5-2012

# Disengaging From Moral Disengagement: Scant Experimental Evidence For a Popular Theory

Lydia Elisabeth Eckstein Jackson
leckstei@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Lydia Elisabeth Eckstein Jackson entitled "Disengaging From Moral Disengagement: Scant Experimental Evidence For a Popular Theory." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

Lowell A. Gaertner, Major Professor

We have read this dissertation and recommend its acceptance:

Michael Olson, C. Daniel Batson, Paula J. Fite, Russell Zaretzki

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Disengaging From Moral Disengagement: Scant Experimental Evidence For a Popular Theory

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Lydia Elisabeth Eckstein Jackson
May 2012

## Dedication

To my parents, Peter and Regina Eckstein, who never clipped my wings, and always encouraged me to fly. The present work and all it represents would not have been possible without their unconditional, unwavering, and indiscriminate encouragement and support of any and all of my personal and academic endeavors.

And to my son Max, for whom I hope to do the same.

## Acknowledgements

Though only my name appears on the title page, the present work is by no means the result of solitary scholarship. It reflects the time and commitment of many mentors, relatives, and friends whose suggestions, criticisms, and encouragements have been invaluable.

I am deeply indebted to my committee members, all of whom have mentored and taught me much along the way, not only about Psychology (and Statistics). I thank Paula Fite for all her support over the years, Michael Olson for always being available for thoughtful talks about research and related matters, and Russell Zaretzki for taking the time to figure out the best way to approach several tricky categorical data analyses.

I am particularly grateful for Lowell Gaertner, who has been an exceptionally patient, available, and supportive advisor all around. He modeled not only skilled and serious scholarship, but also the importance of not taking oneself too seriously. His humility and upbeat nature made all the difference on some days!

There is no question that the present work would not be what it is without the guidance and expertise of Dan Batson, who adopted me as a mentee even as he began his retirement. He spent countless (!) hours of his free time discussing matters of morality with me and navigating me out of sophomoric dead ends of thought and design. I hope I can honor his time and kindness by espousing Galilean rather than Aristotelean science as I continue on my academic path. I have learned a great deal from Lowell and Dan about

the value of patient and involved mentoring. I know I will take it with me, hoping to pay it forward.

I wish to thank my parents, my husband Patrick, and countless friends (especially Melissa, Caroline, Ali, Carolyn, Ben & Elly, Sarah, Erin, Joe, Dana and Aga) for keeping me centered when I feared losing balance. They all made sure that I remembered there was life outside the lab, especially before Max's arrival. I have many great memories of our talks, shared trips, meals, walks, and music and laughter by the fire or on the porch on warm summer nights.

Last, but not least, I wish to thank the research assistants and students who have worked with me over the years. They, too, have taught me lots. Moreover, without their dedication to my work, and their help collecting and entering data I would still be in the pilot-testing phase, no doubt!

## Abstract

Moral disengagement theory (Bandura, 1999) is a popular theory widely used to explain people's ability to violate their moral convictions without incurring self-condemnation. Assuming the internalization of moral standards in socialization, the theory suggests that sufficient enticement may motivate people to disengage their moral standards so as to violate them without negative consequences for the self. Thereby moral disengagement theory is proposed to be distinct from cognitive dissonance theory (Festinger, 1954) in that disengagement is assumed to be an antecedent to injurious behavior. This temporal assumption has been all but ignored by extant research and presents a gap in the literature that the current work seeks to address by use of an allocation paradigm for the experimental study of moral disengagement. Using a slightly altered dictator game, four studies showed that a fairness standard was clearly endorsed and recognized in the abstract (Study 1), but easily violated when behaving unfairly could benefit the self (Study 2). Furthermore, though pre-decisional adjustment of the fairness norm was evident, participants violated the norm even when no pre-behavioral justification took place (Study 3). Lastly, time to think decreased, not increased, self-favoring behavior (Study 4). Together, these studies provide scant evidence of moral disengagement and suggest that processes other than moral disengagement may be at work in the execution of (relatively benign) immoral behavior. Implication and future directions are discussed.

Keywords: moral disengagement, immoral behavior, justification of immoral behavior

## Table of Contents

## List of Tables

## List of Figures

*True! Nervous, very, very dreadfully nervous I had been and am; but why will you say that I am mad? […] Hearken! And observe how healthily, how calmly, I can tell you the whole story.*

*- Edgar Allen Poe, A Tell-Tale Heart*

## Introduction and General Information

Even a cursory scan of the newspaper on any given day suffices to remind us all that immoral behavior, large and small, is alive and well. People pilfer office supplies, cheat on income taxes, lie about reasons they stay home from work or can't meet friends; students take tests for others or copy results; corporations dump hazardous waste into landfills, sickening entire communities; priests abuse children whose grievances are exacerbated by the church's concealment of these crimes; politicians the world over get caught in corruption scandals (among others). There also seems to be no limit to the atrocities and brutalities humans are able to inflict on each other under the cloak of personal and religious freedom, power, or other benefits for oneself or one's group. And yet many of those perpetrators, many of us, are able to return home and sleep just fine. Few of us leave the daily battlefield of (im)moral behavior so scarred by what we've done that we are unable to look at ourselves in the mirror. How do we negotiate the demands of social life with our and our group's desires, particularly when these conflict with the rights of others? Surely, there are many paths to immoral behavior and its justification (e.g., Batson, in prep), but the one most relevant to the present work is the process of moral disengagement.

### *Moral Disengagement*

The concept of moral disengagement originates in Social Cognitive Theory (Bandura, 1991), which proposes that moral behavior is regulated by moral standards and self-sanctions.

Moral standards are necessary consensual codes of conduct that regulate collective life so as to respect everyone's rights and wellbeing. These moral standards that prohibit lying, stealing, cheating, unfair or harmful behavior, etc., according to the theory, are initially externally dictated, but eventually internalized in the process of socialization. Behaviors are then judged in relationship to these standards and anticipated rewards and punishments. Importantly, those rewards and punishments are not just offered by the social environment, but also by oneself. Acting in line with one's standards results in feelings of self-worth and approval, whereas violations of the standards threaten to incur not only social, but also self-punishment (e.g., in the form of guilt or shame). Anticipated self-rewards and punishments are particularly important since external deterrents to immoral behavior are not always immediately present and one's behavior can go undetected. Hence, acts that violate personally held beliefs are avoided so as to not result in self-condemnation.

However, life presents many opportunities to reap benefits from precisely the kind of behavior that violates our standards. Insofar as an individual endorses a certain moral principle (e.g., not to cheat or lie), a sufficiently strong enticement to behave immorally (e.g., to get a bigger tax return if one doesn't report all income, to cheat on a test to make an A, not a B, in the class, etc.) presents a conflict that can be resolved by either abstaining from the immoral behavior (e.g., not cheating) or by disengaging moral self-censure from immoral conduct (i.e., employ self-exonerative reasoning in the interest of self-protection). In the latter case, moral self-censure is disabled, thereby allowing inhumane and immoral behavior without self-condemnation (Bandura, 1999). Such a process leads to the prediction that an action should be judged differently depending on whether the actor has an interest in a particular outcome, or whether the actor has no interest in the outcome. Indeed, research on past behavior supports such

divergent judgments: participants judged favoring themselves to be more fair than an uninvolved observer judged the same action (Valdesolo & DeSteno, 2007). Moreover, this self-favoring bias was eliminated under cognitive constraint, indicating that it is not an automatic affective bias, but rather resembles a more controlled process in the service of self-protection (Valdesolo & DeSteno, 2008). Moral disengagement, too, is proposed to be a controlled process: as both social and self-sanctions act anticipatorily, so do the processes of disengagement.

More specifically, action can be disengaged from moral self-control through four mechanisms: Morally justifying the act, diminishing responsibility, minimizing consequences, and dehumanizing or blaming the victim.

Moral justification entails the depiction of injurious behavior as serving virtuous purposes. Through moral justification, for example, engaging in violence may be viewed as "protecting cherished values, preserving world peace, saving humanity from subjugation or honoring the country's commitments" (Bandura, 2002, p. 103). The very behavior that is usually condemned, in essence, is reified as a duty or moral imperative that serves a greater good.

Diminishing personal responsibility reduces a person's culpability for the execution and consequences of harmful acts. This may be achieved by attributing responsibility to a legitimate authority (e.g., Milgram, 1974) or by diffusing responsibility within a group through perceptions of collective action. Soldiers, for example, "merely" follow orders or act in groups, in which individual contributions can either not be identified or the accountability rests on so many that one alone does not feel incriminated (Darley & Latané, 1968; Schopler et al., 1995).

Minimizing the consequences of detrimental behavior is achieved by creating physical distance to the harm done, and by misconstruing or by outright ignoring the adverse effects of one's actions. The most powerful way to achieve this is to remove oneself from the victims'

agony or other negative results of one's actions (as is often the case with aerial bombings or other forms of modern warfare) – a fact that Milgram (1974) observed when he required the "teacher" to make physical contact with the "learner" in his famous obedience studies. Obedience rates dropped markedly when victim and perpetrator were not removed from each other anymore.

Lastly, the burden of bringing harm and suffering to other human beings may be reduced by blaming or dehumanizing the victim: negative consequences may be construed as befalling members of deserving, subhuman groups and therefore do not need to raise moral concerns (cf. Opotow, 1990; Staub, 1990). In intergroup conflict situations, it is common for the ingroup to be perceived as the only group of people characterized by typically human or exclusively human attributes, whereas such secondary emotions as affection, admiration, pride, conceit, or nostalgia are not attributed to the outgroup (Haslam, 2006; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007). In extreme cases, such devaluation can lead to moral exclusion (Opotow, 1990). So long as people feel some sense of collective responsibility, the awareness that one's group has harmed an outgroup in the past may be enough to reduce associations of uniquely human characteristics with the outgroup (Castano & Giner-Sorolla, 2006).

Bandura (1991, 1999) proposed that mechanisms of moral disengagement are of relevance in cases of competing motivations where the implicit cost-benefit-calculations preceding behavior return that it would be more costly to be moral than to bend one's standards in the pursuit of one's wants. Highlighting the driving force of self-interest, moral disengagement may therefore be particularly relevant to instances in which one's own interests are pitched against those of another (cf., interest morality; Batson, 2011) – in cases of extraordinary immoral behavior as much as in instances of everyday immorality.

Indeed, research on moral disengagement has linked self-exonerative reasoning consistent with moral disengagement to behaviors like interpersonal aggression and delinquent conduct, war support, carrying out the death penalty, animal cruelty, and (un)ethical decision making (Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Cohrs & Moschner, 2002; Detert, Trevino, & Sweitzer, 2008; Grussendorf, McAlister, Sandstrom, Udd, &Morrison, 2002; Jackson & Gaertner, 2011; McAlister, Bandura, & Owen, 2006; Osofsky, Bandura, & Zimbardo, 2005; Vollum, Buffington-Vollum, & Longmire, 2004). However, most studies cited are based on correlational data and self-report measures of moral disengagement, some of which were taken after people had already engaged in the immoral action (Osofsky et al., 2005, for example, measured reports from officers having worked in the death row system anywhere from 1-31 years).

### *The research gap: A dearth of experimental research on moral disengagement and neglect of temporal specifications of the theory*

For as much research as has been generated by the theory of moral disengagement, there is little or no published experimental research. Two notable exceptions are recent studies by Shu, Gino, & Bazerman (2011) and Hartmann & Vorderer (2010). The latter somewhat successfully manipulated disengagement mechanisms in the context of violent video games. The authors showed that violence against a virtual opponent led to less negative affect and guilt if the opponent had committed a previous condemnable action and was portrayed as less human. The same results were true of manipulations that depict a player's violence as serving a good purpose and as having no audible and visible negative consequences. This research is informative in that it shows that the kind of cognitive restructuring proposed by Bandura has the potential to reduce

negative affect in the context of violent behavior. However, the study does not allow for testing the idea that participants actively generate cognitions in line with disengagement mechanisms when faced with the moral dilemma of wanting to uphold a standard while also wanting to reap rewards as a result of violating the standard. Shu et al. (2011) provide such a situation that highlights personal rewards from potential immoral action in the context of academic dishonesty and cheating, however, they conceptualize moral disengagement as a consequence of immoral action. Although it is very plausible (and evident in their research) that self-exonerative cognitive restructuring can be a consequence of immoral behavior, the theory suggests that disengagement happen in anticipation of such behavior, hereby enabling immoral conduct. A standard that is ordinarily embraced is selectively "deactivated" when the behavior proves more beneficial than costly for the self:

> People often experience conflicts where behavior they personally devalue can serve as the means for securing valued benefits. As long as self-sanctions override the force of external inducements, behavior is kept in line with personal standards. However, in the face of strong external inducements, such conflicts are often resolved by selective disengagement of self-sanctions. This enables otherwise considerate people to perform self-serving activities that have detrimental social effects. (Bandura, 1990, p. 28)

Assessing such endorsement of a standard in the absence of self-interest, and the temporal aspect of disengagement seem to have largely been dismissed as a formality by extant research on moral disengagement. Post-hoc justification of attitude-inconsistent behavior is well-documented in psychological research (e.g., Cooper & Fazio, 1984; Festinger, 1957; Festinger & Carlsmith, 1959) and so it seems prudent to pursue the question of whether justification

mechanisms may, in fact, operate anticipatorily, as well. However, this question of whether the anticipation of injurious behavior leads to self-exonerative reasoning and the disengagement of moral standards has yet to be tested.

### *Suggesting a new paradigm for the experimental study of moral disengagement*

In light of this, a research paradigm is needed that (a) shows that a moral standard is endorsed in the abstract, but (b) anticipated self-benefits provide sufficient enticement to violate the standard (c) can assess pre-behavioral disengagement of the standard, and (d) permits manipulation of conditions necessary for disengagement. More specifically, a standard is needed that is widely held and can easily be activated. In the abstract (when participants cannot benefit from violating the standard), this standard should be endorsed. However, a sufficiently strong enticement to violate the standard should create exactly the kind of internal dilemma that Bandura suggests can motivate moral disengagement.

To create such a situation, I borrowed from dictator-game-paradigms to create an allocation paradigm that provides an opportunity to violate a fairness standard. Participants believed they and another participant were partaking in a study on group decision-making in which the members didn't know each other and couldn't interact with each other (in reality, only one participant was in the lab). The participant was told that the session involved an allocation task and eventually every participant was assigned to be the decision maker (for the participant this was framed in such a way that made it less obvious that every participant would be the decision maker, of course): 10 raffle tickets needed to be allocated between the (ostensibly) two participants. This could be done in *any* possible combination and neither the other participant nor

the experimenter would find out how the allocation was made. The tickets were entered into a raffle for a $100 Target gift card. A drawing was conducted a few times during the semester.

### *The Present Research*

The goal of the present research is to test the tenets of moral disengagement theory using the allocation paradigm. Specifically, the present research seeks to address the following questions:

1. Are participants aware of the fairness standard and do they endorse it in the abstract (i.e., when no personal gains are expected from violating the standard)?
2. Is the gift card perceived as a desired end (if it is not desired it is unlikely that participants would risk violating a moral standard to get the card)?
3. When put in the actual role of the decision-maker, do participants violate the fairness standard?
4. Assuming that some violate the fairness standard, do they disengage this standard ahead of their behavior?
5. How do participants feel about their behavior afterwards?

Predictions derived from moral disengagement theory state that a well-recognized and endorsed standard is either adhered to (i.e., no moral violation takes place), or disengaged when personal benefits can be expected from violating the behavior. Disengagement should be expressed as pre-behavioral restructuring of a moral code, making unfair behavior seem fair. Moreover, if the standard is disengaged, the feelings following one's immoral behavior should show a lack of guilt, shame, and remorse and a sense that the individual didn't do anything wrong.

The above questions and predictions were tested across four studies: Study 1 examined whether participants perceived the gift card to be desirable and whether they identified the fairness standard in the present paradigm as giving 5 tickets to self and 5 tickets to other. Study 2 investigated whether participants would deviate from the fair allocation when put in the role of the decision makers themselves. Study 3 manipulated whether participants were asked to indicate their understanding of a fair allocation either before or after learning that they would be the allocator. With self-interest introduced early for half the participants, this design allowed me to test whether the participants who knew of their role would adjust the fairness judgment in their favor in anticipation of behavior and to further test whether these pre-behavioral fairness judgments would predict actual behavior. If disengagement was a process at work, those who had an opportunity to adjust the fairness standard ahead of time should favor themselves more than those who did not have an opportunity to adjust the standard. Because moral disengagement is a deliberate process that takes time, Study 4 manipulated the amount of time participants had to disengage before making their decision, and assessed subsequent behavior. To the extent to which the desirability of the gift card motivated self-exonerative reasoning, time to think should allow for disengagement and more self-favoring behavior. To see how participants felt about their behavior afterwards, Studies 2-4 additionally assessed participants' post-behavioral feelings and ratings of the fairness and morality of their actions.

## Study 1

To test whether the allocation paradigm is suitable for studying moral disengagement processes, it is important to know that participants are aware of and apply the fairness standard.

If so, when participants are not actually in the position of making the allocation, most should indicate that the fair way to assign ten tickets among two participants is giving each participant five tickets. Furthermore, it is important to know that participants identify the gift card as desirable, even if they are not in a position to win it.

## *Method*

*Participants and procedure.* Seventy-six undergraduate students enrolled in Abnormal Psychology (*n*=69, 55 females, 12 males, 2 no gender identified) and Statistics (*n*=7, 4 females, 3 males) were presented with the allocation procedure as a hypothetical decision and asked to indicate the number of tickets allocated to self and other that, in their view, would be moral and fair. In this study, participants were told that the gift card was for $10 and that a winner would be drawn at the end of each session (i.e., 5 tickets would directly translate into a 50% chance of winning the gift card). Participants were asked to rate the desirability of the gift card on a scale from 1 (not at all desirable) to 7 (extremely desirable). Three students were excluded from analyses because their ticket allocations did not add up to 10.

## *Results*

*Fairness and morality judgments*. Of the remaining 73 students, 60 (82%) said the moral and fair way to allocate the tickets was 5 for self, 5 for other; 5 (7 %) said 5-5 was fair, but the moral thing was to give the other more tickets than the self; 2 students (3%) said it was moral and fair to give the other more tickets; 4 (6%) said 5-5 was moral, but it was fair to favor the self; and 1 person each (1%) said favoring the self was fair and moral or 5-5 was fair, but favoring the self was moral.

*Desirability of gift card*. In this study, the gift card was seen as rather desirable ($M$=5.3, $SD$ = 1.1; 1=not at all desirable, 7=extremely desirable). Two participants did not indicate how desirable the card was to them. Responses of the remaining 71 participants ranged from 2 to 7 with 61 (86%) of the participants indicating that the card was desirable by choosing 5 or higher. Two males (2.8%) indicated the gift card was not desirable (assigning a desirability rating of 2). Overall, females indicated that the gift card was more desirable ($M$=5.4) than males ($M$=4.7), $F(1,69)=5.08$, $p$=.03.

## *Discussion*

Study 1 was designed to assess the awareness of the fairness standard and the desirability of the gift card. In order for disengagement to occur, a standard has to be endorsed in the abstract, so that it may be disengaged when a sufficiently enticing opportunity promises to yield more personal benefits than costs. To assess whether these conditions had been met by the present paradigm, participants were presented with the allocation procedure as a hypothetical scenario and asked to identify the fairness standard and whether the prospect of immediately winning a $10 gift card would be perceived as desirable. At least in the current sample, the procedural fairness standard was easily accessible. Only two of 73 participants thought it was moral to prefer the self; the vast majority (71 participants or 97%) indicated that a moral choice would be to act fairly or to favor the other. Similarly, 65 participants (89%) indicated that giving five tickets to each participant would be fair – a clear indication that the fairness standard is salient and easily recalled.

However, the desirability of the gift card ranged from 2 to 7 (2 males and 1 female indicated the gift card was not desirable) and it was more appealing to females than males. Although overall the gift card was rather desirable, it seems that there is potential to make the

card even more appealing. Study 2 addressed this problem in addition to providing participants with an opportunity to adhere to (or violate) the fairness standard by putting them in the actual decision making situation.

## Study 2

Moral disengagement theory posits that we endorse moral standards in the abstract, but if people can expect considerable personal gains, they may adjust their standard so as to enable them to violate the standard in order to increase personal benefits. Study 1 demonstrated that the fairness standard is easily accessible when no personal gain is expected. Will the same endorsement of the standard be observed when participants can benefit from violating it? Study 2 set out to address this by putting participants directly into the position of making the allocation themselves. In addition, I tried to increase he desirability of the gift card in Study 2.

### *Method*

*Participants and procedure*. 21 undergraduate students (10 females, 11 males) participated in exchange for course credit in an Introductory Psychology class. Participants were brought into the lab and run one-at-a-time by a same-sex research assistant to minimize cross-gender presentational concerns. Participants believed they were taking part in a study on how group members make decisions that have the potential to affect others in the absence of an opportunity to interact with another group member. With the exception of attaining informed consents and debriefing the participants, the experiment was run on MediaLab (for a copy of all instructions and prompts in order of presentation please see the Appendix) and participants didn't interact with a research assistant. Before making the allocation, all participants were asked how much

they would like to win the gift card (1=not at all; 7=extremely). Because participants in Study 1 had indicated that a $10 gift card paid out at the end of every session was only moderately desirable, this study offered a $100 gift card to be drawn once a month. Participants were then asked to make their allocation decision by assigning themselves anywhere from 0 to 10 tickets; the other participant would automatically get however many of the 10 tickets the allocator did not give to self. Participants were told that neither the other participant not the experimenter would learn of the decision.

Once participants had made their allocation, they were asked to rate how moral, then how fair their decision was (both on a scale from 1=not at all fair/moral; 7= extremely fair/moral), and how they felt about their decision (sample items include justified, guilty, ashamed, happy; all on a scale from 1=not at all to 7=extremely). Finally, they were asked to provide some demographic information before signaling to the experimenter that they had concluded the computer portion of the study. The experimenter then entered the room and began the debriefing interview.

### *Results*

*Desirability of gift card.* Ratings of the desirability of the gift card ranged from 4 to 7. The mean desirability rating of winning the gift card was 6.4 (*SD*=.93). The desirability did not vary by sex ($F(1,19)=1.2$, $p=.30$). Comparing the desirability of the gift card across Study 1 and Study 2 confirmed that the $100 gift card in Study 2 was seen as more desirable than the $10 gift card in Study 1, $F(1,88)=26.2$, $p<.001$.

*Mean number of tickets allocated to self.* Out of our 21 participants, 10 (48%; 4 females, 6 males) allocated 5 tickets to self and 5 tickets to other, 10 (48%; 5 males and 5 females) favored

themselves, and 1 female (4%) favored the other (for histograms displaying the ticket allocations by gender, please see Figure 1). These numbers are in stark contrast to the percentages of participants in Study 1 who indicated that an equal allocation was fair and moral. More specifically, when participants were put into the position of the decision maker, they allocated an average of 6.5 tickets to themselves. Allocations did not vary by sex ($F(1,19)=.02$, $p=.88$).

However, the mean allocation to self in this study was significantly higher than what participants in Study 1 indicated would be fair and moral in the abstract, $t(20)=3.24$, $p=.002$. When self-interest was at stake, the participants in this study found it difficult to adhere to the fairness norm endorsed in the abstract.

*Post-behavioral ratings of fairness and morality*. How did participants rate their own actions? On average, the participants in this study perceived their actions to be more fair and moral than unfair or immoral (average ratings of post-behavioral fairness: $M=4.8$, $SD=2.5$; average rating of post-behavioral morality: $M=5.0$, $SD=2.1$). Post-behavioral ratings of both fairness and morality did not vary by sex (both $F$s$<.60$). However, those who acted fairly or favored the other self-identified their behavior as both more moral ($M=6.6$) and more fair ($M=6.2$) than those who favored themselves ($M_{moral}= 3.1$, $M_{fair}=3.2$), $F(1,19)=58.04$, $p=.001$ (moral) and $F(1,19)=11.8$, $p<.01$ (fair). It seems participants rightly identified their actions as immoral and unjust if they favored themselves. This is also reflected in the correlation of numbers of tickets awarded to self with morality ratings ($r=-.79$, $p<.001$) and fairness ratings ($r=-.61$, $p<.01$). Generally, ratings fairness and morality were highly correlated, indicating that participants' idea of fairness and morality overlap, but not to the extent to which they are synonymous with each other, $r=.62$, $p<.01$.

*Emotional ratings*. How did participants feel about their behavior? Those who favored themselves felt slightly more ashamed ($M$=2.0) and more guilty ($M$=2.6) afterward than those who acted fairly (means of both ashamed and guilty for those who acted fairly were 1.0 with no variability; hence, the Wilcoxon Rank Sum Test was performed: $Z_{ashamed}$= 3.01, $p$=.01; $Z_{guilty}$ = 4.57, $p$<.001). Those who favored themselves ($M$=1.7) did not feel more sorry than those who acted fairly or favored the other ($M$=1.2), $F(1,19)$=1.51, $p$=.24. The same is true of happy: Favoring oneself ($M$=4.6) did not make participants any more or less happy than acting fairly ($M$=5.2), $F(1,19)$=1.1, $p$=.32. However, those who favored the other or acted fairly felt more justified ($M$=5.5) than those who favored themselves ($M$=4.1), $F(1,19)$=5.43, $p$=.03. The more tickets participants allocated to self the more ashamed ($r$=.39, $p$=.08) and guilty (r=.41, p=.07) they felt. Number of tickets allocated to self did not reliably predict feeling happy, sorry, or justified ($r_s < |.26|$). For all correlations, see Table 1.

## *Discussion*

Study 2 was designed to see whether participants would deviate from the easily identified and widely held fairness norm when violating the norm would provide an opportunity to reap personal benefits. Specifically, participants were put into the role of the decision maker and given full control over how many tickets to allocate between themselves and an ostensible other (unknown) participant. The tickets would then be entered into a raffle for a $100 gift card to be drawn at several times during the semester. Allocating more tickets to oneself translated into entering more tickets in one's name into the raffle.

Though participants did not have full control over the outcome of winning the gift card (as was the case in Study 1, because 10 tickets in that situation would have translated into an

automatic win), the $100 gift card was seen as more desirable than the $10 gift card in Study 1. The desirability of the gift card was incentive enough for half of the participants in this study to violate the widely-held fairness norm in their favor and for participants in this study to allocate, on average, 1.5 more tickets to themselves than would be permissible by the fairness standard. Consistent with the tenets of moral disengagement theory, participants violated the moral standard that was endorsed in the abstract when they were the ones who could benefit from such a violation while incurring relatively little costs (they were told the other participant and the experimenter would not learn of their decision so as to minimize presentational concerns).

Inconsistent with moral disengagement theory, however, is the fact that participants who favored themselves also identified their behavior as less moral and less fair. Though disengagement suggests that participants think of their behavior in favorable terms, it is possible that the current situation did not provide much of a "way out." Perhaps participants saw little room to "spin" the fact that giving more than five tickets to self was not fair. Still, it is noteworthy that post-behavioral ratings of feeling justified, of morality and fairness were not at the bottom of the scale. In fact, they all were closer to the middle of the scale, indicating that participants did not condemn their behavior altogether. In line with recognizing their behavior was unfair, those who favored themselves felt worse about their previous behavior than did those who acted fairly or favored the other: they indicated feeling more ashamed and guilty and less justified. However, all of these ratings are rather low, indicating that even those who recognized their actions as immoral did not experience strong feelings of shame or guilt at all. Similar to the ratings of morality and fairness, this may be a byproduct of the situation (participants may have reasoned that favoring oneself on tickets is hardly as condemnable as some other immoral actions), or it may be a result of participants having justified their behavior. If such justification

occurred, the current study cannot address whether participants, in fact, reasoned that their unfair behavior was justified and fair in anticipation of their behavior, or whether they did so as a means of reducing possible dissonance. Study 3 tried to address this question by investigating whether participants "adjusted" their understanding of what was fair and moral in anticipation of the behavior.

## Study 3

*Overview and predictions*. A central idea of moral disengagement theory is that standards that are endorsed in the abstract may be "shifted" in one's favor if perceived benefits of immoral behavior promise to outweigh the costs. Moreover, this is proposed to happen in anticipation of the behavior. Studies 1 and 2 demonstrated that a fairness standard that was easily accessible and endorsed in the abstract was less likely to be endorsed when participants were given an opportunity to act in a way that could uphold the standard or violate it in the interest of maximizing potential personal gains. If moral disengagement is plausible, participants who are informed of their role as decision maker should proceed to employ self-exonerative reasoning leading to disengagement of a standard insofar as the action promises to be more beneficial than costly to the self. To test this, study 3 asked participants to indicate their understanding of what would be moral and fair *before* making the decision. I manipulated whether participants already knew of their role as decision maker at the time of this pre-behavioral judgment or not. Thus, there were two conditions to which participants were randomly assigned:

In the *role-unknown-condition*, participants were introduced to the task at hand, but asked what the moral and fair allocation would be *before* learning that they would be the allocator (and thus before making the allocation), and again after having made their decision.

In the *role-known-condition*, participants were asked what the moral and fair allocation would be *after* learning that they would be the allocator (but before making the allocation), and again after having made their decision.

It was expected that the pre-allocation ratings of morality and fairness of an allocation would differ across the two conditions in so far as it is true that people start to disengage a standard before the behavior, given that they expect the behavior will be more beneficial than costly. More specifically, it was expected that participants in the role-known-condition would "move the standard" in their favor *before* proceeding to the actual behavior, possibly allowing for more self-favoring allocations. Thus, relative to participants in the role-unknown-condition, those in the role-known-condition should show pre-allocation ratings that resemble self-favoring (i.e., indicate that more than 5 tickets to self are fair). Given the absence of self-interest at the time of pre-allocation ratings, participants in the role-unknown-condition should have no reason to shift the standard of fairness in their favor. As a result, those in the role-unknown-condition should subsequently adhere to the standard and act fairly, whereas the opportunity to adjust the standard ahead of time should allow for more self-favoring allocations in the role-known-condition.

## *Method*

*Participants and procedure*. 42 undergraduate students (19 females, 23 males) participated in exchange for course credit in an Introductory Psychology class. Participants were brought into the lab and run one-at-a-time by a same-sex research assistant to minimize cross-gender presentational concerns. The basic procedure was the same as that of Study 2. However, participants in the present study were asked to indicate what the fair and moral way to allocate the tickets would be before making their decision. They did so either before they knew they

would get to make the allocation (*role-unknown-condition, n=22*) or after learning about their role as the allocator (*role-known-condition, n=20*). They indicated what would be fair and moral by choosing a combination of possible ticket allocations for each item (combinations in each case ranged from 0 to self and 10 to other to 10 to self and 0 to other).

### *Results*

*Desirability of gift card.* Ratings of the desirability of the gift card ranged from 4 to 7. The mean desirability rating of winning the gift card was 6.6 (*SD*=.79). The desirability did not vary by condition or sex (all main effects and the interaction, $F$s < .43, $p$s > .52). Apparently, the $100 gift card was clearly desirable to the participants in this study.

*Ratings of fairness and morality prior to ticket allocation*. Given that the gift card was desirable, what did participants indicate would be the fair and moral way to allocate the tickets? In the *role-unknown-condition* participants made their ratings of pre-behavioral morality and fairness before learning that they would be the decision-maker. Believing that they had a 50-50 chance of being on the receiving end, ratings by participants in this condition should be very much in line with the fairness norm. On the other hand, participants in the *role-known-condition* made these ratings *after* they had learned of their assigned role. Enticed by the pull of the gift card, their judgments of what is fair and moral should already be tilted in their favor.

Indeed, on the fairness item, a main effect of condition emerged: participants in the role-unknown-condition deemed an average number of 4.6 tickets to self fair, while participants in the role-known-condition deemed an average number of 6.2 tickets to self fair, $F(1,38) = 9.61$, $p<.01$. Though participants in the role-unknown condition also indicated that giving fewer tickets to self would be moral (*M*=4.2) relative to role-known-condition (*M*=4.9), this difference was not

reliable, $F(1,38)=1.45$, $p=.24$. However, there was a condition x sex interaction such that females in the role-known-condition thought more tickets to self ($M=5.6$) were moral compared to the role-unknown-condition ($M=3.6$), $F(1,38)=5.94$, $p<.05$. For males, there was no reliable difference between the role-unknown-condition ($M=4.8$) and the role-known-condition ($M=4.2$), $F(1,38)=.80$, $p=.38$. Overall, pre-decisional judgments of fairness and pre-decisional judgments of morality were highly correlated in the role-unknown-condition ($r=.73$, $p<.001$), but not in the role-known-condition ($r=.09$, $p=.71$; for all correlations see Table 4). It seems participants differentiated between what was fair and moral when they knew they were going to have control over the ticket allocation, but not when there was a chance that someone else might make the allocation decision. Though in the abstract recognizing that 50-50 is the fair and moral way to allocate 10 tickets, when faced with the opportunity to favor oneself, the participants' perceptions of fairness, but not morality, shifted in their favor. Moreover, females in particular thought that giving an average of 2 additional tickets to self was moral in the role-known-condition (however, since females also indicated that it would be moral to give the self fewer tickets in the role-unknown-condition, this difference amounts to a deviation from the fairness norm by roughly one ticket).

Another way to look at these judgments is to tally the pre-behavioral indications of what is fair and moral by condition. Out of 42 participants in the two conditions, 31 (74%) indicated that the fair and moral way to allocate the tickets would be 5-5. The only exceptions were 1 male in the role-unknown-condition who indicated the fair way to be 5-5 whereas the moral decision would be 3 tickets to self and 7 to other (exactly the decision he proceeded to make). Also in this condition, 3 females who acted fairly indicated that the fair and moral way to make this decision would have been to favor the other (only one said the fair decision would have been 5-5).

Nobody in the role-unknown-condition indicated before making the decision that it would be fair to favor oneself. Interestingly, however, 6/8 participants (75%) who later favored themselves in the role-known-condition indicated that the fair way to allocate the tickets would, in fact, be to favor oneself. 3 out of those 6 indicated that the moral way to allocate would be 5-5, 2 indicated that the moral way would be to favor oneself, and 1 indicated that the moral decision would be to favor the other. It appears as if 6/20 participants (30% ) in the role-known-condition adjusted the widely held fairness standard in their favor, knowing they would be the ones to allocate the tickets momentarily.

*Ticket allocation predicted by pre-behavioral ratings*. Did the pre-behavioral ratings for fairness and morality predict the decision participants proceeded to make? As expected, indicating that more tickets to self would be fair predicted actually giving oneself more tickets in the role-known-condition ($r$=.83, $p$<.001), but not in the role-unknown-condition ($r$=.10, $p$=.65). For both conditions, there was no association between indicating what the moral decision would be and the actual behavior (both $r_s$ <.20, $p_s$>.38).

When looking at the correlations by condition and sex, pre-allocation judgments of morality positively predicted ticket allocations for females ($r$=.66, $p$=.05), but not males in the role-known-condition ($r$=-.16, $p$=.65), but not in the role-unknown-condition (both $r_s$=.32, $p_s$>.30). So, how exactly did people allocate tickets?

*Mean number of tickets allocated to self*. The mean number of tickets allocated to self did not vary by condition[1] ($F$(1,38)= 1.30, $p$=.27), sex ($F$(1,38) = 0.05, $p$=.82) or condition x sex ($F$(1,38) = .68, $p$=.42). For mean allocations to self by condition and sex, please see Table 2.

---

[1] When treating tickets allocated to self as a count variable and using Poisson regression to estimate the same model, the results do not change.

Among those who indicated in their pre-rating that favoring the self would be fair (i.e, giving more than 5 tickets to oneself would be fair; $n$=6), the pre-rating was perfectly correlated with the number of tickets allocated to self ($r$=1.0, $p$<.001; for all correlations among those who favored self, see Table 5, but note that these are based on a very small sample). Among those who indicated that the fair way to allocate would be to give 5 tickets or less to oneself ($n$=36), the pre-decisional rating of fairness did not predict actual ticket allocations ($r$=.08, $p$=.64).

*Percentages of other-favoring, fair, and selfish behavior.* Another way to look at the allocation of tickets is to categorize the behavior as favoring the other (fewer than 5 tickets to self), fair (5 tickets to self), or selfish (more than 5 tickets to self) (also see Table 3). Calculating the percentages of other-favoring, fair, and selfish behavior by condition and sex yielded the following results: Among females, 80% (role-unknown-condition) and 78% (role-known-condition), respectively acted fairly. Among males, 33% (role-unknown condition) and 45% (role-known condition) acted fairly. For all percentages, please see Table 3. Although it seems that there is a sex effect on behavior, estimating a cumulative logit model indicated that neither condition, sex, nor the interaction reliably predicted the three-level outcome (all Wald $\chi^2$ < 2.3, $p$>.13). The same is true of treating the outcome variable as binary (favoring self vs. being fair or favoring other; all Wald $\chi^2$ < 2.5, $p$>.12). Figure 2 depicts the number of tickets allocated to oneself in histograms by condition and gender.

Using behavior as a three-level outcome variable, it is also possible to ask whether the odds of giving more tickets to oneself increased from what people said to what people did, and whether this varied by condition. Indeed, using a generalized estimating equation (GEE) for multinomial models revealed that such a difference existed (Wald $\chi^2$ = 3.5, $p$=.06 for the interaction of time of estimate and condition). Specifically, the odds of making a selfish

allocation were 9 times the odds of saying it was fair to favor oneself before the allocation in the role-unknown condition. In the role-known condition, the discrepancy between what people said and what they did was smaller: the odds of giving oneself more than five tickets were only 1.5 times the odds of saying it was fair to give oneself more than five tickets.

In sum, in both conditions, eight participants (or 36% of participants in the role-unknown-condition and 40% of participants in the role-known-condition) favored themselves. In other words: even though not one of the participants in the role-unknown-condition said the fair or moral thing to do would be to allocate more tickets to self, eight participants proceeded to do so, going against the judgment they made only minutes prior. This behavior is not consistent with moral disengagement, which would suggest that one should adhere to a standard that was not disengaged.

*Ratings of fairness and morality after the ticket allocation*. When comparing the post-behavioral ratings of fairness and morality of one's actions across the two conditions, no reliable differences between the conditions emerged ($F$s for both measures <.60, $p_s$>.50). A sex effect emerged on ratings of fairness of one's actions, such that females rated their behavior as more fair ($M$=6.1) than males ($M$=4.4), $F(1, 38)=5.96$, $p$=.02. The same was true of morality judgments, though the sex effect was marginal, $F(1,38)=3.66$, $p$=.06. These sex effects remain when controlling for participants' behavior by including the number of tickets allocated to self as a continuous covariate in the model. In addition, a condition main effect emerges on the fairness item indicating that participants in the role-known condition thought their behavior was more fair ($M$=5.8) than did participants in the role-unknown-condition ($M$=4.7), $F(1,37)=6.5$, $p$=.02.

Perhaps more informative is a comparison of post-behavioral ratings of morality and fairness based on actual behavior: do those who favored themselves differ from those who favored other

or those who acted fairly? Indeed, such differences emerged: those who favored themselves rated their behavior as less moral ($M$=3.3) than those who acted fairly or favored the other ($M$=6.6), $F(1,40)=73.1$, $p<.001$ (also see Table 6). The same was true of fairness: Again, favoring oneself was clearly seen as less fair ($M$=2.8) than allocating tickets evenly or giving more to the other ($M$=6.6), $F(1,40)=73.1$, $p<.001$. This clearly indicates an understanding by our participants that allocating resources evenly is fair and moral and that favoring oneself is neither fair nor moral.

Looking at post-behavioral judgments of morality and fairness by pre-behavioral ratings of fairness indicates that those who prior to making their allocation said giving more to the self would be fair (all of whom proceeded to favor themselves) indicated that their behavior was less fair ($M$=3.0) than did those who initially said giving 5 tickets or less to oneself was fair and who did not violate the fairness norm ($M$=6.6), $F(1,30)=31.01$, $p<.001$. The same was true of ratings of post-decisional morality: those who said initially favoring the self would be fair (and proceeded to do so), indicated afterwards that their behavior was less moral ($M$=3.5) than did those who initially said giving 5 tickets or less to self was fair (and adhered to that) ($M$=6.6), $F(1,30)=31.6$, $p<.001$. Furthermore, comparing those who initially said favoring the self would be fair (all of whom went on to favor themselves) to those who initially said giving five or more tickets to the other would be fair, but who also went on to favor themselves, revealed no differences in their post-behavioral ratings of fairness ($M_{priorjustification}$=3.0, $M_{nopriorjustification}$=2.6; $F(1,14)=.38$, $p=.55$; also see Table 7) or morality ($M_{priorjustification}$=3.5, $M_{nopriorjustification}$=3.2; $F(1,14)=.15$, $p=.70$). Thus, even though ahead of time six participants indicated that giving oneself more than five tickets could be fair, their post-behavioral judgments do not reflect such lenient fairness standards. Plausibly, these self-favorable judgments enabled these participants to favor themselves in their allocations. Still, given their post-behavioral judgments it appears as if

the fairness norm re-emerged in their thinking soon after the decision and there was no way for these six men and women to escape the fact that they had, in fact, not acted fairly. Had the standard been disengaged (in the role-known participants), this should have been reflected in their post-decisional ratings, particularly in comparison to those who did not have an opportunity to disengage the standard ahead of time (participants in the role-unknown-condition).

Do allocations predict post-behavioral judgments of fairness and morality? For both conditions, allocating tickets to oneself was negatively related to post-behavioral judgments of morality ($r_{cond1}$=-.73, $p$<.001; $r_{cond2}$=-.66, $p$<.01) and fairness ($r_{cond1}$=-.60, $p$<.01; $r_{cond2}$=-.90, $p$<.001). Among those who indicated before making their decision that it is fair to favor oneself, post-decisional fairness ratings seemed to not be related to tickets allocated to self ($n$=6; $r$=-.30, $p$=.56; but this is based on a very small sample). However, when comparing those two correlations, they did not reliably differ from each other, $Z$=.63, $p$=.53. For those who indicated that it would be fair to give 5 tickets or less to oneself before making the decision, however, giving more tickets to self predicted lower post-decisional ratings of fairness ($n$=36; $r$=-.64, $p$<.001).

*Emotional responses.* Participants were asked to rate their current emotional state for several adjectives on a scale from 1 (not at all) – 7 (extremely). Emotions of interest in this particular setting were feelings of shame, guilt, justification, happiness, and feeling sorry. These ratings did not differ by condition (all $F$s<1).

A marginal sex effect emerged for ashamed such that males ($M$=2.0) were slightly more ashamed of their behavior than females ($M$=1.3), $F$(1,38)=3.43, $p$=.07. The same was true of guilty: men felt more guilty ($M$=2.4) than women ($M$=1.4), $F$(1,38)=5.1, $p$=.03. The interaction with condition was not significant for any outcome (all $F$s<1.7). When controlling for actual

behavior by including tickets allocated to self as a continuous covariate in the model, the sex main effects remain for guilty and ashamed, and a sex x condition interaction emerges for justified, $F(1,37)=5.5$, $p=.03$. Decomposing this interaction indicates that females felt marginally more justified than males in the role-unknown-condition ($F(1,38)=2.94$, $p=.10$), but not the role-known-condition, $F(1,38)=.03$, $p=.87$.

Predicting affective responses from behavior[2] indicated that those who favored themselves felt more guilty ($F(1,40)=25.2$, $p<.001$), ashamed ($F(1,40)=3.7$, $p=.06$), sorry ($F(1,40)=14.6$, $p<.01$), and less justified ($F(1,40)=45.6$, $p<.001$). There were no differences for happy ($F(1,40)=.35$, $p=56$). For all means and standard deviations, see Table 6.

Instead of predicting emotions from a full crossing of sex, behavior, and condition (see footnote 2), it is possible to estimate separate 2-factor models crossing behavior with condition and sex. Doing so supported the main effects of behavior reported above. In addition, a condition x behavior interaction emerged for justified, $F(1,38)= 4.0$, $p=.05$: though in both conditions those who were fair felt more justified than those who favored themselves, this difference was more pronounced in condition 1 ($M_{unfair}= 3.3$, $M_{fair}=6.1$) than in condition 2 ($M_{unfair}=4.0$, $M_{fair}=5.6$).

Comparing the emotional ratings of those who favored themselves and adjusted the fairness standard ahead of time to those who favored themselves but didn't adjust the standard ahead of

---

[2] Predicting emotional reactions from full crossings of sex, condition, and behavior (2-level variable) is problematic due to several cells with no variance (SD=0, e.g. there was no variability in the role-known condition among females who divided the tickets evenly on ashamed and guilty)

time showed no differences in any of the emotional ratings (all $F$s<1.8, $p$>.20; for means and standard deviations, see Table 7).

Although those who acted selfishly clearly indicate that they feel more negative afterwards than those who acted fairly or favored the other, given that the scale ranged from 1-7 it is important to note that these ratings are still very low overall and thus don't suggest very strong negative reactions by the participants to their own transgressions. That said, there appears to be a tendency for those who did adjust the fairness norm ahead of time to feel even less negatively about their behavior than their counterparts who favored themselves a few minutes after indicating that an even allocation would be fair. Perhaps understandably, the latter participants' reactions resemble dissonance whereas the former participants don't have to deal with the discrepancy of what they said and what they did.

*Correlations of behavior with emotions*. Correlations support these findings: the more tickets participants allocated to themselves, the more sorry ($r$=.46, $p$<.01), ashamed ($r$=.26, $p$=.09), and guilty ($r$=.51, $p$<.01), but the less justified ($r$=-.60, $p$<.001) participants felt. Tickets allocated to self were not related to feelings of happiness ($r$=-.17, $p$=.29). Among those who indicated before making their allocation that it would be fair to give more than 5 tickets to oneself (all of whom were in condition 2), behavior is not correlated with feeling guilty, ashamed justified, happy, or sorry (see Table 5; all $p_s$>.19; however, the values of some correlations suggest rather substantial relationships). Among those who indicated initially that it would be fair to give 5 tickets or less to oneself, allocations do predict emotional responses such that more tickets allocated to self predicted feeling less justified ($r$=-.56, $p$<.001), more guilty ($r$=.55, $p$<.001), more ashamed ($r$=.36, $p$=.03), and more sorry ($r$=.46, $p$<.01). Behavior did not predict feelings of happiness ($r$=-.22, $p$=.22).

*Discussion*

Similar to Study 2, this third study brought participants into the lab and put them into the "hot seat" of actually having to make a decision in which they could maximize their personal chances of winning a gift card (at the expense of violating a widely held moral norm), or they could act in accordance with the standard and, in turn, reduce their chances of winning a highly desirable gift card.

In contrast to Study 2, however, some participants were asked to indicate what the fair and moral decision would be before learning of their role (role-unknown-condition), while others were asked after they had learned of their role (role-known-condition). All were asked to rate the fairness and morality of their actions after making an allocation.

Though no differences in the number of tickets allocated to self emerged across the two conditions, the pre-allocation ratings of morality and fairness did, indeed, differ across conditions. On average, participants in the role-known-condition thought it was fair to give themselves 6.2 tickets, whereas participants in the role-unknown-condition thought the fair way to allocate tickets would be to keep slightly less than half of the tickets (4.6) for themselves. More specifically, six out of eight participants who favored themselves in the role-known-condition clearly adjusted their definition of fairness and morality *before* allocating tickets in their favor. Assuming that "thinking is for doing", these participants may, in fact, have momentarily disengaged their standard to enable them to favor themselves.

Inconsistent with moral disengagement is the fact that these participants showed in their immediate post-behavioral ratings of fairness and morality that they knew they had violated a standard for they judged their own actions more harshly than did those who acted fairly or

favored others. One would assume if the standard had been disengaged participants wouldn't think of what they did as morally wrong. Also inconsistent with the idea of disengagement is the fact that the rate of those who favored themselves in the role-unknown-condition did not differ from that in the role-known-condition, although there was no shifting of the fairness norm in one's favor in the role-unknown-condition. Not knowing on which side of the "allocation fence" they would fall, these participants clearly indicated that the fair thing would be to divide the tickets evenly (if not slightly in favor of the other), only to override these judgments once they found out they would allocate the tickets and proceed to violate the fairness standard at the same rate as those who had the opportunity to "adjust" the standard ahead of time. Once self-interest entered the situation participants started to act in their own favor – regardless of whether the standard was disengaged or not. Of course, it is possible that these participants very briefly disengaged the standard once they found out they were the allocator (which would not be captured by the current design). However, there was little time and opportunity for them to do so and the standard is clearly active in their post-behavioral judgments. It seems more plausible that these participants favored themselves despite the knowledge that they were about to violate the very norm they had just reiterated. Perhaps it was a matter of "following the gut" now and dealing with the aftermath, if necessary, later. Alternatively, it is possible that participants experienced negative affect in anticipation of their moral violation, but then overrode this negative affect with higher order thinking, motivated by self-protection. Work by Valdesolo & DeSteno (2008) is inconsistent with the former, but consistent with the latter possibility: when their participants were hindered from engaging in reasoning after favoring themselves, they rated their own behavior no fairer than that of another. Without cognitive constraint, however, a self-favorable bias emerged in post-behavioral judgments of participant's own and someone else's

identical action. If self-favorable attitudes towards one's own transgressions had been activated automatically, a cognitive load should not have eliminated the effect, Valdesolo & DeSteno argue. Elimination of self-favorable biases in post-decision judgments seems more consistent with an explanation that focuses on volitionally-guided post-behavioral justifications. As alluded to previously, the current data cannot rule out the possibility that such motivated reasoning may not only alter post-decision judgments, but may also enter into the pre-decisional period. Such higher order thinking necessitates cognitive resources and time. Indeed, moral disengagement is predicted to be a controlled process in which people actively restructure the behavior they are motivated to employ. By that logic, providing participants with the temptation of the gift card in light of a clearly accessible standard and providing ample time to think should have the potential to result in self-exonerative reasoning consistent with moral disengagement and, hence, more self-favoring. Study 4 explicitly addressed this time component.

## Study 4

*Overview and predictions*. Even though Study 3 provided little support for moral disengagement, it is possible (though unlikely) that participants in the role-unknown-condition adjusted their judgments once they learned of their role, enabling them to favor themselves. Similarly, it is possible that (some) participants in Study 3 sped through the study, not allowing time for disengagement. Therefore, Study 4 aimed to manipulate the amount of time participants had to deliberate. Also, to avoid locking people into a decision by making a pre-decision rating of fair and moral behavior, Study 4 briefly embedded mention of the fairness standard in the pre-decisional instructions or not. This resulted in a 2 (standard salient: yes vs. no) x 2 (time to think: yes vs. no) between-subjects design. Insofar as time is needed to disengage a salient standard

before a behavior, the time manipulation should have provided participants with opportunity to generate self-exonerative reasoning. In this situation, a moral ought was made salient, a situation was provided that lead most participants to consider violating this standard, and participants were provided the opportunity to spend their time and cognitive resources on self-exonerative reasoning in line with moral disengagement to pave the way for self-favoring behavior. In other words: Study 4 was designed as another opportunity for moral disengagement to reveal itself.

If these conditions really provide fertile ground for moral disengagement and disengagement is a process likely to be seen in those who want to favor themselves, a time main effect would be expected such that more selfish behavior should result from having time to think. Alternatively, since a standard can only be disengaged if it is activated, an interaction may be expected such that disengagement occurs when people have time to think, but only once a standard has been made salient. This is especially true if the standard is not clear. However, given the very transparent and unambiguous situation, it is likely that the standard is obvious, even without activation. If the latter is true, an interaction is less likely since the standard should be salient regardless of our manipulation. It is also possible that highlighting the standard encourages more moral behavior (it may be difficult to ignore or turn off something that is likely fairly obvious to begin with and was further highlighted), which should manifest itself in a main effect of salience.

## *Method*

*Participants and procedure*. 96 undergraduate students (48 females, 48 males) participated in exchange for course credit in an Introductory Psychology class. The cover story and basic procedure were the same as in Studies 2 and 3. The instructions were altered slightly to reflect the change in manipulations. Specifically, to make the standard salient, participants were told in

the pre-allocation instructions that "most people think the fairest way to divide the tickets is evenly, giving 5 tickets to oneself and 5 to the other person, but how you choose to divide them is entirely up to you." This statement was omitted from instructions for participants who did not receive a salience manipulation. To manipulate time to think, participants in the time conditions were told:

> *Some of our participants have found it helpful to be given some time to think about their decision. You may take the next few moments to do so. The screen will automatically prompt you to make your allocation in a few minutes. However, after you see the prompt, you may take as much time as you need before entering your allocation.*
>
> *So, take a few moments to think about how you want to divide the tickets, then, after you see the prompt and are ready, enter your allocation.*

The screen then "froze" for 2 minutes (participants were never told of the exact time). Once the time to think had passed, the screen changed and automatically directed participants to the next screen, prompting them to make their allocation (all instructions and prompts in order of presentation can be found in the Appendix). Participants in the no-time conditions did not read the above instructions. They were prompted directly to indicate their decision. The post-decision questionnaire was identical to that in Studies 2 and 3 with the exception of an added item, prompting participants to write a few sentences, telling us "what thoughts went through your mind between (a) when you learned that you would allocate the tickets and (b) when you indicated your allocation decision on the computer. Which of these thoughts affected your decision? How?"

## *Results*

*Desirability of gift card.* Ratings of the desirability of the gift card ranged from 4 to 7. The mean desirability rating of winning the gift card was 6.5 (*SD*=.97). The desirability did not vary by condition or sex (all main effects and the interaction, *F*s ≤ 1.08, *p*s > .30). Hence, the gift card was again clearly desirable to the participants in the present study.

*Effects of time and salience manipulations on mean number of tickets allocated to self.* Predicting the number of tickets allocated from time, salience, and participant sex indicated that there was a marginal effect of time such that those who had time to think allocated fewer tickets to themselves (*M*=6.0) than did those who had no time to think (*M*=6.7), *F*(1,88)=3.51, *p*=.06. There were no main effects of standard salience *(F*(1,88) = .13, *p*=.73), or sex (*F*(1,88) = 0.93, *p*=.34). None of the 2-way interactions reached statistical significance, either (all *F*(1,88) < .93, *p*>.34). However, a 3-way interaction emerged *(F*(1,88) = 7.21, *p*<.01). Decomposing this interaction indicated that the time x salience interaction was significant for males *(F* (1,88) = 4.93, *p*=.03), but not females *(F* (1,88) = 2.48, *p*=.12). Decomposing the 2-way interaction for males indicted that when the standard was salient, males allocated fewer tickets to self when they had time to think (*M*=4.8) than when they did not have time to think (*M*=6.9), *(F* (1,88) = 6.4, *p*=.01). No such difference occurred when the standard was not made salient for men *(M*s= 6.2 and 6.9; *F*(1,88) = .37, *p*=.55)[3]. Although this effect was not reliable in the current sample, a trend in the current study suggests that time to think actually *decreased* number of tickets allocated to self. For men, this was only true when the standard was made salient. In both cases,

---

[3] When treating tickets allocated to self as a count variable and estimating the model using Poisson regression, p-values are higher. However, the results do not change.

this is contrary to what we would expect if participants needed (and used) time before making a decision to disengage moral standards, enabling them to act more selfishly. For mean allocations to self by condition and sex, please see Table 8.

Another way to look at the behavior is to predict selfish, fair, or other-favoring behavior from time to think, salience of standard, and participants' sex in an ordered logistic regression model. Consistent with the previous, time predicted behavior (*Wald* $\chi^2$ (1, *N*=96) = 5.8, *p*=.02) such that the less time participants had, the more they allocated to self. Also, the three-way interaction previously reported emerged, (*Wald* $\chi^2$ (1, *N*=96) = 5.1, *p*=.02). No other effects reached significance.

Lastly, I regressed the two manipulations and participant sex onto a binary outcome (combining those who acted fairly and those who favored the other as having acted morally). The results were consistent with the previous analyses.

*Percentages of favoring other, fair, and selfish behavior.* Similar to Study 3, it may be informative to tally for each condition the percentage of male and female participants who made a decision to favor the other, act fairly, or favor the self. In each of the two conditions in which participants were given time to think 33% of participants favored themselves. In the two conditions in which participants did not have time to think, 50% (without a salience manipulation) and 58% (with a salience manipulation) favored themselves, reflecting the trend of time to *reduce* ticket allocations to self (for all percentages please see Table 9). Figure 3 displays histograms for the number of tickets allocated to self by condition and sex.

*Ratings of fairness and morality of behavior.* Does the number of tickets allocated to self predict post-behavioral judgments? Indeed, the more tickets participants allocated to themselves, the less moral (*r*=-.76, *p*<.001) and the less fair (*r*=-.74, *p*<.01) they rated their own actions

afterwards. Ratings of morality and fairness were highly correlated, $r$=.88, $p$<.001 (for all correlations of ticket allocation, ratings of morality, fairness, and emotions, please see Table 11). In line with these findings, participants who favored themselves rated this behavior as less moral ($M$=3.2) and less fair ($M$=2.7) than those who acted fairly or favored the other ($M_{moral}$=6.4; $M_{fair}$=6.5), $F_{moral}(1, 94)=168.01$, $p$<.001, $F_{fair}(1,94)=160.8$, $p$<.001.

Predicting participants' ratings of post-behavioral morality and fairness from a full crossing of the two manipulations and participants' sex revealed a marginal main effect of time to think for moral judgments such that those who did have time to think rated their behavior as more moral ($M$=5.3) than those who did not have time to think ($M$=4.6), $F(1,88)=2.88$, $p$=.09. There were no effects on the fairness item. When controlling for participants' actual behavior[4] by including behavior as a continuous covariate, however, the time effect on the morality item completely disappeared, $F(1,87)=0.2$, $p$=.70. However, a 3-way interaction emerged, $F(1,87)=4.42$, $p$=.04. Decomposing this interaction revealed a significant 2-way interaction of time x salience for males ($F(1,87)=5.66$, $p$=.02), but not females ($F(1,87)=.41$, $p$=.52). Further decomposing the 2-way interaction for males indicated that they rated their behavior as more moral when they had had time to think ($M$=5.6) rather than no time to think ($M$=4.4), but only if the standard had not been made salient, $F(1,87)=5.44$, $p$=.02. The same was true for the fairness item: a marginal 3-way interaction emerged, $F(1,87)=3.31$, $p$=.07. Decomposing this interaction revealed a significant 2-way interaction of time x salience for males ($F(1,87)=5.85$, $p$=.03), but not females ($F(1,87)=.03$, $p$=.85). Further decomposing the 2-way interaction for males indicated

---

[4] As in Study 3, including behavior as a 2 or 3-level variable in the model led to data-analytical problem because some cells were either empty or had no variability at all.

that they rated their behavior as fairer when they had had time to think ($M$=5.1) rather than no time to think ($M$=4.7), but only if the standard had been made salient, $F(1,87)=4.91$, $p=.03$.

Because a full crossing of behavior with all manipulations and sex is not possible, additional models were estimated crossing behavior (fair /unfair) with each of the independent variables separately. In each of these 2-factor analyses, only behavior predicted fairness and morality judgments in line with previously reported findings (those who acted fairly or favored the other rated their behavior as more fair and moral than did those who favored self). Across analyses, neither salience, time, sex or any of the 2-way interactions were predictive, all $F_s<2.5$. Looking at all possible 3-factor models including behavior (time, sex, behavior and salience, sex, behavior and time, salience, behavior) in the same way revealed the same patterns: only behavior predicted post-behavioral fairness and morality; none of the 2-way or 3-way interactions were significant, all $F$s< 2.3 It was not possible to estimate effects for post-behavioral fairness for a 2 (salience: high vs. low) x 2 (behavior: fair vs. unfair) x 2 (sex: male vs. female) model due to females who acted fairly and for whom the standard was made salient (n=11) all indicating their behavior was extremely fair ($M$=7.0).

*Predicting emotional reactions from behavior.* The more tickets were allocated to self the more participants reported feeling sorry ($r$=.52, $p$<.001), ashamed ($r$=.46, $p$<.001), and guilty ($r$=.61, $p$<.001) and the less they reported feeling justified ($r$=-.41, $p$<.001). Ticket allocations did not predict feeling happy ($r$= -.13, $p$=.22). These patterns hold across all conditions with one exception: When participants had neither time to think, nor the standard made salient to them (condition 4), allocating more tickets to self predicted feeling less happy, $r$=-.51, $p$<.02. Overall, this reiterates that participants recognized that favoring themselves was a violation of the fairness standard and they felt worse the further from the standard they deviated in their own interest.

Further reiterating these findings, there were reliable mean differences on all emotional ratings (except for happy) between those who favored themselves and those who acted fairly or favored the other, indicating that favoring oneself left participants feeling more guilty, ashamed sorry, and less justified (see Table 10 for means and standard deviations).

*Effects of manipulations on emotional reactions*[5].

---

[5] As in previous analyses, including behavior in these analyses as a factor resulted in problems. 8/16 cells in a time (yes vs. no) x standard salience (yes vs. no) x behavior (self vs. fair/other) x sex (male vs. female) between-subjects design, had several items without any variance. These items were:

Condition 1 (time and standard salient): all 6 females who acted fairly indicated that their behavior was extremely fair (M=7.0, SD=0); all 10 males who acted fairly indicated that they felt no shame whatsoever (M=1.0, SD=0). 2 males who acted selfishly indicated that they didn't feel any shame or guilt (M=1.0 and SD=0 on both items)

Condition 2 (no time, standard salient): 5 females and 5 males who acted fairly indicated their behavior was extremely fair (M=7.0, SD=0 for both males and females); 5 men who acted fairly additionally indicated that they were not at all guilty, ashamed, or sorry (all Ms=1.0, SD=0).

Condition 3 (time, standard not salient): 9 females who acted fairly indicated that they felt not at all ashamed (M=1.0, SD=0) and 7 males who acted fairly indicated that their actions were entirely fair (M=7.0, SD=0) and they did not feel guilty (M=1.0, D=0).

Condition 4 (no time, standard not salient): 5 females who acted fairly indicated that their behavior was extremely moral (M=7.0, SD=0) and they felt not at all guilty or ashamed (M=1.0, SD=0).

*Justified.* A time effect emerged indicating that those who had time to think (*M*=5.4) felt more justified than those who did not have time to think (*M*=4.7), *F*(1,88)= 5.3, *p*=.02. Furthermore, males (*M*=5.4) reported feeling marginally more justified than females (*M*=4.8), *F*(1,88)=3.7, *p*=.06. When including behavior as a continuous covariate, both effects were reduced: time to think, *F*(1,87)=2.74,*p*=.10; sex, *F*(1,87)=2.72, *p*=.10. No other reliable predictors emerged.

When predicting feelings of justification from all possible two- and three-factor models as described above, the same behavior, time and sex main effects emerged. Additionally, a time x behavior interaction consistently emerged indicating that among those who favored themselves participants with time to think felt more justified (*M*=4.8) than participants who did not have time to think (*M*=3.7), *F*(1,27.3)=4.6, *p*=.04. No such difference occurred for those who were fair or favored the other (both *M*s=5.8), *F*(1, 41.3)=.04, *p*=.85. However, this was further qualified by a 3-way interaction with standard salience *F*(1,88)=3.3, *p*=.08, indicating that the former only held for those for whom the standard was salient, *F*(1,34.8)=7.8, *p*=.01 and not for those for whom the standard was not made salient, *F*(1,24.1)=0, *p*=.98.

---

The problem was dealt with by controlling for behavior by entering it as a continuous covariate and also by running all possible 2- and 3-factor models including behavior and the independent variables (2-factor models: time and behavior, salience and behavior, sex and behavior; 3-factor models: time, salience, behavior, time, sex, behavior, salience, sex, behavior). None of the latter analyses encountered the problem of zero variance in a cell. To account for heterogeneity of variance, the Welch-Satterthwaite adjustment was used in all follow-up tests.

*Guilty.* Consistent with the previous analyses, those who had time to think felt slightly less guilty ($M$=1.9) than those who had no time to think ($M$=2.5), $F(1,88)=3.20$, $p$=.08. Females reported feeling more guilty ($M$=2.6) than males ($M$=1.8), $F(1,88)=4.6$, $p$=.04. No other effects reached significance. When controlling for behavior, the time effect was no longer reliable, $F(1,87)=.72$, $p$=.40. Predicting guilty feelings from behavior, sex, and the manipulations in separate models further supported the behavior and sex main effects previously reported. Additionally, a sex x behavior interaction emerged $F(1,88)=3.8$, $p$=.05, indicating that females who had acted unfairly had a tendency to feel more guilty ($M$=4.0) than males who had favored themselves ($M$=2.9), $F$ (1, 22.6)=3.5, $p$=.07. Females who acted fairly did not differ from males who acted fairly ($M$s=1.2 and 1.1, respectively), $F(1,24.3)=.21$ $p$=.70.

*Ashamed.* With the exception of a marginal time x salience interaction $F(1,88)=2.91$ $p$=.09, there were no reliable predictors of feelings of shame. Decomposing the interaction revealed that those who had no time to think felt more ashamed ($M$=2.2) than those who did have time to think ($M$=1.5), but only when the standard was made salient ($F(1,88)=3.8$, $p$=.05). Controlling for behavior did not change these results. Again, including behavior in several different analyses supported the main effect of behavior across all analyses and, additionally, revealed a 3-way interaction of behavior, salience, and time, $F(1, 88)=3.4$, $p$=.07. Decomposing this interaction revealed that participants who acted immorally felt more ashamed when they had time to think ($M$=3.2) than when they did not have time to think ($M$=1.9), $F(1,16.6)=3.97$, $p$=.06, but only when the standard had not been made salient, $F(1,22.2)=4.5$, $p$=.05.

*Sorry*. Before entering participants' actual behavior, a marginal time effect emerged ($F(1,88)=3.4$, $p$=.07), suggesting that those who had no time to think felt more sorry ($M$=6.7) than those who did have time to think ($M$=6.0). Also, a salience x sex interaction emerged

($F(1,88)=4.5$, $p=.04$), suggesting that females felt more sorry ($M=2.7$) when the standard was made salient than males ($M=1.6$), $F(1,88)=5.5$, $p=.02$. This gender difference did not emerge when the standard was not made salient, $F(1,88)=.43$, $p=.51$. When including behavior as a covariate in the model, the time effect was no longer significant, $F(1,87)=1.13$, $p=.30$. Predicting sorry from all possible 2- and 3-factor models that include behavior in addition to sex, time, and salience replicated the previously reported main effect of behavior (also see Table 8) and salience x sex interaction.

*Happy*. Happiness was not predicted by sex, time, salience, or any of the interactions, including behavior as a factor or covariate, all $F$s$<2.2$, $p=.15$.

### *Discussion*

Study 4 manipulated whether or not participants had time to think about their allocation before making the decision, either after a standard had been made salient to them or not. Once more, there was evidence that the participants perceived the $100 gift card as highly desirable. Although this effect was not reliable in the current sample, a trend emerged for those who had time to think to allocate *fewer* tickets to self than those who did not have time to think. For men, this was only true when the standard had also been made salient, possibly indicating that men were otherwise able to ignore the standard. However, a combination of a salient standard and time to reflect on the standard for men seems to have provided a condition of high self-awareness in which it was difficult to favor oneself without negative implications for the self. Only two men favored themselves and indicated feeling justified and happy, but not at all sorry, guilty, or ashamed about their behavior; six females favored themselves in the same condition, but clearly indicated feeling bad about their decision afterwards. Overall, moral behavior seems to have

been strengthened, rather than attenuated, by providing participants with time to think, which is inconsistent with the idea that participants generate self-exonerative arguments in anticipation of acting immorally.

## General discussion

### *Recap of studies' rationale and findings*

The goal of the present research was to provide a first test of the temporal aspect of moral disengagement theory. Though moral disengagement has enjoyed much popularity as a theory when explaining immoral behavior, many of its predictions have not actually been tested. Rather, moral disengagement appears to have become somewhat of an umbrella term for the justification of immoral behavior, particularly in the context of (intergroup) violence.

Arguably the most crucial of the untested assumptions is that of temporal order. Is it that people restructure their behavior in anticipation of violating a standard (as the theory predicts), or are other processes equally, if not more, plausible? Captured by correlational research, other processes may resemble disengagement on the outside. One may recognize the various disengagement mechanisms in the justifications of immoral behavior (e.g., in scale items intended to measure moral disengagement), yet the processes underlying such justifications remain unclear. Are they post-hoc justifications by those who violated the standard (for whatever reason) and were then put on the spot, or are they deliberate attempts at disengaging a valued standard in the interest of circumventing self-sanctions?

Using a new paradigm, the present research set out to address this problem and provide a first test of the theory as stated by Bandura (e.g., 1999). Specifically, the research sought to test the predictions that when a standard was endorsed in the abstract, but easily violated when personal

gains were expected, restructuring of the behavior as fair and justified, even moral, would precede violation of the standard. Further, there would be an absence of negative emotions in response to the behavior as a result of such disengagement. Adapted from a dictator game, a ticket allocation paradigm was used to show that the fairness standard was an easily activated and widely endorsed moral code (Study 1) and that a $100 gift card was desirable enough to provide a sufficient enticement to violate the fairness standard (Studies 2-4). Moreover, when participants were put in the position of the allocator (i.e., they could enter more tickets into the raffle in their own name, increasing their chances of a win), behavior was greatly at odds with the fairness standard that was easily endorsed in the abstract. Participants here on average allocated 1.5 tickets more to themselves than would be indicated by the fairness norm (Study 2). While self-favoring behavior and its post-decisional justification has been identified in the literature (e.g., Valdesolo & deSteno, 2008), it was unclear whether participants would adjust the fairness standard in anticipation of behaving immorally. For that reason, participants in Study 3 were randomly assigned to make pre-behavioral fairness and morality judgments before or after learning of their personal roles. Though average ticket allocations to self did not differ by condition, 75% of those who favored themselves had indicated ahead of time that such behavior would be fair - if they made these judgments after learning of their role. Even though nobody justified self-favoring ahead of time in the role-unknown condition, participants favored themselves at a comparable rate. In other words, motivated by personal gain, participants in the role-known condition shifted the moral standard in their favor to enable behavior that maximized personal outcomes. However, among participants in the role-unknown condition, *not* shifting the standard in their favor ahead of time was no deterrent to violating the fairness norm. Moreover, regardless of pre-decisional judgments, participants who unfairly favored themselves recognized

their behavior was not fair and moral in the aftermath of their decision. Therefore, the moral "shifting" was not long-lasting, nor do these results seem to indicate post-behavioral adjustment of the standard as a way of justifying one's previous behavior. In fact, those who favored themselves felt comparatively worse about having favored themselves, although overall the emotional reactions were rather mild. Lastly, providing participants with time to generate justifications in light of a salient standard and enticement for violating it (Study 4) provided another test of moral disengagement theory which suggests that disengagement is an effortful rather than an automatic process, requiring time and cognitive resources. Results indicated that participants had a tendency to allocate *fewer* tickets to themselves when they had time to think – another deviation from predictions. Having time to think seems, if anything, to have given students an opportunity to engage the standard. For men this was only true once the standard had been made salient, possibly because they were otherwise able to ignore the fairness standard. Furthermore, there was evidence that those who favored themselves knew their behavior was wrong and that they felt relatively more guilty, ashamed, and sorry than those who acted fairly or favored the other.

In sum, there was little evidence of moral disengagement across the four studies. Although participants were prone to violating a standard that they had endorsed in the abstract when they could gain benefits from doing so, there was little evidence that it was pre-behavioral justification that enabled people to do so. Self-favoring occurred even in the face of clearly violating a previously endorsed norm. Providing participants with time to deliberate didn't encourage the kind of self-exonerative reasoning suggested by moral disengagement either – in fact, people were less likely to violate the norm when they had time to think. Moreover, participants who did violate the fairness norm across the board indicated knowing that they had

done so and that they felt relatively worse about it than their fair counterparts (although not to a degree to which they seemed to suffer as a result of their action.)

### *If not moral disengagement, then what?*

Among those who favored themselves, there was no evidence of moral disengagement in the studies presented here. In the context of this paradigm the failure to show disengagement of the moral standard begs the question of what, then, allowed people to violate the standard. Rather than disengaging the standard, it may have been that participants never *engaged* the standard.

It is possible that participants avoided thought of the moral standard altogether – however, that seems unlikely given the frequent reminders of it in every study, either in the form of asking participants to indicate the standard, or by telling them the standard. Moreover, Studies and 1 and 2 indicate that the fairness standard in the current situation is very salient and hence unlikely to be easily ignored.

Perhaps, when asked, students identified knowledge of, rather than endorsement of the standard. It seems that participants did what they desired to do and took into account that they may feel guilty. Still, the feelings of guilt were mild and it is, of course, a possibility that participants reported feeling guilty because it seemed the appropriate (and learned) response to the situation. In the terms of Deci (1995), participants did not seem to have integrated the fairness standard to the point where they genuinely embraced and valued it for its own sake. If anything, they may have introjected it; they may have learned to adhere to it to avoid punishment from self and others. However, little speaks to the fact that participants tried to adhere to the standard in order to avoid self-punishment in the form of guilt (external punishment was explicitly removed as a possibility by telling participants that neither the experimenter nor the other participant would learn of their decision). Rather, it seems participants knew of the fairness

rule (they indicated so), but broke it anyway, and then, for good measure and because that's what they have learned to do, they nodded to the standard but without conviction. Perhaps they have learned that the fairness standard is one that is broken frequently by themselves and others without many repercussions. In other words: they shrugged off their moral violation. Even if so, shrugging may be an after-effect. It is possible that participants, rather than avoiding thought of the moral standard, avoided thinking of their behavior before engaging in it. Study 4 in particular indicates that the time participants were provided was not so much a welcome opportunity for disengagement, but rather an unwelcome interruption to a more automatic desire to favor the self (or, as mentioned, an opportunity, if not an unintended but unmistakable hint, to them to engage the standard). Counter to moral disengagement, participants may not have *wanted* to think too much about their impending action and its implications. In light of the incentive of the gift card, they may have been driven by a desire to *avoid* engaging or disengaging the fairness norm altogether until they get to the desired behavior, at which point they would deal with the effects of their actions (however, some participants in Study 3 clearly engaged the fairness norm and still went on to break it, full-well knowing they did so). If anything, participants quickly altered the standard to serve their purposes, but soon thereafter admitted to having done wrong. Perhaps participants have learned that the rewards of acting against the "oughts" can exceed the punishments. And even if punishment is to come, it can often be reduced by appealing to ignorance, need, or other extenuating circumstances. This idea that participants may "act now, think later" is reminiscent of cognitive escape models that have been proposed in the context of coping, clinically relevant eating patterns, or risk-taking behaviors in health settings (e.g., unprotected sex among homosexual males; Heatherton & Baumeister, 1991; Krohne, 1993; McKirnan, Ostrow, & Hope, 1996; Nemeroff, Hoyt, Huebner, & Proescholdbell, 2008). A

similar process has been suggested by Batson, Thompson, Seuferling, Whitney, & Strongman (1999) who found that participants ignored a clearly-labeled coin when flipping it assigned another participant (rather than themselves) to a positive task. The authors argue that participants avoided linking their own behavior to the very salient standard and show that making participants self-aware can eliminate this avoidance of one's standard-behavior discrepancy.

Lastly, it is possible that participants disengaged neither a standard nor thought, but affect. Perhaps thinking of one's own (impending) behavior as immoral arouses anxiety but also anticipatory guilt and shame. What if it is those feelings that participants try to avoid; what if it is those feelings that they cannot escape from when they are made self-aware, as was the case in Batson et al.'s study? Indeed, altering or avoiding aversive experiences (rather than "just" thought) has been discussed as an important factor underlying many of the psychopathologies mentioned above (Chawla & Ostafin, 2007). It seems plausible that feeling the negative consequences of one's behavior (quite literally) is what encourages participants to behave morally. Perhaps it's not thoughts that are punishing, but the accompanying feelings. Knowing that one is about to violate a standard without the pang of anticipatory guilt as a warning sign to oneself might not be a sufficient or effective means of discouraging immoral behavior.

Two studies hint at the importance of affect in immoral behavior, and the possibility of affective disengagement: Dienstbier & Munter (1971) found that participants who had an opportunity to attribute negative arousal to a pill subsequently engaged in more cheating. "Knowing" the aversive affect they experienced resulted from something other than one's intended behavior paved the way to that behavior. Similarly, Schachter & Latané (1964) found that participants were more likely to cheat when they had been given the tranquilizer chlorpromazine rather than a placebo. Anecdotal evidence, too, speaks to the possibility that

depressing arousal and numbing oneself enables people to commit even the worst of crime: police officers in Battalion 101 were charged with the task of exterminating the village of Józefów in Poland in July 1942. Among those who stayed to carry out the massacre, heavy alcohol consumption was reported, both during and after this unspeakable carnage (Browning, 1992).

In sum, further inquiry into cognitive and affective disengagement as ways in which people deal with the anticipation of immoral behavior seems warranted.

### *Why no moral disengagement? Is moral disengagement dead?*

The fact that the present data are inconsistent with moral disengagement does not, of course, mean that it is time to lay the idea and study of moral disengagement to rest. If anything, it seems the story may be more complicated than thus far assumed, and it is quite possible that disengagement occurs in situations other than the one created in the present study. One may argue that allocating raffle tickets and giving oneself a ticket or two extra is hardly comparable to some other moral violations, examples of which were mentioned at the beginning of this paper. The present paradigm and paper are merely one way of addressing the question of how and when people justify immoral acts. Of course, it is possible that the prospect of other, more severe, violations is more threatening and harder to "shrug off", more difficult to not think about, impossible to not feel. Bandura (1999) discusses several processes of moral disengagement in more serious circumstances, like the workings of corporations or moral disengagement in the context of war. It is possible that moral disengagement processes are activated in such contexts, in which repeated exposure to transgressions by others, acceptance of responsibility by authority figures, mention of the "other" as subhuman or deserving of harmful treatment, etc. provide the fiber out of which individuals may spin their immoral actions. Still, the reports of the Józefów

massacre and the high rates of PTSD among those who have killed in the context of war (and who have been exposed to ample rhetoric about the good purpose of the act and the bad character of the "enemy") indicate that these justifications do not necessarily suffice to wipe out the embarrassment, the horror, the shame, and the guilt of one's actions either (Maguen, Metzler, Litz, Seal, Knight, & Marmar, 2009; Maguen et al., 2010).

Many immoral actions, no matter how desired, righteously perceived, and carefully calculated do seem to leave their mark not only the victim, but also on the perpetrator. Edgar Allen Poe's character was not able to calmly tell his story, as promised, after all. His guilt so famously manifested itself in the reverberations of the victim's still-beating heart from under the floorboards, making him increasingly fearful and ultimately leading himself to confess the crime.

Do we shrug off the small transgressions, while the big(ger) ones eventually come back to haunt us - not just in form of the law, but also in our conscience, making justifications fleeting and fickle? Clearly, there remains a need and plenty of room for further research on how, when, and why people are enabled to behave immorally, and to what extent the various strategies are successful at reducing the kinds of negative thoughts and feelings that so haunted Edgar Allen Poe's character.

# List of References

Bandura, A. (1990). Selective activation and disengagement of moral control. *Journal of Social Issues, 46*, 27-46.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W.M. Kurtines & J.L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol.1, pp. 45-103). Hillsdale, NJ: Erlbaum.

Bandura, A. (1999). Mechanisms of moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review, 3*, 193-209.

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education, 31*, 101-119.

Bandura, A., C. Barbaranelli, G. V. Caprara, & C. Pastorelli (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71*, 364-374.

Batson, C.D. (2011). What's wrong with morality? *Emotion Review, 3*, 230-236.

Batson, C.D. (in prep.). *What's wrong with morality?* Unpublished manuscript.

Browning, C.R. (1992). *Ordinary men: Reserve Police Battalion 101 and the Final Solution in Poland*. New York: HarperCollins.

Castano, E., & Giner-Sorolla, R. (2006). Not quite human: Infra-humanization as a response to collective responsibility for intergroup killing. *Journal of Personality and Social Psychology, 90*, 804-818.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology, 63*, 871-890.

Cohrs, J.C., & Moschner, B. (2002). Zur kognitiven Konstruktion von (Un)Gerechtigkeit militärischer Gewalt: Die moralische Beurteilung des Kosovo-Kriegs [Cognitive

construction of the (un-)justness of military force: The moral evaluation of the Kosovo War]. *Zeitschrift für Sozialpsychologie, 33*, 13-24.

Cooper, J., & Fazio, R.H. (1984). A new look at dissonance theory. *Advances in Experimental Social Psychology, 17*, 229-266.

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

Deci, E. L. (1995). *Why we do what we do. Understanding self-motivation*. New York: Penguin.

Detert, J.R., Trevino, L.K., & Sweitzer, V.L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology, 93*, 374-391.

Dienstbier, R.A., & Munter, P.O. (1971). Cheating as a function of the labeling of natural arousal. *Journal of Personality and Social Psychology, 17,* 208-213.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson, and Company.

Festinger, L. & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology, 58*, 203-210.

Grussendorf, J., McAlister, A., Sandström, P., Udd, L., & Morrison, T.C. (2002). Resisting moral disengagement in support of war: Use of the "Peace Test" scale among student groups in 21 nations. *Peace and Conflict: The Journal of Peace Psychology, 8*, 73-83.

Hartmann, T., & Vorderer, P. (2010). It's ok to shoot a character: Moral disengagement in violent video games. *Journal of Communication, 60*, 94-119.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*, 252-264.

Heatherton, T.F., & Baumeister, R.F. (1991). Binge eating as escape from self-awareness. *Psychological Bulletin, 110*, 86-108.

Jackson, L.E., & Gaertner, L.A. (in press). Mechanisms of moral disengagement and their differential use by Right-wing authoritarianism and Social Dominance Orientation. *Aggressive Behavior*.

Krohne, H. W. (1993). Attention and avoidance: Two central strategies in coping with aversivesness. In: H. W. Krohne (Ed.), *Attention and avoidance: Strategies in coping with aversiveness* (pp. 3–15). Seattle: Hogrefe and Huber.

Leyens, J.P., Demoulin, S., Vaes, J., Gaunt, R., & Paladino, M.P. (2007). Infrahumanization: The wall of group differences. *Social Issues and Policy Review, 1*, 139-172.

Maguen, S., Metzler, T.J., Litz, B.T., Seal, K.H., Knight, S.J., & Marmar, C.R. (2009). The impact of killing in war on mental health symptoms and related functioning. *Journal of Traumatic Stress, 22*, 435-443.

Maguen, S., Lucenko, B.A., Reger, M.A., Gahm, G.A., Litz, B.T., Seal, K.H., Knight, S.J., & Marmar, C.R. (2010). The impact of report direct and indirect killing on mental health symptoms in Iraq War veterans. *Journal of Traumatic Stress, 23*, 86-90.

McAlister, A., Bandura, A., & Owen, S. (2006). Mechanisms of moral disengagement in support of military force: The impact of September 11. *Journal of Social and Clinical Psychology, 25*, 141-165.

McKirnan, D.J., Ostrow, D.G., & Hope, B. (1996). Sex, drugs, and escape: A psychological model of HIV-risk sexual behaviours. *AIDS Care, 8*, 655-669.

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

Nemeroff, C.J., Hoyt, M.A., Huebner, D.M., & Proescholdbell, R.J. (2008). The cognitive escape scale: Measuring HIV-related thought avoidance. *AIDS and Behavior, 12*, 305-320.

Opotow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues, 46*, 1-20.

Osofsky, M., Bandura, A., & Zimbardo, P. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior, 29*, 371-393.

Schachter, S., & Latané, B. (1964). Crime, cognition and the autonomic nervous system. In D. Levine (Ed.), *Nebraska symposium on motivation*. Lincoln; Univ. of Nebraska Press, 12, 221–273.

Schopler, J., Insko, C. A., Drigotas, S. M., Wieselquist, J., Pemberton, M. B, & Cox, C. (1995). The role of identifiability in the reduction of interindividual-intergroup discontinuity. *Journal of Experimental Social Psychology, 31*, 553-574.

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

Staub, E. (1990). Moral exclusion, personal goal theory and extreme destructiveness. *Journal of Social Issues, 46*, 47-65.

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

Valdesolo & DeSteno (2008). The duality of virtue: Deconstructing the moral hypocrite. *Journal of Experimental Social Psychology, 44*, 1334-1338.

Vollum, S., Buffington-Vollum, J., & Longmire, D.R. (2004). Moral disengagement and attitudes about violence toward animals. *Society & Animals, 12*, 209-235.

# Appendix

*Table 1 Correlations of allocations, ratings of morality and fairness, and emotions in Study 2 (n=21)*

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|
| 1. Number of tickets allocated to self | 1 | -.61** | -.79*** | -.26 | .39$^{++}$ | .41$^{+}$ | .13 | .24 |
| 2. Fair | | 1 | .62** | .40$^{+}$ | -.43* | -.43* | .04 | -.25 |
| 3. Moral | | | 1 | .44* | -.61** | -.49* | .11 | -.38$^{+}$ |
| 4. Justified | | | | 1 | -.52* | -.57** | .67*** | -.39$^{+}$ |
| 5. Ashamed | | | | | 1 | .84*** | -.41$^{+}$ | .62** |
| 6. Guilty | | | | | | 1 | -.36 | .40$^{+}$ |
| 7. Happy | | | | | | | 1 | -.08 |
| 8. Sorry | | | | | | | | 1 |

*Note*. + $p<.10$, *$p<.05$, ** $p<.01$, ***$p<.001$

*Table 2. Mean number of tickets allocated to self (standard deviations in parentheses) by condition (Study 3)*

| | Women | Men | Total |
|---|---|---|---|
| Role-unknown-condition (*n*=22) | 5.9 *(1.9)* | 5.5 *(2.4)* | 5.7 *(2.2)* |
| Role-known-condition (*n*=20) | 6.1 *(2.2)* | 6.8 *(2.0)* | 6.5 *(2.1)* |

*Table 3. Frequencies of favoring other, acting fairly, or favoring self in ticket allocations by gender, condition (Study 3)*

| | Other-favoring allocation (more than 5 tickets to other) | | Fair allocation (5 tickets to self, 5 tickets to other) | | Selfish allocation (more than 5 tickets to self) | |
|---|---|---|---|---|---|---|
| | Females | Males | Females | Males | Females | Males |
| Role-unknown-condition (n=22) | 0 | 2 (17%) | 8 (80%) | 4 (33%) | 2 (20%) | 6 (50%) |
| Role-known-condition (n=20) | 0 | 0 | 7 (78%) | 5 (45%) | 2 (22%) | 6 (55%) |

*Table 4. Correlations of allocations, ratings of morality and fairness pre and post, and emotions in Study 3 (n=42)*

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. | 10. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Condition 1 (role-unknown, *n*=22) | | | | | |
| 1. Number of tickets to self | 1 | .10 | .20 | -.60** | -.73*** | -.62** | .39$^{+}$ | .46* | -.31 | .54** |
| 2. Fair (pre) | | 1 | .73*** | -.27 | -.12 | -.26 | .18 | .21 | -.16 | .20 |
| 3. Moral (pre) | | | 1 | -.27 | -.01 | -.39$^{+}$ | .06 | .20 | -.03 | -.12 |
| 4. Fair (post) | | | | 1 | .71*** | .65** | -.18 | -.32 | .40$^{+}$ | -.40$^{+}$ |
| 5. Moral (post) | | | | | 1 | .72*** | -.31 | -.43* | .33 | -.70*** |
| 6. Justified | | | | | | 1 | -.26 | -.45* | .33 | -.55** |
| 7. Ashamed | | | | | | | 1 | .82*** | -.26 | .60** |
| 8. Guilty | | | | | | | | 1 | -.18 | .57** |
| 9. Happy | | | | | | | | | 1 | -.51* |
| 10. Sorry | | | | | | | | | | 1 |
| | | | | | Condition 2 (role-known, *n*=20) | | | | | |
| 1. Number of tickets to self | 1 | .83*** | .10 | -.90*** | -.66** | -.56** | .14 | .61** | .05 | .37$^{+}$ |
| 2. Fair (pre) | | 1 | .09 | -.75*** | -.55** | -.52* | -.02 | .34 | .10 | .27 |
| 3. Moral (pre) | | | 1 | .01 | .40$^{+}$ | .12 | -.40$^{+}$ | -.10 | -.04 | -.50* |
| 4. Fair (post) | | | | 1 | .83*** | .65** | -.30 | -.80*** | -.13 | -.56* |
| 5. Moral (post) | | | | | 1 | .69*** | -.67** | -.76*** | -.13 | -.85*** |
| 6. Justified | | | | | | 1 | -.50* | -.66** | .02 | -.68*** |
| 7. Ashamed | | | | | | | 1 | .44* | -.01 | .85*** |
| 8. Guilty | | | | | | | | 1 | .13 | .60** |
| 9. Happy | | | | | | | | | 1 | .09 |
| 10. Sorry | | | | | | | | | | 1 |

*Note*. Fair and moral pre refer to the number of tickets participants indicated would be fair and moral to allocate to themselves; . + p<.10, *p<.05, ** p<.01, ***p<.001

*Table 5. Correlations of number of tickets allocated to self, ratings of morality and fairness, and post-behavioral emotions for those who favored self by pre-behavioral justification (Study3)*

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9.[#] | 10. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Adjusted fairness standard ahead of time ($n$=6) | | | | | |
| 1. Number of tickets to self | 1 | 1*** | -.08 | -.30 | .11 | -.37 | -.61 | -.29 | -.39 | .17 |
| 2. Fair (pre) | | 1 | -.08 | -.30 | .11 | -.37 | -.61 | -.29 | -.39 | .17 |
| 3. Moral (pre) | | | 1 | .77$^{+}$ | .68 | .11 | .04 | -.46 | -.74$^{+}$ | -.59 |
| 4. Fair (post) | | | | 1 | .72$^{+}$ | .41 | .00 | -.30 | -.55 | -.91** |
| 5. Moral (post) | | | | | 1 | .30 | -.62 | -.62 | -.92** | -.83 |
| 6. Justified | | | | | | 1 | .00 | -.62 | .00 | -.60 |
| 7. Ashamed | | | | | | | 1 | .44 | .63 | .32 |
| 8. Guilty | | | | | | | | 1 | .61 | .50 |
| 9. Happy | | | | | | | | | 1 | .60 |
| 10. Sorry | | | | | | | | | | 1 |
| | | | | | Did not adjust fairness standard ahead of time ($n$=10) | | | | | |
| 1. Number of tickets to self | 1 | .08 | .17 | -.64*** | -.69*** | -.56*** | .36* | .55*** | -.21 | .46** |
| 2. Fair (pre) | | 1 | .63 | -.16 | -.05 | -.21 | .13 | .14 | -.15 | .14 |
| 3. Moral (pre) | | | 1 | -.18 | .18 | -.20 | -.20 | .15 | -.01 | -.30$^{+}$ |
| 4. Fair (post) | | | | 1 | .73*** | .61*** | -.26 | -.48** | .20 | -.41* |
| 5. Moral (post) | | | | | 1 | .71*** | -.50** | -.54*** | .18 | -.75*** |
| 6. Justified | | | | | | 1 | -.37* | -.48** | .24 | -.58*** |
| 7. Ashamed | | | | | | | 1 | .73*** | -.16 | .73*** |
| 8. Guilty | | | | | | | | 1 | -.07 | .57*** |
| 9. Happy | | | | | | | | | 1 | -.29$^{+}$ |
| 10. Sorry | | | | | | | | | | 1 |

*Note*. Fair and moral pre refer to the number of tickets participants indicated would be fair and moral to

allocate to themselves; . + $p$<.10, *$p$<.05, ** $p$<.01, ***$p$<.001, # 5 out of 6 participants rated their happiness with a “5”, only one person indicated happiness as a “4”. Hence, correlations with happy may be spurious.

*Table 6. Means and standard deviations (in parentheses) of post-behavioral ratings of fairness, morality, and emotions by behavior (Study 3)*

| | Fair (post) *** | Moral (post) *** | Guilty *** | Ashamed + | Sorry ** | Happy | Justified *** |
|---|---|---|---|---|---|---|---|
| Acted fairly or favored other (n=26) | 6.6 *(1.5)* | 6.6 *(1.0)* | 1.2 *(1.0)* | 1.4 *(1.3)* | 1.4 *(1.1)* | 4.9 *(1.4)* | 5.9 *(1.0)* |
| Favored self (n=16) | 2.8 *(1.2)* | 3.3 *(1.4)* | 3.1 *(1.5)* | 2.2 *(1.2)* | 2.8 *(1.4)* | 4.6 *(.81)* | 3.6 *(1.2)* |

*Note*. *** those who acted fairly differed from those who acted unfairly at $p<.001$; ** $p< .01$; = $p=.06$

*Table 7. Means and standard deviations (in parentheses) of post-behavioral ratings of morality, fairness, and emotions among those who favored self and either adjusted fairness standard before behavior or not (Study 3, n=16)*

| | Fair (post) | Moral (post) | Guilty | Ashamed | Sorry | Happy | Justified |
|---|---|---|---|---|---|---|---|
| Adjusted fairness standard before behavior (n=6) | 3.0 *(.90)* | 3.5 *(1.9)* | 2.8 *(1.5)* | 1.6 *(.52)* | 2.5 *(1.2)* | 4.8 (.41) | 4.0 *(1.1)* |
| Did not adjust fairness standard before behavior (n=10) | 2.6 *(1.4)* | 3.2 *(1.2)* | 3.3 *(1.5)* | 2.5 *(1.4)* | 3.0 *(1.5)* | 4.5 *(.97)* | 3.4 *(1.2)* |

*Note.* there were no mean differences for any of the outcome variables between those who adjusted the fairness standard ahead of time and those who did not

*Table 8. Mean number of tickets allocated to self (standard deviations in parentheses) by condition (Study 4)*

| | Time to think | Standard made salient | Women (n=12 per condition) | Men (n= 12 per condition) | Total |
|---|---|---|---|---|---|
| Condition 1 (*n*=24) | Yes | yes | 6.8 *(2.0)* | 4.8 *(1.9)* | 5.8 *(2.1)* |
| Condition 2 (*n*=24) | | no | 5.6 *(1.6)* | 6.7 *(2.2)* | 6.1 *(2.0)* |
| Condition 3 (*n*=24) | No | yes | 6.6 *(1.7)* | 6.9 *(2.2)* | 6.8 *(1.9)* |
| Condition 4 (*n*=24) | | no | 7.3 *(2.5)* | 6.2 *(2.0)* | 6.7 *(2.3)* |

*Table 9. Frequencies of favoring other, acting fairly, or favoring self in ticket allocations by gender, condition (Study 4)*

| | Time to think | Standard made salient | Other-favoring allocation (more than 5 tickets to other) | | Fair allocation (5 tickets to self, 5 tickets to other) | | Selfish allocation (more than 5 tickets to self) | |
|---|---|---|---|---|---|---|---|---|
| | | | Females | Males | Females | Males | Females | Males |
| Condition 1 (n=24) | Yes | Yes | 0 | 3 (25%) | 6 (50%) | 7 (58%) | 6 (50%) | 2 (17%) |
| Condition 2 (n=24) | | No | 1 (8%) | 0 | 8 (67%) | 7 (58%) | 3 (25%) | 5 (42%) |
| Condition 3 (n=24) | No | Yes | 0 | 0 | 5 (42%) | 5 (42%) | 7 (58%) | 7 (58%) |
| Condition 4 (n=24) | | No | 0 | 1 (8%) | 5 (42%) | 6 (50%) | 7 (58%) | 5 (42%) |

*Table 10. Means and standard deviations (in parentheses) of post-behavioral ratings of fairness, morality and emotions by behavior (Study 4)*

| | Fair (post) *** | Moral (post) *** | Guilty *** | Ashamed *** | Sorry *** | Happy | Justified *** |
|---|---|---|---|---|---|---|---|
| Acted fairly or favored other (*n*=54) | 6.5 *(1.4)* | 6.4 *(1.1)* | 3.5 *(1.9)* | 1.2 *(.6)* | 1.3 *(.8)* | 5.1 *(1.5)* | 5.8 *(1.2)* |
| Favored self (*n*=42) | 2.7 *(1.5)* | 3.2 *(1.3)* | 1.2 *(.7)* | 2.6 *(1.5)* | 3.1 *(1.8)* | 4.8 *(1.2)* | 4.1 *(1.6)* |

*Note*. *** those who acted fairly differed from those who acted unfairly at $p<.001$

*Table 11. Correlations of allocations, ratings of morality and fairness, and emotions in Study 4 (n=96)*

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|
| | | | | Condition 1 (time and salience, n=24) | | | | |
| 1. Number of tickets allocated to self | 1 | -.50* | -.74*** | .14 | .50* | .59** | .24 | .55** |
| 2. Fair | | 1 | .73*** | .10 | -.38$^{+}$ | -.51* | -.07 | -.50** |
| 3. Moral | | | 1 | .10 | -.45* | -.66*** | -.18 | -.63** |
| 4. Justified | | | | 1 | -.15 | -.31 | .42* | -.27 |
| 5. Ashamed | | | | | 1 | .80*** | -.07 | .77*** |
| 6. Guilty | | | | | | 1 | -.18 | .88*** |
| 7. Happy | | | | | | | 1 | -.21 |
| 8. Sorry | | | | | | | | 1 |
| | | | | Condition 2 (no time and salience, n=24) | | | | |
| 1. Number of tickets allocated to self | 1 | -.90*** | -.90*** | -.60** | .39$^{+}$ | .49* | -.09 | .35$^{+}$ |
| 2. Fair | | 1 | .97*** | .70*** | -.56** | -.64*** | .16 | -.55** |
| 3. Moral | | | 1 | .74*** | -.56** | -.63** | .14 | -.52** |
| 4. Justified | | | | 1 | -.51* | -.53** | .46* | -.53** |
| 5. Ashamed | | | | | 1 | .87*** | -.44* | .90*** |
| 6. Guilty | | | | | | 1 | -.33 | .92*** |
| 7. Happy | | | | | | | 1 | -.49* |
| 8. Sorry | | | | | | | | 1 |

*Note*. + $p<.10$, *$p<.05$, ** $p<.01$, ***$p<.001$

Table 11 continued.

| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|
| Condition 3 (time and no salience, n=24) | | | | | | | | |
| 1. Number of tickets allocated to self | 1 | -.91*** | -.80*** | -.32 | .77*** | .69*** | -.13 | .54** |
| 2. Fair | | 1 | .91*** | .36$^{+}$ | -.79*** | -.71*** | .09 | -.63** |
| 3. Moral | | | 1 | .48* | -.85*** | -.86*** | .09 | -.70*** |
| 4. Justified | | | | 1 | -.48* | -.51* | .33 | -.22 |
| 5. Ashamed | | | | | 1 | .90*** | -.21 | .72*** |
| 6. Guilty | | | | | | 1 | -.12 | .73*** |
| 7. Happy | | | | | | | 1 | -.21 |
| 8. Sorry | | | | | | | | 1 |
| Condition 4 (no time and no salience, n=24) | | | | | | | | |
| 1. Number of tickets allocated to self | 1 | -.73*** | -.65*** | -.62** | .19 | .64*** | -.51* | .58** |
| 2. Fair | | 1 | .89*** | .54** | -.26 | -.55** | .36$^{+}$ | -.51* |
| 3. Moral | | | 1 | .61** | -.27 | -.43* | .42* | -.42* |
| 4. Justified | | | | 1 | -.48* | -.52** | .66*** | -.50* |
| 5. Ashamed | | | | | 1 | .67*** | -.25 | .81*** |
| 6. Guilty | | | | | | 1 | -.33 | .83*** |
| 7. Happy | | | | | | | 1 | -.33 |
| 8. Sorry | | | | | | | | 1 |

*Note*. + $p<.10$, *$p<.05$, ** $p<.01$, ***$p<.001$

*Figure 1. Histogram of ticket allocations to self by gender, Study 2*

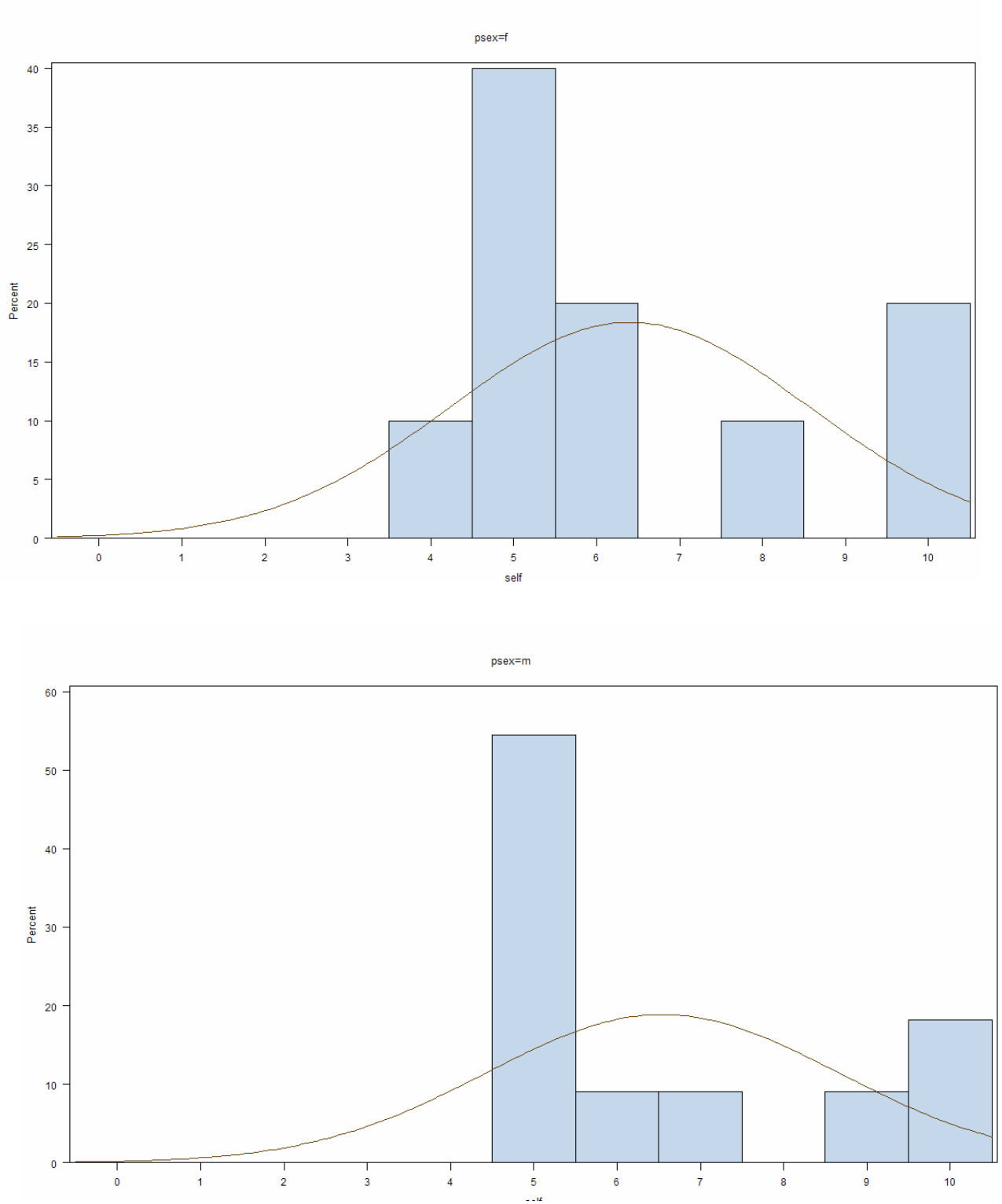
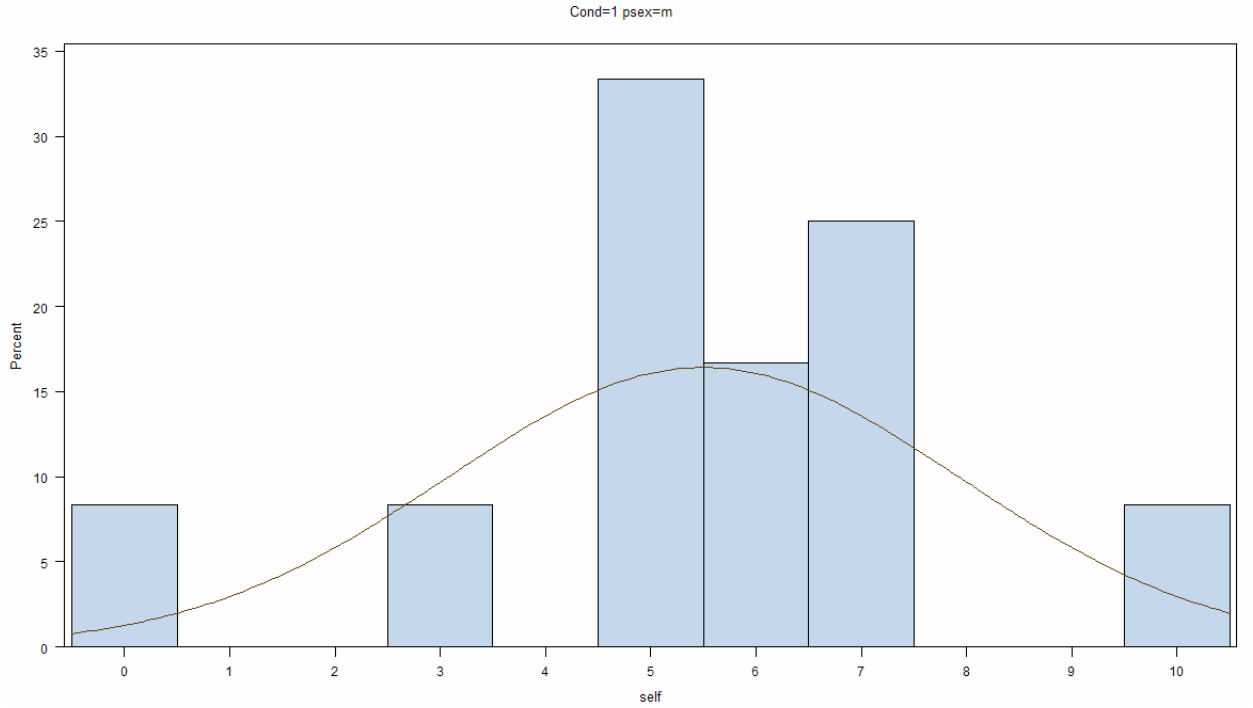


*Figure 2. Histogram of ticket allocations to self by condition and gender, Study 3*



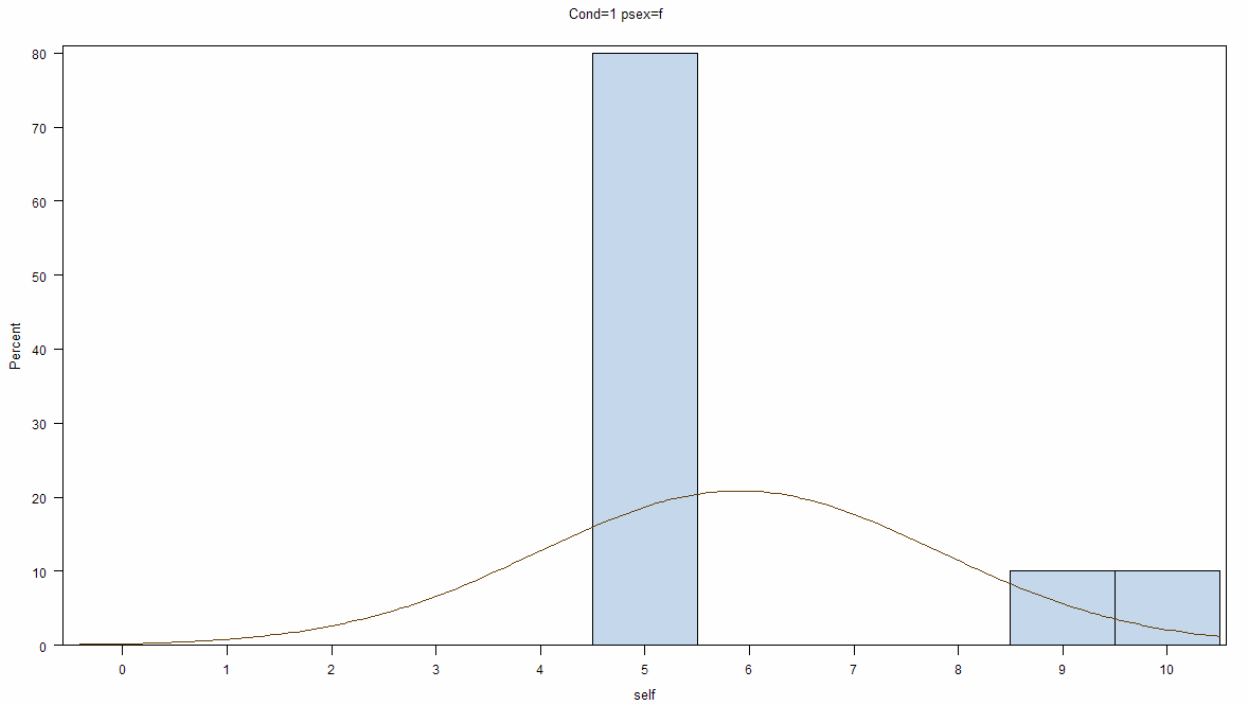
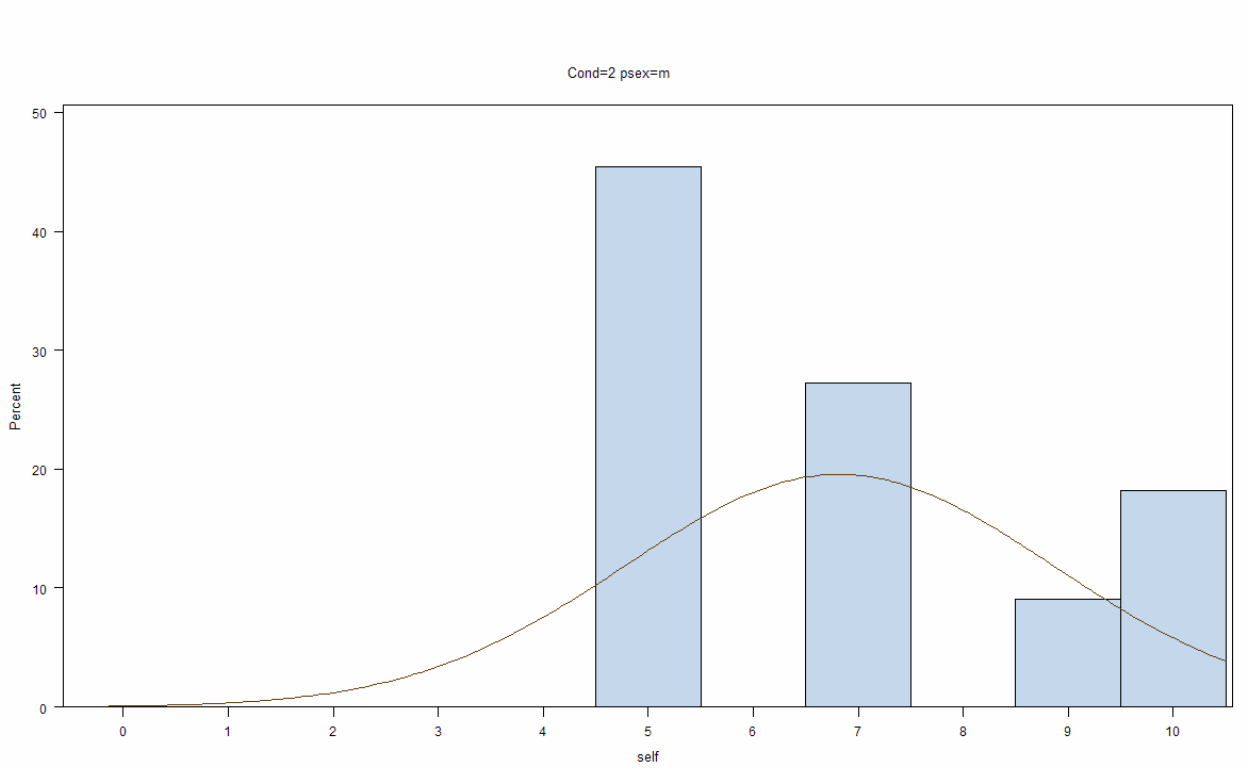



Figure 2 continued.



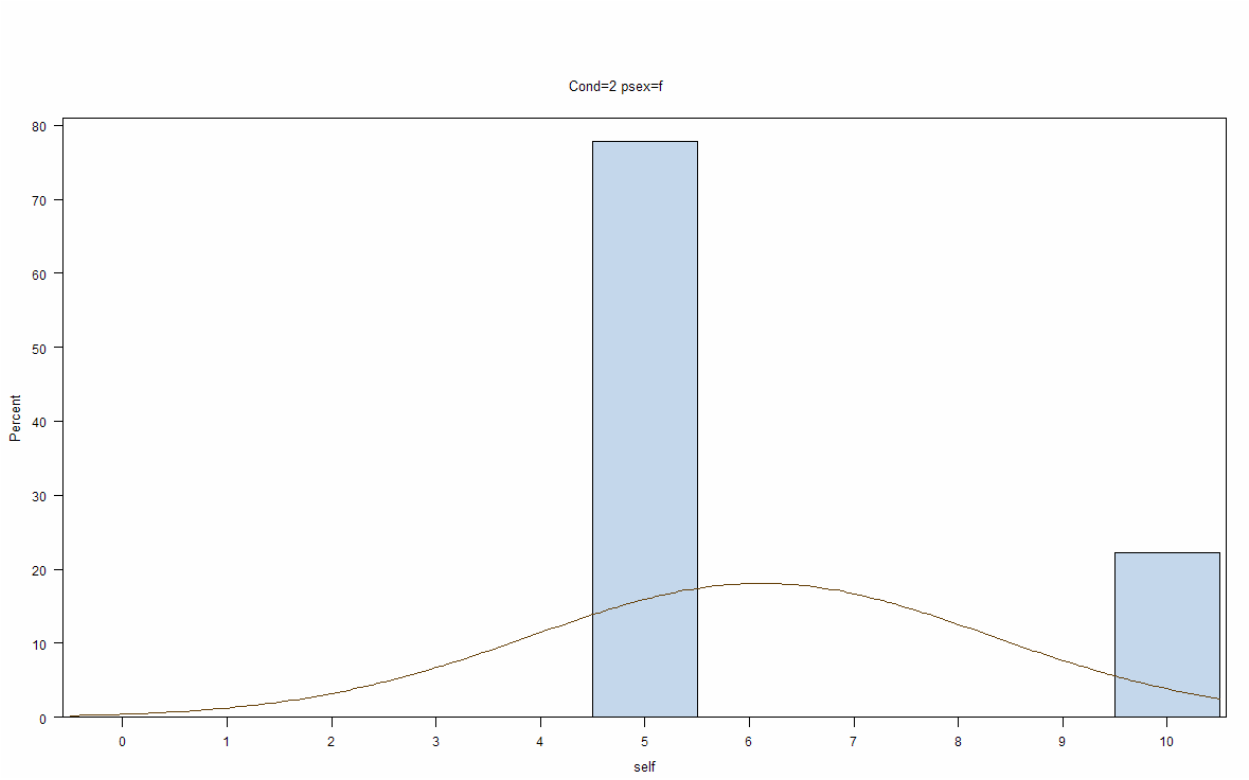



*Figure 3. Histogram of ticket allocation to self by condition and gender, Study 4*

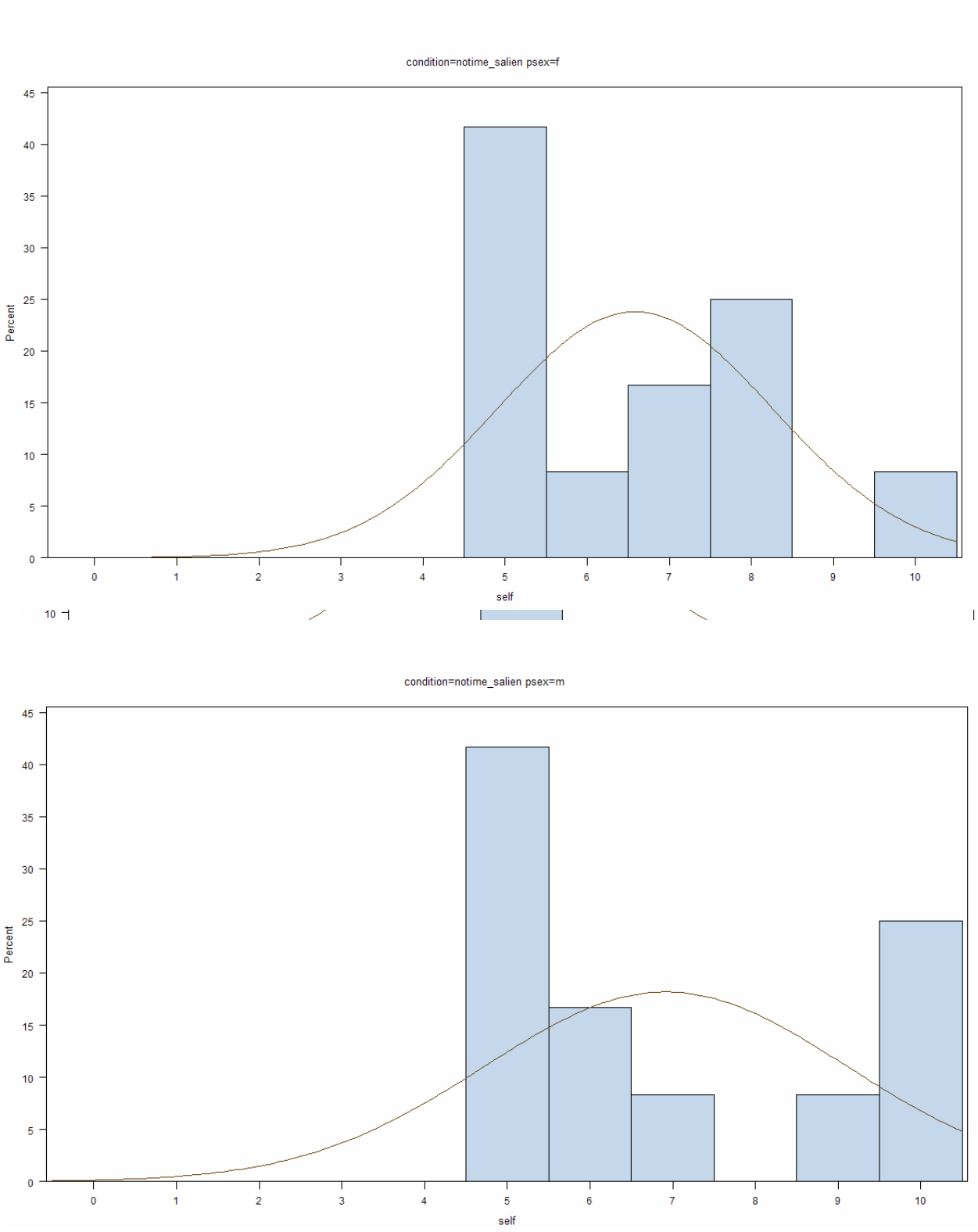
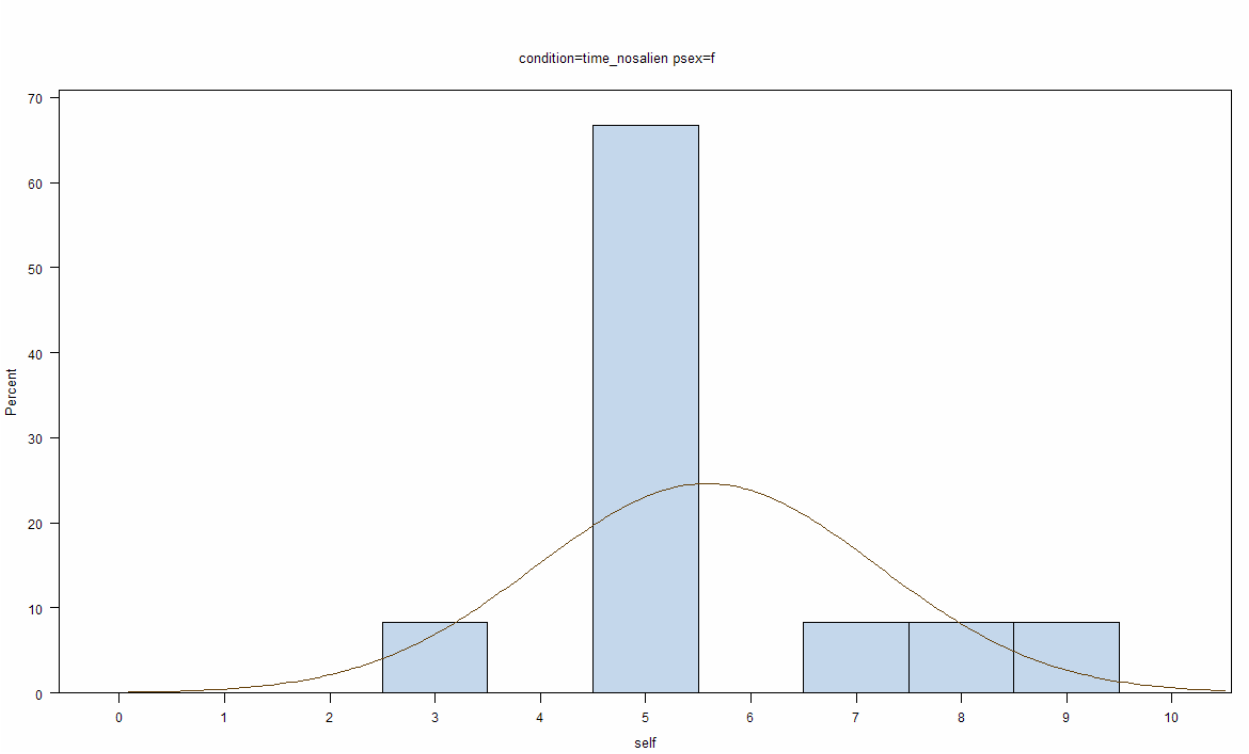


Figure 3 continued.



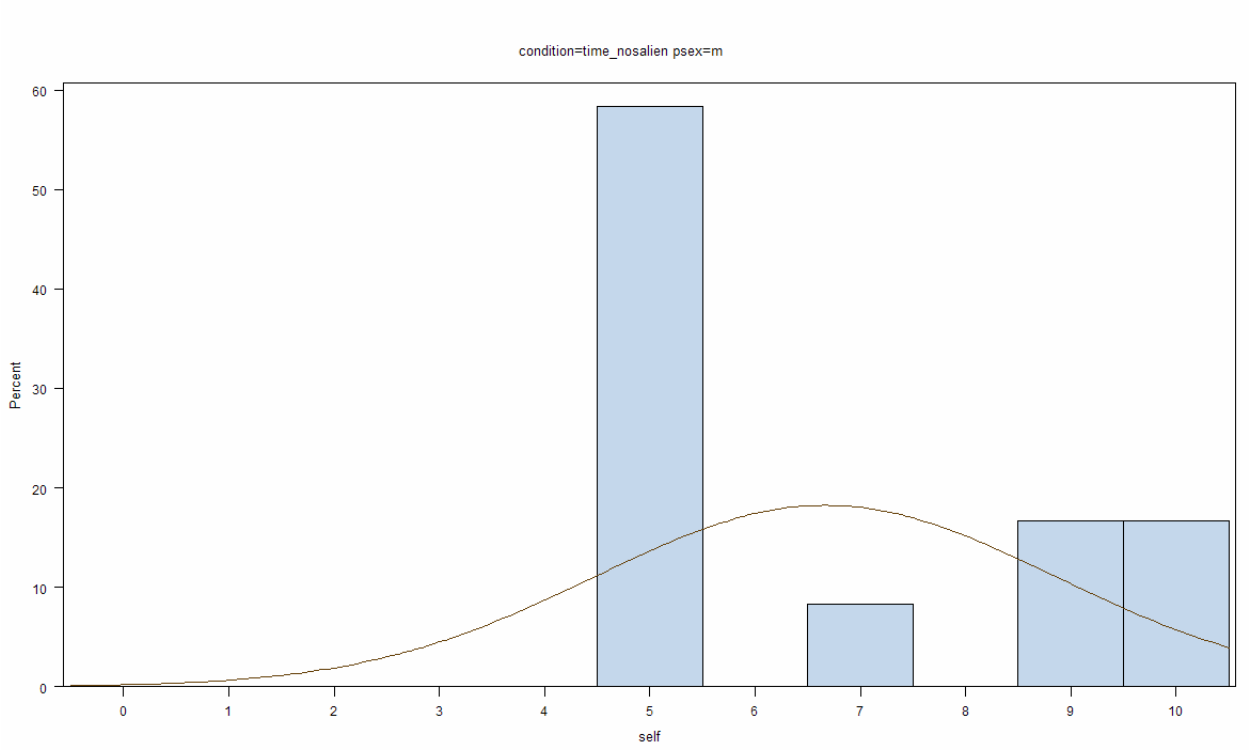
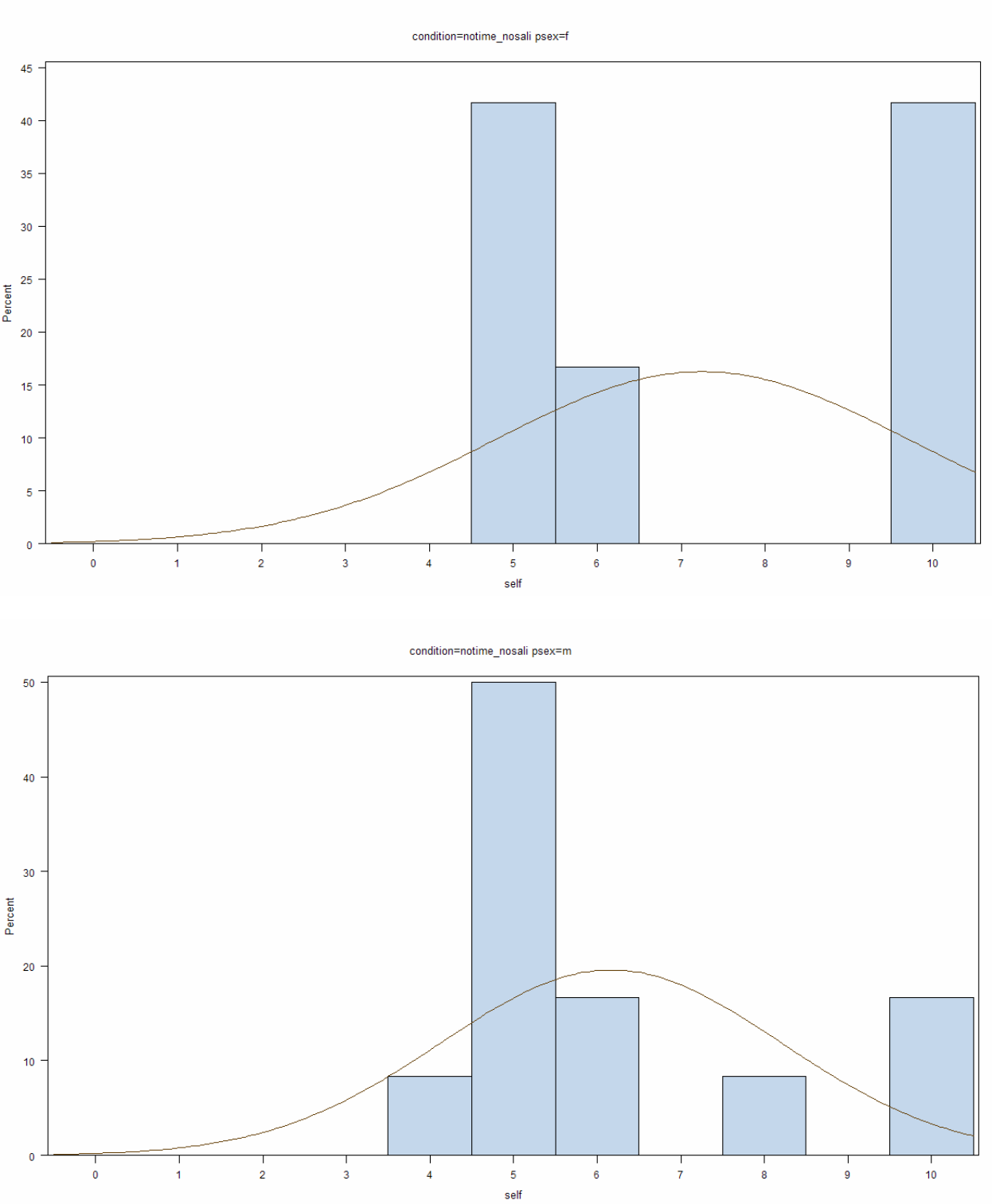



Figure 3 continued.



*Study Materials*
*(informed consent, lottery consent, and Introduction were identical for Studies 2 and 3)*

1. Informed Consent

Informed Consent

The following is a general description of the present study and a reminder of your rights as a potential participant. It is important to keep in mind that this study is part of a research project. As in any study, (a) participation is voluntary, (b) refusal to participate will involve no penalty or loss of benefits to which participants are otherwise entitled, and (c) participants can withdraw from the study at any time without penalty of loss of benefits to which they are otherwise entitled. This study is being conducted in the Psychology Department.

The present study was designed to assess psychological responses to decisions made in groups. You and another same-sex participant will be exposed to a decision that has consequences for you both. One of you will be randomly assigned to make the decision and both of you will be asked for your thoughts and feelings in several questionnaires. All tasks will be completed alone and there will be no interaction between the participants. All of your responses will be confidential and will be coded and securely stored. There are no foreseen risks in participating. The study will take approximately 60 minutes to complete, and you will receive one hour of experimental credit for participating. In addition, you will gain direct experience of psychological research. Your participation will benefit us in furthering our understanding of how people respond to decisions in groups that have the potential to affect several members of the group.

We will try to answer any and all questions you may have about the research. If your questions have not been adequately addressed, you can contact the principal investigators (Lydia E. Jackson at leckstei@utk.edu or her faculty advisors Dr. Dan Batson at dbatson@ku.edu or Dr. Lowell Gaertner at gaertner@utk.edu). This research study has been reviewed and approved by the Institutional Review Board - Human Subjects in Research. For research-related problems or questions regarding subjects' rights, the Institutional Review Board may be contacted through the Compliance Office at 974-3466.

**Consent**

I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study. I have been provided a copy of this consent form.

Printed Name _____

Signature and Date: _____

2. Introduction

## Introduction

Welcome and thank you for participating in our study!

In this study, we are interested in how people think and feel about decisions made in groups. As you know, group members have to make many different decisions that may affect themselves and other members of the group. Some examples of such decisions may be assigning tasks and roles, responsibilities, or resources to group members. How do people who make such a decision feel? How do people who do not make the decision but are still affected by it feel?

One of many open questions is how people feel about such decisions when they don't know each other and don't interact with each other.

This research tries to address that question:

You and the other same-sex participant in the study will be asked for your thoughts and feelings about a decision that has to be made. One of you will be randomly assigned to make that decision, but it will have consequences for each of you. The two of you will not see each other or interact at all during this session, and you will be dismissed separately, so you will not even meet afterward.

If you have no questions at this point and agree to participate, please carefully read, then date and sign the informed consent forms in front of you. When you are finished, please go ahead and crack the door.

3. Lottery consent form

**Lottery consent for tax-exempt organizations**

The University of Tennessee, being a public institution, is a tax-exempt non-profit organization. Such organizations can sponsor lotteries and raffles so long as (a) there is a winner, (b) participants have been informed that there will be a winner, and (c) winning participants are notified.

I hereby consent to participate in the raffle for a $100 Target gift card, should my ID number be entered. I have been informed that there will be a winner who will be contacted in case of a win.

Print name: _____Date: _____

Signature: _____

Form T-43 (Rev 3-01)

*MediaLab Screens Studies 2-4*
*(each paragraph corresponds to a screen)*

**Study 2**

Please carefully read the information on each of the next few screens. It is crucial that you only click "continue" once you have read and understood all the information on a screen. When you are ready to begin, please click "continue".

In this particular session, we are studying how two people think and feel about decisions regarding the allocation of resources. The two of you won't interact; only one of you will make the allocation decision.

Specifically, the decision to be made in this session is this: There are 10 tickets to be entered into a raffle for a $100 Target gift card. There will be several drawings throughout the semester and only participants of this study are eligible to enter, so the chances of winning a card are fairly high.

In a few moments, the computer will randomly assign one of you to decide how to divide the 10 tickets between the two of you. The other person (i.e., the one not making the decision) will be given a different task.

At several points during the study you will both be asked to complete questionnaires concerning your thoughts and feelings. In addition, we would like to talk with you about your reactions to the study. To ensure anonymity and confidentiality of your responses, your identity cannot be linked to your ID number and you will never meet, or interact with, the other participant during or after the study (we ensure that you are dismissed separately).

The computer will now assign each of you to tasks. Please allow a few seconds for this process.

You have been chosen to be the DECIDER in this session, which means you will be the one who makes the decision of how the 10 raffle tickets are divided between the two of you. For more detailed instructions, please click "continue" on the bottom right of the screen.

As the decider, you can give yourself anywhere from 0 to 10 tickets; the other person gets any tickets you do not give to yourself. So, THE TICKETS ALLOCATED TO YOURSELF AND THE OTHER MUST ADD TO 10! Remember that the tickets will enter an actual raffle for a $100 gift card, and a winner will be drawn several times during the semester. Of course, the more tickets a participant has, the better his or her chances of winning the raffle. Thus, your allocation decision has a direct influence on the outcome for yourself and the other participant.

Specifically, the more tickets you keep, the higher your chance of winning the gift card. The more tickets you give to the other participant, the higher the other participant's chance of winning the gift card.

Any tickets entered for "self" will automatically enter the raffle under your participant ID number, any tickets entered for "other" will automatically be entered into the raffle under the other participant's ID number. To ensure that conditions match those of real-world big group settings, the other participant has been assigned a task that is unrelated to you and does not affect you. You will not meet or interact with the other participant at any point during the study and the exact nature of your decision will only be known to you.

Your allocation will remain entirely anonymous and confidential. Your decision will only be known to you. The other person only knows that there is a possibility his or her ID may be entered into a raffle for a $100 Target gift card. Neither the other person nor the experimenter will know how many tickets you entered for yourself and the other person. The exact nature of your decision will therefore only be known to you!

Before you make the decision, please answer a few questions about the task on the next few pages. When you're ready, please click "continue".

Below, please enter your email address. Should your ID number be entered into the raffle and drawn as a winner, you will automatically be notified of your win by email. All information entered today is confidential and cannot be accessed by the research assistant running today's session (this is to ensure that he or she cannot learn of your decision). Should you be drawn as a winner, you will be sent an electronic receipt that you need to fill out and bring back to the department to claim your gift card.

Please take a moment and imagine that you have won the $100 gift card to Target. Think about what you may buy with this card. In the space below, please write down the things you would like to spend it on.

How much would you like to win the gift card?

How difficult do you consider the allocation decision to be?

At this time, do you have any other thoughts about the allocation decision? If not, please type "none".

Ok, now you will make the allocation. Indicate below how many tickets you want to give to yourself and the other participant. Remember, you can give anywhere from 0-10 tickets to yourself - any tickets not allocated to yourself will go to the other person.

Now, please answer a few questions about your experience as the decision maker.

How fair was the way you divided the 10 raffle tickets between yourself and the other participant?

How moral was the way you divided the 10 raffle tickets between yourself and the other participant?

How much did you like being the decider?

The next few screens will list a number of adjectives that describe different feelings and emotions. Indicate by clicking a number the degree to which you are feeling each of these emotional reactions RIGHT NOW, that is, AT THIS VERY MOMENT.

Right now, I feel pleased

Right now, I feel happy

Right now, I feel justified

Right now, I feel sad

Right now, I feel surprised

Right now, I feel upset

Right now, I feel mistreated

Right now, I feel guilty

Right now, I feel satisfied

Right now, I feel annoyed

Right now, I feel irritated

Right now, I feel disappointed

Right now, I feel hurt

Right now, I feel ashamed

Right now, I feel lucky

Right now, I feel proud

Right now, I feel mad

Right now, I feel sorry

Right now, I feel offended

Right now, I feel angry

In the space below, please tell us what thoughts went through you mind between (a) when you learned that you would allocate the tickets and (b) when you indicated your allocation decision on the computer. Which of these thoughts affected your decision? How?

Finally, some questions about you...

Are you ...

How old are you?

What is your current or intended major here at UT?

Where on the following political spectrum do you place yourself?

Have you ever participated in a study in these lab quarters (AP 405) before?

Thank you! You have completed this portion of the study. Please slightly crack the door to give the experimenter a signal.

**Study 3**

Condition 1 (pre-behavioral judgment of behavior before role was known)

Please carefully read the information on each of the next few screens. It is crucial that you only click "continue" once you have read and understood all the information on a screen. When you are ready to begin, please click "continue".

In this particular session, we are studying how two people think and feel about decisions regarding the allocation of resources. The two of you won't interact; only one of you will make the allocation decision.

Specifically, the decision to be made in this session is this: There are 10 tickets to be entered into a raffle for a $100 Target gift card. There will be several drawings throughout the semester and only participants of this study are eligible to enter, so the chances of winning a card are fairly high.

In a few moments, the computer will randomly assign one of you to decide how to divide the 10 tickets between the two of you. The other person (i.e., the one not making the decision) will be given a different task.

At several points during the study you will both be asked to complete questionnaires concerning your thoughts and feelings. In addition, we would like to talk with you about your reactions to the study. To ensure anonymity and confidentiality of your responses, your identity cannot be linked to your ID number and you will never meet, or interact with, the other participant during or after the study (we ensure that you are dismissed separately).

First, please answer a few questions about the task on the next few pages. When you're ready, please click "continue".

Below, please enter your email address. Should your ID number be entered into the raffle and drawn as a winner, you will automatically be notified of your win by email. All information entered today is confidential and cannot be accessed by the research assistant running today's session (this is to ensure that he or she cannot learn of the decision maker's decision). Should you be drawn as a winner, you will be sent an electronic receipt that you need to fill out and bring back to the department to claim your gift card.

Please take a moment and imagine that you have won the $100 gift card to Target. Think about what you may buy with this card. In the space below, please write down the things you would like to spend it on.

How much would you like to win the gift card?

**What, in your view, is the fair way to allocate the tickets?**

**What, in your view, is the moral way to allocate the tickets?**

How difficult do you consider the allocation decision to be?

At this time, do you have any other thoughts about the allocation decision? If not, please type "none".

The computer will now assign each of you to tasks. Please allow a few seconds for this process.

You have been chosen to be the DECIDER in this session, which means you will be the one who makes the decision of how the 10 raffle tickets are divided between the two of you. For more detailed instructions, please click "continue" on the bottom right of the screen.

As the decider, you can give yourself anywhere from 0 to 10 tickets; the other person gets any tickets you do not give to yourself. So, THE TICKETS ALLOCATED TO YOURSELF AND THE OTHER MUST ADD TO 10! Remember that the tickets will enter an actual raffle for a $100 gift card, and a winner will be drawn several times during the semester. Of course, the more tickets a participant has, the better his or her chances of winning the raffle. Thus, your allocation decision has a direct influence on the outcome for yourself and the other participant.

Specifically, the more tickets you keep, the higher your chance of winning the gift card. The more tickets you give to the other participant, the higher the other participant's chance of winning the gift card.

Any tickets entered for "self" will automatically enter the raffle under your participant ID number, any tickets entered for "other" will automatically be entered into the raffle under the other participant's ID number. To ensure that conditions match those of real-world big group

settings, the other participant has been assigned a task that is unrelated to you and does not affect you. You will not meet or interact with the other participant at any point during the study and the exact nature of your decision will only be known to you.

Your allocation will remain entirely anonymous and confidential. Your decision will only be known to you. The other person only knows that there is a possibility his or her ID may be entered into a raffle for a $100 Target gift card. Neither the other person nor the experimenter will know how many tickets you entered for yourself and the other person. The exact nature of your decision will therefore only be known to you!

Ok, now you will make the allocation. Indicate below how many tickets you want to give to yourself and the other participant. Remember, you can give anywhere from 0-10 tickets to yourself - any tickets not allocated to yourself will go to the other person.

*Post-behavioral questionnaire was identical to that in Study 2*

<u>Condition 2 (pre-behavioral judgments were made after role was known) and Condition 3 (identical to condition 2, but no pre-behavioral judgments)</u>

Please carefully read the information on each of the next few screens. It is crucial that you only click "continue" once you have read and understood all the information on a screen. When you are ready to begin, please click "continue".

In this particular session, we are studying how two people think and feel about decisions regarding the allocation of resources. The two of you won't interact; only one of you will make the allocation decision.

Specifically, the decision to be made in this session is this: There are 10 tickets to be entered into a raffle for a $100 Target gift card. There will be several drawings throughout the semester and only participants of this study are eligible to enter, so the chances of winning a card are fairly high.

In a few moments, the computer will randomly assign one of you to decide how to divide the 10 tickets between the two of you. The other person (i.e., the one not making the decision) will be given a different task.

At several points during the study you will both be asked to complete questionnaires concerning your thoughts and feelings. In addition, we would like to talk with you about your reactions to the study. To ensure anonymity and confidentiality of your responses, your identity cannot be linked to your ID number and you will never meet, or interact with, the other participant during or after the study (we ensure that you are dismissed separately).

The computer will now assign each of you to tasks. Please allow a few seconds for this process.

You have been chosen to be the DECIDER in this session, which means you will be the one who makes the decision of how the 10 raffle tickets are divided between the two of you. For more detailed instructions, please click "continue" on the bottom right of the screen.

As the decider, you can give yourself anywhere from 0 to 10 tickets; the other person gets any tickets you do not give to yourself. So, THE TICKETS ALLOCATED TO YOURSELF AND THE OTHER MUST ADD TO 10! Remember that the tickets will enter an actual raffle for a $100 gift card, and a winner will be drawn several times during the semester. Of course, the more tickets a participant has, the better his or her chances of winning the raffle. Thus, your allocation decision has a direct influence on the outcome for yourself and the other participant.

Specifically, the more tickets you keep, the higher your chance of winning the gift card. The more tickets you give to the other participant, the higher the other participant's chance of winning the gift card.

Any tickets entered for "self" will automatically enter the raffle under your participant ID number, any tickets entered for "other" will automatically be entered into the raffle under the other participant's ID number. To ensure that conditions match those of real-world big group settings, the other participant has been assigned a task that is unrelated to you and does not affect you. You will not meet or interact with the other participant at any point during the study and the exact nature of your decision will only be known to you.

Your allocation will remain entirely anonymous and confidential. Your decision will only be known to you. The other person only knows that there is a possibility his or her ID may be entered into a raffle for a $100 Target gift card. Neither the other person nor the experimenter will know how many tickets you entered for yourself and the other person. The exact nature of your decision will therefore only be known to you!

Before you make the decision, please answer a few questions about the task on the next few pages. When you're ready, please click "continue".

Below, please enter your email address. Should your ID number be entered into the raffle and drawn as a winner, you will automatically be notified of your win by email. All information entered today is confidential and cannot be accessed by the research assistant running today's session (this is to ensure that he or she cannot learn of your decision). Should you be drawn as a winner, you will be sent an electronic receipt that you need to fill out and bring back to the department to claim your gift card.

Please take a moment and imagine that you have won the $100 gift card to Target. Think about what you may buy with this card. In the space below, please write down the things you would like to spend it on.

How much would you like to win the gift card?

**What, in your view, is the fair way to allocate the tickets?**

**What, in your view, is the moral way to allocate the tickets?**

How difficult do you consider the allocation decision to be?

At this time, do you have any other thoughts about the allocation decision? If not, please type "none".

Ok, now you will make the allocation. Indicate below how many tickets you want to give to yourself and the other participant. Remember, you can give anywhere from 0-10 tickets to yourself - any tickets not allocated to yourself will go to the other person.

*The post-allocation questionnaire was identical to that administered in Study 2.*

**Study 4, conditions 1-4 (time and salience manipulations are highlighted in bold)**

Please carefully read the information on each of the next few screens. It is crucial that you only click "continue" once you have read and understood all the information on a screen.

In this particular session, we are studying how two people think and feel about decisions regarding the allocation of resources. The two of you won't interact; only one of you will make the allocation decision.

Specifically, the decision to be made in this session is this:
There are 10 tickets to be entered into a raffle for a $100 Target gift card. There will be several drawings throughout the semester and only participants of this study are eligible to enter, so the chances of winning a card are fairly high.

Below, please enter your email address. Should your ID number be entered into the raffle and drawn as a winner, you will automatically be notified of your win by email. All information entered today is confidential and cannot be accessed by the research assistant running today's session. Should you be drawn as a winner, you will be sent an electronic receipt that you need to fill out and bring back to the department to claim your gift card.

Please take a moment and imagine that you have won the $100 gift card to Target. Think about what you may buy with this card. *(participant writes what he/she would like to purchase)*

How much would you like to win the gift card? *(1=not at all; 7=extremely)*

In a few moments, the computer will randomly assign one of you to decide how to divide the 10 tickets between the two of you. The other person (i.e., the one not making the decision) will be given a different task.

At several points during the study you will both be asked to complete questionnaires concerning your thoughts and feelings. In addition, we would like to talk with you about your reactions to the study.

The computer will now assign each of you to tasks. Please allow a few seconds for this process. *(computer takes a few seconds to "work" on random assignment)*

You have been chosen to be the DECIDER in this session, which means you will be the one who makes the decision of how the 10 raffle tickets are divided between the two of you.

As the decider, you can give yourself anywhere from 0 to 10 tickets; the other person gets any tickets you do not give to yourself. So, THE TICKETS ALLOCATED TO YOURSELF AND THE OTHER MUST ADD TO 10!

Specifically, the more tickets you keep, the higher your chance of winning the gift card. The more tickets you give to the other participant, the higher the other participant's chance of winning the gift card.

Any tickets entered for "self" will automatically enter the raffle under your participant ID number, any tickets entered for "other" will automatically be entered into the raffle under the other participant's ID number.

Your allocation will remain entirely anonymous and confidential. Your decision will only be known to you.

***[salience manipulation]***
**Most people think the fairest way to divide the tickets is evenly, giving 5 tickets to oneself and 5 to the other person, but how you choose to divide them is entirely up to you.**

The other person only knows that there is a possibility his or her ID may be entered into a raffle for a $100 Target gift card, but they will not know whether it actually is entered or with how many chances.

Neither the other person nor the experimenter will know how many tickets you allocated to yourself and how many to the other person. How you allocate the tickets will be known only to you.

***[time manipulation]***
**Some of our participants have found it helpful to be given some time to think about their decision. You may take the next few moments to do so. The screen will automatically prompt you to make your allocation in a few minutes. However, after you see the prompt, you may take as much time as you need before entering your allocation.**

So, take a few moments to think about how you want to divide the tickets, then, after you see the prompt and are ready, enter your allocation. *(participant is given 2 minutes time, then the screen changes to ...)*

Ok, now you will make the allocation. Indicate below how many tickets you want to give to yourself and the other participant. Remember, you can give anywhere from 0-10 tickets to yourself - any tickets not allocated to yourself will go to the other person.

Now, please answer a few questions about your experience as the decision maker.

How fair was the way you divided the 10 raffle tickets between yourself and the other participant?

How moral was the way you divided the 10 raffle tickets between yourself and the other participant?

The next few screens will ist a number of adjectives that describe different feelings and emotions. Indicate by clicking a number the degree to which you are feeling each of these emotional reactions RIGHT NOW, that is, AT THIS VERY MOMENT. ... *(our list of adjectives follows)*

How difficult do you consider the allocation decision to be?

In the space below, please tell us what thoughts went through you mind between (a) when you learned that you would allocate the tickets and (b) when you indicated your allocation decision on the computer. Which of these thoughts affected your decision? How?

At this time, do you have any other thoughts about the allocation decision? If not, please type "none".

Finally, some questions about you...

What is your current or intended major here at UT?

How old are you?

Are you ...*(male or female)*

Where on the following political spectrum do you place yourself?

Have you ever participated in a study in these lab quarters (AP 405) before?

Thank you! You have completed this portion of the study. Plesae slightly crack the door to give the experimenter a signal.

## Vita

Lydia Eckstein Jackson was born in Berlin, Germany in 1979. She graduated with her *Diplom* (equivalent to Master's degree) in Psychology from Humboldt University in Berlin, Germany in 2005. She subsequently moved to Nashville, TN, and taught high school for two years before returning to graduate school for her doctoral work in 2007. She will be an Assistant Professor of Psychology at Allegheny College in Meadville, PA starting in August 2012.