Lange et al., 1997, Study 4). If this assumption is correct, then inducing no explicit mindset (as we did in Studies 1–4) should yield a benign disinhibition effect among our participants, whereas inducing a proself mindset should not yield a benign disinhibition effect. We induced the proself mindset in Study 7 with a scrambled sentence task that was independent of the advantageous unfair outcome that participants experienced. Another aim of Study 7 was to obtain evidence for our predictions in a nonstudent population in a study that involved participants' reactions to a real overpayment. Our dependent variable again was a measure of displeasure with obtaining the unfair outcome. Following this line of reasoning, we predicted that those in the neutral mindset conditions would be more displeased with receiving the overpayment following reminders of behavioral disinhibition than not following these reminders. In line with the results obtained in Studies 5 and 6, these effects were expected to reverse in the proself mindset conditions such that these participants would be less displeased with obtaining the overpayment following reminders of behavioral disinhibition than not following these reminders.

Method

Participants and design. Seventy-seven train passengers (29 men and 48 women) were randomly assigned to one of the conditions of the 2 (mindset: neutral vs. proself) $\times$ 2 (behavioral disinhibition: disinhibition vs. no disinhibition) design. Participants were riding trains in the province of Utrecht in the Netherlands and were promised €1 for their participation in a short psychological study. The mean age of the participants was 33.67 years ($SD = 18.18$), with the youngest participant being 16 years old and the oldest participant being 79 years old. Between 18 and 20 passengers were assigned to each cell of the design.

Procedure and materials. This study started with the experimenter sitting herself down reasonably close to a passenger in a quiet train compartment. After this, another experimenter and a confederate (who acted as a train passenger) entered the train compartment. The first experimenter invited the real train passenger to take part in a short psychological study. The other experimenter similarly invited the confederate to participate in the study. Both the real participant and the confederate were informed that they would get €1 for their participation in the study and that on top of that they could earn another €2 if they performed very well.

The study was then presented to the participants as consisting of three unrelated parts. In the first part, participants completed a scrambled sentence task (Srull & Wyer, 1979) consisting of 16 to-be-completed sentences. In the proself mindset condition, the sentences to be constructed described self-related behaviors pertaining to assertiveness (e.g., "I stand up for myself"). In the neutral mindset condition, the sentences to be constructed described self-related behaviors related to neutral behaviors (e.g., "I drink coffee with milk").

After this, the first part of the experiment ended and the second part started. In this part, we induced the disinhibition manipulation in the same way as in Studies 2 and 5 (excluding the PANAS). The second part of the experiment then ended and the third part started. In this part, participants were asked to complete a 12-item quiz. After completing the quiz questions, one experimenter evaluated which participant (the real participant or the confederate) had performed best. At this moment, the confederate received a call on her cell phone. To answer this call she walked outside the compartment, out of sight and hearing distance from the real participant. Then the experimenter announced to the real participant that she had evaluated the quiz answers of the real participant and the confederate and that the confederate had answered more quiz questions correctly. The experimenter continued by saying that she would announce to the confederate that she had in fact answered more questions incorrectly than the real participant. In this way, the experimenter stated, the real participant would get the bonus of €2. After handing over the €2 and a questionnaire measuring the dependent variables, the experimenter left the compartment. The dependent variables were measured amidst filler questions and assessed to what extent participants were angry (1 = *very weakly*, 7 = *very strongly*), sad (1 = *very weakly*, 7 = *very strongly*), and satisfied (1 = *very weakly*, 7 = *very strongly*) with their outcome in this experiment. After recoding the answers to the last question, participants' responses were averaged to yield a reliable index of outcome displeasure ($\alpha = .73$).

After the participant had finished filling out the questionnaire, the experimenter returned and debriefed the participant carefully. During the debriefing procedure, participants indicated that they did not perceive a direct relationship between the various parts of the experiment. Participants were not suspicious about our manipulations and did not object to the experimental procedure used in the experiment.

Results

Outcome displeasure. A 2 (mindset) × 2 (disinhibition) analysis of variance on the scale that assessed to what extent participants were displeased with their outcome showed a significant interaction effect only, $F(1, 42.57) = 9.97$, $p < .01$, $\eta_p^2 = .12$. Figure 4 shows this effect. In correspondence with our predictions, we found that when participants were in a neutral mindset they showed more displeasure with their outcome when they had been reminded about disinhibited behavior ($M = 1.70$, $SD = 1.04$) than when had not been reminded about disinhibited behavior ($M = 1.13$, $SD = 0.33$), $F(1, 21.74) = 5.25$, $p < .04$, $\eta_p^2 = .05$. This replicated the benign disinhibition effect that we found in our earlier studies. In contrast, when participants were in a proself mindset they showed less displeasure with their outcome when they had been reminded about disinhibited behavior ($M = 1.22$, $SD = 0.38$) as opposed to when they had not ($M = 1.88$, $SD = 1.30$), $F(1, 22.21) = 4.86$, $p < .04$, $\eta_p^2 = .07$.

In addition, it can be noted that the effect of the mindset manipulation was marginally significant for participants who had been reminded about disinhibited behavior, $F(1, 22.52) = 3.70$, $p < .07$, $\eta_p^2 = .04$, and was statistically significant when they had

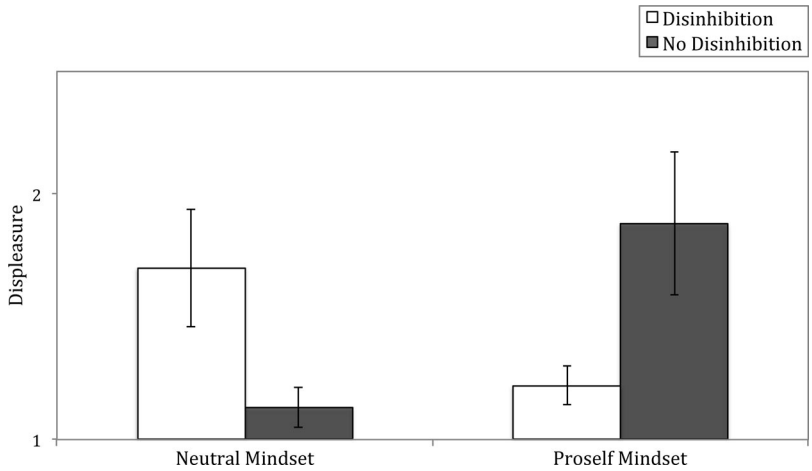

Figure 4. Displeasure with unfairly obtaining a bonus as a function of a neutral or proself mindset and being reminded or not about disinhibited behavior (Study 7). Error bars represent standard errors of the mean.

not been reminded about disinhibited behavior, $F(1, 21.64) = 6.29$, $p < .03$, $\eta_p^2 = .08$.

Specificity of emotions. As in Study 6, we checked whether anger and sadness yielded different responses in Study 7. Again, a mixed analysis of variance with mindset and disinhibition as between-subjects independent variables and type of dependent variable as a within-subject variable yielded no significant main or interaction effects of type of dependent variable (all Fs < 2.22, ps $> .14$, η_p^2s $\leq .03$) and did not qualify the Mindset $\times$ Disinhibition effect. Thus, type of emotion was not an important factor in Studies 6 and 7.

Study 8

One problem with the seven main studies reported thus far is that all seven studies operationalize the construct of behavioral disinhibition with the same manipulation (see, however, Van den Bos & Griffioen, 2011; Van den Bos et al., in press). Of course, evidence for our line of reasoning would be stronger if the same effects were found with a different manipulation of behavioral disinhibition. To this end, we asked participants to complete a scrambled sentence task in which participants in the disinhibition condition constructed sentences that were directly related to our definition of behavioral disinhibition as a state in which people do not care or only weakly care about what others think of their actions. Participants in the no-disinhibition condition constructed sentences that were related to normal behaviors that people can engage in.

Next, in an apparently unrelated part of the experiment, we assigned participants to the prosocial or proself mindset conditions using the manipulation of Study 6. This was followed by informing participants that they won a game of Trivial Pursuit by unfair means, as we did in Studies 5 and 6. In Study 8, we compared this with the nonsurprising version of the scenario used in Pretest 1 in which participants won the game in a fair way. After this we assessed a dependent variable that is directly relevant to our line of reasoning, namely, the extent to which participants wanted to protest the way the game had evolved. On the basis of our line of reasoning and the findings reported thus far, we predicted a three-way interaction effect. That is, we hypothesized that in the unfair win and prosocial mindset conditions, those who had unscrambled the disinhibition-relevant sentences would want to protest more than those who had unscrambled the normal-behavior sentences. In contrast, in the unfair win and proself mindset conditions, those who had unscrambled the disinhibition-relevant sentences would want to protest less than those who unscrambled normal-behavior sentences. Thus, we expected a Disinhibition $\times$ Mindset interaction within the unfair win conditions. Those participants who won the game fairly should not show a Disinhibition $\times$ Mindset interaction.

Method

Participants and design. Two hundred and fourteen students (100 men and 114 women) at Utrecht University were randomly assigned to one of the conditions of the 2 (fairness of winning: unfair vs. fair) $\times$ 2 (mindset: prosocial vs. proself) $\times$ 2 (behavioral disinhibition: disinhibition vs. no disinhibition) design. Between 25 and 29 students participated in each cell of this design.

Procedure and materials. The study was presented to the participants as consisting of two unrelated parts. In the first part, participants completed a scrambled sentence task consisting of 15 to-be-completed sentences. In the disinhibition condition, eight of the sentences to be constructed described behaviors pertaining to not or only weakly caring about what others think of your actions. For example, participants constructed sentences such as "what others think of me is not important," "I do not feel inhibited by other people," and "others do not influence my behaviors." The remaining seven sentences were filler items and consisted of neutral sentences (e.g., "I drink coffee with milk"). The no-disinhibition condition consisted of the same seven filler sentences and eight other sentences that were related to normal behaviors that people can engage in (e.g., "the man greeted the cashier," "grandma put her reading glasses on," and "the patient listened carefully to the doctor").

After this, the first part of the experiment ended and the second part started. In this part, participants were assigned to either the prosocial or the proself mindset condition of Study 6. After this, they read and responded to the scenario used in Studies 5 and 6 to convey the unfair winning of a game of Trivial Pursuit. We compared this with the nonsurprising version of the scenario used in Pretest 1 to convey the fair winning of the game. The dependent variable asked participants to what extent they wanted to protest the way the game evolved, to what extent they wanted to criticize the way the game evolved, to what extent they wanted to protest against the person who was asking the questions, and to what extent they wanted to criticize the person who was asking the questions. All items were answered on 7-point scales (1 = *very weakly*, 7 = *very strongly*), and participants' answers were averaged to form a reliable indicator of their protest intentions ($\alpha = .93$).

Results

Protest intentions. A 2 (fairness of winning) $\times$ 2 (mindset) $\times$ 2 (disinhibition) analysis of variance on participants' protest intentions showed two significant effects. A main effect of outcome indicated that participants wanted to protest more strongly against the unfair outcome ($M = 4.45$, $SD = 1.46$) than the fair outcome ($M = 2.07$, $SD = 1.01$), $F(1, 206) = 193.66$, $p < .001$, $\eta_p^2 = .48$. The other significant effect was the predicted three-way interaction, $F(1, 206) = 6.40$, $p < .02$, $\eta_p^2 = .03$. Figure 5 shows this effect. As predicted, within the unfair condition we found a significant Mindset $\times$ Disinhibition interaction effect, $F(1, 206) = 9.87$, $p < .01$, $\eta_p^2 = .05$, whereas this interaction effect was not significant within the fair condition, $F(1, 206) = 0.38$, $p > .53$, $\eta_p^2 = .00$. In further accordance with our predictions, we found that participants who won unfairly and had adopted a prosocial mindset protested more following the unscrambling of the disinhibition-relevant sentences ($M = 4.76$, $SD = 1.04$) than following the unscrambling of the normal-behavior sentences ($M = 4.08$, $SD = 1.44$), $F(1, 206) = 3.91$, $p < .05$, $\eta_p^2 = .02$. Furthermore, those who won unfairly and had adopted a proself mindset protested less following the unscrambling of the disinhibition-relevant sentences ($M = 4.09$, $SD = 1.66$) than following the unscrambling of the normal-behavior sentences ($M = 4.84$, $SD = 1.49$), $F(1, 206) = 6.20$, $p < .02$, $\eta_p^2 = .03$.

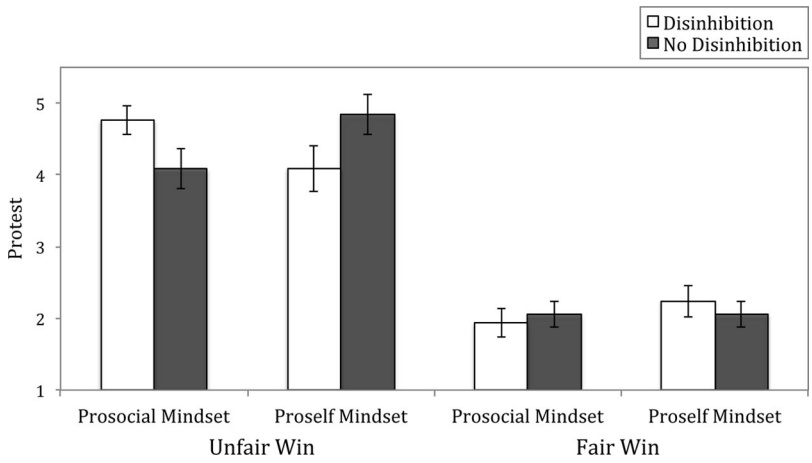

Figure 5. Protest against unfairly or fairly winning a game of Trivial Pursuit as a function of a prosocial or proself mindset and being reminded or not about disinhibited behavior (Study 8). Error bars represent standard errors of the mean.

In addition, we can note that within the unfair condition the effect of the mindset manipulation was significant for participants who had unscrambled sentences about disinhibited behavior, $F(1, 206) = 3.32$, $p = .07$, $\eta_p^2 = .02$, and was statistically significant when they had unscrambled sentences about normal behavior, $F(1, 206) = 7.01$, $p < .01$, $\eta_p^2 = .03$.

General Discussion

The package of studies presented here reveals the benign qualities that behavioral disinhibition can have, such that reminders of disinhibited behavior (Studies 1–7), or other manipulations that can be assumed to trigger behavioral disinhibition (Study 8), can weaken pleasure with getting too much. This effect is robust across different reactions to various arrangements that yield people better outcomes than they deserve (Studies 1–8). The effect can be found among students in the psychology laboratory (Studies 1, 2, 4, 5, 6, and 8) and in nonstudent samples outside the psychology laboratory (Studies 3 and 7). It occurs both in reactions to hypothetical arrangements of overpayment (Studies 4–6 and 8) and in the context of really receiving goods and outcomes that are better than deserved (Studies 1–3 and 7).

Furthermore, in line with our suggestion that the benign disinhibition effect is most prevalent when people are surprised and flabbergasted about how to respond to the situation at hand (Pretest 1 and Study 8), the effect seems to be specific for reactions to arrangements of unfair overpayment and is not found for reactions to situations that are easier to make sense of (Adams, 1965), such as arrangements of underpayment or equal payment (Studies 3–4).

Moreover, Study 8 replicates the patterns of effects that we reported in Studies 1–7 with a different manipulation that is directly related to disinhibited behavior, defined in this article as a state in which people do not or only weakly care about what others think of their actions (Carver & White, 1994; Latané & Nida, 1981). This finding, together with other findings in recent research (Van den Bos & Griffioen, 2011; Van den Bos et al., in press), provides additional confidence in the line of reasoning presented here.

In further accordance with our theorizing that many people, at least in the Utrecht area in the Netherlands (see Pretest 2 and Study 5; see also Van Lange et al., 1997, Study 4), tend to adhere quite strongly to prosocial value orientations and hence are often inhibited about how to respond to advantageous unfair outcomes, we found that social value orientations moderated the benign disinhibition effect. That is, participants who adhered to prosocial values (Study 5) or who had adopted prosocial mindsets (Studies 6 and 8) were less pleased and more displeased with receiving their advantageous unfair outcome in the disinhibition conditions than in the normal day (control) conditions. As expected from the high proportion of people holding prosocial values in our study populations, we found that not inducing an explicit mindset also yielded a benign disinhibition effect among our participants (Study 7). In contrast, in several experiments we found that participants who adhered to proself values, either by individual predisposition (Study 5) or manipulation (Studies 6–8), did not show the benign disinhibition effect.

In fact, in each of our last four studies we found that participants who adhered to proself values (Study 5) or had adopted proself mindsets (Studies 6–8) were more pleased with the advantageous outcomes in the disinhibition conditions than in the normal day conditions. One implication of these findings may be that both prosocial and proself individuals tend to hide their true preferences in the sort of situations we have been examining and that behavioral disinhibition may free them from these inhibitory tendencies. Thus, following the presence (as opposed to absence) of behavioral disinhibition, prosocials are less pleased with receiving advantageous unfair outcomes, whereas proselfs are more pleased with receiving these outcomes.

Furthermore, the findings of Studies 6–8 suggest that this hiding of the true self may be so strong among proselfs who live in a predominantly prosocial culture such as the Netherlands (see Pretest 2 and Study 5; see also Van den Bos et al., 2010) that following no reminders of disinhibition our proself participants indicated that they were more displeased with their unfair outcome than did the prosocial participants. In other words, the control

proselfs in Studies 6–8 may have overdone hiding their true preferences and, as a result, indicated even more displeasure with the unfair outcome than the control prosocials. We hasten to note that this last interpretation of our findings should be treated with caution. One reason for caution is that the effect of the value orientation variable was not statistically significant in the no-inhibition condition of Study 5. On the other hand, this nonsignificant effect may have been caused by the weaker impact of measured (Study 5) as opposed to manipulated (Studies 6–8) value orientations. Our proposition about people hiding their true (prosocial or proself) selves should also be treated with caution because of the cultural component of the explanation. Clearly more research in different cultures and with different populations of participants is needed before we can accept this proposition with high confidence. This noted, we put forward this possible implication of our current findings, as this may further future research into the exciting issues of behavioral disinhibition, social values, and reactions to advantageous unfair outcomes.

On the Concept of Behavioral Disinhibition

In the present article we tried to delineate a mechanism that people sometimes use to overcome social pressures that cause feelings of surprise and confusion. We asked how it might be possible for people to react to these situations (such as situations in which people are overpaid) in terms of their own personal values. The studies we describe here have demonstrated that reminders of behavioral disinhibition (either in the form of three open-ended questions or by means of a scrambled sentence task) appear to engage that mechanism. That is, in line with our propositions that (a) people are somewhat flabbergasted and inhibited regarding how to respond to advantageous but unfair outcomes and that (b) many people are prosocial beings, we found that behavioral disinhibition weakened pleasure with unfairly obtained goods, especially for people who adhere to a prosocial orientation or who have adopted a prosocial mindset, and not among those with a proself orientation or mindset.

More broadly, in the present article we suggested that, although people are generally constrained in their behavior and even in the pleasure they feel when confronted with unfairly advantageous outcomes, there is a dichotomy of psychological reactions available to them. They can follow the social pressure from authorities or peers that pushes them to accept and enjoy, to some extent at least, an unfair win or undeserved benefit, or they can act and feel more on the basis of their personal preferences, rejecting the outcome if they are prosocially inclined or accepting and relishing the outcome if their social orientation is more proself. This suggests that people have a set of social psychological processes that function to help them fit in as members of a group or society, suppressing individual inclinations, and another set of social psychological processes that function to express and feel their individual attitudes and preferences. What our studies demonstrate, if viewed in this light, is that reminders of the possibility of disinhibited behavior can tip or switch a person's feelings and behavior from reflecting one of these process sets to reflecting the other.

The advantageous fairness situation we examined here, like the social situations explored by Asch (1951, 1955, 1956), Milgram (1963, 1974), and Latané and Darley (1968, 1970), is most often resolved in terms of the first, group-fit set of processes; this is seen in the various control conditions used in the studies we report. When a reminder of the possibility of not following social pressures is encountered, as in our disinhibition conditions, then the self-expression processes take over and we see behavior and affective reactions that reflect the personal inclinations of the individual, be they prosocial or proself.

In other words, our "dual-processor" account of social phenomena suggests two ways of reacting to social situations that are sometimes in conflict, for example in our studies in situations of advantageous unfairness, and that when this happens people become flabbergasted and uncertain about how to respond. Experiences that remind them of the possibility of disinhibited behavior can then alter the predominance of one mode of reaction versus another. That finding is certainly interesting as a social phenomenon in its own right—revealing how people can show benign behavioral disinhibition responses—but it is even more interesting because it contributes to the psychological study of how people resolve conflicts that arise from their identity as social actors versus their identity as individuals, arguably one of the core issues in our science (Milgram, 2010).

We note explicitly that although our findings are provocative, more needs to be learned about the concept of behavioral disinhibition, its relationship to the associated concept of behavioral inhibition, and how these concepts play out in various social situations. Future studies could profit from insights from the different literatures that have proposed slightly different perspectives on the concepts of inhibition and disinhibition (see, e.g., Amodio et al., 2008; Carver, 2005; Suler, 2004). We grounded our research on leading perspectives on this issue, namely, the work by Latané and Nida (1981) and Carver and White (1994; see also Gray, 1990), because we thought these behavioral frameworks would be especially useful when studying people's behavioral inhibitory tendencies when reacting to outcome arrangements that are simultaneously unfair and advantageous.

Our work was also inspired by classical studies in social psychology in which participants were flabbergasted and inhibited about how to respond to the situation at hand. Jones and Gerard (1967, see pp. 388–389) addressed this point quite clearly from the point of view of the participant in the Asch experiments on public conformity (see also Tuddenham & McBride, 1959). Similarly, Latané and Nida (1981) conceptualized behavioral inhibition as a major mechanism explaining the bystander effect. Behavioral inhibition in bystander dilemmas manifests itself when a person wants to engage in helping behavior but is restrained from doing so by the presence of others (bystanders) who are not helping. Building on this conceptualization, Van den Bos et al. (2009) argued that inhibition in the bystander situation can be lowered (as evidenced by more and faster helping behavior in the situation) when more general behavioral inhibition (as defined by Carver & White, 1994) is weakened.

Van den Bos et al. (2009) showed that reminding people of having acted without inhibitions (in a manner that is unrelated to the bystander situations participants subsequently experience) is a good manipulation of lowered general behavioral inhibition (as measured by a state version of Carver & White's, 1994, Behavioral Inhibition Scale). The manipulation also leads to more and faster helping behavior in bystander situations. The present article followed the conceptualization developed in Van den Bos et al. (2009) and defined behavioral disinhibition as a state in which

people care only weakly or not at all about what others think of their actions (see also Van den Bos et al., in press).

We encourage future research to examine in more detail the psychological processes that behavioral disinhibition instigates. Future research should also focus on examining both the benign and the less benign or even malignant effects of behavioral disinhibition on people's reactions to advantageous unfair outcomes, as well as the effects of behavioral disinhibition on responses to other situations (see, e.g., Lilienfeld, 1992; Nigg, 2000; F. Peters et al., 2006).

For example, in his study of human behavior on the Internet, Suler (2004) noted that people often say and do things in cyberspace that they would not ordinarily say and do in the face-to-face world. Suler called this the online disinhibition effect. Particularly relevant to the current article is Suler's observation that people "loosen up, feel less restrained, and express themselves more openly" on the Internet (Suler, 2004, p. 321). In our opinion, these liberating effects of disinhibition seem to be related to the effects we have revealed here. In correspondence with our findings, the online disinhibition effect can work in two opposing directions: "Sometimes people share very personal things about themselves on the Internet. They reveal secret emotions, fears, wishes. They show unusual acts of kindness and generosity, sometimes going out of their way to help others" (Suler, 2004, p. 321). Suler called this *benign disinhibition*, and we have followed his lead in using this terminology.

But Suler (2004) also observed that disinhibition on the Internet does not always have salutary consequences. For instance, people sometimes engage in rude language on the Internet, and they ventilate harsh criticisms, anger, hatred, and even threats via e-mail: "Or people visit the dark underworld of the Internet—places of pornography, crime, and violence—territory they would never explore in the real world" (Suler, 2004, p. 321). Suler called this *toxic disinhibition*. Our research raises the possibility that what tips the scale between benign effects of disinhibition and more toxic effects can be the social value orientations the person in question adheres to, in ways similar to how the weakening versus strengthening of people's pleasure with advantageous outcomes depends on the person's social orientation. Future research might do well to examine the similarities and differences between disinhibition on the Internet (cf. Suler, 2004) versus more immediate social interaction contexts (cf. the current article; Van den Bos et al., 2009).

Future research might also consider the inclusion of other manipulations of behavioral disinhibition, examine in detail the effects of finer gradations of disinhibition, and explore other operationalizations of behavioral inhibition and disinhibition. With respect to this latter point, it is noteworthy that in another article (Van den Bos et al., in press) we found that trait BIS (and not trait BAS) led to more interventions in trolley dilemmas. We also found that our behavioral disinhibition manipulation triggered more interventions in footbridge dilemmas. All this suggests, in our opinion, that behavioral disinhibition may well have robust, and conceptually important, effects on human reactions across a variety of social conflicts and dilemmas. We also note that we studied many different reactions to advantageous unfair outcomes in the eight studies of our article. There are certainly other reactions following behavioral disinhibition that could and should be studied in future studies as well. For example, the phenomena we found here might well extend to the feelings of guilt (Adams, 1965) and unease (Jacques, 1961) that can result from being overpaid as well as to attempts to restore the experienced inequity (Adams, 1963a, 1963b) or protest against the perceived injustice (Klandermans, 1997).

Note that our disinhibition manipulation in Studies 1–7 consisted of participants completing only three questions that reminded them of having acted without behavioral inhibitions, and that findings reported by Van den Bos et al. (2009) showed that this weakens state BIS, does not affect state BAS, and does not engender affective responses or strong experimenter demands. When we inspected what participants wrote down when answering the disinhibition manipulation, we found that they described situations in which they did not feel strong public constraints on their behaviors, such as when they were attending big dance parties or other events in which they felt they could do whatever they wanted to do (Van den Bos et al., 2009). This suggests that our disinhibition manipulation is not some kind of action priming or an affect manipulation but rather is a manipulation that, as intended, does lower behavioral inhibition (Van den Bos et al., 2009). Furthermore, by using a scrambled sentence procedure, Study 8 replicated the findings obtained with the disinhibition manipulation of Studies 1–7.

Moreover, our participants indicated no suspicion of the procedures employed during the disinhibition manipulations used here, nor did they suspect a direct relationship between the manipulations and their subsequent reactions in other parts of the experiments in which they were taking part (see also Van den Bos et al., 2009). Furthermore, Van den Bos et al. (in press) showed that the disinhibition manipulation yields comparable effects as Carver and White's (1994) measure of trait BIS. And the reminders of behavioral disinhibition that we used weaken behavioral inhibition, do not influence behavioral activation, do not influence affective states, and do not trigger strong experimenter demands (Van den Bos et al., 2009, in press). Moreover, the findings reported in footnote 2 suggest that the disinhibition manipulation used in our Studies 1–7 does not affect self-awareness, self-monitoring, or experienced accountability. This noted, we explicitly acknowledge here that although it seems likely that some level of disinhibition was involved in our manipulations (Van den Bos et al., 2009), in all likelihood it was not a huge amount, and our manipulations quite possibly did not involve very strong levels of real disinhibited behavior. Therefore, an important issue is of course whether severe levels of disinhibition would still trigger benign effects.

On Being Good Natured

Philosophers, sociologists, economists, psychologists, and others have discussed for ages what determines human reactions (Beauchamp, 2001; Van den Bos, 2003). Parts of this discussion involves the issue of whether people are genuine social or even prosocial beings (e.g., Aronson, 1999; Baumeister & Leary, 1995; Van Lange et al., 1997) who are motivated positively toward fairness and justice (e.g., De Waal, 1996; Lind & Tyler, 1988) or, in contrast, whether we are fundamentally selfish individuals (e.g., Walster, Walster, & Berscheid, 1978) who prefer outcome arrangements that are advantageous to us (e.g., Rivera & Tedeschi, 1976). We do not have the illusion or the pretention that we solve this issue in the current empirical article.

This noted, we worked from the assumption that quite often people are social animals (Aronson, 1999) who are oriented toward prosocial values (e.g., Van Dijk et al., 2004; Van Prooijen et al., 2008). In fact, the findings we reported in Pretest 2 and Study 5 suggest that a large number of our participants were indeed adhering to prosocial values. Perhaps the empirical finding that quite often the largest group of participants filled out the Social Value Orientation Measure in a prosocial manner (as opposed to a proself manner; see, e.g., Pretest 2 and Study 5; Van Dijk et al., 2004; Van Prooijen et al., 2008) is a somewhat overlooked empirical observation in the literature. Implications of this observation may yield interesting findings in future research, as they did in the current article. Future research could also explore the impact of other measures and manipulations of prosocial values.

Furthermore, we hold the opinion that fairness and justice are typically very real concerns to people (Lind & Tyler, 1988; Van den Bos et al., 2006). Again, we work from a strong empirical basis here. For example, Tyler and Blader (2005) revealed that fairness concerns have quite an impact even in contexts that are commonly assumed to trigger selfish reactions. Furthermore, people typically prefer equal divisions in ultimatum and other games and tend to react negatively toward those who violate this fairness norm (see, e.g., Fehr & Gächter, 2000, 2002; Güth, Huck, & Müller, 2001). Moreover, Miller (1999) showed that, in fact, the norm of self-interest is often a myth and that when you inspect what people really do, fairness concerns may play a more substantial role in many human behaviors. This is because people find unfair behaviors genuinely aversive (Miller & Ratner, 1998).

But do not get us wrong. We also have stated explicitly that human reactions may typically involve a combination of both selfish tendencies and prosocial fairness concerns (Van den Bos et al., 2006). This is one of the reasons that we think studying reactions to mixed-motive arrangements such as the outcome arrangements examined here can yield interesting, indeed exciting, insights into human nature. Our findings may shed some light on this issue such that the results we have presented suggest that when people adhere to prosocial values they tend to react in more benign ways to disinhibition, whereas when they adhere more to proself values they tend to react to disinhibition in less benign, more self-centered ways. Our basic argument in this article holds that quite commonly people tend to adhere to prosocial values, and our findings yield supportive evidence for this assumption, but the results of Studies 6–8 also show that even rather subtle manipulations of proself mindsets can considerably affect people's reactions to disinhibition and advantageous outcomes.

In this way our findings converge with Suler's (2004) suggestion that there may not be one true, stable self. In fact, we would argue that people may shift between their multiple selves (cf. Brewer, 1991), with many (but not all) of them typically being predisposed toward their more fairness-oriented, prosocial self. Furthermore, following De Waal's (1996) line of reasoning, we argue that selfish and fairness concerns are not easily reduced to each other and that a hidden but quite noticeable aspect of humans and other primates may be that they are essentially good natured. As De Waal noted, this notion runs counter to

> the proposition that deep down, concern for others remain selfish. By denying the existence of genuine kindness, however, these theories miss out on the greater truth emerging from a juxtaposition of genetic self-interest and the intense sociability and conviviality of many animals, including ourselves. Instead of human's nature being either fundamentally brutish or fundamentally noble, it is both—a more complex picture perhaps, but an infinitely more inspiring one. (De Waal, 1996, p. 5)

Coda

One of the things that we found especially interesting in our finding evidence that behavioral disinhibition can weaken pleasure with unfairly obtained goods lies in the contrast between these benign qualities and earlier insights about the detrimental effects of disinhibition on human behavior. We think there is a certain counterintuitive quality or perhaps even an elegance to the insight that people may in some cases be inhibited from showing their prosocial orientations, and thus that reminders of having acted without inhibitions, wherein they did not care that much about what others thought of their reactions (behavioral disinhibition), can weaken people's pleasure with outcome arrangements that are to their own advantage but are unfair. It is fascinating that the very processes that others have suggested might be responsible for some of the worst of human behavior can also sometimes make people feel bad about unfair advantage.

We have also reported results that qualify the benign disinhibition effect such that those who adhere to proself values are more pleased with advantageous unfair outcomes following behavioral disinhibition. Our findings also have an ironic implication of people being social beings. That is, many humans may be (pro)social beings, but they do not always act prosocially. Why not? Our research suggests that they do not act prosocially precisely because they are social beings: They care too much about what others think of their actions and hence are inhibited in their actions and reactions. Thus, one of the interesting things about our findings is the suggestion that the social motive (to fit in and be in tune with one's social milieu) can actually inhibit people from acting on their prosocial motives. A possible implication of this line of reasoning could be that most (not all) of us should care less about what others think of us. That is, disinhibition conditions may help to free people from behavioral constraints that prevent them from disliking profitable but unfair goods.

References

Adams, J. S. (1963a). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology, 67,* 422–436. doi:10.1037/h0040968

Adams, J. S. (1963b). Wage inequities, productivity, and work quality. *Industrial Relations, 3,* 9–16. doi:10.1111/j.1468-232X.1963.tb00805.x

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). New York, NY: Academic Press.

Adams, J. S., & Jacobsen, P. R. (1964). Effects of wage inequities on work quality. *Journal of Abnormal and Social Psychology, 69,* 19–25. doi:10.1037/h0040241

Adams, J. S., & Rosenbaum, W. B. (1962). The relationship of worker productivity to cognitive dissonance about wage inequities. *Journal of Applied Psychology, 46,* 161–164. doi:10.1037/h0047751

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology, 45,* 11–19.

Aronson, E. (1999). *The social animal* (8th ed.). New York, NY: Worth/Freeman.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership, and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193,* 31–35. doi:10.1038/scientificamerican1155–31

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs, 70*(Whole No. 416).

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117,* 497–529. doi:10.1037/0033-2909.117.3.497

Beauchamp, T. L. (2001). *Philosophical ethics: An introduction to moral philosophy* (3rd ed.). Boston, MA: McGraw-Hill.

Blau, P. M. (1964). *Exchange and power in social life.* New York, NY: Wiley.

Brewer, M. B. (1991). The social self: On being the same and different at the same time. *Personality and Social Psychology Bulletin, 17,* 475–482. doi:10.1177/0146167291175001

Carver, C. S. (2005). Impulse and constraint: Perspectives from personality psychology, convergence with theory in other areas, and potential for integration. *Personality and Social Psychology Review, 9,* 312–333. doi:10.1207/s15327957pspr0904_2

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333. doi:10.1037/0022-3514.67.2.319

Collins Cobuild Dictionary. (1992). *English language dictionary.* London, England: HarperCollins.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 8,* 377–383. doi:10.1037/h0025589

De Waal, F. (1996). *Good natured: The origins of right and wrong in humans and other animals.* Cambridge, MA: Harvard University Press.

Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90,* 980–994. doi:10.1257/aer.90.4.980

Fehr, E., & Gächter, S. (2002, January 10). Altruistic punishment in humans. *Nature, 415,* 137–140. doi:10.1038/415137a

Folger, R. (Ed.). (1984). *The sense of injustice: Social psychological perspectives.* New York, NY: Plenum Press.

Folger, R., Rosenfield, D., Grove, J., & Corkran, L. (1979). Effects of "voice" and peer opinions on responses to inequity. *Journal of Personality and Social Psychology, 37,* 2253–2261. doi:10.1037/0022-3514.37.12.2253

Fowles, D. C. (1993). Biological variables in psychopathology: A psychobiological perspective. In P. B. Sutker & H. E. Adams (Eds.), *Comprehensive handbook of psychopathology* (2nd ed., pp. 57–82). New York, NY: Plenum Press.

Govern, J. M., & Marsh, L. A. (2001). Development and validation of the Situational Self-Awareness Scale. *Consciousness and Cognition: An International Journal, 10,* 366–378. doi:10.1006/ccog.2001.0506

Gray, J. A. (1972). The psychophysiological basis of introversion–extraversion: A modification of Eysenck's theory. In V. D. Nebylitsyn & J. A. Gray (Eds.), *The biological bases of individual behaviour* (pp. 182–205). New York, NY: Academic Press.

Gray, J. A. (1987). *The psychology of fear and stress.* Cambridge, England: Cambridge University Press.

Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. *Cognition & Emotion, 4,* 269–288. doi:10.1080/02699939008410799

Gray, J. A., & McNaughton, N. (2000). *The neuropsychology of anxiety: An enquiry into the functions of the septo-hippocampal system.* Oxford, England: Oxford University Press.

Greenberg, J., Solomon, S., & Pyszczynski, T. (1997). Terror management theory of self-esteem and cultural worldviews: Empirical assessments and conceptual refinements. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 61–139). New York, NY: Academic Press.

Güth, W., Huck, S., & Müller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior, 37,* 161–169. doi:10.1006/game.2000.0829

Jacques, E. (1961). *Equitable payment.* London, England: Heinemann Educational Books.

Jones, E. E., & Gerard, H. B. (1967). *Foundations of social psychology.* New York, NY: Wiley.

Kant, I. (1959). *Foundation of the metaphysics of morals.* Indianapolis, IN: Bobbs-Merrill. (Original work published 1785)

Klandermans, B. (1997). *The social psychology of protest.* Oxford, England: Blackwell.

Kramer, R., McClintock, C. G., & Messick, D. M. (1986). Social values and cooperative response to a simulated resource conservation crisis. *Journal of Personality, 54,* 576–582. doi:10.1111/j.1467-6494.1986.tb00413.x

Kuhlman, D. M., Camac, C., & Cunha, D. A. (1986). Individual differences in social orientation. In H. A. M. Wilke, D. M. Messick, & C. G. Rutte (Eds.), *Experimental social dilemmas* (pp. 151–176). New York, NY: Verlag Peter Lang.

Latané, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology, 10,* 215–221. doi:10.1037/h0026570

Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* New York, NY: Appleton-Century-Crofts.

Latané, B., & Nida, S. (1981). Ten years of research on group size and helping. *Psychological Bulletin, 89,* 308–324. doi:10.1037/0033-2909.89.2.308

Lennox, R. D., & Wolfe, R. N. (1984). Revision of the Self-Monitoring Scale. *Journal of Personality and Social Psychology, 46,* 1349–1364. doi:10.1037/0022-3514.46.6.1349

Lilienfeld, S. O. (1992). The association between antisocial personality and somatization disorders: A review and integration of theoretical models. *Clinical Psychology Review, 12,* 641–662. doi:10.1016/0272-7358(92)90136-V

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice.* New York, NY: Plenum Press.

Loseman, A., Miedema, J., Van den Bos, K., & Vermunt, R. (2009). Exploring how people respond to conflicts between self-interest and fairness: The influence of threats to the self on affective reactions to advantageous inequity. *Australian Journal of Psychology, 61,* 13–21. doi:10.1080/00049530802607605

McClintock, C. G. (1978). Social values: Their definition, measurement, and development. *Journal of Research & Development in Education, 12,* 121–137.

McClintock, C. G., & Liebrand, W. B. G. (1988). Role of interdependence structure, individual value orientation, and another's strategy in social decision making: A transformational analysis. *Journal of Personality and Social Psychology, 55,* 396–409. doi:10.1037/0022-3514.55.3.396

Messick, D. M., & McClintock, C. G. (1968). Motivational basis of choice in experimental games. *Journal of Experimental Social Psychology, 4,* 1–25. doi:10.1016/0022-1031(68)90046-2

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67,* 371–378. doi:10.1037/h0040525

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York, NY: Harper & Row.

Milgram, S. (2010). *The individual in a social world: Essay and experiments* (3rd ed.). London, England: Pinter & Martin.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist, 54,* 1053–1060. doi:10.1037/0003-066X.54.12.1053

Miller, D. T., & Ratner, R. K. (1998). The disparity between the actual and assumed power of self-interest. *Journal of Personality and Social Psychology, 74,* 53–62. doi:10.1037/0022-3514.74.1.53

Newman, J. P., MacCoon, D. G., Vaughn, L. J., & Sadeh, N. (2005). Validating a distinction between primary and secondary psychopathy with measures of Gray's BIS and BAS constructs. *Journal of Abnormal Psychology, 114,* 319–323. doi:10.1037/0021-843X.114.2.319

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin, 126,* 220–246. doi:10.1037/0033-2909.126.2.220

Peters, F., Perani, D., Herholz, K., Holthoff, V., Beuthien-Baumann, B., Sorbi, S., . . . Salmon, E. (2006). Orbitofrontal dysfunction related to both apathy and disinhibition in frontotemporal dementia. *Dementia and Geriatric Cognitive Disorders, 21,* 373–379. doi:10.1159/000091898

Peters, S. L. (2005). *The social psychology of being better off than others* (Unpublished doctoral dissertation). Utrecht University, Utrecht, the Netherlands.

Peters, S. L., Van den Bos, K., & Karremans, J. C. (2008). On the psychology of the advantaged: How people react to being overpaid. *Social Justice Research, 21,* 179–191. doi:10.1007/s11211-008-0063-2

Reeder, G. D., Monroe, A. E., & Pryor, J. B. (2008). Impressions of Milgram's obedient teachers: Situational cues inform inferences about motives and traits. *Journal of Personality and Social Psychology, 95,* 1–17. doi:10.1037/0022-3514.95.1.1

Rivera, A. N., & Tedeschi, J. T. (1976). Public versus private reactions to positive inequity. *Journal of Personality and Social Psychology, 34,* 895–900. doi:10.1037/0022-3514.34.5.895

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2,* 110–114. doi:10.2307/3002019

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology, 37,* 1660–1672. doi:10.1037/0022-3514.37.10.1660

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior, 7,* 321–326. doi:10.1089/1094931041291295

Tuddenham, R. D., & McBride, P. D. (1959). The yielding experiment from the subject's point of view. *Journal of Personality, 27,* 259–271. doi:10.1111/j.1467-6494.1959.tb01834.x

Tyler, T. R., & Blader, S. L. (2005). Can businesses effectively regulate employee conduct? The antecedents of rule following in work settings. *Academy of Management Journal, 48,* 1143–1158.

Tyler, T. R., & Lind, E. A. (1992). A relational model of authority in groups. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 115–191). San Diego, CA: Academic Press.

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. doi:10.1037/0022-3514.85.3.482

Van den Bos, K. (2005). What is responsible for the fair process effect? In J. Greenberg & J. A. Colquitt (Eds.), *Handbook of organizational justice: Fundamental questions about fairness in the workplace* (pp. 273–300). Mahwah, NJ: Erlbaum.

Van den Bos, K. (2010). Self-regulation, homeostasis, and behavioral disinhibition in normative judgments. In D. R. Bobocel, A. C. Kay, M. P. Zanna, & J. M. Olson (Eds.), *The psychology of justice and legitimacy: The Ontario Symposium* (Vol. 11, pp. 205–228). New York, NY: Psychology Press.

Van den Bos, K., Brockner, J., Stein, J. H., Steiner, D. D., Van Yperen, N. W., & Dekker, D. M. (2010). The psychology of voice and performance capabilities in masculine and feminine cultures and contexts. *Journal of Personality and Social Psychology, 99,* 638–648. doi:10.1037/a0019310

Van den Bos, K., & Griffioen, O. (2011). *Behavioral disinhibition and the failure of real-life intervention video clips: Behavioral intervention techniques may in fact inhibit prosocial behavior.* Unpublished manuscript.

Van den Bos, K., Lind, E. A., Vermunt, R., & Wilke, H. A. M. (1997). How do I judge my outcome when I do not know the outcome of others? The psychology of the fair process effect. *Journal of Personality and Social Psychology, 72,* 1034–1046. doi:10.1037/0022-3514.72.5.1034

Van den Bos, K., Müller, P. A., & Damen, T. (in press). A behavioral disinhibition hypothesis of interventions in moral dilemmas. *Emotion Review.*

Van den Bos, K., Müller, P. A., & Van Bussel, A. A. L. (2009). Helping to overcome intervention inertia in bystander's dilemmas: Behavioral disinhibition can improve the greater good. *Journal of Experimental Social Psychology, 45,* 873–878. doi:10.1016/j.jesp.2009.03.014

Van den Bos, K., Peters, S. L., Bobocel, D. R., & Ybema, J. F. (2006). On preferences and doing the right thing: Satisfaction with advantageous inequity when cognitive processing is limited. *Journal of Experimental Social Psychology, 42,* 273–289. doi:10.1016/j.jesp.2005.04.003

Van den Bos, K., Poortvliet, P. M., Maas, M., Miedema, J., & Van den Ham, E.-J. (2005). An enquiry concerning the principles of cultural norms and values: The impact of uncertainty and mortality salience on reactions to violations and bolstering of cultural worldviews. *Journal of Experimental Social Psychology, 41,* 91–113. doi:10.1016/j.jesp.2004.06.001

Van den Bos, K., Wilke, H. A. M., Lind, E. A., & Vermunt, R. (1998). Evaluating outcomes by means of the fair process effect: Evidence for different processes in fairness and satisfaction judgments. *Journal of Personality and Social Psychology, 74,* 1493–1503. doi:10.1037/0022-3514.74.6.1493

Van Dijk, E., De Cremer, D., & Handgraaf, M. (2004). Social value orientation and the strategic use of fairness in ultimatum bargaining. *Journal of Experimental Social Psychology, 40,* 697–707. doi:10.1016/j.jesp.2004.03.002

Van Lange, P. A. M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77,* 337–349. doi:10.1037/0022-3514.77.2.337

Van Lange, P. A. M., & Kuhlman, D. M. (1994). Social value orientations and impressions of partner's honesty and intelligence: A test of the might versus morality effect. *Journal of Personality and Social Psychology, 67,* 126–141. doi:10.1037/0022-3514.67.1.126

Van Lange, P. A. M., Otten, W., De Bruin, E. M. N., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73,* 733–746. doi:10.1037/0022-3514.73.4.733

Van Prooijen, J.-W., De Cremer, D., Van Beest, I., Stahl, T., & Van Lange, P. A. M. (2008). The egocentric nature of procedural justice: Social value orientation as moderator of reactions to decision-making procedures. *Journal of Experimental Social Psychology, 44,* 1303–1315. doi:10.1016/j.jesp.2008.05.006

Van Prooijen, J.-W., Van den Bos, K., & Wilke, H. A. M. (2002). Procedural justice and status: Status salience as antecedent of the fair process effect. *Journal of Personality and Social Psychology, 83,* 1353–1361. doi:10.1037/0022-3514.83.6.1353

Walster, E., Walster, G. W., & Berscheid, E. (1978). *Equity: Theory and research.* Boston, MA: Allyn & Bacon.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54,* 1063–1070. doi:10.1037/0022-3514.54.6.1063

Welch, B. L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika, 34,* 28–35.

Received October 26, 2010
Revision received March 1, 2011
Accepted March 2, 2011 ■