THE UNIVERSITY OF CHICAGO

THE INTUITIVE JURISPRUDENCE OF PUNISHMENT

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE UNIVERSITY OF CHICAGO

BOOTH SCHOOL OF BUSINESS

AND

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

DEPARTMENT OF PSYCHOLOGY

IN CANDIDACY OF THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

JESSICA LYNN BREGANT

CHICAGO, ILLINOIS

AUGUST 2018

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES xii

ACKNOWLEDGMENTS xiii

ABSTRACT xiv

Introduction 1

Intuitive jurisprudence 2

1 Anticipating Punishment and Understanding the "Spirit of the Law" 6

Study 1.1: Children are More Lenient towards Rule-Breakers when the Spirit of the Rule is Intact 11

Method 13

Results 15

Discussion 19

Study 1.2: Children's Lenience does not Depend on Differing Assumptions about Outcome. 20

Method 21

Results 22

Discussion 24

Study 1.3: Children Distinguish between Identical Actions When the Spirit of the Rule Differs 25

Method 26

Results 27

Discussion 29

Study 1.4: Children's Lenience does not Depend on the Rule-Breaker's Cautiousness 30

Method 31

Results 32

Discussion 34

General Discussion 36

Study 1.5: Adult Evaluations Distinguish Between the Letter and Spirit of the Law 40

Method 41

Results 43

Discussion 47

2 Interpreting Punishment's Expressive and Informational Value 51

Introduction 52

The Expressive Function of Punishment 53

Normative Messages, Laws, and Punishment 56

Moral Psychology and Punishment 58

Study 2.1: Punishment, Harm, and Moral Wrongness 63

Methods 64

Results 65

Discussion 68

Study 2.2: Harm and Disgust 69

Method 71

Results 72

Discussion 74

Study 2.3: Inferences of Harm in the Real World 75

Methods 76

Results 77

Discussion 77

Study 2.4: Inferences about Punishment in the Real World 78

Methods 79

Results 80

Discussion 82

General Discussion 82

Study 2.5(a): Children's Inferences about Punishment and Deterrence 90

Methods 92

Results 93

Discussion 94

Study 2.5(b): Children's Inferences about Punishment Policy and Action Frequency 95

Method 95

Results 96

Discussion 96

3 Avoiding Punishment by Asking for Permission 98

Study 3.1: Permission and Taking 101

Methods 101

Results 103

Discussion 111

Study 3.2: Permission and Taking Decisions 111

Methods 111

Results 113

Discussion 119

Study 3.3: The Permission Game 119

Method 122

Results 123

Discussion 128

Study 3.4: Why Grant Permission? A Comparison to Other Paradigms 129

Method 131

Results 133

Discussion 139

General Discussion 139

4 Conclusion: The Future of Intuitive Jurisprudence 141

Intent, Moral Judgment, and Blame 141

Fairness and Procedural Justice 145

Evidentiary Issues and Trust in Testimony 148

Social Biases and Equal Protection Under the Law 151

Conclusion 154

REFERENCES 157

LIST OF FIGURES

Figure 1.1. Example images from stimuli for Study 1.1. (a) The teacher announces the rule; (b) the potential consequences of violating the rule are illustrated (i.e., the books could get dirty) (c) the protagonist remembers his sick friend. 14

Figure 1.2. Mean responses for whether the target action violated the rule (Rule Broken?) and mean disapproval index (Disapproval) for the action in Studies 1.1, 1.2, and 1.4. Error bars represent standard error of the mean. 16

Figure 1.3. Scatterplots of disapproval index scores for each participant in Study 1.1, arranged by age, with Pearson correlations (r) shown. In the Spirit Intact condition, the parenthetical indicates the correlation when the outlier is excluded. * p < .05, **p < .01, ***p < .001. 17

Figure 1.4. Average disapproval among only children who responded that the rule was "definitely" broken in Studies 1.1, 1.2, and 1.4, separated by condition. Error bars represent standard error of the mean. 18

Figure 1.5. Scatterplots of disapproval index scores for each participant in Study 1.2, arranged by age, with Pearson correlations (r) shown. Parentheticals indicate the correlation when the outliers are excluded. *p < .05, **p < .01, ***p < .001 24

Figure 1.6. Scatterplots of disapproval index scores for each participant in Study 1.3, arranged by age, with Pearson correlations (r) shown. *p < .05, **p < .01, ***p < .001 29

Figure 1.7. Scatterplots of disapproval index scores for each participant in Study 1.4, arranged by condition and age, with Pearson correlations (r) shown. *p< .05, **p < .01, ***p < .001. 33

Figure 1.8. Illustration of parking space marker, provided to participants in Study 1.5. 42

Figure 1.9. Mean ratings of agreement in each condition for whether the ordinance was violated (left) and whether punishment should be imposed (right). Error bars represent standard error of the mean. 44

Figure 1.10. Mean responses by condition for the amount of the fine that should be paid (top left), intentionality of the violation (top right), harmfulness of the violation (bottom left) and fairness of the ordinance (bottom right). Error bars represent standard error of the mean..... 47

Figure 2.1. Ratings of harmfulness and moral wrongness, conflicting information condition. Error bars represent standard error of the mean..... 66

Figure 2.2. Ratings of harmfulness and moral wrongness for Study 2.2, conflicting information condition. Error bars represent standard error of the mean. 73

Figure 2.3. Difference between punished and non-punished averages for (a) harmfulness ratings and (b) moral wrongness. Error bars represent standard error of the mean..... 80

Figure 2.4. World choice by age year (reproduced from Bregant et al., 2016). 91

Figure 2.5. Proportion of children at various ages who chose each world when asked which has more bad things (left) and which they would prefer to live in (right). 94

Figure 2.6. Proportion of children choosing each world across various age groups (ages separated for ease of visualization)..... 96

Figure 3.1. Diagram of the Permission Game provided to participants. NOTE: In this diagram, Player 1 is the Permitter, and Player 2 is the Taker. 101

Figure 3.2. Taker's view of the outcome in Study 3.1..... 106

Figure 3.3. Takers' views of the Permitters in Study 3.1..... 106

Figure 3.4. Takers' desire to punish the permitter in Study 3.1..... 107

Figure 3.5. Permitters' reactions to outcome in Study 3.1. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from. 108

Figure 3.6. Permitters' evaluations of takers. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from. 109

Figure 3.7. Permitters' desire to punish takers. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from. 110

Figure 3.8. Taker decisions in Study 3.3 for the punishment condition (left) and no punishment condition (right). 123

Figure 3.9. Takers' reactions across condition and permission outcome. Error bars represent standard error of the mean. 125

Figure 3.10. Taker ratings of the permitter in Study 3.3. Error bars represent standard error of the mean. 126

Figure 3.11. Permitter ratings of satisfaction and unhappiness in Study 3.3. Error bars represent standard error of the mean. 127

Figure 3.12. Permitter evaluations of the taker in Study 3.3. Error bars represent standard error of the mean. 128

Figure 3.13. Participant choices across the conditions of Study 3.4. 134

Figure 3.14. Feelings of control and generosity across the conditions of Study 3.4. Error bars represent standard error of the mean. 136

Figure 3.15. Additional participant evaluations from Study 3.4. Error bars represent standard error of the mean. 138

LIST OF TABLES

Table 3.1. Taker decisions in Study 3.1. 104

Table 3.2. Distribution of Permitters by outcome in Study 3.1. 105

Table 3.3. Distribution of permitters by outcome in Study 3.2. 113

Table 3.4. Taker decisions in Study 3.2. 114

ACKNOWLEDGMENTS

This dissertation, and the work it represents, would not have been possible without the help, mentorship, and support of many people. I have been fortunate to work with incredible and inspiring people throughout my academic career, and I am grateful to all of them. In particular, I need to thank my advisors: Eugene Caruso, for encouraging me to think deeply and methodically about my research; Alex Shaw, for challenging and inspiring me to be precise and productive; and Katherine Kinzler, without whom I never would have considered working with children, and whose unwavering support continues to push me in exciting new directions.

This work also benefitted immensely from the community of scholars at the University of Chicago. I am particularly grateful to George Wu, Jane Risen, Heather Caruso, Amanda Woodward, and the members of the PBJ Lab, DIBS Lab, and DSC Lab for their thoughtful feedback on various incarnations of this research. I also want to thank the many lab managers and research assistants who collected data for these projects, especially Bella Wellberry, Emily Gerdin, and Shun Wang, who collaborated with me on various parts of this work. Finally, thank you to the Museum of Science and Industry in Chicago and the Center for Decision Research.

I also could not have completed this work without the love and support of my friends and family, and I owe them so much more than thanks. To my mom, Jo-An, and my grandparents, Naomi and Bob, who have always given me their unconditional support: thank you. To my sister, Hilary, who has counseled me and consoled me, and who inspires me every day: thank you. To all the friends who have celebrated with me and commiserated with me, played games and talked late into the night, and just generally been there: thank you. And to my husband, Bob, whose love and support got me through the hardest parts of this process, and without whom I could not have succeeded: thank you. I love you all.

ABSTRACT

Although dialogue between psychology research and legal theory has been productive for both fields, significant gaps in communication between the fields remain. In this dissertation, I highlight an emerging direction for law and psychology research—an "intuitive jurisprudence" approach. Applying insights and methods from developmental psychology, intuitive jurisprudence research addresses to questions of *adult* legal psychology. Research at the intersection of developmental psychology and law can guide psychological studies of law and, reciprocally, inform our understanding of human thinking and cognitive development. By drawing clear lines from developmental science to law (and back again), research in both fields can benefit.

Throughout this dissertation, I argue that research in law and psychology can benefit from a broad methodological and theoretical base. In particular, I show the promise of an intuitive jurisprudence approach that borrows from developmental science to craft interesting questions and clear experiments, and to aid in our understanding of how lay people interact with the law. In Chapter 1, I show that children and adults consider the spirit of the law when evaluating rule breakers. In Chapter 2, I show that punishment has a powerful ability to communicate information about the state of the world; in adults, the message seems to be particularly tied to inferences of harm, but children also show many signs of complex reasoning about punishment's place in society. Finally, in Chapter 3, I present complementary research on the role of seeking permission to avoid punishment.

Introduction

Few institutions have as much influence on daily life as the law. For those involved in the criminal justice system, the importance of law is clear and direct, but regulations and legal policies shape lives in many indirect ways as well. Virtually all major milestones in life, including adulthood, marriage, childbirth, and death, are accompanied by legal recognition and responsibilities. On a day-to-day basis, laws shape many private and commercial relationships, including employment. The structure of the workday, working conditions, and wages are generally defined or protected by laws.

Given the importance of law to daily experience, it is unsurprising that psychological researchers across disciplinary subfields have sought to understand how humans interact with the law. Since the psychological study of law began in earnest in the 1960's, the field has transformed the way scholars and jurists think about the courtroom, including research on the nature of jury decision making (Diamond et al., 2003; Hans & Jehle, 2003; Hans & Reyna, 2011; Pennington & Hastie, 1992), the (un)reliability of eyewitness memory (Loftus, 1974; Wells et al., 1998; Wells & Olson, 2003), and community attitudes toward the death penalty (Dillehay, 1996; Ellsworth & Gross, 1994; Hans et al., 2015). More broadly, psychological research has also provided important insights into how legal structures gain and maintain public support and legitimacy (Sevier, 2014; Tyler, 2006; Tyler & Rasinski, 1991; Tyler & Sevier, 2013).

The relationship between law and psychology is not a one-way street; legal processes and theories have contributed to psychology in return. In particular, law has provided a context and comparative framework for "folk psychology" to explore the way people think about a number of weighty psychological topics. These frameworks have been employed in varied areas of psychology, including moral judgment (Cushman, Knobe, & Sinnott-Armstrong, 2008;

Goldberg, Lerner, & Tetlock, 1999), causal reasoning (Hastie et al., 2015; Malle & Knobe, 1997), blame processes (Cushman, 2008; Pimentel, Arndorfer, & Malloy, 2015), and the quantification of harm (Hans & Reyna, 2011; Kelly, Stich, Haley, Eng, & Fessler, 2007).

Although dialogue between psychology research and legal theory has been productive for both fields, significant gaps in communication between the fields remain. In this dissertation, I highlight an emerging direction for law and psychology research—an "intuitive jurisprudence" approach. Applying insights and methods from developmental psychology, intuitive jurisprudence research addresses to questions of *adult* legal psychology. Research at the intersection of developmental psychology and law can guide psychological studies of law and, reciprocally, inform our understanding of human thinking and cognitive development. By drawing clear lines from developmental science to law (and back again), research in both fields can benefit.

Intuitive jurisprudence

Developmental psychologists have historically made great contributions to the study of children and the law, but this past research is in the service of different conceptual goals from what I advance here. Namely, past research has primarily focused on the interests of children as actors in and recipients of legal processes (see, e.g., Brainerd & Reyna, 1996; Goodman, Batterman-Faunce, Schaaf, & Kenney, 2002; Grisso et al., 2003; London, Bruck, Ceci, & Shuman, 2005; Perry et al., 1995; Quas, Qin, Schaaf, & Goodman, 1997; Reyna & Lloyd, 1997; Woolard, Reppucci, & Redding, 1996). Similarly, developmental psychology has long led the study of morality, which is itself highly relevant to the law (see, e.g., Bloom, 2010; Hamlin, 2013b; Hamlin, Mahajan, Liberman, & Wynn, 2013; Killen, Lynn Mulvey, Richardson, Jampol, & Woodward, 2011; Kohlberg & Kramer, 1969; McAuliffe, Jordan, & Warneken, 2015; Olson &

Spelke, 2008; Piaget, 1932/1997; Shweder, Turiel, & Much, 1981; Smetana, 1985; Smetana et al., 2012; Thompson & Newton, 2013; Tisak & Turiel, 1984; Warneken, 2015; Warneken & Tomasello, 2008), but that work is not often explicitly linked to legal reasoning or legal rules (but see, e.g., Friedman & Neary, 2009; Levine & Leslie, 2015; Nobes, Panagiotaki, & Pawson, 2009; Van de Vondervoort & Friedman, 2015). Intuitive jurisprudence is complementary to this past work, but it is meaningfully different. Rather than focusing solely on the treatment of children or on their social and moral development, intuitive jurisprudence demonstrates that research with children and using insights from developmental science may also be used to understand the intuitive psychology of adult legal actors.

Taking the synthesis of developmental science and law one step further, intuitive jurisprudence broadens the aims of research with children in the law, and there are many ways in which this breadth can be valuable. First, an intuitive jurisprudence approach can contribute directly to psychological research. Much as research in moral development has already done, intuitive jurisprudence can shed light on basic human intuitions and reasoning. The law provides both an applied domain and a general model for the study of moral and social judgments. In some ways, the advantage of using developmental methods to the study of basic moral and social reasoning is obvious; after all, children grow into adults, and as Piaget noted, "[i]n a sense, child morality throws light on adult morality. If we want to form men and women, nothing will fit us so well for the task as to study the laws that govern their formation" (1932/1997, p. 12). Beyond that, however, developmental research can address critical questions about how our early intuitions about legal and moral matters emerge, what specific cognitive capacities are required to understand and endorse legal principles, and how maturity and experience may change these

ideas over time. Comparing children across development can also teach us more about the deepest psychological influences on beliefs and behavior.

A second benefit of intuitive jurisprudence research builds on the tradition in law and psychology of speaking to the application of psychology to the legal system. When reactions to legal processes and outcomes are studied from an intuitive jurisprudence perspective, researchers can ask deeper questions about lay legal intuitions and how they change across the lifespan. In a broader sense, children may even serve as models for the way a society comes to understand and agree on social and moral rules; their understanding evolves over time as they develop more nuanced and comprehensive ideas.

Third, intuitive jurisprudence research adds directly to the study of the law by providing new tests of the psychological assumptions that underlie many broad legal arguments. Intuitive jurisprudence can also help policymakers identify and understand ways in which the law reflects or fails to reflect psychological reality. Here, I want to emphasize that I am not making a normative claim; I am not suggesting that the law should, in all cases, reflect lay intuitive psychology. However, where law and intuition diverge, understanding and acknowledging the divide can help policymakers find ways to increase the perceived legitimacy of the legal system in the eyes of "intuitive jurists" – i.e., lay people affected by the law.

An intuitive jurisprudence approach has the potential to contribute to many substantive areas of law and psychology. In this dissertation, I focus on three such areas, all related to the study and understanding of punishment. In Chapter 1, I present research on how children and adults reason about the "spirit of the law" when making punishment evaluations. In Chapter 2, I turn to intuitive understandings of punishment and its role in the world, borrowing from developmental studies to design a clean and simple test for adult participants. I then present

complementary work with children that explores the way ideas about deterrence develop. Finally, in Chapter 3, I present a new experimental economic game, designed to investigate how and why people ask for permission, including two studies that measure permission-seeking under the threat of punishment. The work in this chapter illustrates the potential synergy between intuitive jurisprudence research and other, more traditional work in social psychology and behavioral economics. Finally, in the conclusion, I review some other areas of research that have the strong potential to benefit from intuitive jurisprudence.

1 Anticipating Punishment and Understanding the "Spirit of the Law"

To effectively navigate the social world, with its complex and sometimes contradictory norms and laws, people must often discern when different rules apply and when they do not. One way that adults approach this task is by distinguishing between what a rule or law explicitly states (the "letter of the law") and why a law was created or what end it serves (the "legislative intent" or "spirit of the law") (Garcia, Chen, & Gordon, 2014). For example, consider a city ordinance that forbids the use of any motorized vehicles in a public park. Clearly, a person who drives a motorcycle or a golf cart in the park would be in violation of this ordinance, but what about someone with a motorized remote-controlled toy car? While most people would concede that the toy is technically in violation of a strict textual reading of the ordinance, they would be unlikely to think that the person with the toy ought to be punished. To explain this discrepancy, most adults will draw on their intuitions about the goals of the law (e.g., concerns about pedestrian safety within the park, noise levels, exhaust fumes, etc.), and they will conclude that playing with a toy electric car should be exempt because it does not violate these goals. As the above example illustrates, being a competent norm-follower necessarily entails an understanding of what legal scholars call "legislative intent" (Garcia et al., 2014). The example also demonstrates the importance of this reasoning when evaluating the actions of others—and especially when judging the blameworthiness of others' behaviors. The "spirit of the law" is therefore a key component of normative reasoning. This raises the question: how would a child respond to this case?

Here we investigate when and how humans develop this important ability. A child, viewing the same motorized vehicle ordinance, might adopt a formalistic and highly literal view of the rule, believing that any vehicle with a motor is prohibited; examples of children who

rigidly adhere to such literal readings abound.(see, e.g., "I used to believe: Drinking and driving applies to all drinks," n.d. (on kids who think that "drinking and driving" means any drink)). If it is true that reasoning about the "spirit of the law" is a key component to making judgments about norms and rule-breakers, and if we want to understand humans' earlier normative reasoning, then it seems important to understand how and when children develop such a capacity. Although there has been substantial work on children's understanding of norms and rules (e.g., Kalish, 1998; Kalish & Sabbagh, 2007; Nucci, 1981; Rakoczy, Warneken, & Tomasello, 2008, 2009; Turiel, 2008; Vaish, Missana, & Tomasello, 2011; Van de Vondervoort & Friedman, 2015), we are aware of no work that has explored whether children distinguish between violations of the letter of a rule and violations of the spirit of a rule. Here, we explore when children begin to share the adult intuition that it can be acceptable to *technically* break the letter of a law, if one's actions do not violate its spirit. Before presenting these studies we first review existing literature on children's understanding of norms and intentions.

Past research has demonstrated that infants and young children appear to make normative judgments about others. For example, infants prefer helpers to hinderers (Hamlin, Wynn, & Bloom, 2007), 2-year-olds think that hinderers should be punished (Hamlin, Wynn, Bloom, & Mahajan, 2011), and 3-year-olds protest when they see someone harming a third party (Vaish et al., 2011). These early objections to those who harm or hinder others hint at a nascent sense of morality, and while such moral judgment does not require an understanding of normative rules, related work has shown that children have a robust and flexible understanding of the nature and application of rules (Kalish, 1998; Nucci, 1981; Nucci & Turiel, 1978; Riggs & Kalish, 2016; Schmidt & Tomasello, 2012; Smetana, 1981a; Smetana, Killen, & Turiel, 1991). Indeed, in a

simple game-like interaction, 2- and 3-year-old children raise normative objections to the violation of novel rules (Rakoczy, Warneken, & Tomasello, 2008).

Contrary to Piaget's (1932/1997) view that children's early understanding of rules is rigid and immutable, however, children make nuanced determinations about the importance of following certain rules. For example, by the time they are in preschool (i.e., 3- to 5-years old), children already treat violations of moral rules and conventional rules quite differently (Nucci & Turiel, 1978). They believe that moral violations (e.g., hitting someone) should be punished more severely than merely conventional violations (e.g., wearing pajamas to school) (Smetana, 1981a). Like adults, children also do not believe that moral and conventional rules are equally authority-independent; that is, while children report that a teacher or parent could make a conventional violation acceptable, they are more reluctant to say that the same authority figure could make a moral violation acceptable (Richardson, Mulvey, & Killen, 2012; Smetana et al., 1991; Turiel, 1978). Even when a moral rule is in place, children between the ages of 4 and 8 years increasingly think it is acceptable to violate the rule if doing so would have positive consequences (Heyman, Sweet, & Lee, 2009; Neary & Friedman, 2014; Riggs & Kalish, 2016). Further, they have intuitions about who can *change* rules, thinking that children can make up and change rules of a game when there is consensus among those playing a game (Zhao & Kurshnir, in press). Children understand that different rules and obligations apply to different people based on their group status (Rhodes & Chalik, 2013). Relatedly, children as young as 3 years old seem to believe that violating conventional rules of a game is only wrong for in-group members, even though they think it is unacceptable for *either* in-group or out-group members to violate moral rules (Schmidt, Rakoczy, & Tomasello, 2012). That is, young children appear to think that moral rules apply to everyone, but conventional rules apply only to in-group members. The above

research demonstrates that children have complex intuitions about norms and rules by the time they are 4 years old.

Building on this previous work, which demonstrates how different kinds of rules can provoke different kinds of reactions and evaluations from children in different contexts, the present work is concerned with children's evaluations of rule-breakers who break *identical* rules—that is, rule-breakers whose actions differ in the extent to which they violate the underlying concern the rule is meant to address (i.e. the spirit of the rule). Distinguishing between two apparent rule violations—one that contravenes the intended spirit of the rule and one which does not—is a complicated task that requires a relatively sophisticated understanding of the surrounding context. In particular, it requires an understanding not only of how rules work, but of what it means to violate the spirit of that rule. To believe that an actor who violates only the spirit of a rule should be treated with leniency, children have to understand that a rule has been broken and to simultaneously not think that the person should be punished for breaking the rule. Not because the rule is bad, but because the person's actions violate one feature of the law (its letter) without violating the real goal of the law (its spirit). From a social and cognitive standpoint, performing such reasoning is no simple exercise. It requires the individual to understand a potentially complex situation and its elements, using the rule-maker's goals to shape one's own condemnation. Given the inherent potential difficulty of this kind of reasoning, children may be expected to struggle with it, even as they strive to learn the rules and norms of their social contexts.

Although there has been no specific work on children's use of the intentions (i.e. spirit) behind a rule, there is extensive work on their use of intentions in making evaluations of rule breakers. In the domain of moral judgment, previous work has demonstrated that children

readily take transgressor intentions into account. Consistent with a number of normative theories of morality (for a review, see Cushman, 2015), children's moral judgments appear to be influenced by whether an action was intentional or unintentional (e.g., Darley & Shultz, 1990). Although children (and adults) are sometimes overly focused on outcomes as compared to intentions (Armsby, 1971; Cushman, Dreber, Wang, & Costa, 2009; Gino, Moore, & Bazerman, 2009; Piaget, 1932/1997; Young & Saxe, 2011), even 3- to 5-year-old children judge intentional acts more harshly than accidents (Armsby, 1971; Cushman, 2015; Cushman, Sheketoff, Wharton, & Carey, 2013; Darley & Shultz, 1990; Yuill & Perner, 1988). As children grow older, they place increasing weight on transgressor intentions when making evaluations of others, reporting that it is wrong to *try* (i.e., intend) to hurt someone, even if one does not succeed in doing so (Cushman et al., 2013).

While these previous results suggest that the intentions of a transgressor do play an important role in children's moral judgments, the present chapter explores children's understanding of an entirely different intention: the intentions and goals of the rule-maker(s). As we argue above, to make nuanced judgments of whether and to what extent a rule applies in a given situation, individuals must not only understand what the rule explicitly states (the letter of the rule), but they must also have some understanding of *why* the rule exists (the spirit of the rule). This question of rule-*maker's* intent is essentially orthogonal to the question of the rule-*breaker's* intent; a person may break a rule for any number of possible reasons that have little or nothing to do with the intended aim of the rule itself. To return to the motorized vehicle rule example that we introduced above, a person may break the rule by driving a car across the park for selfish reasons (e.g., to avoid waiting in heavy traffic) or prosocial reasons (e.g., to get an injured person to a hospital more quickly) that have nothing to do with the reasons that

motorized vehicles are prohibited in the first place (e.g., for the safety of park visitors, or to reduce noise levels around the park). So, while existing research shows that children reason about the intentions of rule-breakers, that does not end our inquiry. For children to become truly competent social agents they must develop some understanding of legislative intent, and it is important to understand when and how children develop this critical ability for making normative judgments.

The current project investigates if and when children begin to use legislative intent in their evaluations about rule-breakers. In three studies, children between the ages of 4 and 10 years were told about a boy who violated the *letter* of a rule. The rule in each case was that children may not take more than four books home from the library at one time; the focal boy took six books home at one time. Between conditions, we varied whether the boy violated only the letter of the rule (spirit intact condition) or whether his actions violated the spirit of the rule as well (spirit violated condition). We then asked children whether the boy in the story violated the rule and whether the boy's action should be condemned by others. If children distinguish between the letter of the law and spirit of the law, we predicted they would agree in both conditions that the boys *technically* broke the rule, but that they would be far more lenient in their moral judgments of the boy who did not violate the spirit of the rule.

Study 1.1: Children are More Lenient towards Rule-Breakers when the Spirit of the Rule is Intact

In all four studies in this chapter, participants heard about a boy named Timmy who violates only the letter of the rule, while not violating the spirit of the rule, or he violates the spirit of the rule along with the letter of the rule. In Study 1.1, the rule is identical across conditions, both in terms of the letter and the spirit of the rule, but the character's actions vary slightly so that he either complied with or violated the spirit of the rule. We then asked children

both whether the rule breaker violates the rule and if they should be punished and condemned for violating this rule.

If children incorporate the spirit of a rule when judging a rule breaker, then we should see a difference in how they evaluate the focal actor between conditions. In both conditions, the rule breaker has violated the literal text of the rule, but in one condition he has not violated the spirit of the rule. Thus, while children should report that the rule was broken in both cases, children may be significantly less likely to condemn the violation when the spirit of the rule remains intact.

However, there are at least two alternative possibilities. One is that children, especially young children, focus exclusively on whether the letter of the law was violated. If this is true, then we would expect to see high condemnation in both conditions since the rule was broken by both boys. A second possibility is that children may simply be unable to contend with conflicting information about the letter and spirit of a rule, in which case they may not distinguish between whether the rule was broken and whether the rule breaker should be condemned. If this is true, then children may say that the person who did not violate the spirit of the law also did not violate the rule at all. They may reason that because the person should not be condemned, that he or she must not have violated the law. The data from Study 1.1 will be informative in differentiating these possibilities.

We predict that children will be more forgiving of violations of the letter of the law that do not violate the spirit of the law and that children will become more forgiving of such violations as they get older. Previous research suggests that children making increasingly sophisticated moral inferences as the mature from age 4 to 10 years-old (Cushman et al., 2013; Darley & Shultz, 1990; Richardson et al., 2012) and thus we expect that children will

increasingly differentiate between of whether the rule was "broken" and whether the rule breaker should be condemned as they get older.

Method

Participants. We recruited a sample of 76 children ages 4 to 10 years old (M_{age} = 7.0, 40 females). Participants were randomly assigned to either the spirit violated condition (n = 39, 20 females) or the spirit intact condition (n = 37, 19 females). One participant was dropped for not completing the study. For all of the studies in this paper, children were recruited and tested at a local science museum in the Midwest. Due to the great variability in the daily number of visitors to the museum, we set our stopping criterion for data collection as the end of the day on which we had at least 20 participants per age group (4- to 5-year-olds, 6- to 8-year-olds, and 9- to 10-year-olds). Children were recruited in a public area of the science museum where they ran in a series of brief studies, including this one; the present study took about 5 minutes to complete, and children generally spent no more than 10 minutes in total at the table. No demographic data beyond age and gender was collected from the families. After participation, children were allowed to select a few stickers of their choice.

Procedure. Both parental consent and child assent preceded participation in every study. All children watched a video vignette presented on the laptop. Children were randomly assigned to either the spirit violated condition or the spirit intact condition. The narration for each condition, with accompanying images described in parentheticals, is given below. Figure 1 also shows some of the key images from the video slideshow. In both conditions, children first heard a description of a rule:

> Here are all the kids at the library having fun. The teachers really, really don't want the books to get dirty or lost and that's why they came up with a new rule for the library. And the rule goes like this: Every little kid is only allowed to take four books home with them (*teacher shown describing rule*). They are only allowed to

> take four books home with them because it's impossible to fit so many books into one backpack (*image shows books overflowing from backpack*)! And that's why, when kids try to take the books home, they sometimes lose books or get them dirty! If any kid takes more than four books home then they have to stay inside for recess.

In the spirit violated condition, children then heard:

> This is Timmy and he wants to take home more than four books home because his best friend is home sick. (*Timmy shown remembering sick friend*) So, he wants to take four books home for himself and two books for his best friend. So, Timmy takes home more than four books.

In the spirit intact condition, children instead heard (emphasis added):

> This is Timmy and he wants to take home more than four books home because his best friend is home sick. So, he wants to take four books home for himself and two books for his best friend. **He brought an extra big backpack to school today that can fit all of the extra books too**. So, Timmy takes home more than 4 books.

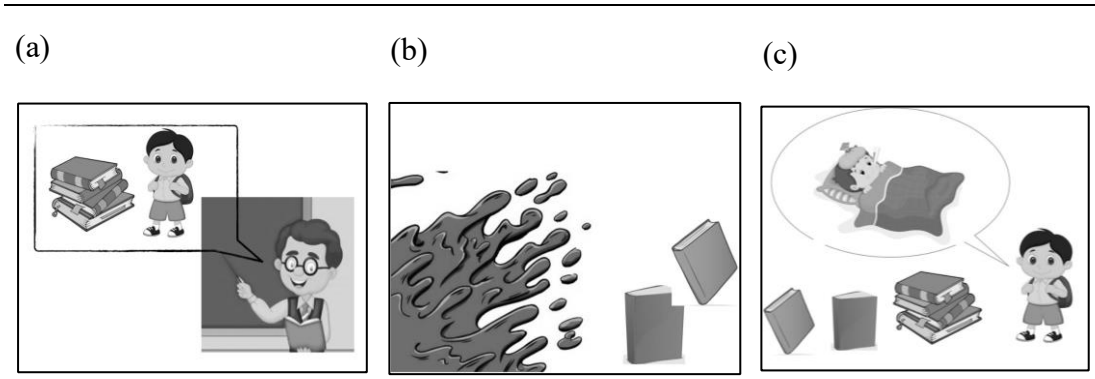


Figure 1.1. Example images from stimuli for Study 1.1. (a) The teacher announces the rule; (b) the potential consequences of violating the rule are illustrated (i.e., the books could get dirty) (c) the protagonist remembers his sick friend.

After watching the video, children were asked a series of questions by the experimenter in a fixed order, and the experimenter told participants that there was no right or wrong answer to any of the questions. Children were asked, "Do you think Timmy should have to stay inside at recess?" (Punishment), "Did Timmy break the rule in the story?" (Letter of the rule), "Was it wrong what Timmy did in the story?" (Wrongness), and then, "Will the teachers who made the rule be mad at Timmy?" (Affective response)." As our primary interest was in how children answered our first question regarding punishment, we always started with that question, to avoid

any consistency effects of asking about the rule violation first. After each question, the experimenter waited for the child's initial "yes" or "no" response and then asked them to clarify whether it was "*maybe* yes/no" or "d*efinitely* yes/no." Participants' responses were coded numerically on a scale from 0 to 3: An answer of "definitely no" was scored as a 0, "maybe no" was scored as a 1, "maybe yes" was scored as a 2, and "definitely yes" was scored as a 3.

We predicted that in both conditions children would say that the boy in the story broke the rule, but we predicted that children would make different evaluations by condition on the three other questions. Preliminary analyses showed very high correlations (all rs > .85, $p < .001$) among the other three questions (on whether the boy should be punished, whether his actions were wrong, and whether his teachers would be mad); we therefore combined these three measures into a single "disapproval" index by taking the average of the three responses (Cronbach's $\alpha = .95$). This disapproval index was then reverse coded, so that higher numbers indicate *more* disapproval.

After answering all other questions, children were asked two recall questions to ensure that they understood the vignette fully. They were asked "How many books did Timmy take from the library?" and subsequently, "How many books was Timmy allowed to take?" No children were excluded with these attention checks.

Results

We fit a multiple linear regression to the disapproval index using condition (spirit violated vs. spirit intact) and participant age (continuous) as predictors. For the disapproval index, the linear regression model explained a significant portion of the variance in children's responses, $F(3,72) = 235$, $p < .001$, $R^2 = .91$. The condition manipulation did have a significant effect in the predicted direction, $\beta = -1.30$, $t(95) = -4.23$, $p < .001$. When the spirit was violated,

children were significantly more likely to report disapproval (M = 2.62, SD = 0.31) than when only the letter of the rule was broken (M = 0.34, SD = 0.48).

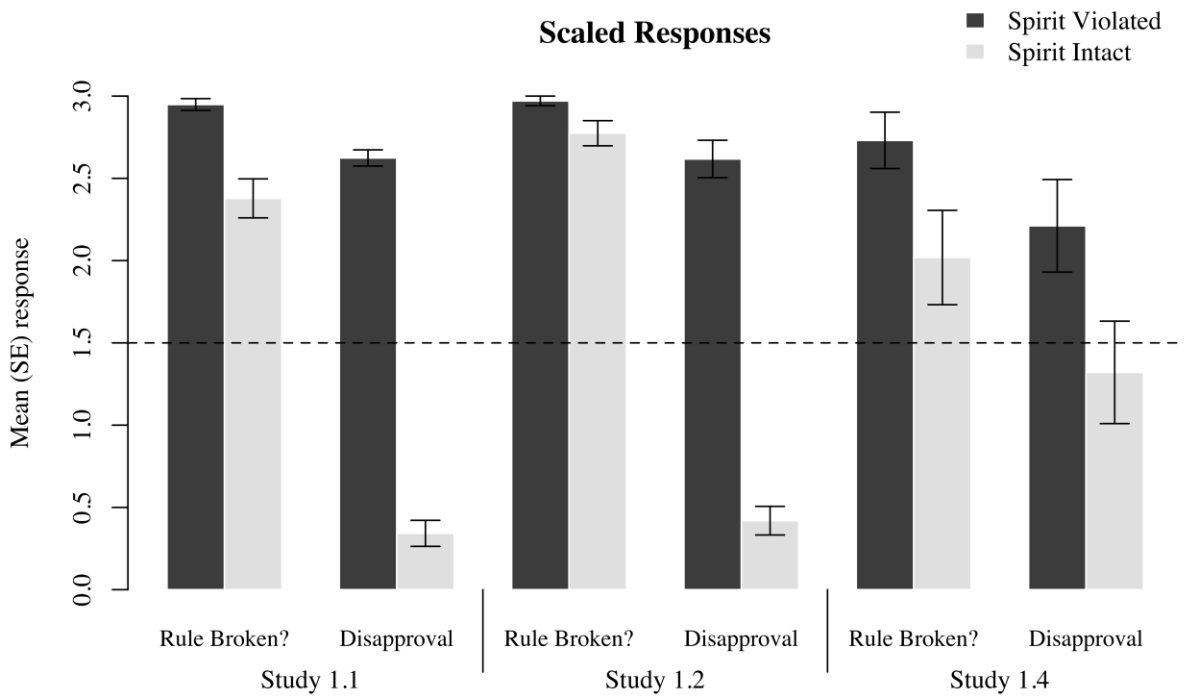


Figure 1.2. Mean responses for whether the target action violated the rule (Rule Broken?) and mean disapproval index (Disapproval) for the action in Studies 1.1, 1.2, and 1.4. Error bars represent standard error of the mean.

Age also independently predicted responses, β = 0.09, $t(72)$ = 2.98, $p < .001$, such that older children were slightly more likely to report disapproval. The interaction between age and condition was also a significant predictor of disapproval, β = -0.14, $t(72)$ = -3.23, $p < .001$. As Figure 3 illustrates, this interaction reflects the fact that older children were generally more disapproving than younger children when the spirit of the rule was violated, but they were generally less disapproving than younger children when the spirit of the rule was intact.

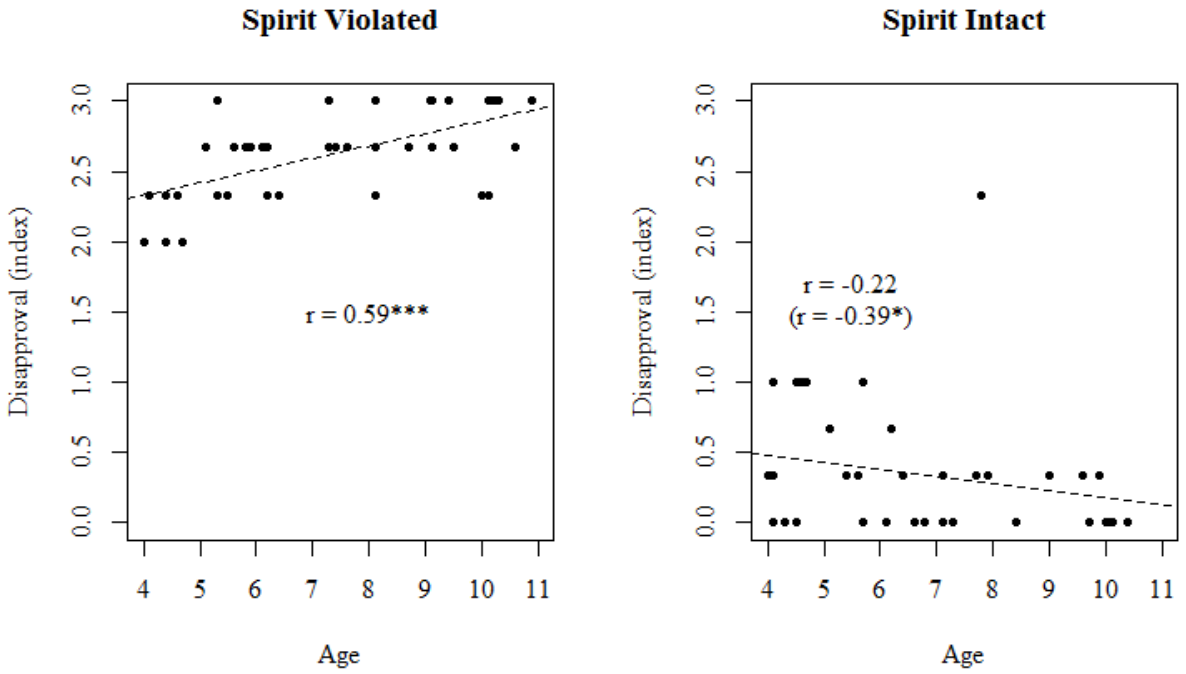


Figure 1.3. Scatterplots of disapproval index scores for each participant in Study 1.1, arranged by age, with Pearson correlations (r) shown. In the Spirit Intact condition, the parenthetical indicates the correlation when the outlier is excluded. * p < .05, **p < .01, ***p < .001.

Of course, lower rates of disapproval in the spirit intact condition could be due to the children who did not think the rule was broken at all; if they did not think the rule was broken, it follows that they would not impose punishment. To ensure this was not driving the difference between conditions, we also did a separate analysis of only those children who agreed that the rule was "definitely" broken (n = 55). Here we again found that children showed more disapproval in the spirit violated condition (M = 2.57, SD = 0.28), as compared to the spirit intact condition, M = 0.24, SD = 0.32), $t(29.89) = 25.36$, $p < .001$. Figure 4 shows the disapproval responses for only these children, across all three studies in this paper.

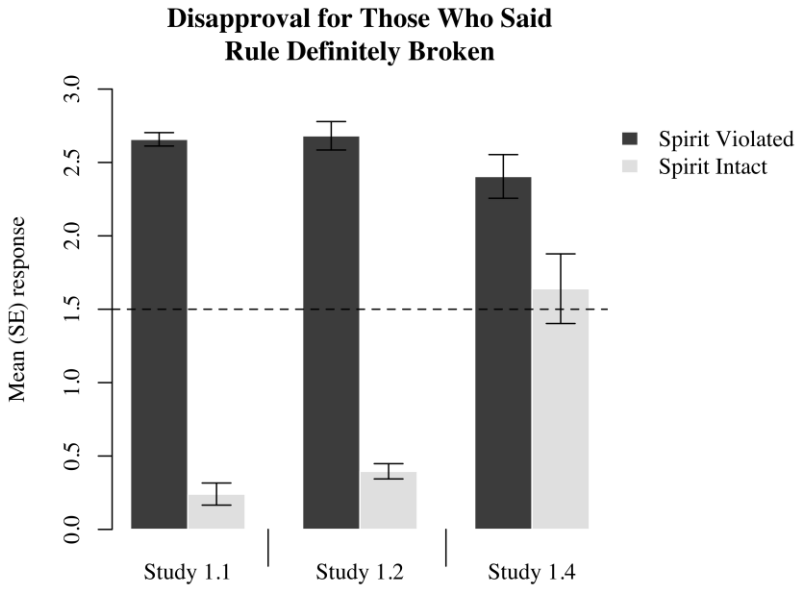


Figure 1.4. Average disapproval among only children who responded that the rule was "definitely" broken in Studies 1.1, 1.2, and 1.4, separated by condition. Error bars represent standard error of the mean.

We also fit multiple linear regression models for whether the rule had been broken using condition (spirit violated vs. spirit intact) and participant age (continuous) as predictors. For the question of whether the rule was broken, the linear regression model explained a significant portion of the variance in children's responses, $F(3,72) = 10.64$, $p < .001$, $R^2 = .31$. While the age of the child did not significantly predict their response, $\beta = 0.03$, $t(72) = 0.85$, $p = .40$, the condition manipulation did have a significant effect, $\beta = -1.06$, $t(72) = -2.60$, $p = .01$. Children in the spirit violated condition were more convinced that the rule had been broken ($M = 2.95$, $SD = 0.22$) than were children in the spirit intact condition ($M = 2.38$, $SD = 0.72$). The interaction between age and condition for the rule-breaking question was not significant, $\beta = 0.07$, $t(72) = 1.35$, $p = .18$.

Finally, to get a clearer sense of children's overall responses, we looked at each participant's initial "yes" or "no" response to the questions, regardless of whether they responded

"maybe" or "definitely" to the experimenter's follow up. In the spirit violated condition, 100% of children (n = 39) answered that "yes", Timmy broke the rule, and 100% said that Timmy should have to miss recess. In the spirit intact condition, 91.9% (n = 34) of children agreed that Timmy broke the rule, but only 2.7% (n = 1) said that Timmy should be punished. In other words, children appear to think that violations in the spirit violated condition should be condemned, but they are significantly more lenient in the spirit intact condition; almost all children thought that violating only the spirit of the law should not lead to condemnation.

Discussion

These results suggest that children as young as 4 years old are more lenient in their evaluations of rule-breakers who violate the letter of a rule, but do not violate the spirit (or intention) of the rule. Children in both conditions overwhelmingly agreed that both target actors broke the rule, and their scaled responses indicate that they were very confident in their judgments. However, they showed a very different pattern in their evaluations of the rule breakers: children in the spirit intact condition were much more lenient than were children in the spirit violated condition. Indeed, children overwhelmingly answered "no" to the question of whether the action should be punished in the spirit intact condition, whereas in the spirit violated condition, they all said "yes" to punishment.

When the spirit of the rule is intact, children at all ages agree that the rule was violated while also declining to disapprove of the rule breaker—that is, advocating leniency. In contrast, when both the letter and the spirit of the rule are violated, children at all ages agree both that the rule has been broken and that the person who broke merits disapproval. However, we also observed a slight developmental increase in children's tendency to forgive rule-breakers when they do not violate the spirit of the rule. As children age, they appear to be increasingly certain

that such rule-breakers should not be condemned. This is not a byproduct of children simply becoming more lenient as they get older; the opposite pattern emerges in the spirit violated condition. There, the older children appear to be even more harsh than the younger children.

Overall, these results indicate that children think it is much less bad to violate the letter of the law if one does not also violate the spirit of the law. Although our participants overwhelmingly indicated that the rule breaker should not be punished for violating only the letter of the rule, we cannot conclude that children only care about violations of the rule's spirit when they are asked to determine punishment; in our paradigm, the rule breaker was always trying to do something fundamentally prosocial: bringing home books to a sick friend. Importantly, however, this prosocial motivation was present in both of our conditions. This therefore cannot explain the differences between conditions (i.e., why children evaluate violations of the spirit and letter much more harshly than violations of the letter alone), but it might explain why the rates of condemnation were especially low here. Importantly, we find that children are much more lenient when the rule is broken, but the spirit of the law is intact.

Study 1.2: Children's Lenience does not Depend on Differing Assumptions about Outcome

Study 1.1 suggests that children make much less harsh evaluations of rule breakers who violate only the letter without violating the spirit of the rule. However, there is another possible explanation for children's apparent lenience; perhaps children are just making different inferences about what *outcomes* will occur in each scenario, and the differences in their judgments follows from that, rather than from any concern for the spirit of the rule. For example, it could be that children in the spirit violated condition were more likely to expect a negative *outcome* (i.e., that the books would be ruined) than were children in the only letter condition. That is, children in the spirit violated condition may have inferred that the books in fact would

get ruined because the teacher had suggested that this would happen, whereas children in the spirit intact condition may have inferred that the books would not be ruined because Timmy circumvented the teacher's fear by bringing an extra-large bag. Outcomes can certainly influence people's moral judgments—and especially children's judgments (for review, see Cushman et al., 2013). Thus, if children made different assumptions about whether the books were ruined in each case, they could have adopted a "no harm, no foul" mentality, where the protagonist in the latter story should not be blamed because no damage was done.

Study 1.2 attempts to control for this possibility. Here, we explicitly tell children in both conditions that the books were not damaged. If the difference we observed in Study 1.1 was based on children's guesses about the outcome of Timmy's actions, then the difference should disappear in this study; that is, both actions should be equally morally acceptable. However, if, as we hypothesize, children's determinations are based on evaluations of the actions and rules themselves, we should see the same distinction between the conditions that we found in Study 1.1: children should provide more lenient evaluations of behaviors that violate the letter, but not the spirit of a rule.

Method

Participants. We recruited a sample of 65 participants, ages 4 to 10 years old (M_{age} =7.2, 32 females). Participants were randomly assigned to either the spirit violated condition (n = 34, 17 females) or the spirit intact condition (n = 31, 15 females).

Procedure. As in previous studies, children watched a vignette video on a laptop at the research desk in a public area at the museum. The vignettes were adapted from Study 1.1 and remained the same, except for the addition of information about the outcome of Timmy bringing the books home.

In the spirit violated condition, the video explains that Timmy slipped in a puddle, nearly getting the books dirty and damaging them, but thanks to good luck, the books miss the puddle and do not get dirty. In the spirit intact condition, Timmy similarly slips in a puddle on his way home, nearly getting the books dirty. However, because the books are securely stored in his backpack, they do not become dirty. Importantly, in both conditions participants are reassured that the books did not get dirty and were delivered safely to the sick friend.

As in Study 1.1, after the video, children were asked whether Timmy should have to stay inside at recess, whether Timmy's actions were wrong, and whether Timmy broke the rule. Children's responses about punishment and wrongness were again highly correlated ($r(65) = .82$, $p < .001$, Cronbach's $\alpha = .90$), so the disapproval index was an average of those two responses. The scaled responses and difference score were calculated in the same manner as in Study 1.1.

Results

All analyses were the same as those performed in Study 1.1. We fit multiple linear regression models for both the scaled rule broken responses and the disapproval index, using condition (spirit violated vs. spirit intact) and participant age (continuous) as predictors. The disapproval model explained a significant amount of variance in children's responses, $F(3,61)=106.1$, $p < .001$, $R^2 = 0.84$. For this measure, age significantly predicted participant responses, $\beta = 0.21$, $t(61) = 4.11$, $p < .001$, such that older children were more disapproving, regardless of condition. There was also a statistically significant interaction between age and condition, $\beta = -0.31$, $t(61)= -4.29$, $p < .001$. As shown below, this interaction reflects a developmental change in children's disapproval by condition; older children were more disapproving in the spirit violated condition and less disapproving in the spirit intact. When

controlling for the interaction between age and condition, the main effect of condition was not statistically significant, β = - 0.01, $t(61)$= -0.03, p = .98.

Although all of the children in this study agreed that the rule was broken, we again wanted to be sure the difference in disapproval between the conditions was not driven by children in the spirit intact condition being less certain that the rule was broken, and therefore less certain that punishment should be imposed. To do so, we compared disapproval ratings among only those children who said the rule was "definitely" broken (n = 57); this difference was still significant, $t(47.58) = 20.80, p < .001$, such that children in the spirit intact condition (M = 0.40, SD = 0.25) were still more lenient than those in the spirit violated condition (M = 2.68, SD = 0.56). See Figure 4 above for a summary of these results.

For the question of whether the rule was broken, our model explained a significant portion of the variance in children's responses, $F(3,61) = 11.31$, $p < .001$, R^2 =.36. Participant age did not predict responses to this question, β = 0.02, $t(61) = 0.76$, p = .45, but the condition manipulation did have a significant effect, β = -1.13, $t(61)$= -3.95, $p < .001$. Children in the spirit intact condition were more certain that the rule had been broken (M = 2.97, SD = 0.17) than were children in the spirit violated condition (M = 2.77, SD =0.43). The interaction between age and condition was significant, β = 0.12, $t(61) = 3.09$, $p < .001$.

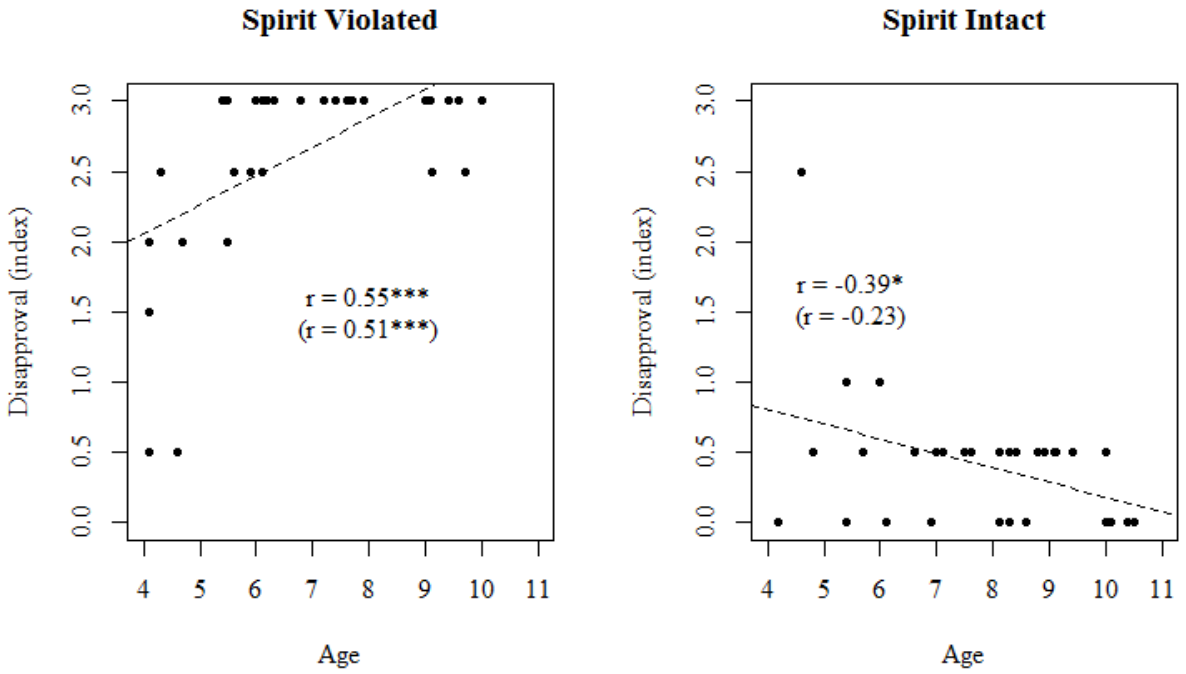


Figure 1.5. Scatterplots of disapproval index scores for each participant in Study 1.2, arranged by age, with Pearson correlations (r) shown. Parentheticals indicate the correlation when the outliers are excluded. *p < .05, **p < .01, ***p < .001

Discussion

Consistent with Study 1.1, children continue to distinguish between the letter and the spirit of the rule in Study 1.2—children think it is less wrong to violate the letter of the rule when one does not violate the spirit of the rule. This replication demonstrates that the effect in our first study was not driven solely by anticipated differences in outcome. Here, when outcomes were explicitly made equivalent (i.e., children were told that there was no harm done to the books), children still report much less disapproval when only the letter is violated as compared to when both the letter and the spirit of a rule are violated.

Just as in Study 1.1, when the letter and spirit are both violated, children of all ages strongly agree both that the boy broke the rule and that his actions merit disapproval. When the spirit of the rule remained intact, children again show increasing lenience as they age. We also

found again that children in the spirit violated condition become significantly *harsher* in their moral evaluations; here even more strongly than in Study 1.1. However, this effect appears to be most strongly driven by the youngest children's ratings of disapproval being more lenient than in our previous study. That we saw this age effect more starkly in Study 1.2 than in Study 1.1 is likely a result of the outcome information we provided; because we explicitly told children that no harm came about due to the rule violation, the youngest children may have been more likely to dismiss the violation and excuse the violator. This is consistent with other research indicating that young children's evaluations are particularly sensitive to outcome (see Cushman et al., 2013; Heyman et al., 2009).

Study 1.3: Children Distinguish between Identical Actions When the Spirit of the Rule Differs

Studies 1.1 and 1.2 found that children are more lenient when someone violates only the letter of a rule without violating its spirit by presenting children with identical rules, based on identical concerns, and varying the actions taken by the children. However, varying the actions for which the target children are being judged can make it more difficult to interpret children's differing evaluations. Study 1.2 controls for (and rules out) one source of this difficulty: the possibility of differing outcomes. But it does not control for any differences that children may perceive between the actions *themselves*. To control for any such differences in the way children perceive the actions, Study 1.3 manipulates the spirit of the rule rather than the actions of the rule breaker. Rather than presenting identical rules with identical underlying motivation to that rule and varying the focal actions, we instead presented children with identical rules and identical actions and varied the spirit of the rule.

We predicted that, consistent with our previous findings, children in both conditions should agree that the boy in the story broke the rule, because in both conditions the boy's actions

clearly violated the letter of the rule. However, in line with the notion that children are capable of distinguishing between the letter and the spirit of the rule, we predicted that children would be more lenient in their evaluations of rule-breaking that does not violate the spirit of the law.

Method

Participants. We recruited a sample of 99 children ages 4 to 10 years old (M_{age} = 7.1, 51 females). Children were randomly assigned to the spirit violated condition (n = 51, 24 females) or the spirit intact condition (n = 48, 27 females).

Procedure. The procedure was substantially the same as Studies 1.1 and 1.2: children watched one of two vignette videos (between subjects) on a laptop and then answered a series of questions. As in those studies, the rule was that children are not permitted to take more than four books home. However, in this study, the motivation behind the rule was manipulated between conditions.

The spirit violated condition was identical to the spirit violated condition in Study 1.1. Participants heard that, because more than four books will not fit into a single child's backpack, the rule exists to prevent books from getting lost or dirty. "Timmy" then takes home more than four books so that he can give some to a sick friend. No outcome information was provided.

In the spirit intact condition, however, the rule was explained differently. Rather than being based on a fear of dirt or damage, the rule was described as existing "because if the kids take too many books home there might not be enough books left for all the other kids who come to the library." The implication is that Timmy's actions here do not violate the spirit of the rule because he is not taking "too many" for himself in this case; some of the books will be used by another child. Importantly, in both conditions the boy is doing the same action (taking home four books) to achieve the same goal (helping a sick friend).

After watching the video, children were asked the same series of questions as in Study 1.1. We again predicted that in both conditions children would say that the boy in the story broke the rule, but that children would make different moral evaluations by condition. We also re-introduced the question about whether or not the teacher would be mad, despite how closely it mirrored the punishment question in Study 1.1, because we thought children might respond differently in this study, where the teacher had a different state of mind at the outset. However, we again found very high correlations (all *r*s > .76, all *p*s < .001; Cronbach's α = .91) among all three disapproval questions, and we therefore combined all three measures into one index.

Results

The analyses were conducted in the same manner as in Studies 1.1 and 1.2. In the spirit violated condition, 100% of participants (n = 50) responded "yes" when asked if Timmy broke the rule, and 100% of participants also responded that Timmy should be punished. In the spirit intact condition, however, while 93.9% of children (n = 46 of 49) responded that the rule had been broken, only 4.1% (n = 2) thought that Timmy should be punished.

On the scaled measure for rule-breaking, children were well above the midpoint, indicating their confidence that the action in question broke the rule, in both the spirit intact condition (M = 2.31, SD = 0.58, $t(48) = 9.65$, $p < .001$) and the spirit violated condition (M = 2.94, SD = 0.24, $t(49) = 42.44$, $p < .001$). But on the scaled disapproval measure, children on average responded significantly below the midpoint in the spirit intact condition (M = 0.54, SD = 0.38, $t(48) = -17.75$, $p < .001$), and significantly above the midpoint in the spirit violated condition (M = 2.55, SD = 0.37, $t(49) = 19.82$, $p < .001$).

As in Studies 1.1 and 1.2, to further explore these effects, we fit multiple linear regression models for both the disapproval index and children's responses to whether the rule

had been broken, using condition (spirit violated vs. spirit intact) and participant age as predictors.

For the question of whether the rule was broken, the model explained a significant portion of the variance in children's responses, $F(3,95) = 17.79$, $p < .001$, $R^2 = .36$. The age of the child did not significantly predict responses, $\beta = -0.01$, $t(95) = -0.18$, $p = .85$, but the condition manipulation did have a significant effect, $\beta = -1.13$, $t(95) = -3.04$, $p < .01$. Children in the spirit violated condition were more certain that the rule had been broken ($M = 2.94$, $SD = 0.24$) than were children in the spirit intact condition ($M = 2.31$, $SD = 0.58$), although both groups responded on average that the rule was broken. The interaction between age and condition was not significant, $\beta = 0.07$, $t(95) = 1.35$, $p = .18$.

The model for disapproval also explained a significant amount of the variance in children's responses, $F(3,95) = 239$, $p < .001$, $R^2 = .88$. Once again, age did not significantly predict responses, $\beta = 0.03$, $t(95) = 1.04$, $p = .30$, but the condition manipulation did have a significant effect in the predicted direction, $\beta = 1.46$, $t(95) = -4.69$, $p < .001$. When the letter and spirit were violated, children were significantly more likely to report disapproval ($M = 2.55$, $SD = 0.37$) than when only the letter of the rule was broken ($M = 0.54$, $SD = 0.38$). The interaction between age and condition was also a marginally significant predictor of disapproval, $\beta = -0.08$, $t(95) = -1.78$, $p = .08$. Figure 1.5 illustrates the relationship between age and disapproval. When the children who said the rule was not broken were excluded, the interaction was significant, $\beta = -0.09$, $t(92) = -2.16$, $p = .03$, but none of the other results change in direction or significance.

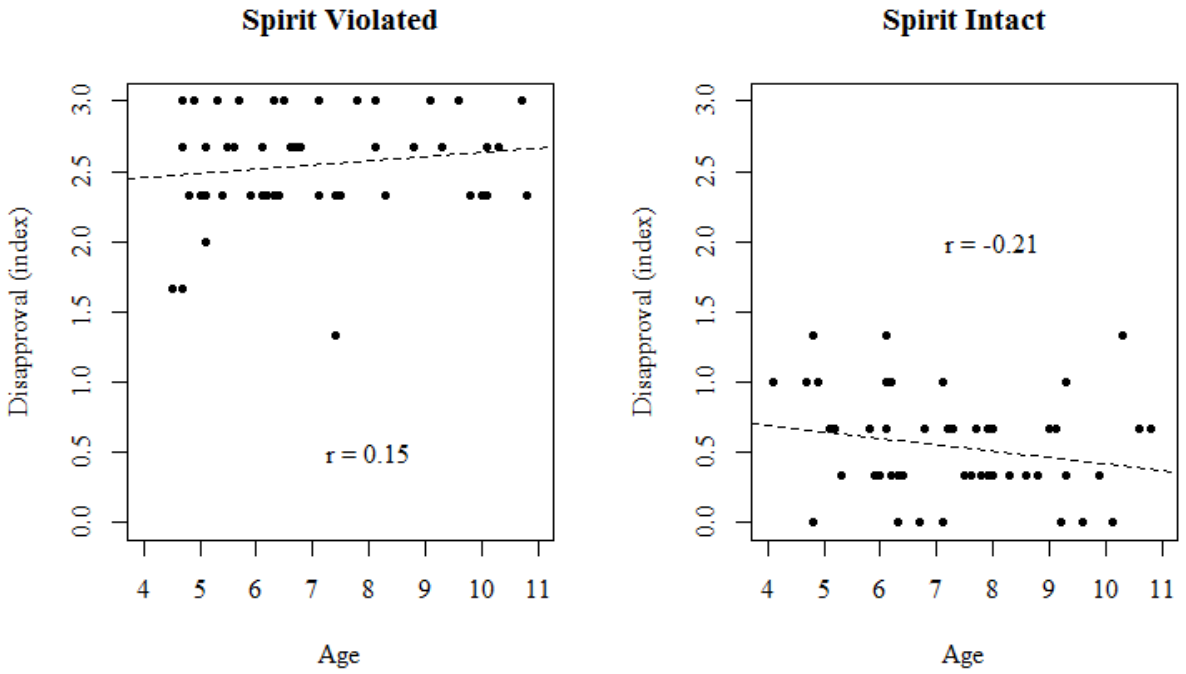


Figure 1.6. Scatterplots of disapproval index scores for each participant in Study 1.3, arranged by age, with Pearson correlations (r) shown. *$p < .05$, **$p < .01$, ***$p < .001$

Discussion

We replicated and extended our previous findings, demonstrating again that children are more lenient towards rule-breakers who do not violate the spirit of the rule, while agreeing that both rule-breakers broke the rule in question. In this study, the actions in question were identical, but children as young as 4 years old were still more lenient when the action violates only the letter of a rule than they are when the action violates the spirit of the rule as well. Taken together, Studies 1.2 and 1.3 help us to rule out a number of alternative explanations for the striking effects of Study 1.1. Neither uncertainty about the outcome of the action, nor the inherent qualities of the action itself (e.g., cautiousness or planning) change the central finding of Study 1.1: children take the intentions behind a rule into account when evaluating those who violate the letter of that rule. Indeed, if taking home books without a backpack is inherently careless, then we should have seen children saying that both boys were wrong for bringing the

books home without an extra large back pack. However, based on a manipulation of the intention behind the rule, children evaluated the same action very differently.

Study 1.4: Children's Lenience does not Depend on the Rule-Breaker's Cautiousness

Although Study 1.3 attempted to controls for any differences that children may perceive between the actions, an argument remains that, given the differences between the scenarios, the child who violates only the spirit of the rule is exercising more care and planning than the child who simply violates the letter and spirit of the rule. From previous work, we know that children think it can be wrong to be careless (Nobes, Panagiotaki, & Pawson, 2009); this could therefore be driving their judgments. Study 1.2 may even have inadvertently exacerbated this problem by increasing the salience of the carelessness, as the planning of the actor in the spirit intact condition actually prevents the books from becoming dirty. In contrast, the child in the letter and spirit condition escapes the damage only through luck. If cautiousness is a virtue, then the reported difference in evaluations in our previous studies could be the result of the difference in cautiousness.

More broadly, we might be wary of generalizing the results of Studies 1.1-1.3 because they involve identical rules and identical contexts. Aside from the above concerns about cautiousness and risk, it is hard to rule out idiosyncratic features of the scenario that might be affecting children's evaluations. To account for these possibilities and test whether children distinguish between the violations of the letter only and violations of the spirit of a rule in a different context, Study 1.4 introduces a totally new rule. In this version, the rule is about when and how often Timmy is allowed to eat dessert. Across conditions, Timmy neither exercises care nor acts carelessly; instead, he simply decides to violate the spirit of the rule or to violate only its letter. We predicted that, consistent with our previous findings, children in both conditions

should agree that the boy in the story broke the rule, because in both conditions the boy's actions clearly violated the letter of the rule. However, in line with the notion that children are capable of distinguishing between the letter and the spirit of the rule, we predicted that children would be more lenient in their evaluations of rule-breaking that does not violate the spirit of the law.

Method

Participants. We recruited a sample of 105 children ages 4 to 10 years old (Mage = 7.4, 51 females). Children were randomly assigned to the spirit violated condition (n = 52, 24 females) or the spirit intact condition (n = 53, 27 females).

Procedure. Participants were presented with a two-dimensional puppet of a boy while the experimenter told a story that varied by condition. Both conditions began with the same setup:

> This is Timmy. Timmy's parents want him to be healthy. Dessert is not very healthy, so Timmy's parents made a rule that Timmy can only have dessert on dessert night, which happens one time each week. Yesterday was dessert night at Timmy's house, and Timmy's dad made ice cream sundaes!

In the spirit violated condition, the story continued:

> After dinner yesterday, Timmy had ice cream for dessert with his parents. Tonight, Timmy is having dinner at a friend's house. They ask Timmy if he wants to have dessert, and he says yes, even though it's not dessert night. Timmy just had dessert last night, so Timmy decides to eat dessert two times this week.

But in the spirit intact condition, the story went a bit differently:

> After dinner yesterday, Timmy did not want any ice cream, so he did not have any dessert. Tonight, Timmy is having dinner at a friend's house. They ask Timmy if he wants to have dessert, and he says yes, even though it's not dessert night. Timmy didn't have dessert last night, so Timmy decides to eat dessert one time this week.

Thus, in both conditions, Timmy violates his family's rule by having dessert on the night after "dessert night". In the spirit violated condition, he does so after eating dessert the night before as well. In the spirit intact condition, however, Timmy skips dessert the night before, so that when

he violates the letter of the rule by eating dessert, he nonetheless preserves the spirit of the rule by having eaten dessert just once in the week.

After hearing the stories, participants were asked just two questions: first, whether Timmy should get in trouble, and second, whether Timmy broke the rule.

Results

The analyses were conducted in largely the same manner as in Studies 1.1 – 1.3. We fit multiple linear regression models for both the disapproval question (i.e., "Should Timmy get in trouble?") and children's responses to whether the rule had been broken, using condition (spirit violated vs. spirit intact) and participant age (continuous) as predictors. The model for whether Timmy should get in trouble explained a significant amount of the variance in children's responses, $F(3,101) = 7.76$, $p < .001$, $R^2 = .19$. Once again, the interaction between age and condition was a marginally significant predictor of disapproval, $\beta = -.020$, $t(101) = -1.72$, $p = .09$. Figure 5 illustrates the relationship between age and disapproval; children in the spirit intact condition were more lenient with age, while children in the spirit violated condition were not. Controlling for this interaction, age did not significantly predict responses, $\beta = 0.02$, $t(101) = 0.17$, $p = .86$, and neither did condition, $\beta = 0.60$, $t(101) = 0.66$, $p = .51$.

As Figure 1.2 illustrates, however, children in Study 1.4 seem to be less sure that the rule was broken than in the earlier studies, especially in the spirit intact condition. In that case, the difference in children's disapproval could be driven by uncertainty; i.e., if children are less sure that the rule was broken, then that could be why they are also more lenient. To examine this possibility, we once again looked at only those children who thought the rule was "definitely" broken (n = 67). Even among those children, disapproval ratings were significantly higher in the

spirit violated condition (M = 2.40, SD = 0.96) than in the spirit intact condition (M = 1.64, SD = 1.19), $t(42.73) = 2.73, p < .01$. These results are also shown in Figure 4 above.

For the question of whether the rule was broken, the model (using the full sample of children) explained a significant portion of the variance in children's responses, $F(3,101) = 8.61$, $p < .001$, $R^2 = .20$. There was a significant interaction between participant age and condition, $\beta = -0.23$, $t(101) = -2.44$, $p = .02$. Older children were less sure than were younger children ($r = -.39$, $p = .04$) that the rule had been broken in the spirit intact condition, but children in the spirit violated condition were equally sure that the rule had been broken, regardless of age ($r = .17$, $p = .23$). After accounting for this interaction, the age of the child did not significantly predict responses, $\beta = 0.07$, $t(101) = 0.90$, $p = .37$, and neither did condition, $\beta = 0.96$, $t(101) = 1.36$, $p = .18$.

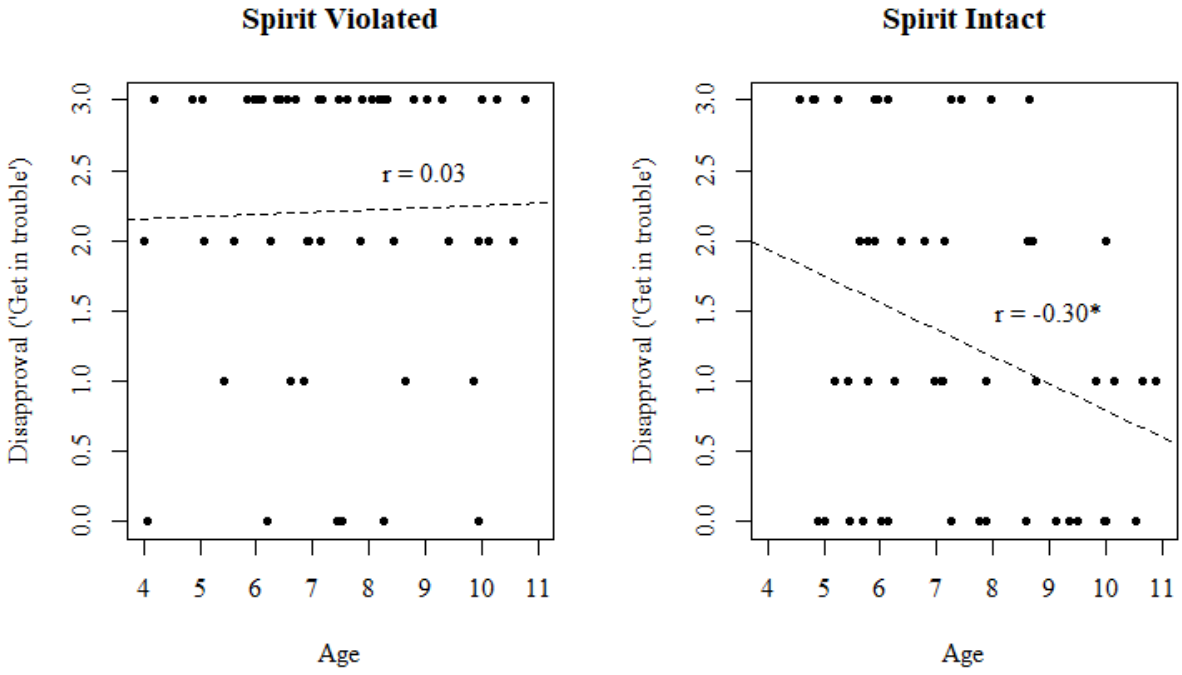


Figure 1.7. Scatterplots of disapproval index scores for each participant in Study 1.4, arranged by condition and age, with Pearson correlations (r) shown. *p< .05, **p < .01, ***p < .001.

In the spirit violated condition, 94.2% of participants (n = 49) responded "yes" when asked if Timmy broke the rule, and 78.8% of participants (n = 41) also responded that Timmy should get in trouble. In the spirit intact condition, however, 64.2% of children (n = 34) responded that the rule had been broken, and only 41.5% (n = 22) thought that Timmy should be punished.

Discussion

Study 1.4 serves as both a conceptual replication and an extension of our previous findings. Once again, children are more lenient towards rule-breakers who do not violate the spirit of the rule, while still agreeing in general that both rule-breakers broke the rule in question. The rule in this study was very different than the rule at issue in Studies 1.1-1.3, involving a different context (home vs. school), different source of authority (parents vs. teacher), and different risks (health vs. property damage), but children were still more lenient when the action violates only the letter of a rule than they were when the action violates the spirit of the rule as well.

Although it did not change the overall results, there was some interesting variation in children's responses in this different rule context. Compared to the previous studies, the children in the spirit intact condition of Study 1.4 seem to have been both less sure that the rule was broken and more disapproving of the rule breaker than were the children in the equivalent condition of the other studies, and there was noticeably more variance in children's responses. We can only speculate about what is driving these differences, but the new scene used in Study 1.4 does change some key aspects of the original scenario that we think are notable. One key difference is that the apparent motivation behind the rule-breaking has changed. In the original stories, we were careful to establish prosocial motives for Timmy's apparently antisocial act (i.e.,

breaking the rule); Timmy wants to bring books home to a sick friend. We did this deliberately and held it constant between conditions, so that we could credibly trace children's disapproval of Timmy to the fact that he violated the rule, rather than to any attribution of selfish motives. Although children were still willing to condemn Timmy's actions when he violated the spirit of the rule, indicating that they were indeed sensitive to the rule's spirit, the prosocial nature of his actions may have made it easier to forgive the rule violation when the spirit of the rule was left intact. In contrast, Timmy's actions in Study 1.4 have no comparably clear prosocial motive; it just seems that Timmy wants to have dessert on the second night. It is possible to construe Timmy's actions as broadly social in Study 1.4, because he eats dessert on the second night with a friend's family, which one could view as an attempt to be polite or cooperative. Even on this interpretation, however, the action is not a selflessly prosocial one; i.e., Timmy still gets to eat dessert.

Another important difference between the scenarios is the changing contexts within Study 1.4; while the rule is described in the context of Timmy's home, the violation of the rule occurs at a friend's house. In comparison, the rule in Studies 1.1-1.3 is both described and violated in the same context—namely, the school library. This change in situation could have led to the comparative uncertainty children had about whether the rule was broken in the sprit intact condition. Despite our attempts to make it clear that Timmy was violating the rule, children may have wondered whether the rule even applied at someone else's house, and this may have been especially true when Timmy's actions did not seem to violate the spirit of the rule.

In spite of these differences, however, it is important to emphasize that the overall pattern of results in Study 1.4 replicates the pattern from the previous two studies. Taken together, Studies 1.2-1.4 help us to rule out a number of alternative explanations for the striking effects of

Study 1.4. Neither uncertainty about the outcome of the action, nor the inherent qualities of the action itself (e.g., cautiousness or planning) change the central finding of Study 1.1: children take the intentions behind a rule into account when evaluating those who violate the letter of that rule.

General Discussion

These studies are the first evidence that children between the ages of 4 and 10 years old employ knowledge of legislative intent to make nuanced evaluations of rule-breaking scenarios. They judge those who violate the spirit of a rule more harshly than those who only violate the letter of the rule without violating its spirit. We found this to be true when the letter and spirit of the rule were held constant, but the actions differed so that in one case the rule-breaker did not violate the spirit of the rule (Study 1.1). This was true even when we made it clear that breaking the rule did not result in a bad outcome (Studies 1.2 and 1.3). Finally, in a completely different social context, with a different rule and potential harm (Study 1.4), we again found that children were less likely to condemn rule-breakers who only violated the letter of the rule.

We also found that children were more lenient toward rule-breakers who only violated the letter of the law as they got older, even when the outcome information was held constant. We found that younger children (4- to 6-year-olds) were less likely to be lenient with rule-breakers who only violated the letter of the law than older children (7- to 10-year-olds). Importantly, this was not simply because children become more tolerant of rule violations as they get older. We found that for violations of the spirit of the rule, the opposite was true: that is, older children provided harsher evaluations for violations that contravene the legislative intent than did younger children. In sum, our results provide evidence that even young children reason about the spirit of a rule when making judgments, and children give more weight to the spirit of the law in their evaluations as they grow older.

The present research extends previous work on children's early reasoning about norms. Researchers have long known that children are far from inflexible in their judgments about rules (Neary & Friedman, 2014; Nucci, 1981; Rhodes & Chalik, 2013); e.g., children evidence much less strong condemnation for rule-breaking that only violates conventional rather than moral rules (Smetana et al., 1991). The present research builds on this previous work on children's understanding of rules and norms in two important ways. First, we demonstrate that children not only distinguish between different types of rules but also between different types of rule-breaking (i.e., rule-breaking that violates only the letter of a rule or violating letter and spirit), and we find that children make very different evaluations of a person who breaks precisely the same rule. Second, our studies provide new evidence for the ways in which a wider consideration of the social efficacy of rules may factor into children's normative evaluations. In demonstrating that children prioritize the spirit of the law over the letter of the law, our findings may suggest that children view the aims of a rule as more important to maintaining order than general obedience (for a related work see Riggs & Kalish, 2016). Children's moral judgments, then, do not only evaluate concrete instances of rule-breaking, but their evaluations may reflect a consideration of the extent to which the wider social aims of a rule have been violated. This research, alongside more recent work on children's intuitive theories of punishment contributes to an understanding of how sophisticated legal ideas and approaches develop alongside more general moral understanding (Bregant, Shaw, & Kinzler, 2016).

Our findings may have important implications for understanding how adults and children reason about rule-breaking in the real world. Reasoning about legislative intent is a necessary cognitive skill for critically examining the complex dynamics of rules specifically and society more generally. Reasoning about legislative intent is also deeply linked to a broader

understanding of society's normative commitments, the framework of intentions behind our rules and laws, and bedrock systems of social values (e.g., "do no harm" or "respect other's property"). We hope this work prompts further investigations into children and adults' understanding of legislative intent.

Although these studies provide evidence that children understand legislative intent, the situations they were exposed to obviously lacked some of the complexities of real world cases. In our present studies, the spirit of the law was always made explicit, but there are very few instances of rule-learning in real life that have the benefit of clear and explicit legislative intent. It is not clear how children learn to infer the intentions of the rule maker, or indeed the extent to which they make such inferences naturally. In children's everyday interactions with rules, is the purpose of a rule relevant at all? If so, do children make their own judgments about a rule's purpose, or do they rely on information received from the rule's enforcers? If children use, either implicitly or explicitly, the purpose of a rule in evaluating rule-breakers, how do they deal with purposes that seem unfair? These questions raise innumerable additional avenues of investigation for future study.

This study also does not address how competing notions of legislative intent— for example, within diverse cultural contexts—interact and compete to create a more complex understanding of what a rule attempts to achieve. A research question that follows naturally from the present study would be to understand how children learn to navigate between and selectively apply competing social rules in diverse contexts. For example, how do children come to understand that it may be important to remove one's shoes at temple in order to show respect, it may be very inappropriate to do so at school (and could even signal disrespect)? Future research

should explore when children can and do reason about the spirit of the law independently in the various domains of their daily lives.

Additionally, although our results demonstrate that children can forgive violations of the letter of the law if the spirit of the law is not violated, these results do not allow us to clearly determine whether children believe one of the two aspects of the rule (letter or spirit) is more important. In all of our studies, we have compared how children evaluate the violation of just one aspect (the letter) as compared to the violation of *both* aspects. Perhaps it is the violation of the rule's spirit that matters most to children, while the technical violation of the letter is viewed as relatively unimportant. Or maybe the two aspects of the rule simply have an additive effect, such that violating both is worse than violating just one. Future research could help resolve this issue, for example, by demonstrating that children think it is worse to violate the spirit of the law while not violating the letter of the law (e.g., when a child puts their hands in another child's face to annoy them and says "I'm not touching you") as compared to violating the letter but not the spirit of a rule.

In demonstrating children's ability to separate the explicit text of a rule from its legislative intent, this study adds to a growing literature on children's ability to reason about the (often implicit) intricacies of social norms (e.g., Riggs & Kalish, 2016). Children's lenience toward rule-breakers who violate only the letter of the law, and not its spirit, suggests that even preschoolers are capable of considering the normative commitments that are left unarticulated by the plain language of the rules. Moreover, these results suggest that children's intuitive moral reasoning incorporates a wide appreciation for the dynamic interactions between rules and contexts.

Study 1.5: Adult Evaluations Distinguish Between the Letter and Spirit of the Law

While children seem to be lenient toward those who violate only the letter—but not the spirit—of the rule, adult evaluations may incorporate more complex concerns. For adults, violations of the letter of a rule might be seen as a harm in and of itself; if rules preserve order, then violations of a rule might represent threats to a broader structure even when no other tangible harm has been done. "Slippery slope" arguments may also compel adults to enforce even technical violations of the rules to prevent future would-be rule violators from taking advantage of any perceived lenience (Burg, 1991; Volokh, 2003). At the extremes, this logic gives rise to zero-tolerance policies and strict liability offenses (Giffin & Lombrozo, 2016).

Structural issues and concerns about procedural fairness may also cause adults to perceive these kinds of violations differently. The origin of the rule, the source of authority, the perceived fairness with which the rule is applied, and other rules or policies that relate to the rule in question may all enter into adult evaluations. Of course, children may share many of these concerns as well, and future research could explore children's reasoning on these points, but adults are likely to be more familiar with the broader legal and social contexts at issue in any given instance of rule enforcement.

Study 1.5 seeks a foothold into this complex area of intersection between psychology and policy. The purpose of this study is not to settle the questions of whether or to what extent each of these factors plays a role in adult moral evaluation; instead, it is simply to provide a place from which to start answering these questions. The basic design mirrors the design used with children in studies 1.1-1.4. Participants read about a rule (in this case a town ordinance), and then one of three people who violated the rule: one who violated only the letter of the rule, one who violated only the spirit of the rule, and one who violated both the letter and the spirit. I predicted

that, like children, adults would be more lenient in their evaluations of those who violate only the letter of the rule. However, when the actor violated only the spirit of the rule, we should see evidence of the conflict discussed above.

Method

Participants. Participants were 161 adults (54 female, 2 other) recruited via Amazon's Mechanical Turk and TurkPrime.com (Litman, Robinson, & Abberbock, 2016).

Procedure. Participants were randomly assigned to one of three conditions: spirit intact (n = 55), letter intact (n = 51), and both violated (n = 55). In all three conditions, participants first read a short description of a city ordinance:

> The City of Springdale has a very popular science museum. People come from all over to visit the museum, and some city residents have complained that it can be very hard for them to find parking spaces at the museum, especially when they have small children with them.
>
> To help with this problem, Springdale has passed a city ordinance that provides for "special access" parking spaces at city building and many shopping centers. These spaces are wider than most normal spaces and much closer to building entrances. The city ordinance states that the special access parking spaces are reserved for "city residents who have limited mobility, who are pregnant, or who have small children" so that "they may receive assistance and easier access to" the buildings.
>
> The special access parking spaces are clearly marked with signs that include the ordinance language and number (as well as the city motto), as shown below. People who violate the ordinance have to pay a fine of up to $200.

Participants then read one of three descriptions of a woman who parked in a designated special access parking space. In the both violated condition, participants read:

> Susan Smith lives in Northwoods, which is about 2 hours away from Springdale. She is not a Springdale resident. Last weekend, Susan visited the Springdale Science Museum alone. It was a busy day at the museum, and there were only a few parking spaces available. Most of the available spaces were far away from the entrance. Susan parked in one of the designated "special access" parking spaces, even though she could have parked further away from the museum's entrance and walked to the door without a problem.

In the spirit intact condition, the woman was instead described as follows:

> Susan Smith lives in Northwoods, which is about 2 hours away from Springdale. She is not a Springdale resident. Last weekend, Susan visited the Springdale Science Museum with her 2-year-old daughter and 3-month-old baby. It was a busy day at the museum, and there were only a few parking spaces available. Most of the available spaces were far away from the entrance. Susan parked in one of the designated "special access" parking spaces, because she was worried that it would be too difficult to park further away from the museum's entrance and walk to the door with her two children and a stroller.

Finally, in the letter intact condition, the description continued:

> Susan Smith lives in Springdale and is a Springdale resident. Susan recently became pregnant, but she is not yet experiencing any symptoms or signs of her pregnancy. Last weekend, Susan visited the Springdale Science Museum alone. It was a busy day at the museum, and there were only a few parking spaces available. Most of the available spaces were far away from the entrance. Susan parked in one of the designated "special access" parking spaces, even though she could have parked further away from the museum's entrance and walked to the door without a problem.

Figure 1.8. Illustration of parking space marker, provided to participants in Study 1.5.

In all three conditions, participants rated their agreement with two statements on a 7-point Likert scale, coded from 0 (strongly disagree) to 7 (strongly agree): “Susan Smith violated the ordinance” and “Susan Smith should have to pay a fine.” On the next page, participants were asked what amount of fine, if any, should be imposed (from $0 to $200). Finally, on another page, participants completed three evaluative inference

measures, where they were asked to rate their agreement (on the same 7-point scale as above) on the following: Susan intended to violate the ordinance, Susan's actions were harmful, and the town ordinance was fair.

Results

Rule broken measure. When asked whether Susan violated the ordinance, an omnibus F-test showed a significant effect of condition on participant responses, $F(2,158) = 23.90, p < .001$. Post hoc comparisons (all post-hoc comparisons in this section are Tukey's Honestly Significant Difference tests) showed that participants in the both violated condition were significantly more sure that the ordinance was violated than were participants in the spirit intact condition, $p = .01$. Participants' responses were significantly above the midpoint of the scale in the spirit intact condition ($M = 3.95$, $SD = 2.16$), $t(54) = 3.27, p < .01$) and both violated condition ($M = 5.13$, $SD = 1.49$), $t(54) = 10.52, p < .001$; in other words, participants in those conditions responded, on average, that the ordinance was violated. In the letter intact condition, the average response was not significantly different than the midpoint of the scale ($M = 2.49$, $SD = 2.16$), $t(50) = -1.69, p = .10$, but it was significantly less than the response in the spirit intact condition, $p = .001$. These results are shown in Figure 1.8.

Punishment measure. With regard to punishment, an omnibus F-test showed a significant effect of condition on participant responses, $F(2,158) = 22.58, p < .001$. Post hoc tests showed that the mean in the both violated condition was significantly more than the mean in the spirit intact condition, $p < .001$, and in the letter intact condition, $p < .001$. Responses did not differ significantly between the spirit intact and letter intact conditions, $p = .45$. Only participants in the both violated condition responded on average ($M = 4.82$, $SD = 1.61$) above the midpoint of the scale, i.e., that Susan should have to pay a fine, $t(54) = 8.37, p < .001$. In the spirit intact

condition, the mean response ($M = 2.87$, $SD = 2.24$) did not differ from the scale midpoint, $t(54) = -0.42, p = .68$. In the letter intact condition, responses were significantly below the midpoint of the scale, $t(50) = -2.09, p = .04$. These results are summarized in Figure 1.8.

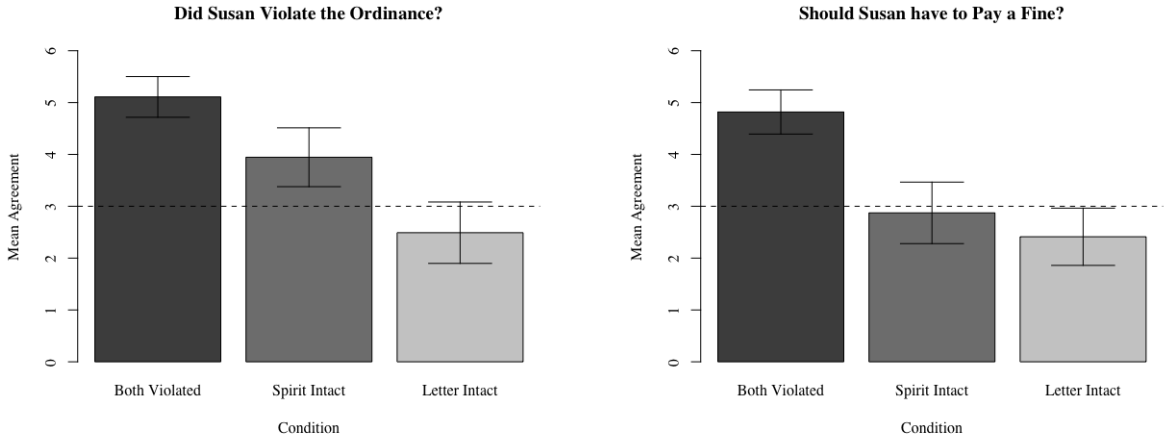


Figure 1.9. Mean ratings of agreement in each condition for whether the ordinance was violated (left) and whether punishment should be imposed (right). Error bars represent standard error of the mean.

Fine amount. When asked what value of fine should be imposed for the ordinance violation (if any, up to $200), the omnibus F-test also showed a significant effect of condition on the amount of the fine, $F(2,158) = 15.08, p < .001$. Participants in the both violated condition chose, on average, $115.73 ($SD$ = 69.50), significantly greater than either the spirit intact condition (M = $59.64, SD = 69.47), $p < .001$, or the letter intact condition (M = $48.96, SD = 64.08), $p < .001$. Fines in the letter intact condition did not differ significantly from those in the spirit intact condition, $p = .70$.

Because participants could answer that no fine should be imposed, I also fit a logistic regression model to the fine data, comparing the likelihood of imposing any fine to the likelihood of answering $0. Participants in the both violated condition selected $0 just 5% of the time, and they were significantly less likely to do so than were participants in both the spirit intact condition (36% selecting $0), $\beta = 2.29, z = 3.49, p < .001$, and the letter intact condition (43%

selecting \$0), $\beta = 2.58, z = 3.92, p < .001$. Participants' rates of choosing \$0 did not vary significantly between the spirit intact and letter intact conditions, however, $X^2(1, n = 106) = 0.26, p = .61$. Excluding participants who chose \$0, the average fine imposed in the both violated condition was \$122.40 (*SD* = 65.43), while it was \$93.71 in the spirit intact condition (*SD* = 66.17) and \$86.10 in the letter intact condition (*SD* = 63.40).

Evaluative inferences. The remaining three measures, intention to violate the rule, harmfulness, and fairness of the rule, are analyzed separately below. Intention to violate the rule ratings and harmfulness ratings were significantly positively correlated, $r(159) = .54$, $p < .001$, as were ratings of harmfulness and whether the law is fair, $r(159) = .23$, $p < .01$. However, ratings of intentionality were not correlated with ratings of whether the law is fair, $r(159) = .13$, $p = .10$. Among all three measures, Cronbach's $\alpha = .58$.

Intention to violate the rule. There was a significant effect of condition on participants' belief that Susan intended to violate the rule, $F(2,158) = 11.57, p < .001$. Participants in the both violated condition were significantly surer that Susan intended to violate the ordinance (*M* = 3.95, *SD* = 1.61) than were participants in the spirit intact (*M* = 2.53, *SD* = 1.96) or letter intact (*M* = 2.47, *SD* = 1.84) conditions, both *p*s < .001. Ratings of intentionality did not differ significantly between the spirit intact and letter intact conditions, $p = .99$. In the both violated condition, the average rating was significantly above the midpoint of the scale, $t(54) = 4.34, p < .001$. Ratings in the spirit intact condition did not significantly differ from the midpoint, $t(54) = -1.79, p = .08$, and the average in the letter intact condition was significantly below the midpoint, $t(50) = -2.06, p = .04$. Participants' belief that the actions were intention was also significantly positively correlated with participants' belief that Susan should be punished, $r(159) = .69$, $p < .001$.

Harmfulness. Condition had a significant effect on how harmful participants believed Susan's actions were, $F(2,158) = 12.69, p < .001$. Participants in the both violated condition were more likely to agree that her actions were harmful (M = 3.33, SD = 1.69) than were participants in either the spirit intact (M = 1.82, SD = 1.62) or letter intact (M = 2.10, SD = 1.69) conditions, both ps < .001. However, the mean response in the both violated condition did not differ significantly from the midpoint of the scale, $t(54) = 1.44, p = .16$. Mean responses were significantly lower than the midpoint of the scale for both the spirit intact condition, $t(54) = -5.40, p < .001$, and the letter intact condition, $t(50) = -3.82, p < .001$, and those means did not differ significantly from one another, p = .66. Harmfulness ratings were also significantly positively correlated with participants' belief that Susan should be punished (i.e., have to pay a fine), $r(159)$ = .62, p < .001.

Fairness of the rule. An omnibus F test also showed a significant effect of condition on fairness ratings, $F(2,158) = 5.08, p < .01$. Post hoc comparisons showed that in the both violated condition (M = 4.18, SD = 1.48), ratings were significantly higher than ratings in the spirit intact condition (M = 3.51, SD = 1.48), p = .05, but not significantly different than ratings in the letter intact condition (M = 4.37, SD = 1.28), p = .78. Ratings in the letter intact condition were also significantly higher than those in the spirit intact condition, p = .01. Finally, participants in all three conditions agreed, on average, that the ordinance was fair; average fairness ratings were above the midpoint of the scale in the both violated condition, $t(54) = 5.92, p < .001$, the spirit intact condition, $t(54) = 2.31, p = .02$, and the letter intact condition, $t(50) = 7.66, p < .001$.

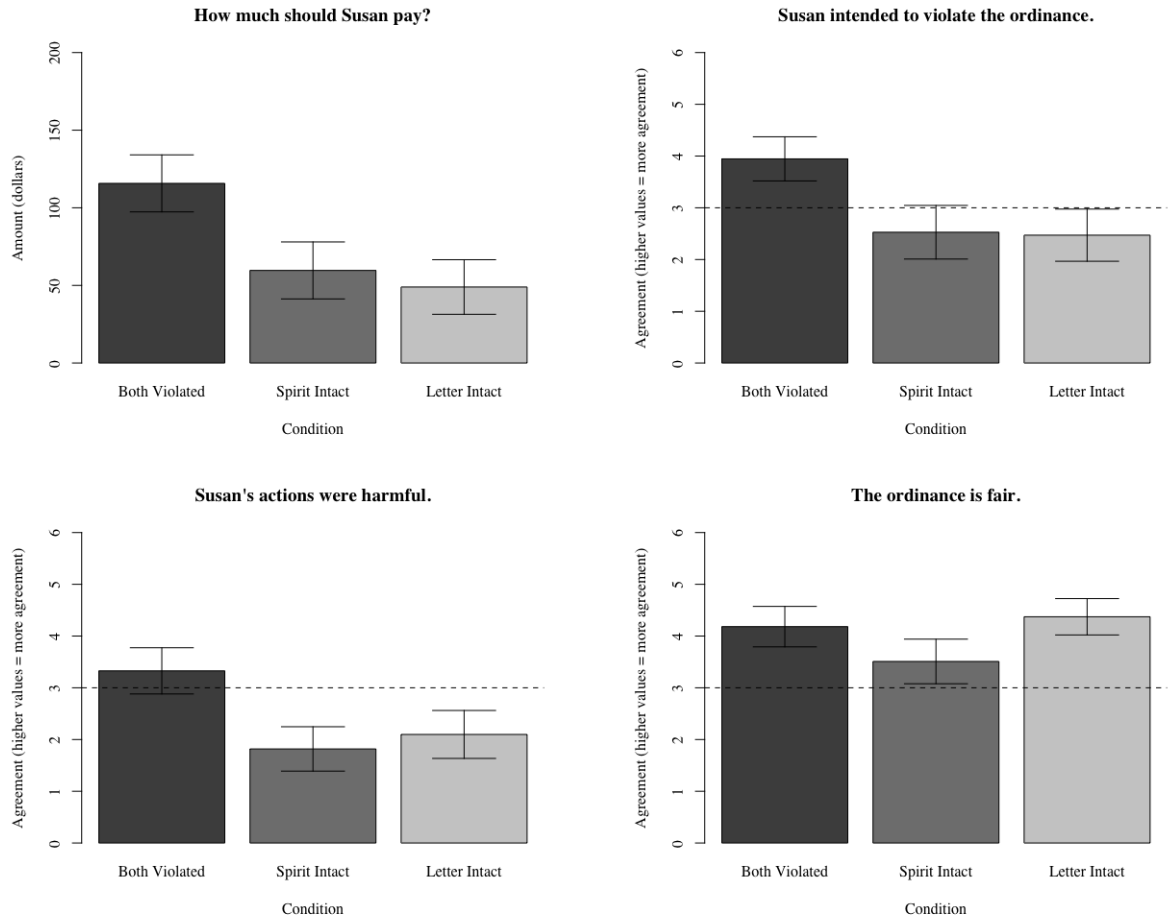


Figure 1.10. Mean responses by condition for the amount of the fine that should be paid (top left), intentionality of the violation (top right), harmfulness of the violation (bottom left) and fairness of the ordinance (bottom right). Error bars represent standard error of the mean.

Discussion

Across all of the measures, adults were harshest toward the person who violated both the spirit and the letter of the ordinance as compared to the people who violated only the letter or only the spirit of the ordinance. While participants in the spirit intact condition were significantly more sure that the ordinance had been violated than were participants in the letter intact condition, those conditions did not differ in terms of punishment or in terms of how intentional or how harmful the actor's behavior was judged to be. In other words, participants seem to treat

the violations similarly when either the letter or the spirit of the rule is preserved, at least in this scenario.

Overall, adults seem to share children's tendency to be more lenient toward rule violations when the spirit of the law is preserved. The results of Study 1.5 also suggest that adults are more lenient when the letter of the law is preserved, while Studies 1.1-1.4 left open this question in children. These results thus provide some evidence that there may be continuity between children's reasoning about the spirit of the law and adults'.

Despite the similarities between the spirit intact and letter intact conditions, the two key differences are telling. Participants were more sure that the ordinance was violated in the spirit intact condition—that is, when the woman in question was not a city resident, but she did need the special access provided by the designated parking spaces—than they were in the letter intact condition. However, participants also rated the ordinance itself as less fair in the spirit intact condition than they did in the other two conditions. These inferences may well be linked; when participants' attention was called to the city resident requirement, as it was in the spirit intact condition, the ordinance is rated as less fair.

As the first step in exploring adult reasoning about the letter and spirit of the law, this study has some limitations that should be addressed by future research. Most importantly, the participants were left to determine the spirit of the ordinance for themselves. While background on the ordinance was provided, the letter and spirit of the law are not spelled out separately. This is most relevant for interpreting the spirit intact condition; the scenario was designed to imply that the spirit of the ordinance was to help parents with young children secure parking close to the museum. Thus, the use of a designated space by a non-resident who needs that help was meant to be a violation only of the letter of the ordinance. However, participants may have

viewed the city residence requirement as part of the spirit of the ordinance as well; the scenario explains that it was the complaints of city residents about limited parking that led to the adoption of the ordinance. If the city residence requirement was seen as part of the spirit of the law, then the comparison between the spirit intact and the letter intact conditions must be treated with some caution. It may be more accurate to think of the spirit intact condition as a less-severe violation of the spirit, rather than as a case in which the spirit was preserved. This may also explain, in part, why participants treated the two conditions so similarly.

In contrast, the actor in the letter intact case clearly preserved the letter of the rule, meeting all of the ordinance's literal requirements, but she also violated the spirit of the ordinance. Participants in that condition were nonetheless more lenient than were those in the both violated condition; adults were not even sure, on average, that the ordinance had been violated, and 43% of participants responded that no fine should be imposed. These results suggest that, as predicted, adults put substantial weight on the letter of the law, even when the spirit is violated.

Of course, the ambiguity in the spirit intact condition is also consistent with the way these kinds of cases operate in the real world. Take, for example, the prohibition against motorized vehicles in a public park discussed in the introduction to this part. Although most adults likely agree that a motorized wheelchair should not be subject to penalty under this law, the question is actually a bit more complicated than that. For example, what about a motorized scooter that helps its user, but which is not medically necessary? And if that, too, is permissible, then what about an electric bicycle (increasingly common in many parts of the world)? Or a scooter being used by someone for fun, rather than for assistance?

The "slippery slope" argument—that one exception may lead to another, and another, and so on, until all is lost—is a familiar trope in law and policy (see, e.g., Burg, 1991; Kirchler, Hoelzl, & Wahl, 2008; Volokh, 2003). While scholars debate the political and legal realities of these arguments (e.g., Schraub, 2013), they carry persuasive weight. In the context of rule and law violations, the specter of the slippery slope haunts spirit intact situations with particular force. In cases where some are tempted to say "this is not what the law was intended to prevent", others will respond that leniency in one case could lead to leniency in other, more questionable, cases. Indeed, "zero tolerance" regimes are designed in part to prevent precisely this kind of disturbance to the rule structure and to minimize uncertainty around whether a rule ought to be enforced in a particular instance.

The results of Study 1.5 provide an interesting starting place for this research in adults. Combined with Studies 1.1-1.4, they begin to paint a picture of how children's early ideas about the spirit of a rule may grow and develop into adult arguments about law and policy. While adult intuitions may be more complicated or nuanced, even young children are more lenient of rule violations when the spirit of the rule remains intact.

2 Interpreting Punishment's Expressive and Informational Value

Throughout the literatures of law, psychology, and philosophy, a great deal of attention has been paid to the question of why people seek to punish one another (Bilz, 2007; Buckholtz et al., 2008; Carlsmith, Darley, & Robinson, 2002; Cushman et al., 2009; Darley & Pittman, 2003; Kurzban, DeScioli, & O'Brien, 2007; Orth, 2003; Price, Cosmides, & Tooby, 2002). Amid all the discussion of what punishment *should* and *can* accomplish or communicate, however, relatively little thought has been given to what punishment actually *does* signal (but see, e.g., Bilz, 2016; J. G. Murphy & Hampton, 1990). Neglect of punishment's signal is no small oversight; many theories of punishment, from deterrence to restorative justice, rely on the assumption that lay people will understand punishment in a particular way that is consistent with normative theory. If this assumption is mistaken, it could undermine the strength and legitimacy of punishment policy.

In this chapter, I present empirical evidence that speaks to the most basic way punishment may be understood by the lay public. We pose a simple research question: What do people infer about an action based on the fact that it is punished? Psychologically speaking, punishment may operate as a special case of social norm information, but we argue that what sets punishment apart from other norms is the moral weight punishment carries. Although norms other than punishment may also communicate moral messages, punishment seems to be unique in its relationship to morality, and especially to judgments of harm. Prior research demonstrates that potential punishers rely heavily on the degree of harm caused by wrongdoing when determining the appropriate level of punishment (Carlsmith, 2006; Cushman et al., 2009). In this paper, we show that the opposite is also true—information about punishment can influence the extent to which an act of wrongdoing is judged to have been harmful.

In the first part of this paper, we discuss the existing research on the message of punishment, drawing on literatures from law, psychology, and philosophy. We also highlight closely related research on social norms and behavior. Our review of the literature concludes with a summary of research on punishment, moral judgment, and harm. In the second part of the paper, we present original experimental evidence that punishment can be an effective cue for moral judgment, influencing such judgments in a way that is similar to social norm information. Interestingly, however, punishment seems to most effectively signal a specific moral concern—harmfulness—especially relative to social normative information. Finally, in part three, we discuss some of the important implications and future directions of this topic.

Introduction

The importance of punishment to law is almost tautological: laws without law *enforcement* mechanisms are little more than aspirations or norms. Though the mechanisms by which laws are enforced vary, most such mechanisms can be broadly described as punishments. Legal punishment therefore marks, at least, the difference between a legal rule and a merely normative one. This difference is psychologically important; as we discuss below, the presence of a legal rule—even one which carries only nominal sanctions—seems to influence behavior. Punishment itself, however, also has a special psychological significance, forming an important building block of human moral reasoning and moral development.

In this section, we briefly review three related bodies of research that each address an important aspect of the present studies. First, we describe the extant research on the so-called expressive function of law, which demonstrates the power of laws to influence behavior. We next examine research on moral and social norms outside of the legal context; although laws undoubtedly provide normative information, our discussion highlights some ways in which the

analogy between norms and laws can break down. Finally, we review the importance of punishment to moral reasoning in particular, paying special attention to the feature of moral judgment that appears to be most related to punishment: harm.

The Expressive Function of Punishment

A single act of punishment can attempt to accomplish many simultaneous ends; for example, the target of punishment may be deterred from future wrongdoing by the threat of future punishment (Darley & Alter, 2013; Darley & Pittman, 2003; Nussbaum, 2013; Rabin, 2006), other members of the community may be deterred from imitating the target's wrongdoing (Byrd, 1989; Carroll, Perkowitz, Lurigio, & Weaver, 1987; Nussbaum, 2013; Orth, 2003; Vidmar & Miller, 1980), the target may be incapacitated (i.e., through incarceration) (Darley, Carlsmith, & Robinson, 2000; Farrington, 1986; Robinson & Darley, 1997) or rehabilitated (i.e., through treatment) (Cullen, Skovron, Scott, & Burton, 1990; McCorkle, 1993; Ward & Durrant, 2011), or restitution may be made to the victims of wrongdoing (Waldman, 2003; Ward & Langlands, 2009; Witvliet et al., 2008). More diffuse retributive interests, such as correcting the moral scales or meting out justice (Cahill, 2007; Fincher & Tetlock, 2015; Grisso, 1996; Hampton, 1992; van Prooijen, Coffeng, & Vermeer, 2014), may also be pursued. Beyond these instrumental ends, however, punishment—or, maybe more precisely what and who we *choose* to punish—carries a communicative weight (Harcourt, 2002; Kahan, 1996; J. G. Murphy & Hampton, 1988; Sunstein, 1995). This communicative aspect of law is often called its "expressive" function (e.g., Bilz, 2016; Gert, Radzik, & Hand, 2004; Mazzone, 1999; Sunstein, 1995).

Though sometimes given as an alternative to retributive or utilitarian theories of punishment, expressive functions of punishment are essentially orthogonal to these aims; the message communicated by a punishment act may itself be retributive, utilitarian, neither, or both of these.

Expressive theories of punishment are theoretically similar to so-called "signaling" accounts that are prevalent in the literatures of evolutionary science and economics (Bulbulia & Sosis, 2011; Connelly, Certo, Ireland, & Reutzel, 2011; FitzGibbon & Fanshawe, 1988; Morris, 1986), because both theories hold that an action can send a message, over and above the immediate consequences of the action itself. However, empirical studies of signaling accounts are generally precise as to the content of the message being sent and received—for example, some gazelles engage in a kind of jumping called "stotting" that appears to send an honest signal to predators about the gazelle's health (and therefore their ability to escape; FitzGibbon & Fanshawe, 1988). In contrast, empirical studies looking at the expressive functions of law tend to be vague about the content of the message sent by punishment (Cooter, 2000a; Funk, 2007; Wittlin, 2011). Even more importantly, the content of the message *received* has been left virtually unexplored by empirical research (but see Bilz, 2016).

To our knowledge, only two experimental studies have examined the message(s) that are communicated by punishment (Bilz, 2016; Bregant et al., 2016); both find support for a particular view of the expressive function that is sometimes called "expressive retributivism" (Gert et al., 2004; Hanna, 2008). Under this theory, crimes are themselves expressive acts that send a message to a victim and to society about the standing of the victim relative to the offender (Bilz, 2016; J. G. Murphy & Hampton, 1990). Punishment, in contrast, sends the opposite message, rejecting the offender's false claim and restoring the victim's position in society. In a set of experiments testing this view, Kenworthey Bilz (2016) found that, across a variety of crimes, punishment decreases the social standing of the offender and—crucially—increases the social standing of the victim. In a study of children aged 5- to 8-years-old, Bregant, Shaw, and Kinzler (2016) similarly found that children liked the victim of a theft more if the thief who

committed the act was punished, compared to when the thief went unpunished. Recent research in social neuroscience further emphasizes the importance of the victim in moral judgments; Patil et al. (2017; see also Rozin & Fallon, 1987; Rozin, Haidt, & Fincher, 2009) find that empathy for the victim contributes to moral blame, even when the harm is accidental.

The expressive retributivism argument centers on condemnation of the bad actor, especially relative to the victim, rather than on condemnation of the act itself. However, if punishment sends a message of condemnation, psychological evidence suggests the condemnation need not be limited to the actor. For example, Bregant, Shaw, and Kinzler (2016) also found that children used punishment as a signal of how "bad" the *act* of stealing is; in a world where those who steal are "never punished," children between the ages of 5 and 8 overwhelmingly reported that stealing was not "bad."

This divergence in children's reactions is, in some ways, a microcosm of the bigger questions surrounding expressive punishment messages, because it highlights two major themes that are relevant: social norms and moral condemnation. There are at least two possible explanations for children's belief that stealing is not "bad" when it is not punished. One possibility, is that punishment information communicates that an action is "bad" in the same way that it is "bad" to eat with one's hands at dinner. That is, punishment may merely be communicating that the action in question is a conventional violation of social norms. A second possibility is that punishment information communicates something about whether the action is immoral. That is, that this action is wrong intrinsically and immutably, which might cause people to infer that the action is harmful or morally disgusting. These two possibilities—social norms and moral judgment—are both cited in the broader literatures as possible messages of punishment, and we explore both below.

Normative Messages, Laws, and Punishment

Although empirical evidence of the messages of punishment is scarce, theories abound. One especially common characterization of the expressive function is that laws express social norms (Cooter, 1998, 2000a; Funk, 2007; Sunstein, 1995). A vast literature in social psychology illustrates the power of social norms to influence behavior across a wide variety of contexts (Borsari & Carey, 2003; Cialdini, 2007; Cialdini et al., 2006; Cole, Mailath, & Postlewaite, 1992; Gerber & Rogers, 2009; Milgram, Bickman, & Berkowitz, 1969; Stok, de Ridder, de Vet, & de Wit, 2014). Experimentally, normative information has been used to reduce self-reported speeding (De Pelsmacker & Janssens, 2007), increase energy conservation (Allcott, 2011), and curb college alcohol use (Borsari & Carey, 2003; Perkins, 2002); when people think that "everyone else" is doing something, they are more likely to engage in that something themselves (Cialdini, Kallgren, & Reno, 1991; Fehr & Fischbacher, 2004; Milgram et al., 1969).

If laws are perceived as the codification of social norms, then information about the legal status of an act could have a similar effect on behavior. Of course, laws may change behavior for other reasons as well; the threat of punishment may deter people from engaging in the illegal act. Nonetheless, a handful of studies have used changes in the law to argue in support of a normative expressive function. One of the clearest is Patricia Funk's (2007) study of Swiss voting laws. Funk analyzed voter turnout in several Swiss cantons during the last half of the 20th century. During that period, five of the cantons repealed long-standing mandatory voting laws that had been accompanied by fines that Funk called "symbolic" – the fines varied from canton to canton but were usually equal to about $1.00 (US) or less. Funk's study found that repeals decreased voter turnout in those cantons by six to ten percent. Because the punishment was so small, Funk argues that this is support for an expressive theory of law; people's behavior seemed to be

influenced by the mere presence of the law even in the absence of meaningful punishment, suggesting that voters were not simply deterred from defecting out of fear of punishment.

Similar studies have documented significant increases in compliance following the adoption of seatbelt laws, dog waste ordinances, and smoking bans, even when the penalty for violating the laws is minimal (Cooter, 1998, 2000b; Dharmapala & McAdams, 2003; Wittlin, 2011). Although these natural experiments generally reveal only the end points of the process—that is, a change in law leads to changes in behavior—researchers often claim (or assume) that the mechanism behind this behavioral change is the expression of social norms (Cooter, 1998; Funk, 2004, 2007; McAdams & Nadler, 2005; Wittlin, 2011).

Of course, laws do carry normative weight. At the very least, legal prohibitions convey injunctive norms against the prohibited actions; for example, a law against speeding suggests that at least the legislature believes one should not speed. But formalizing a social norm through punishment can also lead to unexpected counterintuitive changes in behavior. In a notable field study, for example, Gneezy and Rustichini (2000) introduced a new punishment for late parents at some Israeli day care centers. Parents signed a contract at the beginning of the school year in which they agreed to pick their children up on time, but prior to the study, no enforcement mechanism was specified for the rule. After measuring the number of late parents for four weeks, the experimenters introduced a financial punishment for being late at some of the day cares in the study. The punishment was relatively small—just 10 shekels (worth approximately $2.72 US at the time of the study) per child if the parent was more than 10 minutes late.

The introduction of the monetary punishment did change parental behavior at the day cares in the test group, but the effect was surprising. Rather than decreasing lateness at the selected centers, the fines caused a steady increase in lateness. After 12 weeks, day care centers

where the fine had been introduced reported a near doubling of the number of late parents, and removing the fine at the end of the study did nothing to reduce this new, higher rate of lateness (Gneezy & Rustichini, 2000). The experimenters argued that the introduction of the fine was equivalent to setting a price for late pickup; rather than deterring late parents, the (small) fine changed the prevailing social norm from one of obligation ("Parents should pick up their children on time") to one of transaction ("Parents can pay to pick their children up late").

As the Gneezy and Rustichini study demonstrates, the surface-level similarities between the effects of social norms and enforced laws on behavior may conceal deeper psychological differences. Moreover, punishment can signal a meaningful shift in the nature of the underlying act that colors subsequent behavior. This shift could be one from a social cooperation dynamic to a transactional dynamic, as occurred in the day care centers, but it could also be another kind of shift, such as one from a norm to a moral imperative.

Moral Psychology and Punishment

In contrast to the research noted above, which tends to treat punishment as a simple enforcement mechanism for social norms, philosophical approaches often emphasize the distinctly moral component of punishment (Bilz, 2010; Gert et al., 2004; Hampton, 1992; Hanna, 2008; Kahan, 1996, 1997; J. G. Murphy & Hampton, 1990). Indeed, many legal scholars characterize the message of punishment—rather vaguely—as moral condemnation. Dan Kahan (1996) argues, for example, that "[p]unishment…is a special social convention that signifies moral condemnation." Although moral psychology has not yet approached our question directly, that literature provides many important connections between punishment and moral judgment that may be particularly relevant to understanding what, exactly, punishment signals. Indeed,

amid the vast body of research on moral judgments, one link emerges repeatedly: the link between punishment and harm.

Harm is the central feature of retributive theories of punishment (Bilz & Darley, 2004; Byrd, 1989; Darley & Pittman, 2003; Hampton, 1992; Vaish et al., 2011). Under a retributive view, punishment is morally justified—indeed, morally required—to balance the harm done by the offender (Bilz & Darley, 2004; Byrd, 1989; Cahill, 2007; Carlsmith et al., 2002; Hampton, 1992; McKee & Feather, 2008; J. G. Murphy & Hampton, 1990). In contrast to consequentialist or utilitarian theories of punishment, which advocate punishment only to stem the future risk posed by an offender, retributivism is concerned primarily (or, in the extreme, exclusively) with evaluating the harm already caused and ensuring that perpetrators get what the deserve even if this does not lead to better consequences (Carlsmith, 2006; Darley et al., 2000; Robinson, 2008; Robinson & Darley, 1997).

Research in psychology also demonstrates the close relationship between punishment and harm. Empirical studies designed to compare the degree to which people rely on implicit theories of retributivism or consequentialism have found that the degree of harm caused is one of the most important pieces of information to (mock) punishers (Carlsmith, 2006, 2008; Darley et al., 2000; Darley & Pittman, 2003). Of course, moral psychology research often includes examinations of punishment outside of the retributivism vs. consequentialism debate, and that research also supports the idea that punishment judgments are closely related to harmfulness judgments. For example, studies of the "outcome bias" in moral psychology demonstrate that harm caused can even be more important for judging blame and assigning punishment than the wrongdoer's intent (Cushman et al., 2009, 2013; Finkel, 2000; Gino et al., 2009). Even more tellingly, studies of so-called "moral luck" have demonstrated that when an act causes harm,

judgments of punishment and blame are increased relative to judgments of the same action when it does not cause harm (Cushman, 2008; Martin & Cushman, 2016). In contrast, judgments of moral character and the wrongness of the act itself do not seem to rely as much on whether harm was done. It seems that outcomes matter for harm and punishment more than they matter for wrongness (we will return to this issue in our later studies).

Developmental research has also long recognized the connection between harm, immorality, and punishment. Developmental morality scholars have repeatedly demonstrated that children and adults distinguish between rules that they see as conventional (i.e., social norms), those they see as prudential or safety-related, and those that they see as moral (Ardila-Rey & Killen, 2001; Richardson et al., 2012; Shweder, Turiel, & Much, 1981b; Smetana, 1981b; Smetana et al., 1991; Tisak & Turiel, 1984; Turiel, 1978). Whereas moral rules like "don't hit" are seen as universal and immutable, even fairly young toddlers are more flexible when it comes to rules based in social convention, like "don't wear pajamas to school" (Ardila-Rey & Killen, 2001; Kalish & Sabbagh, 2007; Neff & Helwig, 2002; Piazza, Sousa, & Holbrook, 2013; Smetana, Schlagman, & Adams, 1993; Tisak & Turiel, 1984; Zhao & Kushnir, 2017). Importantly, a key distinction between rules that are perceived as moral and those that are perceived as conventional seems to be that the former—but not the latter—involve harm done to another person or creature (Kelly et al., 2007; Tisak & Turiel, 1984; Zelazo, Helwig, & Lau, 1996). In short, research indicates that moral transgressions—that is, acts that harm others—*demand* punishment, even if the surrounding social conventions are changed.

The previous research makes it clear that harm leads to increased punishment, but we do not know if punishment leads people to infer that an action in question is harmful. Are there other candidates for what punishment could signal about an action? Despite the early theories of

morality in social and developmental psychology that tended to treat moral violations as fairly homogenous and harm-based, many contemporary theories adopt a broader approach. Moral foundations theory, for example, identifies several underlying themes, in addition to harm, that may help to explain why moral violations are perceived as moral in the first place. In response to harm-centric theories of morality, Jonathan Haidt (2001) and others point to apparently harmless scenarios, such as a case of consensual incest with no negative consequences for either party. That such scenarios provoke a negative moral reaction has been used to argue in favor of a "purity" or "sanctity" domain of morality. Analogous hypotheticals led to Haidt's first categorizations of five "moral foundations": harm, fairness, loyalty, authority, and purity (Haidt, 2007, 2008), In addition to generating some of the most memorable hypotheticals for researchers (see, e.g., Gray, Schein, & Ward, 2014; Haidt, 2001; Kelly et al., 2007; Sunstein, 2005), purity violations are often cited as a rebuttal to critics of moral foundations theory.

The purity (or "sanctity") domain also attracts attention because it has the clearest connection to a specific emotional response—namely, disgust. Numerous studies have linked disgust reactions to moral judgments (Capestany & Harris, 2014; Inbar & Pizarro, 2009; Moll et al., 2005; Pizarro, Inbar, & Helion, 2011; Rottman & Kelemen, 2012; Salerno & Peter-Hagene, 2013; Schnall, Haidt, Clore, & Jordan, 2008), and even incidental feelings of disgust (such as those caused by a foul smell in the experiment room) can increase the harshness of moral evaluations and the desire to punish, especially for perceived violations in the purity/sanctity domain (Pizarro et al., 2011). Although these scenarios strongly minimize or eliminate obvious harms, these scenarios are nonetheless viewed as *morally* wrong (and therefore deserving of punishment) by participants.

It is worth noting that critics of moral foundations theory, the most prominent of whom argue that harm can adequately explain moral judgments without the need for other foundations (Gray, 2014; Gray & Keeney, 2015; Gray et al., 2014; Schein, Ritter, & Gray, 2016), have responded with a variety of explanations. Psychologist Kurt Gray has argued that these apparently harmless violations are not really perceived as harmless at all (Gray et al., 2014). Instead, Gray argues, subjective harm is imputed even when the scenarios are written to foreclose the possibility of objective harm.

The present project is not designed to resolve the debate between moral foundations theory and its critics by adjudicating whether morality is driven solely by harm or by other concerns beyond harm, or even to address it directly, though we discuss some possible implications of this research on this debate in the general discussion. However, the close association between harm and punishment led us to predict that punishment would communicate messages of harm particularly well, and the current debate in moral psychology provides us with an interesting alternative possibility. Perhaps the apparent relationship between harm and punishment is not so unique, but instead is an artifact, provoked by researchers who treat harmfulness as synonymous with morality. In that case, the disgustingness (i.e., the lack of purity) of an action might also be communicated by punishment information. Indeed, this possibility also finds support in the literatures of psychology and law. As noted above, disgust can increase the harshness of moral judgments; feelings of disgust have also been associated with more frequent and more severe punishment in vignette studies, mock juries, and economic games (Capestany & Harris, 2014; Inbar & Pizarro, 2009; Olatunji, David, & Ciesielski, 2012). Drawing on the literatures discussed above, we set out to look for evidence of what messages people actually receive from learning about punishment. We adopt a simple experimental

paradigm in which participants we told about a novel action. In the first condition of both studies, participants are told only that the novel act is or is not punished. Participants are then asked to rate action on several dimensions. To test whether punishment signals moral condemnation, for example, participants are asked to judge the "moral wrongness" of the action. Across Studies 2.1 and 2.2, we compare normative information to punishment information on three dimensions that are suggested by our review of the literature: moral wrongness, harmfulness, and disgust. In the last two experiments, we extend our findings from controlled but artificial alien actions to familiar but messier real-world actions.

Study 2.1: Punishment, Harm, and Moral Wrongness

Study 2.1 tests whether information about punishment leads people to make inferences about the moral status of an action and, if so, whether those inferences are specific to a particular moral dimension, such as wrongness or harm. Although it is not obvious that people will make any inferences, especially in such a simplified and artificial context, even if they do, such an inference is not very informative without additional comparisons. Is there anything special about punishment, or would any information about others' negative reactions give rise to the same inferences? To address this issue, our paradigm compares punishment with normative information, which we operationalized as telling participants that an action either causes or does not cause the actor to be disliked by others. By focusing the normative information on the actor, we can keep the information in the "dislike" conditions parallel to the information in the punishment conditions.

In Study 2.1, participants in all conditions were first introduced to the novel actions "blicking" and "gomping". Participants then received limited information about each action; the type of information varied by condition, as described in more detail below. In the punishment

information condition, participants were told that one action was generally punished and one was generally not punished. In the normative information condition, participants were told that one action generally caused the actor to be disliked, and the other action did not. The third condition—the conflicting information condition—pitted the punishment and normative information against each other. One action is described as punished but not likely to cause dislike of the actor, whereas the other action is described as *not* punished but generally causing dislike. This condition allows us to gauge whether punishment of the action or dislike of the action is a stronger signal of moral wrongness or harm.

Methods

Participants. Participants were 270 adults (100 female), ages 19-65 (M = 37.28, SD = 15.63), recruited from Amazon's Mechanical Turk (MTurk) and paid for their participation. We planned to assign 90 participants to each condition, but we allowed the numbers in each condition to vary slightly due to random assignment.

Procedure. Participants were randomly assigned to one of three conditions: punishment information (n = 91), normative information (n = 88), and conflicting information (n = 91).

In all three conditions, participants first read very brief instructions in which they were told to imagine an alien planet populated with aliens. Participants were also told that on this planet, "some things are quite similar to Earth, and some things are quite different." To minimize the degree to which participants incorporated their pre-existing moral beliefs into their responses, we used nonce words to describe unknown and novel actions; participants were told that these were two things that people on Earth "do not do." One was called "blicking," and the other was called "gomping."

In the punishment information condition, participants learned only whether the actions were or were not punished; i.e., they read that while blicking is punished, gomping is not. In the normative information condition, participants were told: "An alien who blicks another alien is generally disliked. An alien who gomps another alien is generally not disliked." Finally, in the conflicting information condition, participants received all of the information provided in the other two conditions: "An alien is disliked when she blicks another alien, but she is generally not punished. An alien is not disliked when she gomps another alien, but she is generally punished." After participants were given the condition-specific information about blicking and gomping, they were asked to choose which action was "morally worse." On a separate screen, participants also rated how morally "good or bad" they believed each action was on a scale ranging from "Very bad" to "Very good." Participants were also asked to rate how harmful each action was, and to choose which act was the more harmful. The moral wrongness questions were always presented together (though on separate screens), and the harmfulness questions were always presented together (though on separate screens), but the order of moral wrongness and harmfulness questions was randomized between participants.

Results

Punishment information condition. In the punishment information condition, 90% (n = 82) of participants reported that the act that was punished was more morally wrong than the act that was not punished; three percent (n = 3) responded that the non-punished act was more morally wrong, and seven percent (n = 6) said the acts were about the same in terms of moral wrongness, $\chi^2(2, \mathrm{N} = 91) = 132.15, \mathrm{p} < .001$ (unless otherwise noted, all reported chi-square results are chi-square tests for goodness of fit). Participants in the punishment information

condition also rated the punished act (M = 5.02, SD = 1.20) as significantly more morally wrong than the non-punished act (M = 2.54, SD = 1.28), $t_{paired}(90) = 12.55, p < .001$.

Similarly, 92% (n = 84) of participants in the punishment information condition reported that the punished act was the more harmful. Two percent (n = 2)

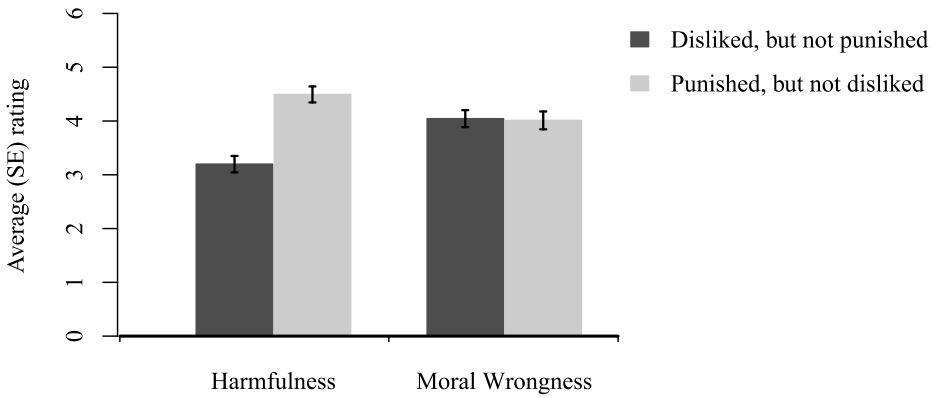


Figure 2.1. Ratings of harmfulness and moral wrongness, conflicting information condition. Error bars represent standard error of the mean.

answered that the non-punished act was more harmful, and six percent (n = 5) said the acts were about the same, $\chi^2(2, N = 91) = 142.57, p < .001$. Participants also rated the punished act (M = 5.16, SD = 1.09) as significantly more harmful than the non-punished act (M = 2.49, SD = 1.17), $t_{paired}(90) = 13.72, p < .001$.

Normative information condition. In the normative information condition, 83% (n = 73) of participants chose the disliked act as the more morally wrong. Another two percent (n = 2) responded that the act which does not cause the actor to be disliked is more morally wrong, and 15% (n = 13) said that they were about the same, $\chi^2(2, N = 88) = 99.57, p < .001$. Ratings of moral wrongness reflected a similar pattern. Participants rated the disliked act (*M* = 4.50, *SD* =

1.58) as significantly more morally wrong than the act that was not disliked (*M* = 2.40, *SD* = 1.44), $t_{paired}(87) = 10.16, p < .001$.

Again, responses about harmfulness were similar. The disliked act was chosen by 88% of participants (n = 77) as the more harmful, while 2% (n = 2) chose the other action and 10% (n = 9) said the actions were about the same in terms of harmfulness, $\chi^2(2, N = 88) = 117.02, p < .001$. Participants also rated the disliked act (*M* = 4.48, *SD* = 1.41) as significantly more harmful than the not disliked act (*M* = 2.27, *SD* = 1.25), $t_{paired}(87) = 10.16, p < .001$.

Conflicting information condition. When participants were given conflicting information about whether an act caused an actor to be disliked and whether an act was punished, 43% (n = 39) of participants chose the act that is disliked but not punished as the more wrong, and 42% (n = 38) chose the act that is punished but not disliked, while 15% (n = 14) responded that the acts were about the same in terms of moral wrongness, $\chi^2(2, N = 91) = 13.21, p = .001$. There was no significant difference in ratings of moral wrongness for the punished (but not disliked) act (*M* = 4.01, *SD* = 1.59) and the disliked (but not punished) act (*M* = 4.04, *SD* = 1.51), $t_{paired}(90) = 0.12, p = .91$.

In the same condition, however, 65% of participants (n = 59) said that the punished (but not disliked) action was more harmful than the disliked (but not punished) action. Of the rest, 19% of participants (n = 17) chose the disliked act as more harmful, and 17% (n = 15) responded that the two acts were about the same. The punished (but not disliked) action was also rated as significantly more harmful (*M* = 4.49, *SD* = 1.42) than the alternative (*M* = 3.20, *SD* = 1.45), $t_{paired}(90) = 5.75, p < .001$.

Discussion

Even in this minimal paradigm, with little context and no additional information, participants in the punishment information condition believed that a punished act was more harmful and less moral than a non-punished act. Our results also reaffirm that normative information—in the form of dislike—can act as a signal of harmfulness and moral wrongness: here too participants believed that a disliked action was more harmful and less moral than a non-disliked action. As a first step, these results confirm a necessary assumption for the current research—i.e., that people are willing to make inferences about an act based solely on information about whether it is punished. This finding is also consistent with prior research demonstrating that children will use punishment as a cue to the moral "badness" of an act (Dungan, Chakroff, & Young, 2017).

More importantly, however, when punishment information conflicted with information about what was disliked by others, participants regarded the punished action as more harmful but *not* more morally wrong than the disliked action. Thus, although punishment and dislike appear to be equally good at expressing that an action is morally wrong, punishment appears to be a better cue that an action is harmful. When asked about harm, the same participants who decline to distinguish between the wrongness of a punished act and a disliked one report that a punished (but not disliked) action is significantly more harmful than a disliked (but not punished) one. This reasoning is also robust to the type of question asked; participants made this distinction in both scaled ratings and forced choice responses. To corroborate these results, we conducted a separate replication of the conflicting information condition only; as in Study 1, participants in the replication were significantly more likely to answer that the punished but not disliked act was

the more harmful act, but they were only marginally more likely to choose the punished but not disliked act as the more morally wrong.

Although we take these findings to be evidence that punishment may contribute uniquely to judgments of harmfulness, another possibility is that punishment is simply a particularly intense variety of dislike or disapproval; when this dislike is strong enough, people assume an action is not only wrong, but also harmful. That is, punished actions are not different from disliked actions in kind, but only in degree. If this is true, then we should always find that punishment is taken as stronger evidence for the negative qualities of an action than is normative dislike. Our finding in Study 2.1 that punishment is not taken as stronger evidence of general moral wrongness casts some doubt on the simplest version of this explanation, but it is nonetheless possible that harmfulness—and not punishment—is the distinguishing factor. In other words, it could be that harmfulness judgments are especially sensitive to the degree of dislike or disapproval expressed, while moral wrongness judgments relatively insensitive, so that it is only harmfulness ratings that pick up the difference in degree between punishment and normative dislike. To test this alternative explanation, we can see whether punishment is also a stronger signal of moral concerns other than harmfulness. As we noted above, harmfulness is just one of several important psychological aspects of morality. In Study 2.2, we turn to another important aspect: disgust.

Study 2.2: Harm and Disgust

Study 2.1 suggests that participants treat both punishment and dislike by others as a cue that an action is immoral, but that when the two types of information conflict (when one action is punished and the other is disliked), punishment is taken as a particularly strong indication that the action is harmful. Participants interestingly think that both punished and disliked actions are

equally morally wrong. At first glance, these results are puzzling; if harmfulness is an important component of moral wrongness, and punishment is a strong signal of moral wrongness, such that the punished act is more harmful than the disliked act, then why isn't a punished act also seen as more morally wrong than a disliked act? The answer, of course, could be that dislike communicates one or more different moral concerns more strongly than punishment does.

To test this possibility, we sought to identify a second moral dimension on which to compare punishment and dislike. As discussed above, the moral psychology literature has largely focused on two primary moral concerns in recent years: harm and purity.(2017) Beyond this focal relevance, however, the literature provides some reasons to think that purity might be a good candidate. Though both harm and purity concerns are often moralized, researchers have demonstrated a number of striking differences between the two. Brain imaging studies suggest that concerns about harm and purity may have significantly different neural origins,(Litman et al., 2016) be influenced by different situational and social factors,(Bregant et al., 2016) lead to different emotional and behavioral reactions,(Gray, 2014; Gray & Keeney, 2015; Haidt, 2007) and ultimately lead to different inferences about the actors involved.(e.g., Moll et al., 2005)

Purity violations are often associated with feelings of disgust.(Young & Saxe, 2011) While the precise nature of the relationship between disgust and moral judgment is unclear, some speculate that moral disgust provides an incentive to reject and distance one's self from the moral offender, just as non-moral disgust prompts one to reject a potential contaminant.(Molho, Tybur, Güler, Balliet, & Hofmann, 2017; Rozin et al., 2009) Indeed, experiments have repeatedly demonstrated that people who feel disgusted will physically distance themselves from the source of the disgust.(Dungan et al., 2017) Dungan, Chakroff, and Young argue that moral purity concerns may have evolved as a way of identifying group members whose behavior does not

conform to group norms. Thus, they note, although harm-based moral violations often seem to signal that an actor is a bad person, purity-based violations seem to signal instead that a person is a bad group member (e.g., Pizarro et al., 2011).

In the current studies, our social normative information that an alien who blicks or gomps is disliked by other aliens essentially implies that aliens seek to put social distance between themselves and the offending alien. In other words, the social norm information we have provided may be signaling a moral concern more akin to disgust than to harm. Thus, in Study 2.2, we measure participant's inferences about the *disgustingness* of the underlying action, in addition to its harmfulness. We predict that we will again find that punishment will be seen as a better indication of harm than will dislike; in contrast, we also posit that dislike may be taken as better evidence than punishment that an action is disgusting.

Method

The paradigm for Study 2.2 was substantially identical to the paradigm used in Study 2.1, with changes noted below.

Participants. Participants were 125 adults (61 female), ages 18-73 (M_{age} = 34.57, *SD* = 10.74), recruited from Amazon's Mechanical Turk and paid for their participation. As in Study 2.1, the exact numbers in each condition were allowed to vary as a function of random assignment. We recruited 125 participants so that, even with this variation, each condition would have at least 40 participants.

Procedure. As before, participants were randomly assigned to one of three conditions: normative information (n = 41), punishment information (n = 43), and conflicting information (n = 41). The conditions were identical to those used in Study 2.1, such that participants in the conflicting information condition read that one of the alien acts causes the actor to be disliked

but not punished, and the other act is punished but does not cause the actor to be disliked. In contrast, participants in the normative information condition read that one act caused dislike and the other did not, whereas participants in the punishment information condition read that one act was punished and the other was not.

In Study 2.2, however, we added a new set of "disgust" measures. Participants still rated the harmfulness of each act (on a seven-point scale) and chose which was the more harmful, but then we asked participants to rate the degree to which each act was "disgusting" and to choose which act was the more disgusting (forced choice, including an option for "about the same").

Results

Punishment information condition. In the punishment information condition, participants overwhelmingly (73%, n = 30) reported that the punished act was more harmful than the non-punished act, $\chi^2(2, N = 43) = 68.98, p < .001$; of those who did not choose the punished act, 5% (n = 2) chose the non-punished act and 2% (n = 1) chose "about the same". On the scale response, participants also rated the punished act as significantly more harmful (M = 4.63, SD = 0.98) than the non-punished act (M = 1.28, SD = 1.32), $t(40) = 7.53, p < .001$.

When asked which was more disgusting, 74% of participants (n = 32) chose the punished act, 9% (n = 4) chose the non-punished act, and 16% (n = 7) chose "about the same" $\chi^2(2, N = 43) = 32.98, p < .001$. Participants also rated the punished act as significantly more disgusting on the scale measure (M = 3.53, SD = 1.65) than the non-punished act (M = 1.53, SD = 1.33), $t_{paired}(42) = 6.01, p < .001$.

Normative information condition. Results in the normative information condition were also as predicted; on the forced-choice measure, 73% (n = 30) of participants chose the disliked act as the more harmful, 12% (n = 5) chose the non-punished act, and 15% (n = 6) said they were

about the same, $\chi^2(2, N = 41) = 29.32, p < .001$. Participants also rated the disliked act as significantly more harmful (M = 3.63, SD = 1.48) than the non-punished act (M = 1.20, SD = 1.33), $t(40) = 7.53, p < .001$.

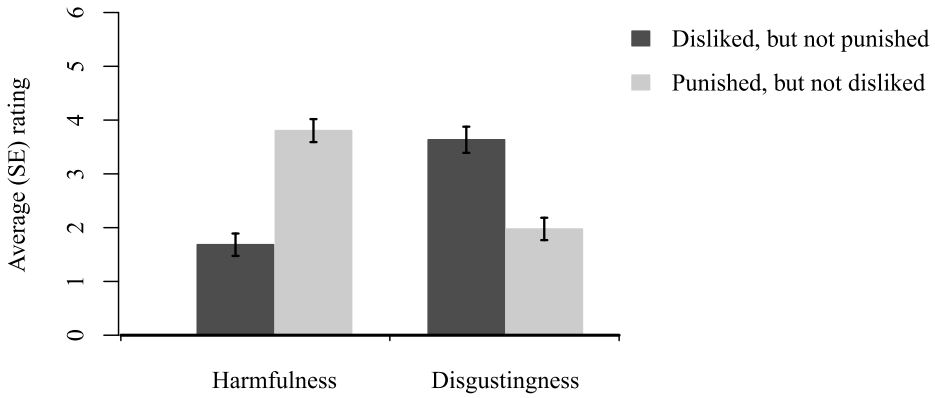


Figure 2.2. Ratings of harmfulness and moral wrongness for Study 2.2, conflicting information condition. Error bars represent standard error of the mean.

Similarly, 83% of participants (n = 34) chose the punished act as the more disgusting, seven percent (n = 3) chose the non-punished act as the more disgusting, and ten percent (n = 4) responded that the acts were about the same, $\chi^2(2, N = 43) = 45.41, p < .001$. On average, participants also rated the punished act as significantly more disgusting (M = 4.49, SD = 1.21) than the non-punished act (M = 1.37, SD = 1.26), $t_{paired}(40) = 10.38, p < .001$.

Conflicting information condition. When asked to compare an action that is punished but not disliked to an action that is disliked but not punished, 56% of the participants in this condition (n = 23) chose the punished act as the more harmful of the two, $\chi^2(2, N = 43) = 9.56, p < .01$. The remaining participants split evenly between the other two choices: 22% (n = 9) chose the disliked act as the more harmful, and 22% (n = 9) responded that they were about the same. Participants also rated the punished act as significantly more harmful (M = 3.80, SD = 1.36) than the disliked act (M = 1.68, SD = 1.33), $t_{paired}(40) = 6.16, p < .001$.

However, the results were different for the disgust measures. A majority of the participants in this condition (63%, n = 26) chose the disliked act as the more disgusting of the two, $\chi^2(2, N = 41) = 17.02, p < .001$, while 15% (n = 6) said the punished act was more disgusting and 22% (n = 9) responded that they were about the same. Participants also rated the disliked act as significantly more disgusting (M = 3.63, SD = 1.56) than the punished act (M = 1.98, SD = 1.33), $t(40) = 4.62, p < .001$.

Discussion

This study replicates several key findings from Study 2.1. First, participants were again willing and able to make inferences about an action based solely on knowing that the action was punished (in the punishment information condition) or that the action was disliked (in the normative information condition); both pieces of information again caused participants to rate the actions as more harmful than the actions that were not punished or not disliked. The same held for an action that was punished but not disliked (in the conflicting information condition); as in Study 2.1, the punished act was viewed as more harmful than the non-punished but disliked act.

The results of Study 2.2 also suggest that punishment information is particularly informative about harm, and that dislike information is particularly informative about disgust. Consistent with Study 2.1, we again found that participants thought a punished action (that is not disliked) is more harmful than a disliked action (that is not punished). Dislike, though a weaker signal of harmfulness than competing punishment information, is a stronger signal of the disgustingness of an action than is punishment information. This fact may also help to explain why the two actions in the conflicting information condition were not seen as differing in terms of moral wrongness even though the punished action was seen as more harmful; if punishment

has a relatively targeted effect on judgments of harm, and dislike has a similar effect on judgments of disgust, then the two effects may effectively cancel each other out in the broader moral judgment. Of course, neither punishment information nor normative information is necessarily limited to influencing a single moral domain; the results from the punishment information and normative information conditions show that both kinds of information are able to influence broad moral judgments in some circumstances. However, as we discussed at the outset, the theoretical landscape corroborates our argument that the harm-punishment relationship is special, and the results of Studies 2.1 and 2.2 further support this view.

Although Studies 2.1 and 2.2 have the advantage of simplicity, allowing us to inquire directly about the moral constructs we are interested in, we can draw only limited conclusions about how information about punishment may influence moral judgments in everyday life. If punishment information does indeed lead to increased inferences of harmfulness, then we should be able to see that effect outside the minimalistic alien worlds that we created for Studies 2.1 and 2.2. In Studies 2.3 and 2.4, we look for evidence of this effect in the real world, asking participants to rate realistic actions—described as being either punished or illegal but unpunished—in terms of their harmfulness. Of course, real world actions often carry with them pre-existing ideas about the morality and harmfulness of the action, as well as increased noise from social context. Nonetheless, we predicted that participants would view actions as more harmful when they were led to believe the actions were punished than when they were not.

Study 2.3: Inferences of Harm in the Real World

Studies 2.1 and 2.2 provide evidence that people will infer the harmfulness of a novel action if they learn that it is punished. These results are interesting from a psychological

perspective, but the artificial nature of the scenario used in the studies leaves open the question of whether people make these inferences in the "real world." In other words, although punishment may lead to inferences of harm (in particular) when no other information about the action or the world in which it occurred is known, we do not yet know whether this carries over into non-novel acts. Study 2.3 tests for inferences of harm on real-world actions.

Methods

In this study, we presented participants with two real-world actions, one of which we claimed was generally punished and the other we claimed was generally not punished. We then asked participants to rate the harmfulness of each act. If, as studies 2.1-2.3 suggest, people infer that a punished act is more harmful than a non-punished act, then participants' ratings of the harmfulness of each act could change, depending on the (purported) presence or absence of punishment.

Participants. Participants were 161 adults (70 female), ages 19 to 72 (M = 34.21, SD = 10.21) recruited from Amazon's Mechanical Turk and paid for their participation.

Design and Procedure. Participants were asked to evaluate the harmfulness of two ostensibly illegal acts: (1) "Bringing firewood from another part of the country into a state park" and (2) "Gambling on professional sporting events (outside a licensed casino or gambling facility)". They were told that the items were drawn from a larger pool of items that were "illegal in most places," but whose enforcement varied. In fact, we pre-tested 38 items to obtain pre-existing beliefs about the harmfulness of each action, as well as pre-existing beliefs about whether the action "should be illegal." We then selected two items that had average and modal ratings near the neutral point of the scale; i.e., these items were chosen because the pretest ratings suggested the harm they cause was ambiguous.

Participants were then randomly assigned to one of two conditions. In one condition, they were told that the Firewood action was "punished in most places", and in the other condition participants were told that the Firewood action was "not punished in most places." Each participant was given the opposite punishment information for the Gambling action; thus, each participant was told that one of the acts was generally punished and one of the acts was generally not punished. The order of the acts themselves was randomized across participants.

For each of the two acts, participants were asked to rate how harmful the act was by moving a slider along a scale marked "Not at all harmful" at one end (coded as 0) and "Extremely harmful" at the other end (coded as 100). The coded numerical value out of 100 was computed by the survey software and not displayed to participants.

Results

Across both actions, participants the punished action as more harmful than the non-punished action, $M_{\text{punished}} = 38.31$, $M_{\text{not punished}} = 28.72$, $t(320.98) = 3.05$, $p < .01$. However, participants' ratings of the individual actions varied. Participants rated the Firewood action as significantly more harmful when they were told it was punished than when they were told it was generally not punished, $M_{\text{punished}} = 45.46$, $M_{\text{not punished}} = 29.25$, $t(158.68) = 3.47$, $p < .001$. Participants did not rate the Gambling action as significantly more harmful when told that it was punished, $M_{\text{punished}} = 31.25$, $M_{\text{not punished}} = 28.35$, $t(158.98) = 0.71$, $p = .48$.

Discussion

These results, though not conclusive, suggest that information about punishment can influence participants' inferences about the harmfulness of an action in the real world. Participants rated transporting firewood as more harmful when they believed the act was punished, although that difference did not occur in the gambling action. Taken together with the

results of Studies 2.1 and 2.2, this is further evidence that punishment can convey unique information about the harmfulness of an act, both in abstract cases and in familiar actions.

However, caution is warranted in interpreting these results. Although participants rated transporting firewood as more harmful when they believed it to be punished, the lack of a difference for gambling is notable. We noted at the outset of this study that real world was likely to be noisier than the artificial alien world used in Studies 2.1 and 2.2; the null result for gambling may reflect this additional noise and complexity. Moreover, prohibitions on gambling are undoubtedly more familiar to many participants than are prohibitions on transporting firewood; prior to the study, participants may have had clearer ideas and preconceptions about gambling and its harmfulness.

Of course, it could also be that something about the firewood prohibition made it particularly susceptible to this effect. If that is the case, then our results would have very limited generalizability. A replication of the effect and a demonstration that it applies to more than just transporting firewood is necessary before making any further conclusions. In Study 2.4, we repeat this experiment using the firewood action and three other new actions. To ensure that this replication is transparent, we also preregistered the planned data collection and analyses for Study 2.4.

Study 2.4: Inferences about Punishment in the Real World

Study 2.3 provides some preliminary evidence that people judge even real-world actions as more harmful if they are punished. The design of that study, using just two real-world actions, has the advantage of simplicity, but the results are far from definitive. The effect of punishment information was consistent in direction, in that a punished act was rated as more harmful than the

same act when not punished, but the difference was only significant for transporting firewood into a state park. As noted above, there are a number of possible explanations for this discrepancy. To address these possibilities, Study 2.4 examines a broader range of real-world actions and employs a larger sample, in a pre-registered replication of Study 2.3. The purpose was two-fold: first, to replicate and confirm the effect of punishment information on harmfulness ratings for the Firewood action; and second, to better assess whether the effect is consistent across a range of actions.

Methods

Participants. Four hundred and four participants (149 female, 1 non-binary, 1 gender fluid), ages 19-77 (M = 34.87, SD = 11.18), recruited from Amazon's Mechanical Turk, participated in exchange for payment.

Procedure. All procedures and analyses for this study were preregistered on AsPredicted.org (https://aspredicted.org/g5xb3.pdf). As in Study 2.3, participants were told that they would see a series of illegal actions, some of which were punished in most places and some of which were not. In fact, each participant saw the same four actions described; participants were then told that two of the acts (assigned at random) were punished and the other two were not. Using a slider scale identical to the measure used in Study 2.3, participants then rated the harmfulness of the act and whether the action was morally wrong on a scale that was coded from 0 (Not at all harmful, Not at all morally wrong) to 100 (Extremely harmful, Extremely Morally Wrong). The four actions were: (1) taking home for personal use something your employer plans to throw away ("Employee act"); (2) carrying a switchblade knife ("Switchblade act"); (3) taking a shortcut through private property, where "no trespassing" signs are posted ("Trespass act"); and

(4) bringing firewood from another part of the country into a state park ("Firewood act"). The order of the four actions was randomized for each participant.

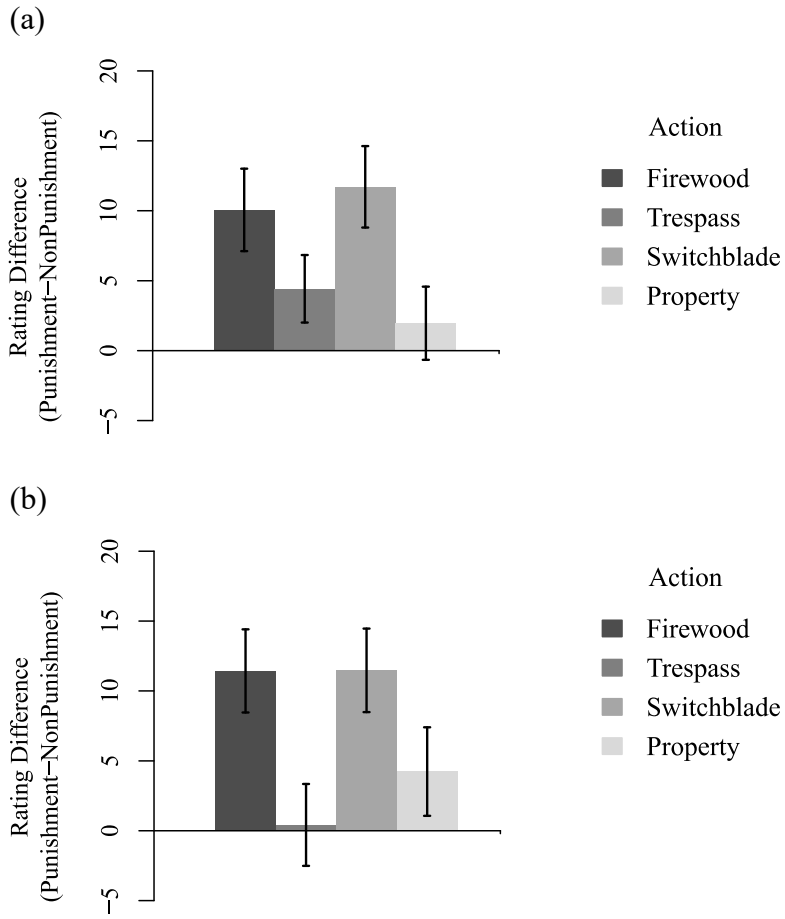


Figure 2.3. Difference between punished and non-punished averages for (a) harmfulness ratings and (b) moral wrongness. Error bars represent standard error of the mean.

Results

Overall, participants rated punished actions (M = 36.18, SD = 29.96) as significantly more harmful than non-punished actions (M = 29.14, SD = 26.64), $t(1592.2) = 4.99, p < .001$. Participants also rated the actions as more morally wrong when they believed the actions were punished (M = 40.25, SD = 31.70) than when the actions were described as not punished (M = 33.39, SD = 29.76), $t(1607.6) = 4.48, p < .001$.

For each individual action, we also conducted an ANOVA to compare the ratings of participants who read that the act was punished to the ratings of those who were not punished.

Employee act. For taking home an employer's discarded property for personal use, there was no significant difference between the harmfulness ratings of participants who were told the action was generally punished (M = 23.32, SD = 26.88) and those who were told the action was generally not punished (M = 21.36, SD = 25.62), $F(1,402) = 0.561, p = .45$. Similarly, the acts were not rated differently on moral wrongness, (M_{punished} = 34.67, SD = 32.63; $M_{\text{not-punished}}$ = 30.43, SD = 30.96), $F(1,402) = 1.79, p = .18$.

Switchblade act. Participants rated carrying a switchblade as significantly more harmful when told that doing so was generally punished (M = 46.57, SD = 31.15) than when they were told it was generally not punished (M = 34.86, SD = 27.26), $F(1,402) = 16.16, p < .001$. Participants also rated the punished version as more morally wrong (M = 40.66, SD = 32.70) than the non-punished version (M =29.18, SD = 27.02), $F(1,402) = 14.76, p < .001$.

Trespass act. Trespassing through private property was rated as marginally more harmful when it was described as punished (M = 32.73, SD = 26.09) than when it was described as not punished (M = 28.31, SD = 22.33), $F(1, 402) = 3.35, p = .07$. However, there was no significant difference between ratings of wrongfulness between the punished (M = 45.33, SD = 29.67) and non-punished (M = 44.92, SD = 29.25) versions, $F(1,402) = 0.02, p = .89$.

Firewood act. As in Study 2.4, participants in Study 2.5 rated transporting firewood across state lines as significantly more harmful when they believed such transportation was punished (M = 42.10, SD = 30.11) than when they believed it was generally not punished (M = 32.03, SD = 29.13), $F(1,402) = 11.66, p < .001$. They also rated the punished version as more

morally wrong (M = 40.41, SD = 30.99) than the non-punished version (M = 28.97, SD = 28.81), $F(1,402) = 14.76, p < .001$.

Discussion

Pooling across all of the actions we studied, we found a main effect of punishment such that the punished action is seen as more harmful and also more wrong. Indeed, for all four acts used in this study, participants rated them as directionally more harmful and more morally wrong when they were described as being punished than when they were described as not punished, even though the actions were described as being illegal in all cases. However, it was clear that the effect of punishment was stronger in some cases than others. With respect to harm, this difference was statistically significant for the switchblade act and the firewood act and marginally significant for the trespass act. In comparison, the difference in moral wrongness was significant only for firewood and the switchblade act.

Taken together with the results of Study 2.3, these data demonstrate the power of punishment to communicate information about morality, and especially about harmfulness. As in Study 2.3, when participants were led to believe that an action is punished, they rated the action as consistently more harmful, at least for a subset of the actions we tested. We replicated the effect of punishment information on harmfulness judgments for transporting firewood, but also found that the effect holds for carrying a switchblade knife and, to a lesser extent, trespassing on private property.

General Discussion

Across four studies, we find that people use information about punishment to make meaningful inferences about the punished act; in particular, our results show that punished acts are viewed as more harmful than identical actions that are not punished. Our results not only

provide strong psychological support for expressive and communicative theories of punishment, but they also add an important new component to our understanding of such theories by shedding light on the *content* of punishment's expressive message. In our studies, harm seems to be the strongest message of punishment, but it is not the only message; in the absence of other information, people also infer that a punished act is more morally wrong and more disgusting than an act that is not punished. Overall, these findings suggest that punishment can serve as an important psychological cue. In this section, we first review the key findings from our empirical studies and then discuss how those findings may inform law and policy and increase our understanding of moral and legal psychology.

In Study 2.1, learning that an act is punished leads people to infer that it is more harmful and more morally wrong than an act that is not punished, even in a minimal and artificial context. When a non-punished action also causes the actor to be disliked, however, people do not make the same distinction between the two acts in terms of moral wrongness; both the punished but not disliked and the disliked but not punished actions are rated as equally morally wrong. However, participants do infer that the punished but not disliked act is more harmful than the disliked but non-punished act.

Study 2.2 confirmed that, in the absence of other information, participants will use the fact that an act is punished as a cue to harmfulness, but it also showed that participants will use the same information to infer that a punished act is more disgusting than a non-punished act. However, Study 2.2 also showed that the special contribution of punishment information to harmfulness judgments that we observed in Study 2.1 does not carry over onto all sub-domains of morality; when punishment and dislike information conflicted in Study 2.2, the punished action was still chosen as the more harmful, but the disliked (and not punished) action was

chosen as the more disgusting. Thus, punishment information seems to lead to inferences that an action is harmful, over and above any inferences that the action is morally wrong.

In Studies 2.3 and 2.4, we extended our findings into more real-world contexts, and we found that people will make the inference that a punished action is more harmful than a non-punished (but illegal) action. Although we found evidence of this inference in only some of the cases we tested, these studies nonetheless demonstrate that the inference is not limited to the bare bones scenarios we used in Studies 2.1 and 2.2.

In both artificial and real-world contexts, punishment seems to lead people to make a number of meaningful inferences about the action that is being punished, at least when other cues are not available. In other words, punishment has informational value. This is consistent with prior research on the expressive function of law (Funk, 2007), and work finding that punishment can convey information about the victims of harm (Bilz, 2016; Bregant et al., 2016), although to our knowledge this is the first evidence that punishment also conveys nuanced information about the morality of the punished action.

Our results also provide an intriguing starting point for a broader discussion about the role of punishment in society. In law and policy, the inference that a non-punished act is somehow less harmful than a comparable punished act may have troubling consequences. When punishment varies in the real world, some crimes or victims of crimes may be perceived as more or less important, especially if the presence and absence of punishment is repeated or systematic. Here, we highlight a few areas where such inferences may be of particular interest.

Following the 2008 financial crisis, many people took a renewed interest in the prosecution of corporate malfeasance. The Securities and Exchange Commission (SEC) and Department of Justice (DOJ) investigated many allegations of criminal activity and breaches of

trust on the part of financial institutions, but the government also developed a number of somewhat unusual ways of dealing with the results of their investigations. Rather than pursuing civil or criminal suits against the (allegedly) offending institutions, the government reached agreements with them that allowed them to avoid official sanctions. Although many SEC settlements required the institutions to submit to increased federal monitoring or pay fines or both to avoid litigation, many also allowed the institutions to agree to such measures while still maintaining that they did nothing wrong (Ferrin, Kim, Cooper, & Dirks, 2007; Kaul, 2015). The so-called "neither admit nor deny" statements came under heavy fire from the public and from judges, though the SEC maintained that they encouraged fast and efficient resolutions to important cases (see, e.g., Bregant & Robbennolt, 2013; Winship & Robbennolt, 2018a). The DOJ has also created a number of ways for corporations to save face while still cooperating with government investigations and oversight; among the most notable is the deferred prosecution agreement (DPA). DPAs, like "neither admit nor deny" settlements, represent an agreement between a corporation or corporate employee and the government. The former avoids a criminal prosecution (at least temporarily), and the latter gets to set terms—often quite stringent—to which the corporation must adhere if it is to remain unprosecuted (Bregant & Robbennolt, 2013; Kaul, 2015). This procedure is quite similar, at least conceptually, to criminal prosecutions of individuals in which the defendant pleads "no contest" (*North Carolina v. Alford*, 1970, note 8 (discussing the "variety of different ways" courts have described nolo contendre pleas)).

Even if the wide use of these non-punishment strategies has allowed the government to tighten corporate oversight and more directly control corporate affairs following malfeasance, our results may suggest that the costs of these agreements could be more than previously believed. The idea that, as some have quipped, a financial institution may be "too big to jail,"

(Packin, 2014; Pontell, Black, & Geis, 2014) even when it is accused of serious wrongdoing, may resonate in the public psyche (see also Winship & Robbennolt, 2018b). When the government declines to prosecute or punish such an institution through a DPA, or when it imposes sanctions but allows those sanctions to be couched in terms that are not condemnation, through a "neither admit nor deny" settlement, the public perception of the institution's actions may change. Our results suggest, moreover, that the change in perception might be predictable: an act which is not punished is viewed as less harmful. In other words, the government's decision not to punish corporate wrongdoing could lead people to infer that the corporation's acts were less harmful than previously believed.

We can only speculate about the further implications of such an inference, but one possibility is that the blame for such acts may be relocated. After all, the reasoning could go, if the actions of the corporations that led to the financial collapse were not actually as harmful as people believed, then perhaps the "real" blame lies more on the victims of the corporate actions (e.g., "Well, they should not have taken out mortgages they could not afford"). Indeed, some prior research demonstrates that failing to punish a wrongdoer can have negative consequences for how a victim is viewed (Bilz, 2016; Bregant et al., 2016). The effects of non-punishment on the perception of victims could be further exacerbated if the failure of punishment is systematically linked to certain victims or certain crimes. The Black Lives Matter movement, for example, reflects a line of thinking that is consistent with our findings; when people perceive that violence by police officers goes unpunished, they may infer that the police action was less harmful—even if that action resulted in someone's death. Thus, as activists argue, if officers are punished less often (or appear to be punished less often) when the injured party is black, it could signal that injuring and killing black people is less harmful than injuring and killing others

(Brodin, 2016; Carter, 2017; Greene, 2015). These concerns also apply in other contexts. For example, the same logic can be applied to crimes against women, including domestic violence and sexual assault, which are often thought to be under-reported and under-punished (Anderson, 2016; Bandes, 1999; Bond & Jeffries, 2014; Steinman, 2002; Wood & Toppelberg, 2017). Such crimes may be viewed as less harmful when they are not punished, which could in turn reflect poorly on victims and lead to even lower rates of reporting and punishment.

Of course, the broader context in which a given example of punishment or non-punishment occurs will be an important factor in how it is interpreted. In our studies, we state the presence or absence of punishment as a descriptive fact, i.e., "Aliens who blick are generally not punished", "[Transporting firewood] is generally punished." In Studies 2.3 and 2.4, when we described apparently real criminal offenses, we told participants that all of the actions were illegal "in most places"; we were careful not to give any explanation for why enforcement and punishment might vary. In contrast, when a high-profile case ends in punishment or non-punishment, the reasons likely matter a great deal to people and to the inferences they make. Very different inferences might arise when the underlying action is not punished because it is simply not illegal, or because it is not reported, or because it is not proven. However, the injustice that people may feel after an instance of non-punishment may have far-reaching effects that go beyond the particular context at hand. That sense of injustice may lead to a kind of unintentional backlash; research shows that when a perceived wrongdoing goes unpunished, people's anger may lead them to act as "intuitive prosecutors," unconsciously transferring their anger and sense of injustice to future, *unrelated* transgressions (Goldberg et al., 1999).

Though not our primary focus in this project, our results do add some interesting new information to the ongoing debate in moral psychology over the centrality of harm to moral

judgments. As we alluded to above, there are debates about whether people truly find actions to be immoral in the absence of demonstrated harm (Gray & Keeney, 2015; Haidt, 2001). Despite the vast body of research showing that information about harm influences punishment judgments (Chakroff, Dungan, & Young, 2013; Chakroff, Russell, Piazza, & Young, 2017; Crockett, Kurth-Nelson, Siegel, Dayan, & Dolan, 2014; Cushman, Young, & Hauser, 2006; Gino et al., 2009; Hampton, 1992; Helwig, Zelazo, & Wilson, 2001; R. Murphy, 1988; Schulhofer, 1974; Shultz, Schleifer, & Altman, 1981; Stern, 1970; Tisak, 1993; Vaish, Carpenter, & Tomasello, 2010), to our knowledge this is the first to show the opposite is also true: information about punishment leads to increased inferences about harm. This relationship between punishment and harm adds further complexity to these ongoing debates. On one hand, this finding underscores the importance of harm judgments in moral reasoning, which may lend some support to the arguments that all moral judgments are, at their core, based on perceived harms (Gray & Keeney, 2015; Gray et al., 2014). On the other hand, these data do present something of a puzzle for an account in which condemnation and punishment are predicated on intuitions about harm. If, as our results demonstrate, this path can also be reversed, then the relationship between harm and punishment must be, at a minimum, bi-directional. Perhaps a kind of over-learning model could account for this discrepancy; such a model could posit, for example, that people rely so completely and automatically on their judgments of harmfulness to intuit on the appropriate level of punishment that they eventually come to associate harm with punishment even when punishment information comes *before* the harm judgment. But more work is necessary to determine whether this can be squared with harm-only models of morality.

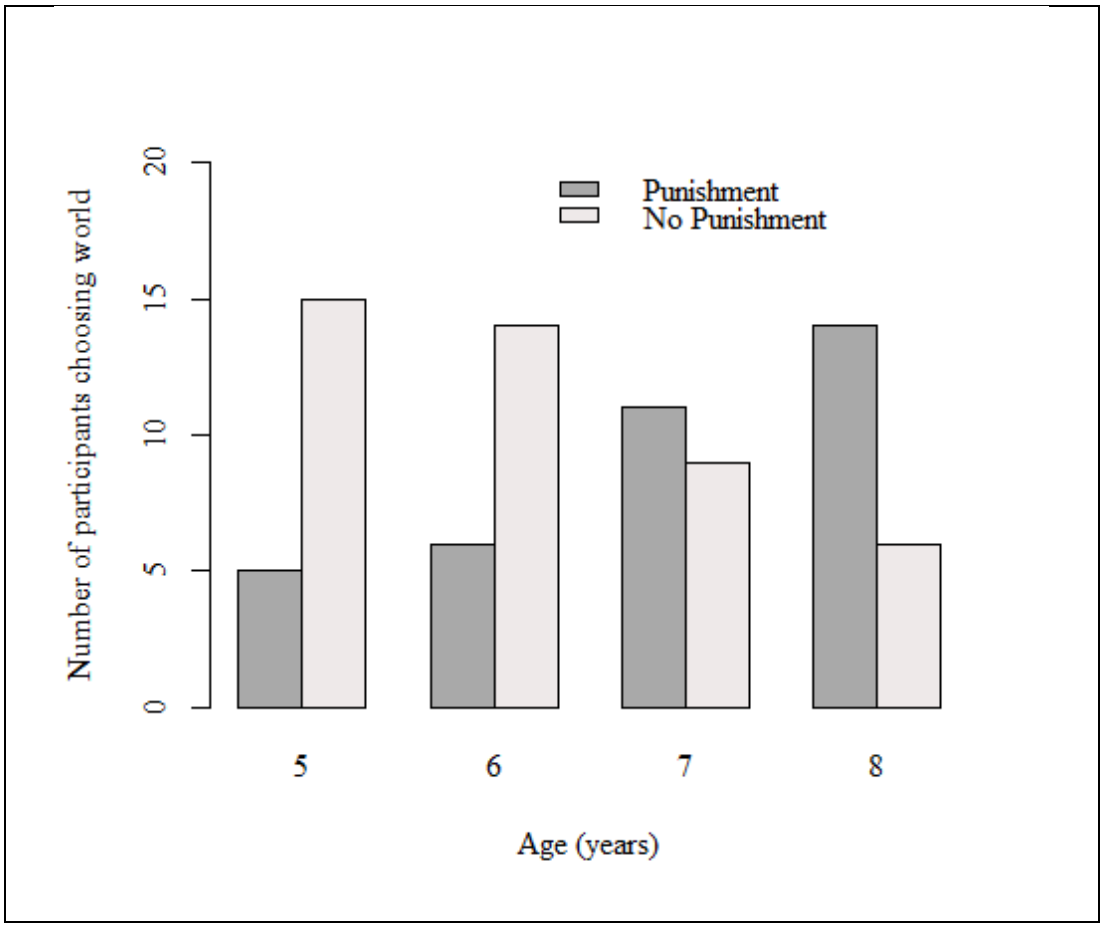
More broadly, these results also add to a growing body of research addressing the intuitive underpinnings of legal thinking (Bregant et al., 2016; Cushman, 2008; Finkel, Liss, &

Moran, 1997; Friedman, Van de Vondervoort, Defeyter, & Neary, 2013; Ginther et al., 2014; Knobe, 2009; Mikhail, 2009; Mull & Evans, 2010; Redding, 1998; Tyler & Boeckmann, 1997). We therefore join others in this field of research who seek to understand when the law aligns and misaligns with human psychology. Identifying the causes and consequences of misalignment is important for understanding how the law operates in people's lives and—where possible—addressing the mismatch. It is perhaps more important, however, in shaping how people react to the law. Although legal rules cannot (and should not) always reflect lay intuitions, research suggests that when policies and procedures make sense to people, they believe the system is more just, more legitimate, and more trustworthy (Boeckmann & Tyler, 1997; Tyler, 1987, 2005; Tyler & Rasinski, 1991). By providing a deeper understanding of people's intuitive beliefs, research like ours can help policymakers find and address the gaps that might otherwise undermine these beliefs.

As I have argued throughout this dissertation, one underutilized tool for understanding intuitive legal beliefs is developmental science. By exploring the origins and development of these beliefs in children, researchers can better understand what cognitions support and contribute to adult beliefs. In prior work, we have shown that children also make inferences about an act based on whether it is punished (Bregant et al., 2016). Moreover, children's inferences seem to be consistent with the adult inferences in Studies 2.1-2.4; children overwhelmingly responded that stealing was not "bad" when it was not punished. However, that study left open several important questions about how children understand the functions of punishment in society. Study 2.5 starts to address some of these questions.

Study 2.5(a): Children's Inferences about Punishment and Deterrence

In addition to providing evidence that children use punishment to infer that an action is "bad", our previous results provide striking evidence of children's emerging understanding of the role of punishment in the social contract (Bregant et al., 2016). When asked which world they would prefer to live in, 5- and 6-year-olds reliably chose the world without punishment, while 8-year-old children overwhelmingly chose the world with punishment. Though it is not surprising that young children might prefer a world in which they cannot be punished, it is remarkable to see such a clear shift in thinking as they age. Anecdotally, many children who chose the world without punishment gave qualitatively different explanations for their choice than those who chose the world with punishment. As one might expect, the former group tended to explain, "Because you don't get punished", while the latter tended to argue that people *should* get punished for stealing. Moreover, when the children referred to themselves in their explanations, those who chose the no punishment world often identified themselves as potential targets of punishment (e.g., "Because I won't get punished"), while those who selected the punishment world tended to identify as potential victims (e.g., "So if someone steals from me, they get punished").



NOTE: For all four age groups, $n = 20$.

Figure 2.4. World choice by age year (reproduced from Bregant et al., 2016).

As adults, it may be tempting to interpret children's explanations as evidence that they expect punishment to deter wrongdoing. Indeed, when adults were presented with the same task used in Bregant et al. (2016), they overwhelmingly (97%) chose the world with punishment, and many cited deterrence in their explanations. As one adult participant put it, "Punishment may be imperfect at times, but you need some sort of deterrent to enforce rules, laws, or social norms, etc.". And although children do seem to expect punishment to deter offenders, we found no evidence that children expect punishment to have a general deterrent effect; the children in our study did not seem to think that bystanders living in a world with punishment were any less likely to steal in the future than those living in a world without punishment. However, as we

noted in that paper, children were fairly reluctant to conclude that a non-thief character would steal in any event. Because of this reluctance, it is difficult to draw any conclusions from that study about how children think about the effect of punishment on behavior.

In Studies 2.5(a) and 2.5(b), we explore children's reasoning about punishment and deterrence. Study 2.5(a) poses the question in a very simple way: if punishment always occurs in one world and never occurs in another, in which world do more punishment-worthy events take place? We also included a question about world preference, in an attempt to replicate the finding, noted above, that children increasingly prefer a world with punishment as they get older.

Methods

Participants. Participants were 77 children (38 female), ages 4-11 (M = 7.13, SD = 2.04). They were recruited and tested at a local science museum. Two additional children's responses were removed prior to analysis due to experimenter error (one child was too young to participate; the other's age was not recorded at all).

Procedure. Children were shown a simple diagram of anthropomorphized squares and triangles. The experimenter then explained that the squares live in "Square World" and the triangles live in "Triangle World." For one of the two worlds, varied randomly between participants, children were told that the characters who lived there (i.e., squares or triangles) who "do bad things are always punished"; in the other world, the experimenter said, those who "do bad things are never punished."

To examine deterrence reasoning, children were asked "Which world do more bad things happen in?", followed by "Which world would you rather live in?" and why. Finally, children were asked a comprehension check question ("Do you know what 'punishment' means? What does it mean?").

Results

Comprehension. Five children could not explain what the word "punishment" meant to the satisfaction of the experimenter. They were three 4-year-olds, one 5-year-old, and one 7-year-old (M_{age} age of remaining participants = 7.27). Three of these participants answered "I don't know", one answered just "mad", and one explained "we have a book at home called the Parsha that talks about punishment" but would not elaborate further. The results below include all participants, and they do not meaningfully change if these children are excluded.

More bad things. A binomial logistic regression model was fitted to the data for each of the questions, using participant age (measured continuously) as the predictor and the child's response as the dependent variable. For the first question, which world has more "bad things" happen, 54% of participants chose the world without punishment. However, age was positively associated with likelihood of choosing the world without punishment, $\beta = 0.32, p = .01$. As Figure 2.5 below shows, the youngest children in this study (ages 4-5) chose the world without punishment 46% of the time, compared to 78% of the oldest children (ages 10-11).

World preference. The preference question showed the opposite pattern; there, 74% of participants across all ages chose the world without punishment. But this was largely driven by the younger children; the logistic regression showed that age was negatively associated with likelihood of choosing the world without punishment, $\beta = -0.60, p < .001$. By way of illustration, 96% of the youngest children chose the world without punishment, but just 22% of the oldest children did so. These results are also illustrated in Figure 2.5. There was a significant negative correlation between children's responses to the two questions, $r = -.36, p < .01$, such that children were likely to prefer the world where they believed fewer bad things happened.

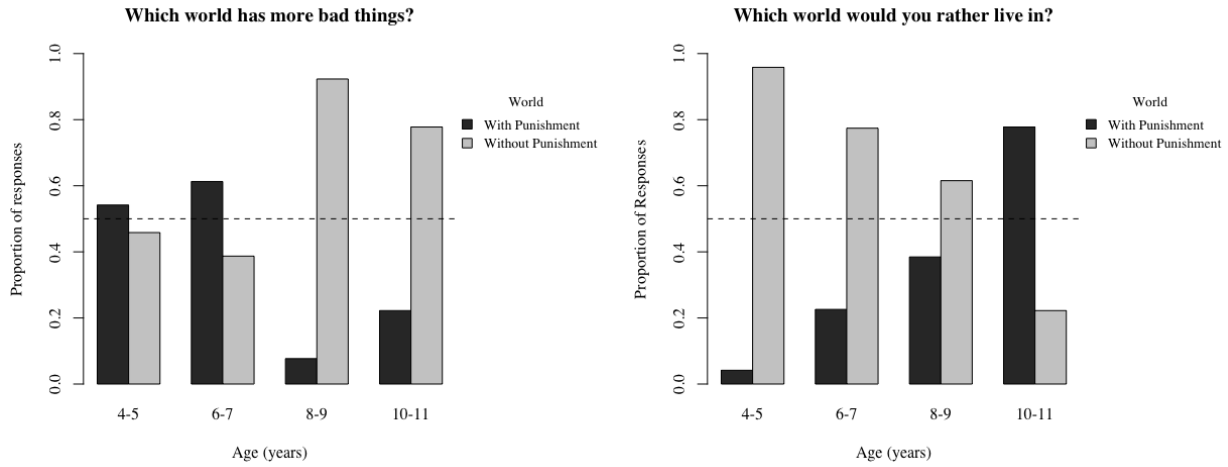


Figure 2.5. Proportion of children at various ages who chose each world when asked which has more bad things (left) and which they would prefer to live in (right).

Discussion

These data show clear and striking age-related patterns. As children get older, they increasingly believe that more bad things happen in would without punishment, and they increasingly reply that they would prefer to live in a world *with* punishment. The correlation between children's responses to these questions also suggests that the two ideas may be linked.

However, the two measures are not perfectly correlated; this is most evident in the younger children. Between the ages of 4 and 7, children show a slight tendency (58%) to choose the world with punishment as the one in which more bad things happen, but on the choice of which world to live in, they overwhelmingly (85%) choose the world without punishment.

One possible explanation for younger children's belief that more bad things happen in the world with punishment is that they view punishment itself as a "bad" thing. This would be consistent with their clear preference to live in a world without punishment, and it may also be consistent with their own limited experiences of punishment. If young children believe punishment is a "bad" thing, then their belief that more bad things happen in the world with punishment is almost tautological; in other words, if this explanation is correct, young children

should answer that the world with punishment has more "bad" things. Study 2.5(b) modifies the procedure to rule out this explanation.

Study 2.5(b): Children's Inferences about Punishment Policy and Action Frequency

To ensure that young children's responses in Study 2.5(a) were not driven by the assumption that punishment is a "bad" thing, Study 2.5(b) specifies particular actions whose punishment status differs (punching and kicking). Then, rather than asking where more "bad things" happen, we can ask specifically about the frequency of punching and kicking. We also changed the set up for children so that only one group of characters was at issue. This change means that we cannot ask which world children would prefer to live in, but it may also simplify the cognitive difficulty of the task for younger children by reducing the comparison from one between groups of people to one between two different acts.

Method

Participants. Eighty-six children (49 female) ages 3-11 (M = 7.17, SD = 2.61) participated at a local science museum.

Procedure. Participants were shown just the triangles from the picture used in Study 2.5(a). Pointing to the triangles, the experimenter introduced the child to the characters as triangles from "Triangle World." The experimenter then explained: "In Triangle World, triangles are [always/never] punished for punching other triangles but are [never/always] punished for kicking other triangles." The bracketed words were varied randomly between participants, such that one action (kicking or punching) was described as always punished and the other was described as never punished. Children were then asked, "What happens more in Triangle World, punching other triangles or kicking other triangles?"

Results

Across all ages, 57% of participants responded that the never punished action happens more in Triangle World. To look at the effect of age on responses, a binomial logistic regression was fitted to the data, using participant age as a continuous predictor. There was a significant positive relationship between participant age and likelihood of choosing the never punished action as the more frequent, $\beta = 0.25, p < .01$. As Figure 2.6 illustrates, this effect amounted to difference between 45% of the youngest children (ages 3-5) choosing the never punished action, compared to 84% of the oldest children (ages 10-11).

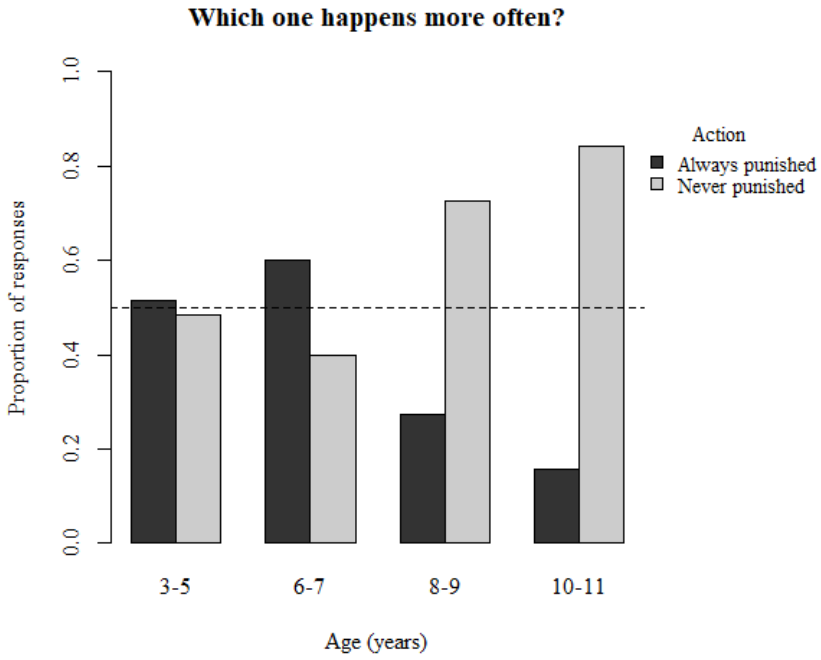


Figure 2.6. Proportion of children choosing each world across various age groups (ages separated for ease of visualization).

Discussion

Study 2.5(b) rules out a relatively uninteresting explanation for the results of Study 2.5(a); this study replicates our prior results in the context of specific, familiar actions—punching and kicking—and it does not rely on children to interpret what a "bad thing" means.

Nonetheless, the results show that as children age, they have an increasing association between a lack of punishment and the frequency of an act. These results suggest that young children to not appear to understand the punishment may act as a deterrent for specific actions. By the time they are 8 or 9 years old, the children in this study show some evidence of being intuitive deterrence theorists, believing that punished actions happen less often than non-punished actions.

In contrast, younger children still seem unsure. Having ruled out a literal belief that punishment is a "bad thing" as the explanation for our previous results, we are left with the question: why do younger children make less clear predictions about where bad things happen? One possibility is that they simply struggle with the cognitive difficulty of the question. Deterrence-based reasoning—that is, concluding that more bad things happen in the world where no one is punished for doing bad things—requires a number of fairly sophisticated leaps in reasoning about the motivations and behaviors of others.

On the other hand, young children may be reasoning in a completely different manner than their older counterparts. If punishment occurs in one world and not in the other, or for one action but not another, children may assume that there is a greater need for punishment in the first instance; that is, *because* more bad things happen there, those who do bad things are always punished. This causal explanation flips adult deterrence logic on its head, but it may be consistent with a kind of "just world" belief (see, e.g., Lerner, 1980); in a fair and just world, children might reason, people would neither harm one another nor be punished. The results of Studies 2.5(a) and 2.5(b) do not allow us to draw strong conclusions about the causal direction of reasoning for children at any age, but future research could examine the issue by introducing a change in punishment policy over time. Such a design would force participants to reason explicitly about how adding or removing punishment might affect behavior.

3 Avoiding Punishment by Asking for Permission

A popular adage, attributed to Rear Admiral Grace Hopper of the U.S. Navy, suggests that "It is often easier to ask for forgiveness than to ask for permission." If this is true, it may be very difficult indeed to ask for permission; while very little psychological literature examines permission, a wealth of research documents the difficulty of obtaining forgiveness (for examples, see e.g., Banerjee, Bennett, & Luke, 2010; Dhami, 2011; Enright & Kittle, 1999; Exline, Worthington Jr., Hill, & McCullough, 2003; Ho & Liu, 2011; McCullough, 2001; J. G. Murphy & Hampton, 1988; Witvliet et al., 2008). This research suggests that forgiveness may be hindered by any number of factors, including feelings of resentment or anger (Hanna, 2008; Witvliet et al., 2008), inadequate or insincere apologies (Banerjee et al., 2010; Kim, Ferrin, Cooper, & Dirks, 2004; Scher & Darley, 1997), or reluctance to give up a claim against the forgiveness-seeker (J. Cohen, 1999; Robbennolt, 2006, 2010).

In contrast, however, almost no empirical work has examined the process of obtaining permission. In part, this may be because of a tendency to think of permission and forgiveness as essentially the same process, separated only by temporal perspective; forgiveness is retrospective, while permission is prospective, but both operate to excuse a transgression. On this view, it may actually be easier to get forgiveness, because research on temporal perspectives suggest that moral transgressions seem less severe in retrospect than they do in prospect (Caruso, 2010). However, some philosophers and, separately, some legal scholars have sought to distinguish between permission and forgiveness on more than just a temporal basis. For example, philosopher Nicolas Cornell recently argued that "preemptive forgiveness" need not entail permission at all (2017). Legal concepts of permission, including consent and license, are

sometimes similarly distinguished from forgiveness (Abel, Becker, & Cunningham-Rathner, 1984; Harman, 1983; Wetzel, 2006).

Intuitively, permission seeking does seem qualitatively different from forgiveness seeking in at least one key respect: intent. Asking for forgiveness for a transgression provides little insight into how intentional the transgression was—a regretful apology, which may be a part of such forgiveness seeking (but which need not be!) may suggest a lack of intentionality, at least with respect to the magnitude of the consequences, but seeking forgiveness itself does not even require that the forgiveness-seeker feel remorse. In contrast, however, asking for permission in advance of an action that might otherwise be a transgression is a clear signal of one's intent to complete the action. The role of intentions in moral judgments is well-documented (e.g., Barrett et al., 2016; Cushman, 2015; Hamlin, 2013a; Young & Saxe, 2011). Asking for permission may therefore be somewhat damning to the permission-seeker, who might prefer to maintain a kind of "plausible deniability" about her intentions. Indeed, research suggests that people do prefer to keep information about their intentions as ambiguous as possible when they are harming another person, such as in a taking game (see, e.g., Cushman et al., 2008; DeScioli, Bruening, & Kurzban, 2011; DeScioli, Christner, & Kurzban, 2011).

In spite of the many reasons to think that asking permission might be a suboptimal strategy for social harmony, at least when compared to just asking for forgiveness after the fact, people continue to ask one another for permission. Why? What is gained by asking for permission that might not be gained by asking for forgiveness? And, perhaps more interestingly, why do people grant permission to others at all? To answer these questions, we created an economic game that can be used to study the many features of asking for and granting permission.

The basic permission game consists of two rounds: the Permission Round and the Taking Round. Before the first round, the participants receive an endowment; in Study 3.1, the taker received $1, and the permitter received $3. The players are told that, later in the game, the taker will have the chance to take money ($1 in Study 3.1) from the permitter. If the taker chooses to take, the money taken will be tripled; thus, in Study 3.1 the permitter would lose $1 but the taker would receive $3. Before any taking can occur, though, there is a Permission Round.

During the Permission Round, the taker is given the chance to ask the permitter for permission to take. If the taker chooses to ask for permission, the permitter is asked to respond by granting or denying permission to take. If the taker chooses not to ask for permission, then the permitter is told that the taker had chosen not to ask. Regardless of the players' decisions in the Permission Round, no money changes hands; the players' decisions in the Permission Round are not binding. In the Taking Round, the taker is given the option to take or not take. The permitter has no actions to complete in the Taking Round. Figure 3.1, which was shown to participants in Study 3.1, summarizes the decision points in the game.

The permission game is therefore fairly straightforward conceptually: one player is given an endowment, and the other player has a chance to ask for permission to take some of the endowment for herself. However, it has many interesting and complex moving parts that can be adjusted, allowing for a nearly infinite range of possible outcomes to explore. Studies 3.1 and 3.2 below demonstrate the kind of information that can be gathered using the permission game, and they shed some light on a basic question of frequency: if given the chance, how often do players ask for and grant permission? These studies also illustrate one key feature of the permission game that can be easily altered—whether or not the player who is asked for permission has an

opportunity to enact punishment once the deed is done. Study 3.3 moves from vignette studies to the lab and puts the game into action, featuring live pairs of participants. Finally, in Study 3.4, we compare the permission game to several similar paradigms.

Study 3.1: Permission and Taking

Methods

Participants. Ninety-seven people (45 male, M_{age}=30.75 years) participated as walk-in volunteers in two lab spaces. Participants were 18-72 years old (mean age = 30.75 years, median age = 26), and they were given a guaranteed cash payment ($2.00) in exchange for their participation; they were also allowed to keep whatever money they earned during the study, as explained below.

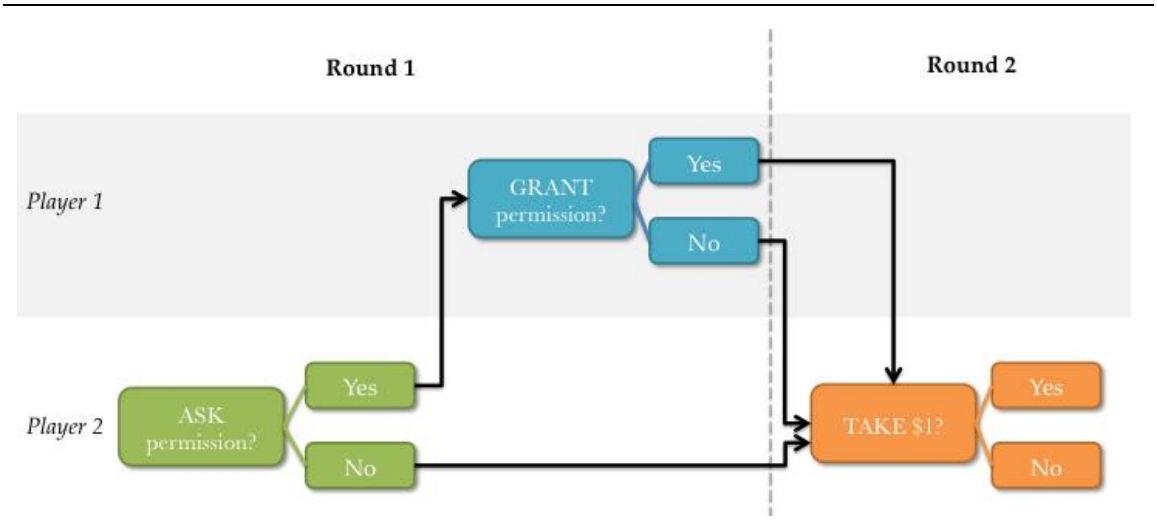


Figure 3.1. Diagram of the Permission Game provided to participants. NOTE: In this diagram, Player 1 is the Permitter, and Player 2 is the Taker.

Procedure. Upon arrival, participants were told that they would be participating in a "two-player game study" with another volunteer. Before beginning, participants were given full written instructions for all phases of the study, and they were given the opportunity to ask

questions. The experimenter then randomly assigned the participant to either the "taker" or the "permitter" role.

The permission game. The game consists of two rounds: a Permission Round, and a the Taking Round. Before the first round, the participants receive an endowment: the taker receives $1, and the permitter receives $3. The players are told that, later in the game, the taker will have the chance to take $1 from the permitter. If the taker chooses to take $1, that money will be tripled, so that the permitter will lose $1 but the taker will receive $3. Before any taking could occur, though, there is a Permission Round.

During the Permission Round, the taker is given the chance to ask the permitter for permission to take $1 from the permitter. If the taker chooses to ask for permission, the permitter is asked to respond by granting or denying permission to take. If the taker chooses not to ask for permission, then the permitter is told that the taker had chosen not to ask.

Regardless of the players' decisions in the Permission Round, no money changes hands; the players' decisions in the Permission Round are not binding. In the Taking Round, the taker is given the option to take or not take. The permitter has no actions to complete in the Taking Round.

Details of this study. In this study, each participant actually played alone, and an experimenter responded on behalf of the fictional second player. The responses for the other "player" were randomly generated by the experimenter from the available options, with a few exceptions. First, when the participant was in the role of the permitter, the "decision" of the taker in the Taking Round was assigned randomly, but rather than a 1/2 chance of asking, the experimenter had a 2/3 chance of asking permission. This was done to ensure that sufficient data could be collected on the permitters' responses to being asked for permission.

Similarly, in cases where the participant was in the permitter role and the experimenter randomly selected to not ask for permission in the Permission Round, the experimenter always chose to take in the Taking Round. Thus, no participant was a permitter in a round in which the taker did not ask and did not take. In addition, no participant was supposed to be a permitter in a round in which the taker asked, received permission, and did not take. However, due to experimenter error, one participant was asked for permission, granted it, and was assigned to the "no take" outcome.

While the participant waited for his or her "opponent" to respond, they were given a word search task to complete. After completing the Taking Round, participants were asked to agree or disagree with a series of questions about the game: how satisfied they were with the way the game went, how unhappy they were with the other player, how nice the other player was, how fair the game was, and how much they would like to punish the other player if they were given the chance (they were not given the chance, but see Study 3.2). Finally, participants were thoroughly debriefed.[1] A full copy of the instructions given to participants in this study is included as Appendix A.

Results

Forty-nine participants were randomly assigned to be takers, and forty-eight were assigned to be permitters.

---

[1] The last 73 participants were all probed for suspicion as well; 16 (22%) of those asked indicated they had a suspicion that they were not playing with a real person.

Taker Decisions. In the Permission Round, 35 takers (71.4%) chose to ask for permission. Of those who asked, 17 (48.6%) were told that the permitter had denied permission, and 18 (51.4%) were told that the permitter had granted permission.

In the Taking Round, 43 takers (87.6%) chose to take, and 6 (12.2%) chose not to take. A chi-square test for equality of proportions showed that takers were equally likely to take in the Taking Round regardless of whether they had chosen to ask for permission in the Permission Round, $\chi^2(1, N = 49) < 0.001, p = 1.00$. For those who chose to ask, takers were also equally likely to take regardless of whether they had been granted permission or denied permission, $\chi^2(1, N = 49) < 0.001, p = 1.00$. Table 3.1 shows the distribution of taker decisions in both rounds, with those who asked for permission split out by whether permission was granted or denied.

	Did Not Ask	**Asked Permission**	
		Permission Granted	Permission Denied
Take	12	16	15
No Take	2	2	2

Table 3.1. Taker decisions in Study 3.1.

Permitter Decisions. In the Permission Round, 29 permitters (60.4%) were randomly assigned (via a random number generator, weighted for a 2/3 chance of asking) to be asked for permission; the remaining 19 (39.6%) were not asked for permission.

Of those who were asked for permission, 18 (62.1%) chose to grant permission. Eleven (38.0%) permitters chose to deny permission. For those who chose to grant permission, 17 (94.4%) were told that the taker had chosen to take; one (5.6%) was told that the taker decided not to take. For the 11 who denied permission, 5 (45.5%) were told that the taker did not take in

the Taking Round, and the other 6 (54.5%) were told that the taker did take. For the participants who were not asked for permission, all 19 were told that the taker chose to take.

	Not Asked	**Asked Permission**	
		Granted Permission	Denied Permission
Taken From	19	17	6
Not Taken From	0	1	5

Table 3.2. Distribution of Permitters by outcome in Study 3.1.

Reactions. Participant reactions were converted to numerical ratings from 0 to 6, where greater numbers indicate more agreement. Mean results are displayed in the figures below.

Taker satisfaction. In a linear model with Permission Round outcome (permission granted, permission denied, permission not asked) and Taking Round outcome (take, no take), there was a main effect of taking on taker satisfaction, $F(1,43) = 6.13, p = .02$, and a marginal effect of permission, $F(2,43) = 2.89, p = .07$, but no interaction between the two, $F(2,43) = 0.26, p = .77$. Follow-up comparisons suggest that the marginal effect of permission is driven primarily by the difference between not asking for permission ($M = 4.43, SD = 1.44$) and being denied permission ($M = 5.36, SD = 1.08$), as shown in Figure 3.2. Given the very small number of takers who chose the "no take" option, it is hard to read too much into their ratings in those cells. Interestingly, note that takers who did not ask and did not take (the darker bar in the right set, $M = 4.50, SD = 2.12$) are just as satisfied as those who did ask, got permission, and took (the lighter bar in the center set, $M = 4.75, SD = 1.13$), even though the latter group ended the game with substantially more money ($4, compared to $1) and at a comparative advantage over the permitter (which the former group did not have).

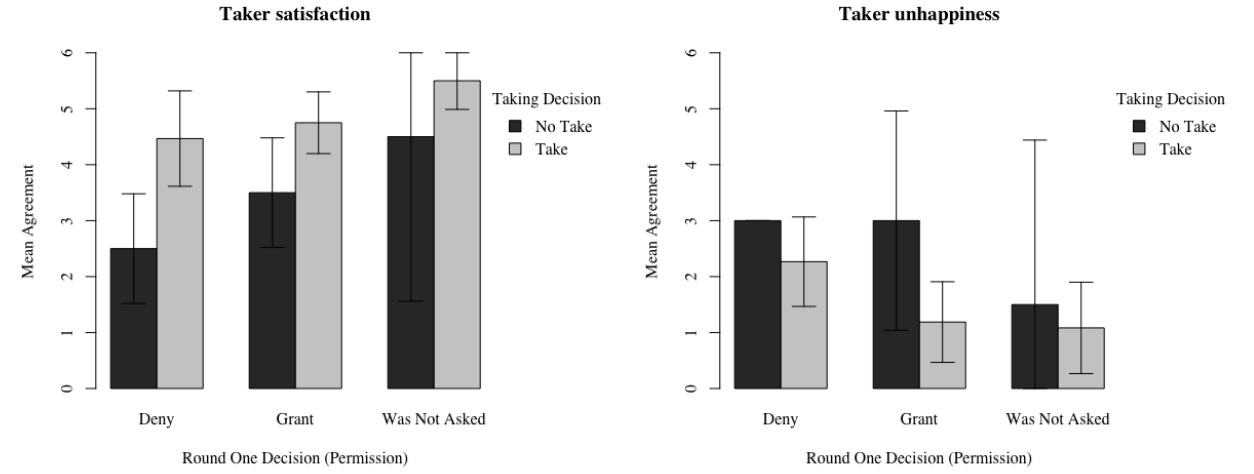


Figure 3.2. Taker's view of the outcome in Study 3.1.

Taker unhappiness. Ratings of unhappiness were largely unaffected by outcomes. In the ANOVA model, takers who took (M = 1.53, SD = 1.56) were slightly, but not significantly, more unhappy than those who did not take (M = 2.50, SD = 1.38), $F(1,43) = 2.31, p = .14$. There was also a marginal main effect of permission round outcome, $F(2,43) = 2.84, p = .06$, which also seems to have been driven by the relatively low unhappiness among those who did not ask for permission. There was no significant interaction between permission and taking on taker unhappiness, $F(2,43) = 0.42, p = .66$.

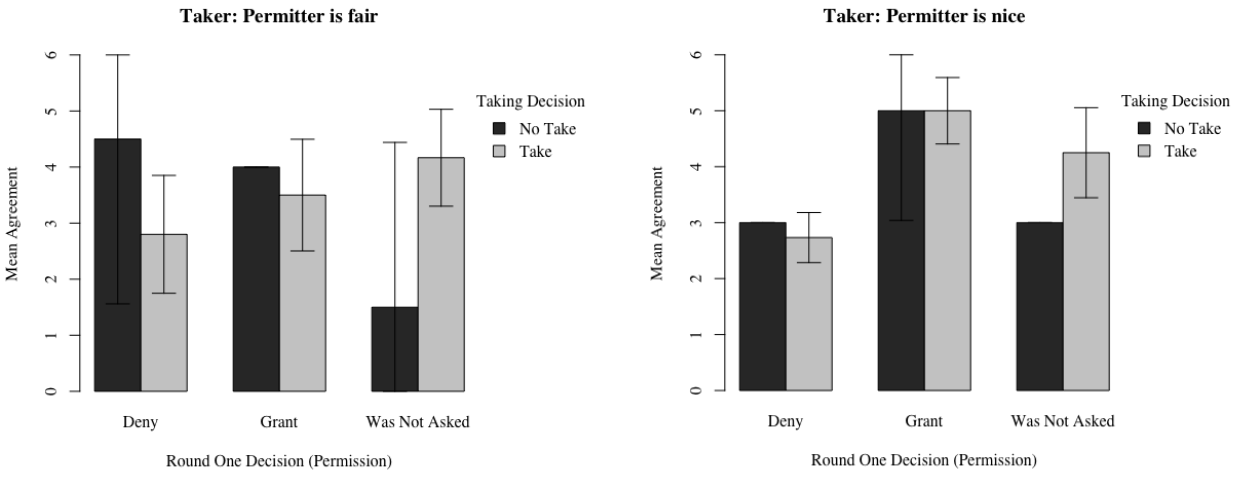


Figure 3.3. Takers' views of the Permitters in Study 3.1.

Taker ratings of permitter fairness. Neither the permission round outcome nor the taking round outcome had a significant effect on ratings of fairness (permission round: $F(2,43) = 0.71, p = .50$; taking: $F(1,43) = 0.02, p = .88$). Directionally, permitters were rated as more fair when they granted permission (M = 3.56, SD = 1.92) than when they denied permission (M = 3.00, SD = 2.09)—and fairer still when they were not asked at all (M = 3.79, SD = 1.81), but those differences were not significant (all ps > .40).

Taker ratings of permitter niceness. Permission outcome had a dramatic effect on takers' ratings of permitter niceness, $F(2, 43) = 16.52, p < .001$. Post hoc comparisons showed that permitters who granted permission were rated as significantly nicer (M = 5.00, SD = 1.19) than those who denied permission (M = 2.76, SD = 0.83, p < .001) and marginally nicer than those who were not asked (M = 4.07, SD = 1.38, p = .07). Those who denied permission were also rated as significantly less nice than those who were not asked, p = .01. Taking decisions had no significant effect on niceness ratings, $F(1, 43) = 0.40, p = .53$, and there was no significant interaction between permission and taking outcomes, $F(2, 43) = 0.85, p = .43$.

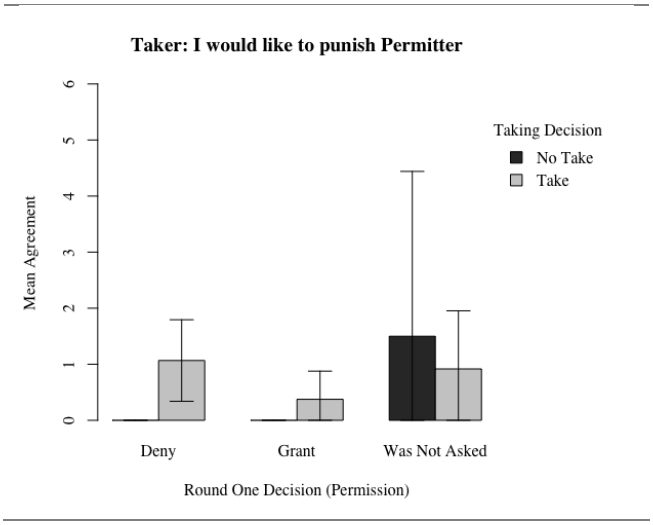


Figure 3.4. Takers' desire to punish the permitter in Study 3.1.

Taker desire to punish permitters. Overall, and somewhat unsurprisingly given the fairness and niceness ratings above, takers reported very little desire to punish the permitter (M = 0.73, SD = 1.40). Neither the permission round outcome, $F(2, 43) = 1.15, p = .33$, nor the taking round outcome, $F(1, 43) = 0.23, p = .64$, significantly predicted takers' desire to punish, and there was no effect of interaction, $F(2, 43) = 0.60, p = .56$.

Permitter satisfaction. In a linear model using round one outcome (i.e., grant permission, deny permission, or not asked for permission) and taking outcome as predictors, only round one significantly predicted satisfaction, $F(2,43) = 5.31, p < .01$. A post hoc model using permission seeking (ask vs. no ask) and taking outcome (take vs. no take) as predictors showed that permitters were more satisfied when they were asked for permission (M = 3.90, SD = 1.65) than when they were not asked (M = 2.32, SD = 1.95), $p < .01$.

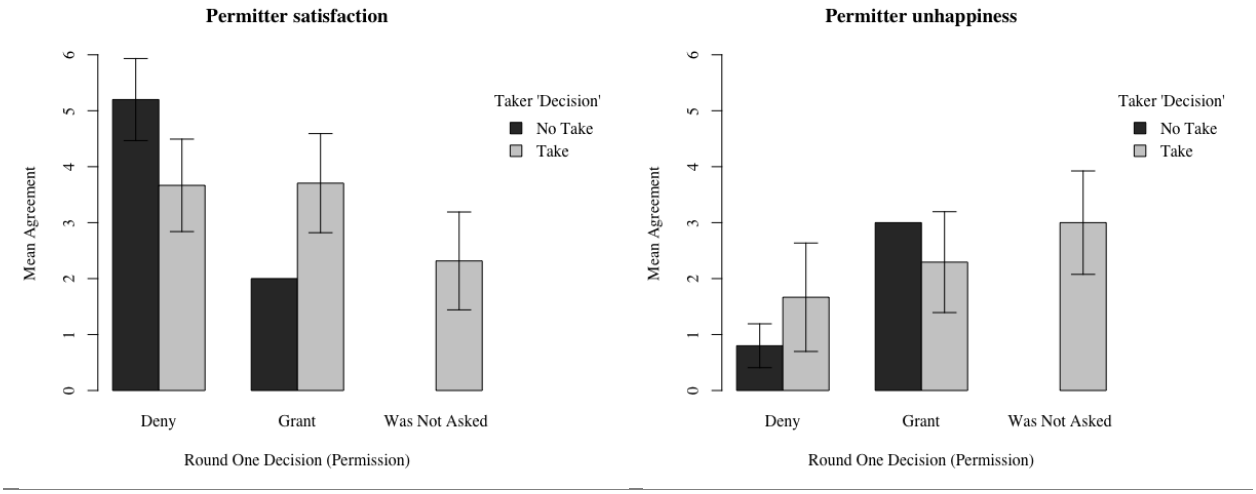


Figure 3.5. Permitters' reactions to outcome in Study 3.1. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from.

Permitter unhappiness. In the linear model, there was a significant effect of permission round outcome on permitter unhappiness, $F(2,43) = 3.16, p = .05$. Participants who were not asked for permission (M = 3.00, SD = 2.05) were significantly more unhappy than those who

were asked and chose to deny permission (M = 1.27, SD = 1.01), $p = .04$. There were no significant differences in unhappiness between permitters who chose to deny permission and those who chose to grant permission (M = 2.33, SD = 1.85), p = .29, or between those who chose to grant and those who were not asked for permission, p = .51. There was no significant effect of taking on permitter unhappiness, $F(1,43) = 0.24, p = .63$, and no significant interaction, $F(1,43) = 0.53, p = .47$.

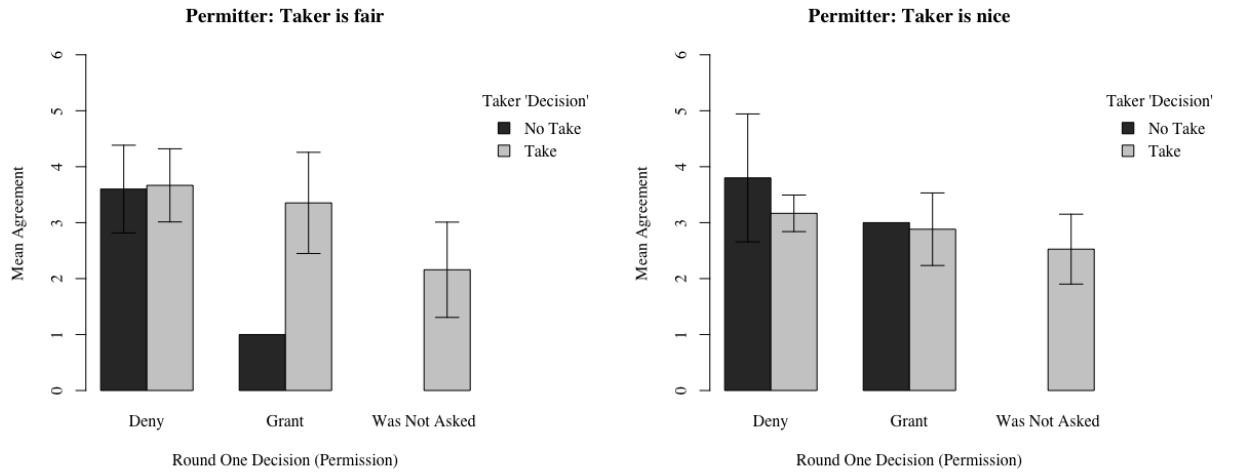


Figure 3.6. Permitters' evaluations of takers. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from.

Permitter ratings of taker fairness. As shown in Figure 3.6, permission round outcome had a marginally significant effect on permitters' ratings of how fair the taker was, $F(2,43) = 3.05, p = .06$. Post hoc comparisons showed those who were not asked for permission (M = 2.16, SD = 1.89) rated the taker as marginally less fair than those who chose to deny permission (M = 3.64, SD = 0.81), p = .07. There was no difference between those who chose to deny permission and those who chose to grant permission (M = 3.22, SD = 1.93), p = .81, or between those who chose to grant permission and those who were not asked (M = 2.16, SD = 1.89), p =

.16. There was also no significant main effect of taking on fairness ratings, $F(1,43) = 0.52, p = .47$, and no significant interaction, $F(1,43) = 1.22, p = .27$.

Permitter ratings of taker niceness. In the ANOVA model, neither asking for permission, $F(2,43) = 1.79, p = .18$, nor taking, $F(1,43) = 0.55, p = .46$, significantly changed permitter ratings of the taker's niceness, and there was no significant interaction, $F(1,43) = 0.11, p = .74$. Across all outcomes, permitters' average rating of taker niceness was 2.88 (*SD* = 1.30).

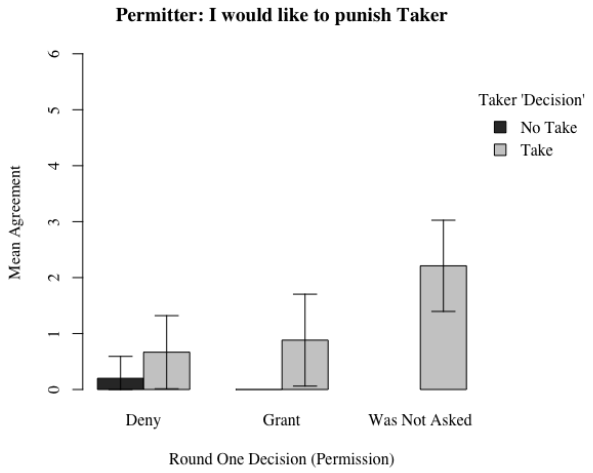


Figure 3.7. Permitters' desire to punish takers. Error bars represent standard error of the mean; only one participant who granted permission was not "taken" from, and no participant who was not asked for permission was not taken from.

Permitter desire to punish. The model fitted to permitters' desire to punish showed a significant main effect of permission round outcome, $F(2,43) = 5.33, p < .01$, but no main effect of taking, $F(1,43) = 0.47, p = .50$, and no significant interaction, $F(1,43) = 0.05, p = .83$. Post hoc comparisons showed that participants who were not asked for permission were significantly more likely to desire punishment (M = 2.21, SD = 1.81) than were those who chose to deny permission (M = 0.45, SD = 0.69), p = .02, or those who chose to grant permission (M = 0.83, SD = 1.69), p = .03. There was no difference in desire to punish between those who chose to grant permission and those who chose to deny it, p = .81.

Discussion

Overall, these results are promising for the permission game, and they demonstrate how prevalent the processes of asking for and granting permission are. The majority of takers chose to ask for permission, even though they did not need permission to take. Additionally, a majority of the permitters who were asked for permission chose to grant it. Interestingly, the takers overwhelmingly chose to take, regardless of whether they had permission or not. This suggests that, although the takers wanted to ask for permission, they were not overly concerned with whether or not they got it. Of course, unlike many situations in real life, here there was no threat that the permitter—a stranger—would impose any costs on the taker for taking without permission. In Study 3.2, we revise the game to change that.

Study 3.2: Permission and Taking Decisions

One reason to ask for permission before imposing on someone is to avoid the threat of punishment. To maximize the usefulness of the permission game, we need to implement the threat of punishment within the confines of the game. To do so, we added a punishment round to the game, in which the permitter has the opportunity to enact costly punishment.

Methods

Participants. Participants were 125 (57 females, 3 non-binary) adults who came in to one of two lab spaces in Chicago to participate in studies. They were again paid a guaranteed rate for their participation ($2) and told that they could take home whatever they earned during the study.

Procedure. As before, participants were randomly assigned to either the permitter or the taker role, and all participants were given full information about the entire game before they

began. The procedure was nearly identical to that used in Study 3.1, except that a third round was added to the game. The first two rounds of the permission game were described and conducted in the same way as in Study 3.1. Round three was called the "Response Round." In the response round, the permitter was given an opportunity to destroy all of the taker's remaining endowment. Regardless of whether the Taker had chosen to "Take" or "Not Take", the permitter could opt to punish in round three. Because the permitters in Study 3.1 reported relatively little desire to punish the takers, I anticipated that punishment might be rare in Study 3.2; to avoid artificially depressing rates of punishment even more, the punishment option here was costless for permitters.

As in Study 3.1, participants actually played the game alone, with an experimenter randomly choosing among the possible outcomes (using a random number generator) where applicable. When the experimenter was acting as the taker (i.e., the participant was randomly assigned to the role of permitter), the choice of whether or not to ask for permission was made with 50/50 odds (i.e., equivalent to a coin toss). When "not ask" was selected in round one, the experimenter always chose the "take" option in round two, to avoid putting participants in a situation where they were neither asked nor taken from. Similarly, when "ask" was selected in round one and the participant permitter chose to "grant" permission, the experimenter always chose to take in round two. When "ask" was selected in round one and the participant permitter chose to deny permission, a 50/50 choice was made so that taking occurred in roughly half of these cases (see results below).

When the experimenter acted as the permitter (i.e., the participant was in the taker role), participants who asked for permission were granted or denied based on 50/50 odds. Participants then decided whether or not to take as usual, and the experimenter then determined whether or

not they should be punished. Based on a short piloting phase, during which time all participants were assigned to the permitter role, the odds of punishment were set at 20% (roughly equal to the rates of punishment among the real pilot participants).

As in Study 3.1, participants were then asked to answer some brief questions by indicating whether the agreed or disagreed with the following statements (on a 7-point Likert scale): "I am satisfied with the way this game went", "I am unhappy with the other player", "The other player is nice", and "The outcome of this game was fair". Permitters who were not asked for permission were also asked to evaluate one additional statement: "If the Taker had asked me for permission, I would have said yes."

Results

Permitter Decisions. Sixty-nine participants (55.2%) were randomly assigned to the role of permitter. Based on random assignment (see methods, above), 36 (52.2%) of permitters were (ostensibly) asked for permission. Of those who were asked for permission, 24 (66.7%) chose to grant it. As shown in Table 3.3 below, 5 (41.7%) of the permitters who denied permission in round one were ultimately taken from anyway.

	Not Asked	**Asked Permission**	
		Granted Permission	Denied Permission
Taken From	33	24	5
Not Taken From	0	0	7
Total	**33**	**24**	**12**

Table 3.3. Distribution of permitters by outcome in Study 3.2.

In round three, 60 (87.0%) permitters chose to do nothing—i.e., not to punish. The remaining 9 (13.0%) opted to punish. Of those 9, 5 (55.6%) had not been asked for permission, 3 (33.3%) had granted permission, and 1 (11.1%) had denied permission. All of them had been taken from.

Taker Decisions. Fifty-six participants (44.8%) acted as takers in this study. In round one, an overwhelming majority (80.4%, n = 45) chose to ask for permission. Roughly half of those who asked for permission were told that the permitter granted permission (53.3%, n = 24).

In round two, 45 participants (80.4%) chose to take from the permitter. The proportion of those who took was lowest for those who did not ask for permission (63.64%, n = 7) and highest for those who asked for and were granted permission (95.83%, n = 23), while those who asked for and were denied permission fell in between (71.43%, n = 15), as shown in Table 3.4 below. A logistic regression model fitted to taking decision, using permission round outcome as a predictor showed that takers were significantly more likely to take if they were granted permission than if they were not, $\beta = 2.22, z(55) = 1.96, p = .05$. They were not significantly less likely to take if they did not ask for permission than if they were denied permission, $\beta = -0.36, z(55) = -0.45, p = .65$.

	Did Not Ask	Asked Permission	
		Permission Granted	Permission Denied
Take	7	23	15
No Take	4	1	6

Table 3.4. Taker decisions in Study 3.2.

Permitter Evaluations. As in Study 3.1, mean agreement ratings were calculated after participant responses were converted into a numerical value ranging from 0 (strongly disagree) to 6 (strongly agree). Each set of responses was then analyzed using a MANOVA model, with the outcome of each round used as an independent variable, so that the overall model was a 2 (round one: permission denied, permission granted, permission not sought) x 2 (round two: take, no take) x 2 (round three: punishment, no punishment). Because all permitters who were not asked permission or who chose to grant permission were taken from, two cells of the model are empty.

As such, the only interaction included in the model is the interaction between rounds one and three—i.e., between permission and punishment. Additional planned and post hoc comparisons were performed as noted.

Permitter satisfaction. In the overall model, the only statistically significant predictor of permitter satisfaction was whether or not they were taken from, $F(1,62) = 5.66, p = .02$. Perhaps unsurprisingly, permitters were more satisfied if they were not taken from (M = 5.00, SD = 1.53) than if they were taken from (M = 3.31, SD = 1.55), but adjusted post hoc comparisons showed a marginal effect, p = .11. Directionally, permitters were less satisfied if they were not asked for permission (M = 3.18, SD = 1.42) than if they granted permission (M = 3.58, SD = 1.69), p = .61, and less satisfied if they denied permission (M = 4.08, SD = 1.93), p = .22. Permitters who opted to punish, however, reported satisfaction levels similar to those who opted to do nothing in round three (M = 3.22, SD = 1.20 and M = 3.52, SD = 1.68, respectively), p = .92.

Permitter unhappiness. Permitters reported relatively low levels of unhappiness across the various outcomes ($M_{overall}$ = 2.23). In the combined model, both round one (permission, $F(2,62) = 4.51, p = .01$) and round two (taking, $F(1,62) = 16.63, p < .001$) outcomes were statistically significant predictors of unhappiness. Participants were more unhappy if they were not asked for permission (M = 2.76, SD = 1.46) than if they were asked for and granted permission (M = 1.63, SD = 1.31), adjusted p = .01, or if they asked for an denied permission (M = 2.00, SD = 2.26), adjusted p = .27. The difference in unhappiness between those who granted permission and those who denied permission was not significant, p = .74. Consistent with the satisfaction ratings, permitters were also much more unhappy if they were taken from (M = 2.42, SD = 1.58) than if they were not taken from (M = 0.57, SD = 1.13), adjusted p = .01. The

permitter's punishment decision in round three did not have a significant effect on unhappiness, $F(1,62) = 0.01, p = .93$; those who chose to punish reported being roughly equally as unhappy (M = 2.56, SD = 1.51) as those who did not (M = 2.18, SD = 1.66), adjusted p = .93.

Permitter ratings of taker fairness. None of the outcome variables had a significant effect on permitters' ratings of how fair the taker was (permission outcome: $F(2,62) = 1.08, p = .35$; taking outcome: $F(1,62) = 0.24, p = .63$; punishment outcome: $F(1,62) = 2.57, p = .11$). Directionally, permitters who were asked for and granted their permission rated the taker as slightly more fair (M = 3.46, SD = 1.56) than either those permitters who denied permission (M = 2.67, SD = 1.61), p = .34, or those who were not asked for permission (M = 3.03, SD = 1.59), p = .78.

Permitter ratings of taker niceness. The taker's taking decision was a significant predictor of permitters' niceness ratings, $F(1,62) = 20.22, p < .001$, and whether or not the taker chose to ask for permission first was a marginally significant predictor, $F(2,62) = 2.87, p = .06$. Niceness ratings were not significantly related to the permitter's own decision on whether or not to punish in round three, $F(1,62) = 0.47, p = .49$. Participants rated takers who took as less nice (M = 2.73, SD = 1.13) than those who did not (M = 4.71, SD = 0.95), adjusted p < .01. They rated takers who asked for and were denied permission (M = 3.50, SD = 1.98) as marginally nicer than those who did not ask (M = 2.64, SD = 0.93), p = .06, but not those who were granted permission (M = 3.04, SD = 1.16), p = .47; there was no significant difference in niceness ratings between those who were granted permission and those who did not ask, adjusted p = .37.

Counterfactual. The last question asked of permitters was whether, had they been asked, they would have granted permission to take. This question is only relevant to one group of participants: those permitters were not asked for permission. Because of the study design, all of those participants were taken from. As a result, the only relevant predictor in the ANOVA model is whether the permitters chose to punish, which did not significantly predict permitter responses, $F(1,35) = 0.12, p = .73$. Permitters generally responded in the affirmative to this question ($M_{\text{overall}} = 4.70$), though those who opted to punish indicated directionally more willingness to grant permission in the counterfactual case ($M = 5.00$, $SD = 0.71$) than did those who did not opt to punish ($M = 4.66$, $SD = 2.21$).

Taker evaluations. Taker evaluations were analyzed using the same MANOVA models as permitter evaluations, described above. However, the final models in this section were more complex, because the design for takers included no empty cells, so each model tests three main effects, three two-way interactions, and one three-way interaction.

Taker satisfaction. For overall satisfaction, there was a significant main effect of punishment on the takers' ratings of satisfaction, $F(1,45) = 32.77, p < .001$. There was also a significant interaction between the takers' round two (i.e., taking) choice and the permitters' round three (i.e., punishment) choice, $F(1,45) = 4.73, p = .03$. Combined, these effects reflect much higher satisfaction rates for takers who were not punished ($M = 4.51$, $SD = 1.53$) than for those who were punished ($M = 1.67$, $SD = 1.76$). The effect of punishment was greater for those takers who chose to take ($M_{\text{punished}} = 1.34$, $M_{\text{not punished}} = 4.56$) than for those who did not choose to take ($M_{\text{punished}} = 3.28$, $M_{\text{not punished}} = 4.18$). No other variables significantly predicted taker satisfaction.

Taker unhappiness. For taker unhappiness, there were no significant interactions among the three round outcomes. However, there was a significant main effect of permission round outcome, $F(2,45) = 17.66, p < .001$, and a significant main effect of punishment, $F(1,45) = 31.43, p < .001$. Predictably, takers were significantly more unhappy when they were punished (M = 3.33, SD = 2.23) than when they were not punished (M = 1.07, SD = 1.33), $p < .001$. They were also more unhappy when they were denied permission (M = 2.90, SD = 1.95) than when they were granted permission (M = 0.67, SD = 1.27), $p < .001$, or when they did not ask for permission (M = 1.55, SD = 1.63), $p = .02$. There was no difference in unhappiness between takers who did not ask for permission and those who were granted permission, $p = .15$.

Taker rating of permitter fairness. None of the outcome variables (permission, taking, or punishment) significantly predicted taker ratings of how fair the permitter was, although there was a marginal main effect of punishment, $F(1,45) = 3.31, p = .08$. This marginal effect reflects lower fairness ratings among takers who were punished (M = 2.13, SD = 1.68) as compared to those who were not (M = 3.22, SD = 1.97), adjusted $p = .08$.

Taker ratings of permitter niceness. In the model fitted to takers' ratings of permitter niceness, there were significant main effects of permission round outcome, $F(2,45) = 5.08, p = .01$, and punishment, $F(1,45) = 44.51, p < .001$. There was also a significant interaction effect between taking and punishment, $F(1,45) = 9.84, p < .01$. Post hoc comparisons showed that permitters were rated as nicer if they granted permission (M = 4.21, SD = 2.13) than if they denied permission (M = 2.86, SD = 1.98), $p = .01$. Niceness ratings did not differ significantly between those permitters who did not ask for permission (M = 3.18, SD = 2.23) and those who received permission, $p = .14$, or those who were denied permission, $p = .82$. Takers also rated the

permitter as nicer when they chose not to punish the taker ($M = 4.32$, $SD = 1.77$) than when they did punish the taker ($M = 1.27$, $SD = 1.39$), $p < .001$.

The post hoc comparisons also suggested that the significant interaction of taking and punishment was driven by differences in niceness ratings between those who took and were punished ($M = 0.92$, $SD = 1.16$) and those who did not take and were not punished ($M = 2.75$, $SD = 1.98$), $p = .01$, as well as those who took and were not punished ($M = 4.70$, $SD = 1.51$), $p < .001$.

Discussion

As in Study 3.1, the majority of takers chose to ask for permission, and the majority of permitters opted to grant it. Takers also generally chose to take, regardless of whether they got permission or not. Given the strikingly low number of permitters who actually chose to punish, it's also possible that takers correctly predicted the low odds of being punished. In that case, however, it is unclear why they would ask permission in the first place.

The evaluations made by participants in this study were also broadly consistent with the evaluations in Study 3.1, which is encouraging for the permission game's use as an experimental tool. The consistency of the results with those of Study 3.1 also suggest that the threat of punishment may not have influenced participants' behaviors, but a direct comparison is necessary before drawing any conclusions.

Study 3.3: The Permission Game

In Studies 3.1 and 3.2, we tested participants under conditions that attempted to simulate a live, two-player game. But simulation was necessarily limited, and a sizable fraction of our

participants indicated that they had some suspicions about the nature of the study.[2] Thus, though the results above are encouraging with regard to the viability and usefulness of the permission game as a tool for behavioral research, it remains to be seen if the responses we have thus far obtained are representative of "real" responses. In Study 3.3, we drop the simulation altogether and have participants actually play the game in pairs in the lab.

To maximize the usefulness of the paradigm, it should mirror real-world social situations as closely as is feasible. The punishment round, which provides the permitter with some recourse against the taker, is an important part of accomplishing this parity; after all, in many cases the threat of negative consequences may be a major impetus for asking permission in the first place. Study 3.3 focuses on the effects of the threat of punishment, comparing a version of the game with a punishment round to a version without.

In Study 3.2, however, we saw strikingly low rates of punishment from permitters. In other pilot studies, permitters were similarly reluctant to engage in costly punishment, preferring to keep what was left of their bonuses. Even in purely hypothetical vignette versions of this paradigm, expressed desire to punish is fairly low, and relatively few permitters say that they would indulge that desire. Although this reluctance to punish might be an accurate indication of "real" consequences for this kind of transgression, it also limits the usefulness of the permission

---

[2] Participants who were probed for suspicion were generally asked after they had been fully debriefed, so the suspicion numbers we report are likely to overestimate actual suspicion.

game as a tool for study. With such low rates of punishment, it becomes difficult to evaluate that aspect of the game with adequate statistical power, for example.

Fortunately, punishment—and more precisely, the desire to punish—is highly context-dependent. One of the advantages of the permission game paradigm is that it is extremely flexible, which allows researchers to adjust the context to suit the research. In Study 3.3, we took advantage of this flexibility to try and increase the desire and willingness to punish. Based on pilot testing, we adjusted the bonus structure and removed the multiplier for taking. In the pilot, which was run as a hypothetical vignette study in which participants had the permitter role, the permitter began the game with $4 and the taker began the game with $3. During the taking round, the taker had the opportunity to take $2, with no multiplier, so that the new totals would be $2 and $5 for the permitter and taker, respectively. In the punishment round, the permitter could then spend $1 to destroy *all* of the taker's bonus.

This adjustment was successful in increasing rates of punishment in the pilot, although the overall rate of punishment remained relatively low. It averaged about 32% across all pilot participants, ranging from roughly 20% among permitters who had granted permission to about 40% among those who did not. The adjustment also substantially reduced the likelihood that the permitter would grant permission in the first place to just over 30% (*c.f.* the nearly 70% who granted permission in Study 3.2).

As intended, the new set of payouts increased both desire to punish and actual (hypothetical) choices to engage in costly punishment in the pilot study. However, if we predicted this change, might takers do the same? If so, a number of changes to taker behavior, when compared to the takers in Study 3.2, are possible. In the taking round, rates of actual taking should decrease across the board, and especially when the permitter has denied permission to

take. Risk-averse takers might prefer to keep $3 rather than risk leaving with nothing, and a taker who has refused permission has sent a signal that they may be at higher risk of punishing the taker.

Method

Participants. Participants were 114 adults (M_{age} = 28.54, 47 females) at the Center for Decision Research (two locations). Participants were paired together by experimenters. Pairs were randomly selected from eligible visitors to the Center who there at the same time, but who did not arrive together.

Procedure. Participant pairs were randomly assigned to one of two conditions: they either played the punishment version (n = 31) or no punishment version (n = 26) of the permission game, as described in Studies 3.1 and 3.2. In both conditions, the payout structure was as described above: the permitter began the game with $4 and the taker began the game with $3. During the taking round, the taker had the opportunity to take $2, with no multiplier, so that the new totals would be $2 and $5 for the permitter and taker, respectively. In the punishment round, if the pair was assigned to the punishment version of the game, the permitter could then spend $1 to destroy *all* of the taker's bonus.

Within each dyad, one participant was randomly assigned to the taker role and the other to the permitter role. The game was carried out just as it was in Studies 3.1 and 3.2. An experimenter acted as go-between, and the bonuses earned during the game were paid out to participants at the end of the study. Participants were separated during the actual study, but they were allowed to see one another before the game begins, to increase the appearance that they were playing with a real participant partner (which, of course, they were).

On completion of the game, participants answered evaluation questions. They rated their satisfaction and unhappiness, as well as the niceness and fairness of the other player.

Results

Taker decisions. Across both conditions, 64.9% of takers chose to ask permission. There was no significant difference between the rates of asking permission in the punishment condition (65.4%) and the no punishment condition (64.5%), $\chi^2(1, n = 57) = 0.00, p = 1.00$.

For the taking decision, I fit a binomial logistic regression model using condition (punishment vs. no punishment) and permission round outcome (deny, grant, or not asked) as predictors. Note that the permission round outcome variable includes both whether the taker asked for permission (not asked vs. grant or deny) and whether the permitter granted permission. There was a significant interaction between condition and not having asked for condition, $\beta = -3.11, p = .05$. As shown in the figure below, this interaction reflects a substantially higher taking rate in the no punishment condition when the taker did not ask for permission (88.9%) than in the punishment condition when the taker did not ask for permission (27.3%).

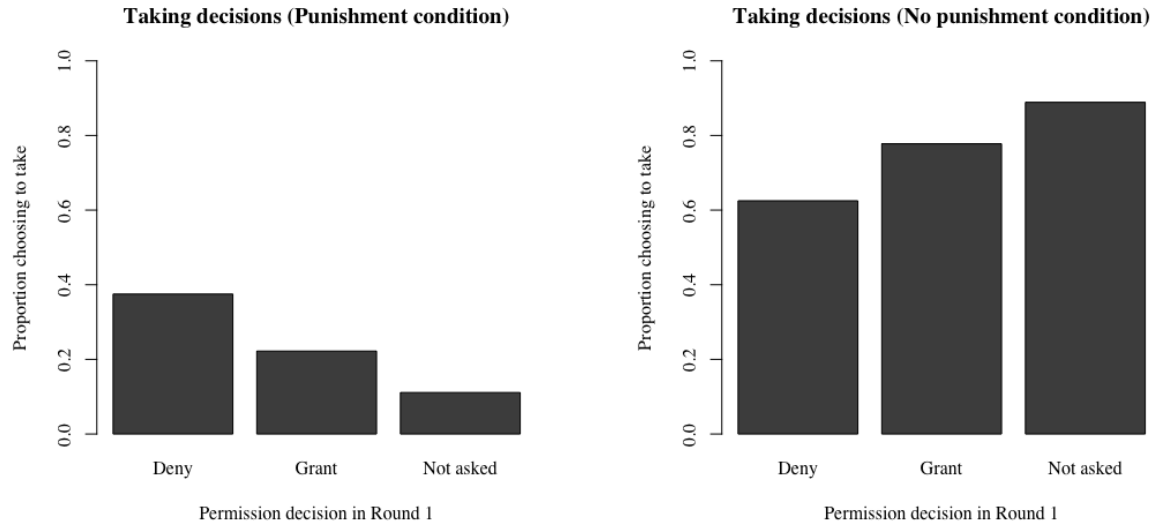


Figure 3.8. Taker decisions in Study 3.3 for the punishment condition (left) and no punishment condition (right).

Looking only at the effect of condition (i.e., without controlling for permission round outcome or the interactions among them all), there was also a simple effect of condition on taking decision, $\chi^2(1, n = 57) = 5.74, p = .02$. Takers were significantly more likely to take in the no punishment condition (41.9%) than in the punishment condition (76.9%).

Permitter decisions. Of those permitters who were asked for permission, 48.6% chose to grant it. There was no significant difference in rates of granting permission between those in the punishment condition (45.0%) and the no punishment condition (52.9%), $\chi^2(1, n = 57) = 0.02, p = .88$.

In the punishment condition, only three permitters (9.7%) chose to punish the taker. All three had been taken from. Two had denied permission, and one had not been asked for it. Unfortunately, there were not enough participants opting to punish to conduct meaningful analyses. Thus, punishment decision is not included as a predictor in the evaluations below.

Taker evaluations. Participant evaluations were done on a seven-point Likert scale, which were converted into numerical ratings ranging from 0 to 6. A MANOVA model was fitted to each of the evaluation measures, using condition, round one permission decision, and round two taking decision as predictors. Post hoc comparisons were completed with Tukey's Honestly Significant Difference; all reported post hoc *p*-values are adjusted for multiple comparisons.

Taker satisfaction. For takers, there were no significant differences in satisfaction. Across both conditions and all outcomes, takers were fairly satisfied ($M = 4.60$, $SD = 1.53$).

Taker unhappiness. Similarly, there were no significant effects on taker unhappiness ($M = 1.21$, $SD = 1.48$).

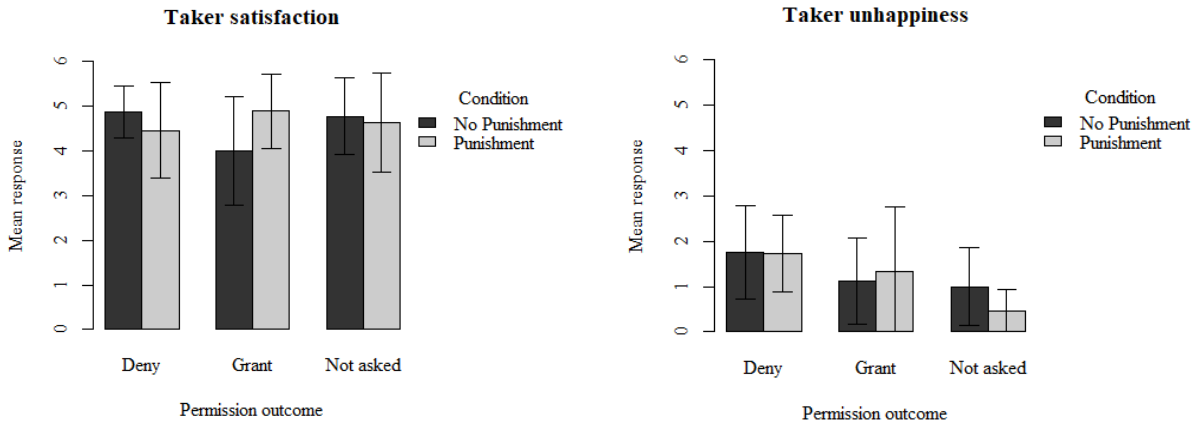


Figure 3.9. Takers' reactions across condition and permission outcome. Error bars represent standard error of the mean.

Taker rating of permitter fairness. For fairness ratings, there was a significant main effect of condition on taker's ratings of how fair the permitters were, $F(1,45) = 6.42, p = .01$. Takers rated the permitters as significantly fairer in the punishment condition ($M = 4.90$, $SD = 1.83$) than in the no punishment condition ($M = 3.77$, $SD = 2.03$), adjusted $p = .01$. There was also a main effect of taking decision on taker's ratings of permitter fairness, $F(1,45) = 19.08, p < .001$; takers who decided not to take rated the permitters as significantly fairer ($M = 5.67$, $SD = 1.46$) than did takers who decided to take ($M = 3.45$, $SD = 1.80$), $p < .001$. No other terms in the model had a significant effect.

Taker rating of permitter niceness. There was a significant interaction between permission outcome and taking decisions on how nice the takers rated their counterparts, $F(2,45) = 3.63, p = .03$. As shown in the figure below, this interaction seems to be driven by takers who asked for permission (whether it was denied or granted) rating the permitters as nicer when they decided to take ($M = 4.18$, $SD = 1.56$) than when they did not decide to take ($M = 3.07$, $SD = 1.39$), while takers who did not ask for permission rated the permitter as nicer

when they opted not to take (*M* = 4.78, *SD* = 1.39) than when they did not take (*M* = 3.36, *SD* = 1.63). No other effects in the model were significant.

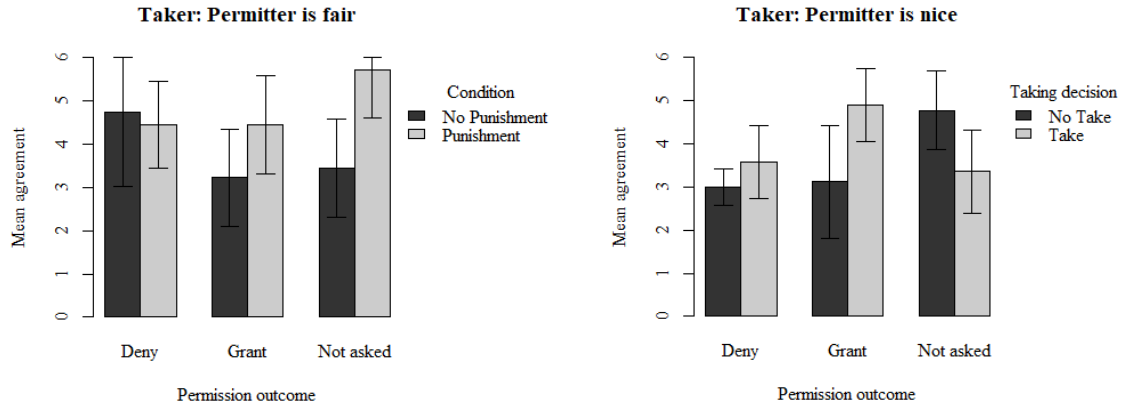


Figure 3.10. Taker ratings of the permitter in Study 3.3. Error bars represent standard error of the mean.

Permitter Evaluations. Permitter evaluations were analyzed in the same manner as taker evaluations. A MANOVA model was fitted to each of the evaluation measures, using condition, round one permission decision, and round two taking decision as predictors.

Permitter satisfaction. There was a significant main effect of condition on permitter evaluations of satisfaction, $F(1,45) = 12.64, p < .001$. Permitters were significantly more satisfied in the punishment condition (*M* = 3.84, *SD* = 1.83) than in the no punishment condition (*M* = 2.31, *SD* = 1.95), adjusted *p* = .001. There was also a significant main effect of taking on permitter satisfaction, $F(1,45) = 17.53, p < .001$; permitters were significantly more satisfied when they were not taken from (*M* = 4.42, *SD* = 1.67) than when they were taken from (*M* = 2.21, *SD* = 1.75), *p* < .001. Permission outcome did not have a significant effect on permitter satisfaction, $F(2,45) = 0.65, p = .53$.

Permitter unhappiness. Permitters reported being significantly unhappier when they were taken from (*M* = 3.24, *SD* = 1.77) than when they were not taken from (*M* = 1.17, *SD* = 1.90),

$F(1,45) = 13.54, p < .001$. There were no differences in unhappiness by condition, $F(1,45) = 2.03, p = .16$, or permission outcome, $F(2,45) = 0.22, p = .80$.

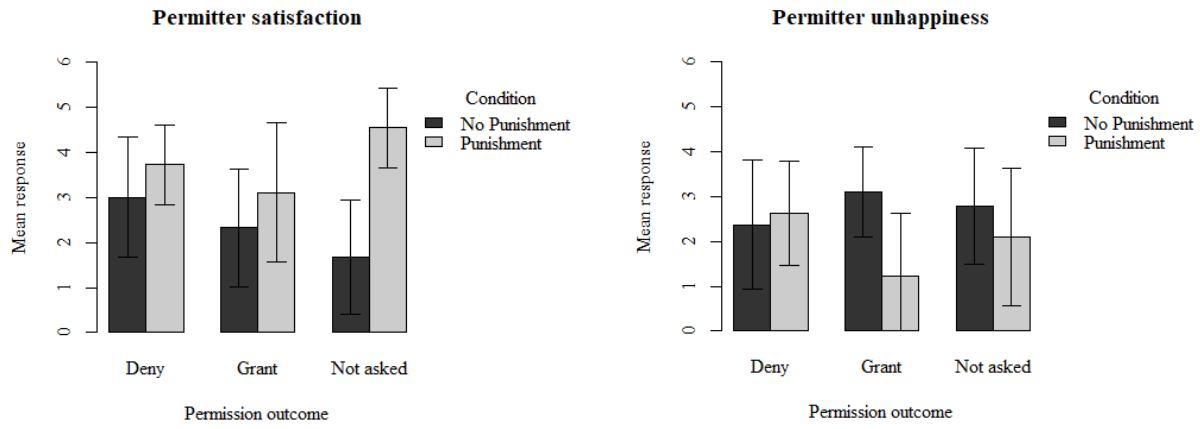


Figure 3.11. Permitter ratings of satisfaction and unhappiness in Study 3.3. Error bars represent standard error of the mean.

Permitter ratings of taker fairness. Permitters rated the taker as fairer when they were not taken from (*M* = 4.79, *SD* = 2.04) than when they were taken from (*M* = 3.27, *SD* = 1.89), $F(1,45) = 6.23, p = .02$. None of the other predictors had a significant effect on fairness ratings.

Permitter ratings of taker niceness. There were significant main effects of condition, $F(1,45) = 6.21, p = .02$, and taking decision, $F(1,45) = 80.31, p < .001$, on permitters' ratings of how nice the taker was. Permitters rated the taker as nicer in the punishment condition (*M* = 3.55, *SD* = 2.16) than in the no punishment condition (*M* = 2.73, *SD* = 1.56), *p* = .02. They also rated the taker as nicer when they did not take (*M* = 4.96, *SD* = 1.16) than when they did take (*M* = 1.88, *SD* = 1.22), *p* < .001. Permission outcome did not have a significant effect on ratings of niceness, $F(2,45) = 0.41, p = .67$.

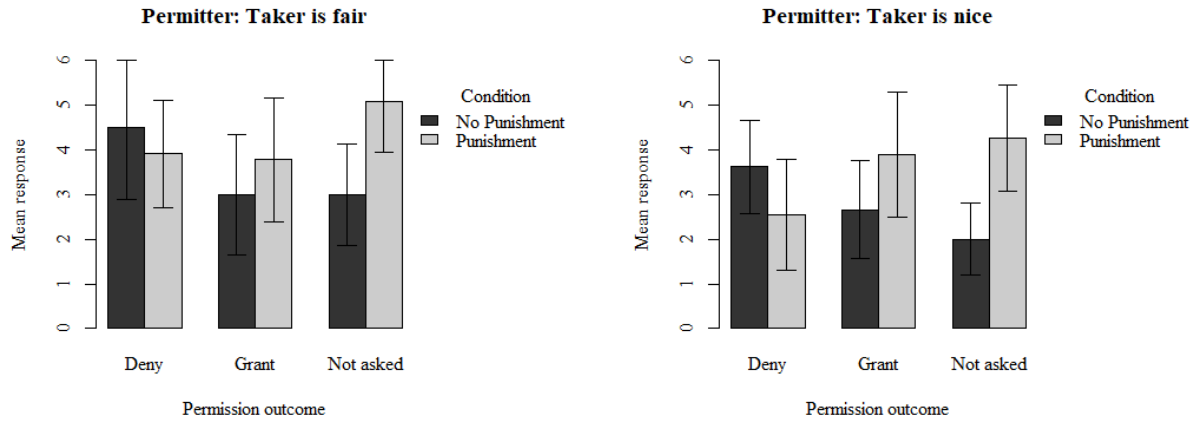


Figure 3.12. Permitter evaluations of the taker in Study 3.3. Error bars represent standard error of the mean.

Discussion

The threat of punishment did not affect takers' decisions to ask for permission; the majority of takers asked for permission in both conditions. Similarly, the possibility of punishment does not seem to have affected whether permitters *granted* permission. However, there were differences in how takers behaved after the permission round; takers who did not ask for permission in the punishment condition overwhelmingly decided to take, while those who did not ask for permission overwhelmingly decided not to take. This suggests that takers in the two conditions may have had different reasons for choosing not to ask; in the punishment condition, participants may have planned not to take, and therefore they did not ask for permission, while in the no punishment condition, participants may have decided to take, and therefore they did not ask for permission.

Unsurprisingly, permitters were happiest when they were not taken from. However, even those who were taken from overwhelmingly declined to punish the takers. It is possible that the apparent discrepancy is driven by money; punishment in this game was costly, and if permitters

were already unhappy with the loss of money they suffered from the taking, they may have wanted to preserve their remaining bonuses rather than enacting punishment.

Study 3.4: Why Grant Permission? A Comparison to Other Paradigms

In the permission game, as in many real-life cases, the permitter seems to gain precious little by granting permission. To the extent that granting permission increases the odds that they will be taken from, the permitter actually loses by choosing to grant. So why do it at all?

One possibility is that social pressure gives permitters little choice. It is often assumed that permission will be granted for relatively small requests. A person who refuses to permit someone to borrow a pen momentarily, or to take an unused chair from a table, for example, is seen as unfriendly at best. Within the permission game, the social pressure is less clear, but it likely still exists. Asking for permission is, at some level, an inherently polite act; refusal can therefore feel like meeting politeness with rudeness, putting pressure on the askee to grant permission instead.

Social pressure is not the only force that could be at work, however. In characterizing permission as an act that does not benefit the permitter, I have glossed over several intangible benefits that permission could convey, including feelings of agency and generosity. The latter is fairly straightforward; granting permission may allow the permitter to imagine herself as a benefactor, even if the permission does nothing to facilitate the actual taking, because she could have withheld permission. Agency, on the other hand, requires a bit more explanation.

In the permission game, particularly in the case where punishment is not an option, the actual agency of the permitter is near zero. The same is often true in real life, where adults are seldom bound by a denial of permission, even if the threat of negative consequences exists if the

permitter is ignored. However, it could be that granting permission strengthens an illusory sense of agency. The logic is similar to those that might undergird feelings of generosity (indeed, it is conceptually difficult to imagine generosity without agency), in that the permitter could reason that taking is more difficult or more unpleasant if they deny permission to the taker. Thus, although they cannot change the outcome of the game if the taker decides to take, the permitter can choose to make the game better for the taker.

Of course, if this is accurate, then the opposite is also true; the permitter can choose to make the game *worse* for the taker. So why might a desire for agency increase permission, rather than decreasing it? I hypothesize that, under the circumstances, the denial of permission is the default or status quo. Both the permitter and the taker likely expect the permitter to deny permission if she is asked, because both players correctly realize that the permitter would prefer to keep her money. As such, the choice to deny permission may not actually feel agentic at all; it merely reinforces the status quo.

Several changes to the structure of the permission game can help test these hypotheses, and in Study 3.4, we compare the permission game to some of these alterations. To increase perceived agency, we give permitters complete control over the outcome of the game. In this case, the permitter would get to choose whether the taker gets additional money (at the permitter's expense) or not. To maximize agency, the permission round can be skipped altogether; the end result is a dictator game in which the permitter selects between two options: $3 for herself and $1 for the other player; or $2 for herself and $4 for the other player. For this study, we have returned to the original endowment amounts and multipliers, to provide some incentive for the permitter to choose the option that gives her less money but benefits the pair overall.

Another clear way to increase permitter agency is simply to make the permission decision binding; if the permitter grants permission, the money is taken, but if the permitter denies permission, the taker may not take. Note that, in effect, this is just a re-framing of the dictator game; the permitter still makes a unilateral decision that determines the final payout. However, this version, in which the permitter is *asked* to make the more generous decision, adds the social pressure of permission. By comparing this version to the dictator game, we can therefore get a real sense of the impact of social pressure on permitter decisions.

To increase perceived generosity in a way that may be less tied to agency, we also create a situation in which the permitter can allow the taker to make the decision; i.e., a chance to abdicate the choice to the other player. Early research on abdication in this kind of scenario (Kardas, Shaw, & Caruso, 2017) suggests that both players view this abdication as a generous act. However, abdication is also a less agentic choice than the alternative, which is for the permitter to make the decision herself. The abdication paradigm thus provides us with a way to test a choice between generosity and agency, as well as to compare abdication to granting permission.

Finally, we can remove any sense of agency and generosity by forcing a decision on the permitters; in other words, we can put the permitter into the other, passive role of the dictator game. This should provide a useful baseline against which to compare the other conditions.

Method

Participants. Participants were 250 adults (114 female, 134 male, 2 others; M_{age} = 35.50) recruited from Amazon's Mechanical Turk.

Procedure. Participants were randomly assigned to one of five conditions, each of which represents a different economic game, as described above: permission (n = 43), dictation

(n = 57), binding permission (n = 46), abdication (n = 48), and passive (n = 56). After reading the instructions for the appropriate game, participants completed a series of comprehension questions; they could not continue with the study until they answered these questions completely. In all conditions, participants were then asked imagine themselves as a player in the appropriate game. Except in the passive condition, the participant was always assigned to the role making the decision equivalent to granting or denying permission; permitter in the permission condition, dictator in the dictation condition, permitter in the binding permission condition, and decider in the abdication condition. In the passive condition, participants were assigned to the recipient role of the dictator game.

To make outcomes as consistent as possible, the "other player's" decisions, where needed, were held constant. In the permission and binding permission conditions, the other player asked for permission, so that the participant would need to make a choice. In the permission and abdication conditions, if the participant granted permission to take or abdicated the decision to the other player, the other player always decided to take from the participant.

At the end of the game, participants answered two more comprehension check questions about the specific scenario they were asked to imagine themselves in; unlike the generic comprehension questions, participants were allowed to continue regardless of their responses to those questions. Participants then completed a short questionnaire. In addition to the satisfaction and fairness questions asked in studies 3.1-3.3, we also measured their feelings of agency, generosity, and social pressure. We also asked participants how generous they thought the other player would think they were, as well as how nice they believed the other player was. In conditions where the "other player" made a decision (permission, binding permission, and when the participant chose to abdicate in the abdication condition), they were asked how surprising

and how confusing they found the other player's actions. Finally, participants were probed for confusion over the rules.

Results

For each condition, the participants' decisions, comprehension, and confusion about the rules are summarized below. Figure 3.13 summarizes the decisions made in each condition (except the passive condition, in which no decisions were made). We then compared participants' questionnaire responses across conditions. All of the evaluations were measured on seven-point Likert scales, and we analyzed them as continuous measures on a scale from 1 to 7.

Permission condition. Thirty-three participants (76.7%) in the permission condition chose to grant permission (classified as the generous choice in Figure 3.13). They expressed very low rates of confusion over the rules ($M = 1.91$, $SD = 1.67$), which is borne out by their comprehension question performances. All of the participants correctly recalled their assigned roles, 39 (90.7%) correctly reported how much money they would make in the hypothetical scenario, and 40 (93.0%) correctly reported the other player's payout.

Dictation condition. Thirty-three participants (57.9%) in the dictation condition chose the more generous distribution (giving the other player $4 and themselves $3, rather than giving the other player $1 and themselves $4). They expressed very low rates of confusion ($M = 1.42, SD = 1.02$). All of the participants correctly recalled their role in the dictation game, and they all correctly reported how much money their opponent would receive. All but one participant (98.2%) correctly reported their own hypothetical winnings.

Abdication condition. In the abdication condition, 4 participants (8.3%) chose to abdicate. Twenty-two (45.8%) chose the generous distribution (giving the other player $4 and themselves $3, rather than giving the other player $1 and themselves $4), and 22 (45.8%) chose

the less generous distribution. Overall confusion was very low ($M = 1.73, SD = 1.33$). All of the participants correctly identified their role in the abdication game. Only one participant (2.1%) incorrectly answered the question about their own hypothetical payout, and two (4.2%) incorrectly answered the question about the other player's payout.

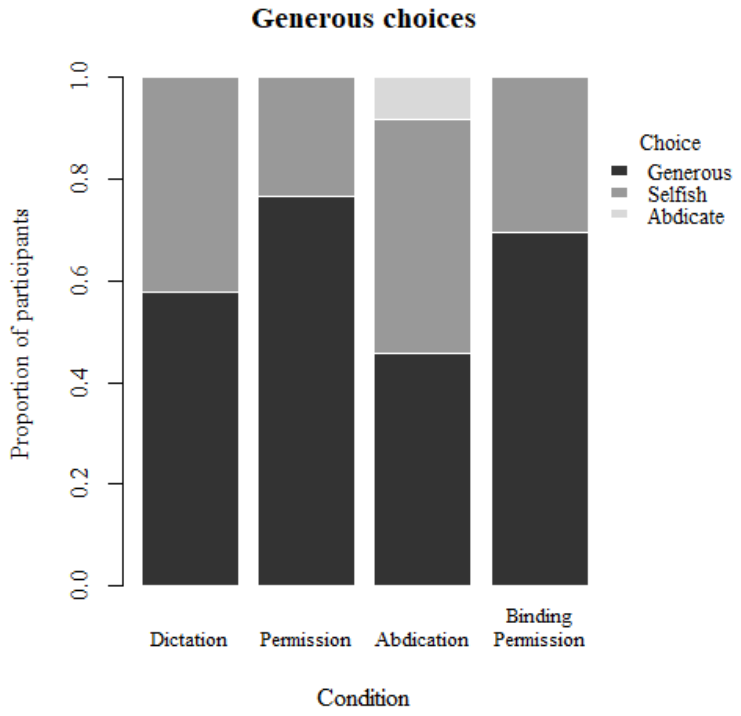


Figure 3.13. Participant choices across the conditions of Study 3.4.

Binding permission condition. Thirty-two participants (69.6%) in the binding permission condition chose to grant the permission (classified as the generous outcome in Figure 3.13). Confusion was low ($M = 1.63, SD = 1.29$), and all of the participants correctly identified their role. Four participants (8.7%) incorrectly reported their own payout, and four (8.7%) incorrectly reported the other player's payout.

Passive condition. In the passive condition, participants had no choices to make. Confusion was low ($M = 1.66, SD = 1.37$), and three participants (5.4%) incorrectly identified

their role in the game. Seven (12.5%) incorrectly reported their own payouts, and four (7.2%) incorrectly reported the other player's payout.

Decision comparison. A binomial logistic regression was fitted to the choice type (generous vs. selfish vs. abdicate), using condition as the predictor and permission condition as the base case. There was a significant negative effect of the abdication condition on likelihood of choosing the generous option, $\beta = -1.36, p < .001$, and a marginally significant negative effect of the dictation condition, $\beta = -0.88, p = .05$.

Perceived control. There was a significant effect of condition on perceived control, $F(4,245) = 95.23, p < .001$. Post-hoc comparisons showed that participants reported feeling significantly less in control in the permission ($M = 2.79, SD = 2.17$) and passive ($M = 1.82, SD = 1.73$) conditions than in the dictation ($M = 6.37, SD = 0.34$), abdication ($M = 6.02, SD = 1.63$), and binding permission ($M = 6.00, SD = 1.23$) conditions. There was no difference in perceived control between the permission and passive conditions, or among the other three conditions. These results are summarized in Figure 3.14 below.

Perceived generosity. An ANOVA showed significant differences in perceived generosity by condition, $F(4,245) = 3.28, p = .01$. Post-hoc comparisons showed this to have been driven by a significant difference between the binding permission condition ($M = 4.91, SD = 2.16$) and the passive condition ($M = 3.48, SD = 2.12$). None of the other conditions varied significantly from one another, as shown in Figure 3.14.

Perceived pressure. There were no significant differences in the amount of pressure the participants reported feeling to take the action they chose ($M = 2.96, SD = 1.91$), $F(3,190) = 0.94, p = .42$ (participants in the passive condition did not make any choices, and so they were not asked about feelings of pressure).

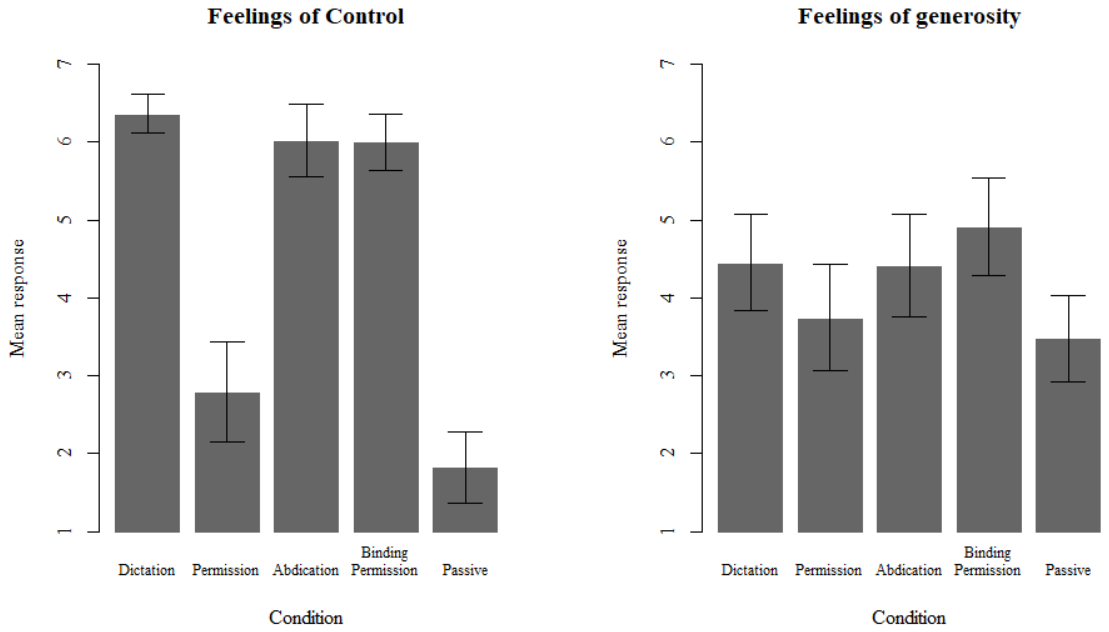


Figure 3.14. Feelings of control and generosity across the conditions of Study 3.4. Error bars represent standard error of the mean.

Other players' perceptions. There was a significant effect of condition on how generous participants thought the other player would believe they were, $F(4,245) = 5.15, p < .001$. Post-hoc comparisons showed this to be the result of significantly lower ratings in the passive condition ($M = 2.88, SD = 1.73$) than in the dictation condition ($M = 4.28, SD = 2.55$), $p < .01$, the permission condition ($M = 4.23, SD = 1.89$), $p = .03$, and the binding permission condition ($M = 4.74, SD = 2.37$), $p < .001$. None of the other conditions differed from one another.

Surprise. In the permission condition, binding permission condition, and abdication condition (when the participant chose to abdicate), participants were asked how surprised they were by the other players' actions (the other player did not take any actions in the other conditions). There was a significant effect of condition on participants' reported surprise,

$F(2,90) = 3.73, p = .03$. Post-hoc comparisons showed a significant difference in surprise between the permission condition ($M = 1.58, SD = 1.31$) and the binding permission condition ($M = 2.48, SD = 1.94$), $p = .04$. Neither permission conditions differed significantly from the abdication condition ($M = 3.00, SD = 2.31$), but the measurement in the abdication condition is based on just four participants.

Satisfaction. There was a significant effect of condition on participant satisfaction, $F(4,245) = 16.30, p < .001$. As shown below, post-hoc comparisons indicated that this effect was driven by significantly lower satisfaction in the permission condition ($M = 4.49, SD = 1.68$) and the passive condition ($M = 4.09, SD = 1.65$) than in the other three conditions (dictation: $M = 5.42, SD = 1.40$; abdication: $M = 5.90, SD = 1.08$; binding permission: $M = 5.83, SD = 1.23$).

Happiness. There was a significant effect of condition on participants' reported happiness, $F(4,245) = 6.56, p < .001$. As shown below, post-hoc comparisons showed that this is because happiness was significantly lower in the passive condition ($M = 3.93, SD = 1.63$) than any of the other conditions (dictation: $M = 4.96, SD = 1.32$; permission: $M = 4.79, SD = 1.57$; abdication: $M = 4.77, SD = 1.24$; binding permission: $M = 5.28, SD = 1.33$). None of the other conditions differed from one another.

Niceness of other player. There was a significant effect of condition on participants' ratings of how nice the other player was, $F(4,245) = 7.11, p < .001$. Participants rated the other player as less nice in the passive condition ($M = 3.96, SD = 1.40$) and more nice in the binding permission condition ($M = 4.35, SD = 1.15$) than they did in the other three conditions, which did not differ from one another (dictation: $M = 4.26, SD = 0.97$; permission: $M = 4.35, SD = 1.15$; abdication: $M = 4.44, SD = 0.85$).

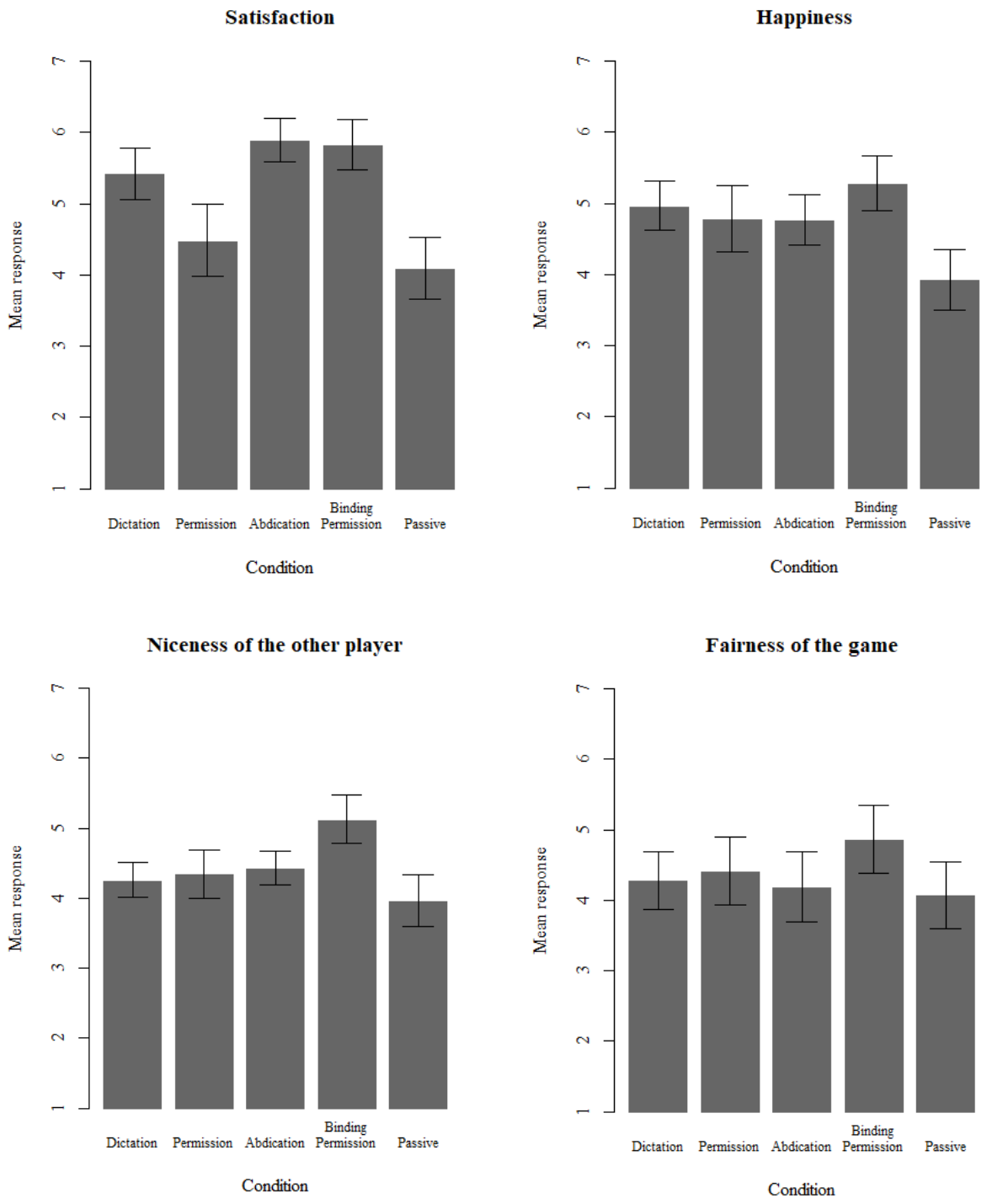


Figure 3.15. Additional participant evaluations from Study 3.4. Error bars represent standard error of the mean.

Fairness of the game. There were no significant differences among the conditions in how fair they rated the game, $F(4,245) = 1.61, p = .17$. Across all five conditions, participants' fairness ratings were near the middle of the scale ($M = 4.35, SD = 1.70$).

Discussion

Participants in the permission condition felt significantly less in control than did participants in the dictation, binding permission, and abdication condition; this is unsurprising, given that they actually were less in control. However, there were no differences in feelings of generosity or pressure between the permission condition and the other games. Despite feeling less in control, participants were more likely to make the generous choice in the permission condition than they were in the abdication or dictation conditions. They were also less satisfied, but they did not think the game was any less fair.

Taken together, these results suggest that the act of granting permission is different than making the generous choice in the other conditions. Participants in the permission condition were more likely to make the generous choice than were those in the dictation condition. Participants in the permission condition may have been hoping that the other player would not take from them, but then why not deny permission? Although the results of this study are far from conclusive, they provide some support for the idea that granting permission is not a purely economic decision.

General Discussion

If it is better to ask for forgiveness than permission, why do we so often ask each other for permission? The permission game detailed in this chapter could prove a useful tool for answering this and other questions related to the social dance of permission asking and granting. Studies 3.1-3.3 demonstrate the flexibility of the permission game as an experimental tool. Many

adjustments can be made to the experimental procedures to allow for the investigation of those questions; adjustments could be made to the payouts, multipliers, punishment opportunities, punishment costs, or social aspects of the game.

4 Conclusion: The Future of Intuitive Jurisprudence

Throughout this dissertation, I have argued that research in law and psychology can benefit from a broad methodological and theoretical base. In particular, I have shown the promise of an intuitive jurisprudence approach that borrows from developmental science to craft interesting questions and clear experiments, and to aid in our understanding of how lay people interact with the law. In Chapter 1, I showed that children and adults consider the spirit of the law when evaluating rule breakers. In Chapter 2, I showed that punishment has a powerful ability to communicate information about the state of the world; in adults, the message seems to be particularly tied to inferences of harm, but children also show many signs of complex reasoning about punishment's place in society. Finally, in Chapter 3, I present complementary research on the role of seeking permission to avoid punishment.

However, the potential of intuitive jurisprudence goes well beyond the work presented in this dissertation. To conclude, I want to identify some ways in which intuitive jurisprudence has already led to valuable insights and outlining some areas that I think are ripe for exploration. This is not meant to be an exhaustive recitation; I seek only to demonstrate the future potential of an intuitive jurisprudence approach. In addition, I acknowledge that the areas of research I discuss are not perfectly (or even mostly) discrete; research and theory frequently cross these boundaries, combining work on basic moral psychology, for example, with analysis of the legal system. I have chosen these distinctions not to emphasize the differences between the areas, but to emphasize the breadth of ways in which intuitive jurisprudence can make powerful contributions to many kinds of research.

Intent, Moral Judgment, and Blame

To determine the hidden mental states of other people, including intentions, knowledge, and desires, humans must rely on inferences drawn from words and actions. Many lines of research in social and developmental psychology converge on questions about how these inferences are made and how they lead us to understand others' mental states. This inferential process, in turn, is centrally important in many legal settings, where the application of law in a particular case or situation often turns on the mental states of the people involved; such mental states must often be inferred *post hoc* by judges and juries.

Intentions, in particular, occupy a special place in both fields. Intent is of critical importance in criminal (and, to a lesser extent, civil) law; the influential Model Penal Code of 1962, which provides suggestions and models for state and local criminal laws, marked out four categorical degrees of intent (or "mens rea") that continue to be used—with some minor variations—in laws across the country: purposeful, knowing, reckless, and negligent (*Model Penal Code § 2.02*, 1962). Laws based on the Model Penal Code assign different levels of legal blame and punishment for similar acts that are judged to have been committed with different degrees of intent; purposeful acts are generally punished most severely, followed by knowing and reckless acts, and negligent acts are punished least severely.

Outside of these formal legal categories, intentions play a central role in moral judgments—at least for adults (see, e.g., Barrett et al., 2016; Cushman, 2015; Lagnado & Channon, 2008; Malle & Knobe, 1997). Compared to adolescents and adults, young children's moral evaluations tend to focus more on the outcome of the act (Cushman, Sheketoff, Wharton, & Carey, 2013; Nobes et al., 2009; Piaget, 1932/1997; Yuill & Perner, 1988; Zelazo, Helwig, & Lau, 1996), and Piaget argued that young children are essentially unconcerned with intentions (1932/1997). However, even preverbal infants seem to weigh the intentions of a bad actor in their

social evaluations (Hamlin, 2013a), and 3-year-olds use the intentions of others to guide their own behavior; they are less likely to help someone who has demonstrated harmful intentions toward another person (Vaish et al., 2010). Thus, despite the relative emphasis on outcomes, intentions seem to play an important role moral judgment even for young children.

Attempts to square legal definitions of intent with psychological intuitions have met with limited success. Adults can reliably distinguish among legal categories of intent at the outer limits, such as between purposeful actions and those that are merely negligent, but they seem to struggle with distinguishing "knowing" conduct and "reckless" conduct (Shen, Hoffman, Jones, Greene, & Marios, 2011). This work takes a narrow approach, relying on mock juries and hypothetical legal cases, which is useful for directly evaluating the law, but which leaves many gaps in our understanding of the underlying moral reasoning. An intuitive jurisprudence approach would broaden the scope, making the work more applicable to moral reasoning at all stages of cognitive development. This broader view would also provide a richer and clearer perspective for legal scholars who want to understand whether and how people make these distinctions in everyday life as well as in legal contexts.

Research with children, on the other hand, has not explicitly considered the relationship between childhood understanding of intentions and legal categories of intent (but see Nobes et al., 2009, examining children's understanding of negligent behavior), but it is well equipped to do so. Developmental psychologists have painstakingly mapped out children's earliest understandings of the minds of others. Beginning in infancy, babies exhibit rudimentary understanding of other people's actions as being guided by their intentions (e.g., Woodward, 2013). With age, children's reasoning about other people's actions becomes more complex. Judgments of intentionality begin to incorporate more information about the actors' beliefs and

desires, and children start to distinguish even more between intentional and accidental acts (Malle & Knobe, 1997; Mull & Evans, 2010). In addition, a great deal of research has been devoted to understanding the origins of theory of mind—that is, people's ability to think and reason about the mental states of others—and the degree to which theory of mind is intuitively present in young humans versus changed across development (Butterfill & Apperly, 2013; Killen et al., 2011; Onishi & Baillargeon, 2005; Wellman, Cross, & Watson, 2001). By applying an intuitive jurisprudence approach in this domain, this rich body of existing work can be leveraged to help determine the mental processes and capacities that contribute to people's ability to distinguish the legal categories of intent.

While developmental insights can help legal scholars understand intuitive moral judgments, legal theory may likewise help psychologists contend with questions of morality and judgment. Consideration of the structure and function of formal legal systems, such as the Model Penal Code, can provide morality research with new testable predictions and applications, and an intuitive jurisprudence approach provides new tools to explore this domain. Moreover, because certain legal mandates—such as the prohibition against murder—seem to reappear across cultures, and because they seem also to be deeply tied to moral intuitions, these laws can provide a window into which elements of morality, if any, are truly universal (Green, 2000).

Laws, like those based on the Model Penal Code, can also provide a meaningful descriptive challenge for fundamental moral theories; as Mikhail notes, "[a]ny normative system purporting to achieve descriptive adequacy must presumably include a set of basic legal prohibitions" (2009, p. 53). Drawing on language acquisition and processing as a model, Mikhail and other proponents of "universal moral grammar" contend that moral judgments can be decomposed into their constituent cognitive building blocks, and that the rules for combining those modular

building blocks are predictably computational (Hauser, 2006; Mikhail, 2007). From these basic elements of moral thinking, they argue, humans construct surprisingly robust moral intuitions, including many that are reflected in common legal principles. Mikhail has used models of moral grammar to describe and analyze psychological and legal categories of intent, integrating legal distinctions among labels like "intentional" and "reckless" with familiar psychological evidence on, for example, people's reactions to trolley dilemmas (Mikhail, 2009).

Mikhail hypothesizes that the relationship between law and morality is apparent early in development, writing, "the intuitive jurisprudence of young children is complex and exhibits many characteristics of a well-developed legal code"[3] (2007). However, so far there has been very little research with children on the development of moral grammar or the extent to which childhood reasoning resembles legal doctrine. An intuitive jurisprudence approach provides exactly the kind of tool that such research would require.

Fairness and Procedural Justice

Identifying the ways in which law and psychology diverge is a major aim of intuitive jurisprudence research. Importantly, I am not suggesting that the law should perfectly match lay intuitive psychology in all cases; instead, I argue only that there is significant value to understanding the conflicts between the two. Doing so can help both fields understand one

[3] Mikhail's concept of "intuitive jurisprudence" is a computational theory of the way human reasoning generally, and computations of moral grammar in particular, may reflect formalized legal ideas. In this sense, fundamental legal conceptions can shed light on a range of human moral intuitions in a variety of cases. Our use of the term is very similar, although we aim to broaden the scope of intuitive jurisprudence research to open a reciprocal dialogue between legal ideas and human reasoning research, and also to shed light on the psychological processes that govern people's interactions with the law.

another, but it can also help policymakers craft and present laws that are understood better by the general public, increasing the perceived fairness of the legal system. Research in psychology has demonstrated that perceptions of procedural fairness—or procedural justice—are centrally important to the perceived legitimacy of the system (MacCoun & Tyler, 1988; Tyler & Sevier, 2013). Evidence suggests that people who interact with the legal system are more satisfied with case outcomes when they feel that they have been treated fairly (Casper, Tyler, & Fisher, 1988; Tyler, 1987); even unpopular decisions by the U.S. Supreme Court seem to gain wider public acceptance when the process by which the decisions were made is perceived as fair (Tyler & Rasinski, 1991).

Given the far-reaching effects that perceptions of procedural justice can have, the value of understanding what contributes to (or detracts from) those perceptions is clear. Because fairness concerns evolve during development and reflect both intuitive components and input from cultural experiences, an intuitive jurisprudence approach could be especially useful for understanding the development of people's interactions with the law. Regarding early intuitions, research has shown that even pre-verbal infants are sensitive to inequity and inequality (Geraci & Surian, 2011). By three years old, children show surprisingly nuanced sensitivities to fairness that go beyond mere self-interest (Blake, McAuliffe, & Warneken, 2014). Indeed, research has shown that children seem to weigh and balance concerns for fairness and preferences for self-serving favoritism across a variety of situations (Shaw, DeScioli, & Olson, 2012). Nonetheless, notions of fairness also evolve differently across different cultures. By middle childhood, children across societies object to disadvantageous inequity (getting less than someone else), whereas objections to advantageous inequity (getting more than someone else) are much more culturally variable (Blake et al., 2015). Likewise, work that compares human behavior across

cultures finds substantial variation in how people choose to divide resources themselves and under what circumstances they will reject unequal distributions (Henrich et al., 2001).

But do fairness concerns in children bear any relationship to procedural justice concerns? Initial research suggests that it does. Evidence suggests that school-age children are sensitive to procedural concerns when evaluating punishments given in hypothetical stories (Gold, Darley, Hilton, & Zanna, 1984). More recently, work by Alex Shaw and Kristina Olson (2014) demonstrates that, as it is for adults, procedural fairness is of paramount importance to children. Children who are given the choice between a fair procedure and an unfair procedure will choose to distribute resources using the fair procedure, and they will even discard a resource when they cannot distribute it with a fair process (Shaw & Olson, 2014).

In the context of procedural justice, some notion of "fairness" may be a universal concern, but different communities and cultures may develop different standards for how fairness is perceived (Henrich, Heine, & Norenzayan, 2010). Similarly, in law, disparate communities may arrive at very different legal processes and rules (see, e.g., Van Hoecke & Warrington, 1998), which may ultimately reflect these differing underlying psychological intuitions. Even within a single society, communities may have wildly different views of the inherent fairness of legal procedures and actors, and these views can have serious consequences (e.g., D. Cohen & Nisbett, 1994). For example, the degree to which citizens trust the police varies greatly across different racial and ethnic groups, and research suggests that these differences are most strongly related to perceptions of procedural justice—specifically, perceptions of how fair the police are in the exercise of their authority (Tyler, 2005). Going forward, intuitive jurisprudence research could continue to engage these questions, bringing to bear both child-focused research and cross-

cultural work, and making more explicit connections between the legal and the psychological work on these topics.

Evidentiary Issues and Trust in Testimony

Another topic of special interest to both lawyers and psychologists is the issue of trust. The legal rules that govern the admissibility of evidence are designed not only to protect the rights of the parties, but also to ensure that the evidence that is presented to judges and juries is trustworthy. The rules establish prohibitions, guidelines, and exceptions that generally serve as proxies for trustworthiness, particularly when dealing with hearsay testimony (e.g., Posner, 2016; Saltzburg, 2016) and expert witnesses (*Daubert v. Merrell Dow Pharmaceuticals,* 1993). The rules implicitly rely on an understanding (or, perhaps more accurately, an assumption) of how fact finders will process and weight the potential testimony; thus, for example, the Federal Rules of Evidence prohibit evidence whose "prejudicial impact" on fact finders outweighs its "probative value" (Fed. R. Evid. 403, 2014).

The question of how people process and evaluate the truth value of information—especially the physically unverifiable testimony of others—is vitally important to psychology as well. Humans learn a remarkable amount from testimony as opposed to from personal observation. If you reflect on all the scientific knowledge or historical facts you know about the world—the earth is round and orbits the sun, George Washington was the first president of the United States, Western Europe has experienced divisive wars—these were learned via testimony from others rather from direct experiences. Developmental research has explored how children evaluate information from others, and how they determine which information is trustworthy and which is not. In this area, developmental psychology and law are raising complementary questions, and intuitive jurisprudence could help provide answers for both fields.

Developmental research suggests that children intuitively trust the testimony of others, and as such they can be quite gullible at times. Preschoolers, for example, trust the testimony of an informant, even when they have knowledge to suggest she is ignorant or even purposefully misleading them (Jaswal, 2010; Jaswal, Croft, Setia, & Cole, 2010; Palmquist & Jaswal, 2012). Yet, despite children's early inclination to see other people as inherently truthful, children can nonetheless use fairly sophisticated reasoning to evaluate and compare the credibility of others' testimony. For example, children credit informants who have a history of providing reliable information over informants who have a history of unreliability (Harris & Corriveau, 2011; Koenig & Harris, 2005; Rakoczy, Warneken, & Tomasello, 2009), and children understand that people can be experts in one domain but not in another (Koenig & Jaswal, 2011). These considerations are consistent with legal rules about the presentation of evidence; courts will often allow evidence that a witness has previously been unreliable, and the expertise of a witness's testimony is often scrutinized (Fed. R. Evid. 608, 2014; Fed. R. Evid. 701, 2014).

But the research on children's evaluations of testimony also raises some troubling implications for the law. A variety of extraneous social information seems to influence children's evaluations. Whether an informant seems *nice* or *mean* to children impacts children's judgments of her credibility (Landrum, Mills, & Johnston, 2013; Lane, Wellman, & Gelman, 2013); these character judgments can even outweigh information about the informant's past accuracy. An informant's social group membership, and her social relations with others, also impacts children's judgments of her trustworthiness. An informant with a foreign-sounding accent is trusted less than one with a native-sounding accent, even when both speakers' words are gibberish (Kinzler, Corriveau, & Harris, 2011). Adults, too, interpret foreign-accented speech as presenting information that is less credible (Lev-Ari & Keysar, 2010). In general, children tend

to trust familiar people over unfamiliar people, a heuristic that is especially prevalent among very young children (Corriveau & Harris, 2009). Children also choose to discredit an informant who has previously violated the norms of his or her social group (Harris & Corriveau, 2011), and they discredit the counsel of an informant whose past testimony deviated from that of the majority opinion (Harris & Corriveau, 2011).

On one hand, this body of research demonstrating the impact of social factors on early judgments of trust could suggest the potential perniciousness of these biases in a legal context. On the other hand, however, research has also shown that with age, children become more sensitive to cues to reliability. In some circumstances, information about reliability can supersede information about social familiarity in guiding children's judgments (Corriveau, Kinzler, & Harris, 2013; Corriveau & Harris, 2009); this research also suggests the possibility of individual and cultural differences pertaining to children's reliance on social familiarity and consensus (Corriveau, Kim, Song, & Harris, 2013; Corriveau et al., 2009). Further research, using an intuitive jurisprudence approach, could provide valuable insight into the degree to which these biases persist into adulthood, which kinds of social information are most resistant to updating, and how children and adults in different social and cultural contexts may differentially evaluate the credibility of evidence.

The law of evidence raises reciprocally intriguing questions for developmental psychology. Many exceptions to the common legal prohibition against hearsay are based on the intuition that statements made in certain circumstances are inherently more trustworthy. So, for example, hearsay rules may exempt statements made that undermine the interests of the speaker or statements made out of apparently uncontrollable excitement (Fed. R. Evid. 803, 2014). Developmental psychology could explore whether these legal intuitions are reflected in early

psychological intuitions. Do children believe that a statement against the speaker's interests is more likely to be true than one that is consistent with her interests? Do they view an excited utterance as more reliable than a deliberate one? If so, when and how do they begin to do so?

Social Biases and Equal Protection Under the Law

A final area that is ripe for intuitive jurisprudence would use research on the development of social cognition to help evaluate law and policy on bias and the Constitution's guarantee of equal protection. For years, legal scholars have documented ways in which social biases seem to creep into the legal system, particularly the criminal justice system, in an attempt to understand how the social category membership of the defendant may impact decisions of guilt and punishment. As an example, one 2001 study of the U.S. Federal Courts found that, holding constant the severity of the offense, nature of the offense, criminal history of the defendant, and location of the crime, people of color received prison sentences between five and twelve percent longer than those of white defendants; the same study found that, again controlling for details of the offense, the sentences of male defendants were about twelve percent longer than those of female defendants (Mustard, 2001) Other research, including work in juvenile courts, has noted significant disparities in the treatment and outcomes of defendants based on the defendant's ethnicity (e.g., Demuth & Steffensmeier, 2004), social class (e.g., Zatz, 1987), and gender (e.g., Bishop & Frazier, 1992; Horowitz & Pottieger, 1991). Racial disparities, in particular, have been extensively studied (Demuth & Steffensmeier, 2004; Everett & Wojtkiewicz, 2002; Mustard, 2001; Steffensmeier & Demuth, 2000; Stevenson & Friedman, 1994; Zatz, 1987), and it is not only the defendant's race that seems to play a significant role in leading to these disparities. Holding many other variables constant, murders of white victims are more likely than murderers

of black victims to be sentenced to death (Baldus, Pulaski, & Woodworth, 1983; Baldus, Woodworth, & Pulaski, 1985; Baldus, Woodworth, Zuckerman, Weiner, & Broffitt, 1998; United States General Accounting Office, 1990).

Research from psychology, too, has sought to understand the mechanisms by which social biases impact legal thinking and decision-making. Using a variety of methods, including mock jury studies (e.g., Gutek et al., 1999; Sommers, 2006; Sommers & Ellsworth, 2000), implicit association tasks (e.g., Levinson, Cai, & Young, 2010; Rachlinski, Johnson, Wistrich, & Guthrie, 2009), and visual processing measures (e.g., Eberhardt, Goff, Purdie, & Davies, 2004; Kleider, Cavrak, & Knuycky, 2012), psychologists have demonstrated that outcomes can be influenced by the social group memberships of all key players—the victim (e.g., George & Martinez, 2002), perpetrator (e.g., Sommers & Ellsworth, 2000), attorneys (Riger, Foster-Fishman, Nelson-Kuna, & Curran, 1995), judge (e.g., Rachlinski et al., 2009), and jury (e.g., Sommers, 2006). In some of the most striking research in this area, Jennifer Eberhardt and colleagues have found that perceptions of a defendant's race impact the responses of legal decision makers (Eberhardt et al., 2004). In their seminal paper titled "Looking deathworthy" (Eberhardt, Davis, Purdie-Vaughns, & Johnson, 2006), Eberhardt and colleagues examined real capital cases; they found that in murder cases involving a white victim and a black defendant, the more stereotypically black the defendant's facial features, the more likely he or she is to be sentenced to death.

Developmental psychology methods provide a profitable tool to further clarify the mechanisms by which humans create social categories and evaluate others based on their social group membership, and how social factors play a role in various decision-making processes. Research with young children has shown that they attend to many of the same social variables

(e.g., gender, race, age) as do adults (e.g., Dunham, Baron, & Banaji, 2008; Heiphetz, Spelke, & Banaji, 2013; Kinzler & Dautel, 2012; Kinzler, Dupoux, & Spelke, 2007; Kinzler & Spelke, 2011; Rhodes & Chalik, 2013). By studying children's early intuitions, and how these intuitions change over time, this research has the potential to ask about the robustness of early social categorization, the degree to which such categorization is dependent on experience, and how social thinking changes across the lifespan. To give an example, research on children's thinking about gender and race showcases multiple ways in which children attend to gender over race. Children automatically encode gender more reliably than they encode race (Shutts, Pemberton Roben, & Spelke, 2013), and they think of gender as a more objective category, whereas they see race as more flexible (Rhodes & Gelman, 2009). Yet, children's perceptions of race change with developmental and cultural context – for instance, children in more conservative environments come to see both race and gender as less essential features of a person's identity than do children in more liberal environments (Rhodes & Gelman, 2009), and minority-race children have an earlier knowledge of race-based categorization than do majority-race children (Kinzler & Dautel, 2012; Roberts & Gelman, 2015, 2016). Recent research with biracial children further highlights the changing nature of race in the United States, and the psychological consequences of this for people of multiple racial groups (Gaither, 2015; Gaither et al., 2014).

Research with children can reveal the priorities with which humans assign weight to social categories, and how this may change across different contexts. Such research may also open up new avenues for thinking about the nature of protected classes, and how social variables may or may not impact legal decision-making. For instance, research suggests that social status, class, and prestige are variables of human societies that infants and young children highly attuned to (e.g., Henrich & Gil-White, 2001; Mascaro & Csibra, 2012; Horowitz, Shutts, &

Olson, 2015). As such, differences in the trustworthiness of informants and victims may vary reliably based on their perceived status.

Likewise, research shows that language—and accent specifically—is a primary marker by which children divide their social worlds, and that in many cases, accent matters more than race (Kinzler, Shutts, Dejesus, & Spelke, 2011). This idea that accent serves as a marker of group membership provides some interesting points of contact with legal issues surrounding accent-based discrimination. As described by legal scholar Mari Matsuda (1991) national origin is a protected category under federal discrimination law, but accented speech is not. If a person is fired from her job due to a failure to communicate effectively because her accent is deemed incompressible, this is permissible. The problem with this approach from a psychological perspective is that communication is two-sided. People can "turn off" when they do not want to listen. Research in sociolinguistics shows that when undergraduate students evaluate their teachers, they rate the *same voice* as sounding accented when they think that the person is Asian rather than White (Rubin, 1992). And, the more conservative people's attitudes, the more they think that they don't understand an accented speaker, even when they actually do (Hansen & Dovidio, in press). Thus, research from psychology, and particularly developmental psychology, may shed light on the processes by which language guides decisions that have legal relevance. Adults evaluate statements read by foreign-accented speakers as being less likely to be true (Lev-Ari & Keysar, 2010), and research with children suggests that, beginning early in life people are less likely to seek out novel information from foreign-accented people (Kinzler, Corriveau, & Harris, 2011). This research could inform the law by shedding light on social variables that, even when people are not aware of them, meaningfully influence decisions.

Conclusion

Increased dialogue between developmental psychology and law could prompt researchers in both fields to consider new questions and perspectives. Leveraging knowledge about cognition and social understanding from developmental psychology may help people understand law in a new way; likewise, research in the law may help formulate new theories and hypotheses in developmental psychology. This approach brings together disparate fields of research, but it also broadens the scope and impact of research in those fields.

Importantly, intuitive jurisprudence is not only—or even primarily—about understanding children. Research with children as participants can illuminate adult psychology as well, and further an understanding of the component parts of adult psychology that are present early in life and may continue to guide cognition and learning across the lifespan. On this point, research into the development of scientific and social knowledge may provide a particularly useful guide. For instance, past research in the cognitive sciences has fruitfully investigated intuitive reasoning and early cognitive capacities in areas such as physics (Baillargeon, 1994; McCloskey, 1983) and biology (Carey, 1985; Inagaki & Hatano, 2006); similar approaches are beginning to investigate the origins of a naïve sociology, or the cognitive processes that underlie everyday thinking about social groups and behavior (Rhodes, 2013; Sperber & Hirschfeld, 2001). This core knowledge approach suggests that the human mind is calibrated to reason about entities differentially across different domains in productive ways (Kinzler & Spelke, 2007). Often, signatures of adulthood reasoning are visible as early as infancy, though these cognitive capacities can shift with age and be recombined to create new forms of knowledge that emerge with development and experience (Carey & Spelke, 1994). Going forward, intuitive jurisprudence will contribute to a related approach by investigating the conceptual beginnings of legally-relevant topics, which will

likewise inform an understanding of the building blocks of human cognition and the ways in which mature legal knowledge reflects and improves on early-emerging constructs.

Capitalizing on approaches from developmental psychology also opens the door to thinking about questions cross-cultural psychology. These fields share a similar quest to understand the nature of human universals and human differences, and intuitive jurisprudence research could benefit from the broad perspective they provide. Cross-cultural evolutionary work has explored a number of legally-relevant topics, including the willingness to engage in costly third-party punishment (Henrich et al., 2006), conceptions of ownership and property (Rochat et al., 2014), beliefs about fairness and resource allocation (Henrich et al., 2001), and emerging ideas about intellectual property (Yang, Shaw, Garduno, & R. Olson, 2014). In addition to the clear contributions these approaches make to broadening our understanding of culture and universality, looking at where the results converge and diverge across cultures could help explain downstream differences and similarities among legal practices and attitudes around the world. Likewise, evolutionarily-focused research with non-human primates reveals some intuitive bases to adults' judgments and decision-making biases (e.g., intertemporal choice and risk preferences, Santos & Rosati, 2015). Intuitive jurisprudence inquiries, combining a range of disciplinary approaches, can contribute to an understanding of the human psychology surrounding legally-relevant decisions.

REFERENCES

Abel, G. G., Becker, J. V., & Cunningham-Rathner, J. (1984). Complications, consent, and cognitions in sex between children and adults. *International Journal of Law and Psychiatry*, *7*, 89–103.

Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, *95*(9–10), 1082–1095. http://doi.org/10.1016/j.jpubeco.2011.03.003

Anderson, M. J. (2016). Campus sexual assault adjudication and resistance to reform. *Yale Law Journal*, *125*(7), 1940–2005.

Ardila-Rey, A., & Killen, M. (2001). Middle class Colombian children's evaluations of personal, moral, and social-conventional interactions in the classroom. *International Journal of Behavioral Development*, *25*(3), 246–255. http://doi.org/10.1080/01650250042000221

Armsby, R. E. (1971). A reexamination of the development of moral judgments in children. *Child Development*, *42*(4), 1241–1248. http://doi.org/10.2307/1127807

Baillargeon, R. (1994). How Do Infants Learn About the Physical World? *Current Directions in Psychological Science*, *3*(5), 133–140.

Baldus, D. C., Pulaski, C., & Woodworth, G. (1983). Comparative review of death sentences: An empirical study of the Georgia experience. *Journal of Criminal Law & Criminology*, *74*(3), 661–753. http://doi.org/10.2307/1143133

Baldus, D. C., Woodworth, G., & Pulaski, C. A. (1985). Monitoring and Evaluating Contemporary Death Sentencing Systems: Lessons from Georgia. *U.C. Davis Law Review*, *18*, 1375–1408.

Baldus, D. C., Woodworth, G., Zuckerman, D., Weiner, N. A., & Broffitt, B. (1998). Racial discrimination and the death penalty in the post-Furman era: an empirical and legal overview, with recent findings from Philadelphia. *Cornell Law Review*, *83*(6), 1638.

Bandes, S. (1999). Victim Standing. *Utah Law Review*, (2), 331–349.

Banerjee, R., Bennett, M., & Luke, N. (2010). Children's reasoning about the self-presentational consequences of apologies and excuses following rule violations. *The British Journal of Developmental Psychology*, *28*(Pt 4), 799–815. http://doi.org/10.1348/026151009X479475

Barrett, H. C., Bolyanatz, A., Crittenden, A. N., Fessler, D. M. T., Fitzpatrick, S., Gurven, M., … Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proceedings of the National Academy of Sciences*, 1–6. http://doi.org/10.1073/pnas.1522070113

Bilz, K. (2007). The puzzle of delegated revenge. *Boston University Law Review*, *87*, 1059–1112.

Bilz, K. (2010). We Don't Want to Hear It: Psychology, Literature and the Narrative Model of Judging. *Illinois Law Review*, *2010*, 429–488.

Bilz, K. (2016). Testing the expressive theory of punishment. *Journal of Empirical Legal Studies*, *13*(2), 358–392. http://doi.org/10.1111/jels.12118

Bilz, K., & Darley, J. M. (2004). What's Wrong with Harmless Theories of Punishment. *Chicago-Kent Law Review*, *79*, 1215–1252.

Bishop, D. M., & Frazier, C. E. (1992). Gender Bias in Juvenile Justice Processing: Implications of the Jjdp Act. *Journal of Criminal Law & Criminology*, *82*(4), 1162–1186. http://doi.org/10.2307/1143718

Blake, P. R., McAuliffe, K., Corbit, J., Callaghan, T. C., Barry, O., Bowie, A., … Warneken, F. (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258–261. http://doi.org/10.1038/nature15703

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge-behavior gap. *Trends in Cognitive Sciences*, *18*(11), 559–561. http://doi.org/10.1016/j.tics.2014.08.003

Bloom, P. (2010, May 5). The Moral Life of Babies. *The New York Times Magazine*. New York.

Boeckmann, R. J., & Tyler, T. R. (1997). Commonsense Justice and Inclusion within the moral community When Do People Receive Procedural Protections From Others ? *Psychology, Public Policy, and Law*, *3*(2), 362–380.

Bond, C. E. W., & Jeffries, S. (2014). Similar punishment?: Comparing sentencing outcomes in domestic and non-domestic violence cases. *British Journal of Criminology*, *54*(5), 849–872. http://doi.org/10.1093/bjc/azu034

Borsari, B., & Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: a meta-analytic integration. *Journal of Studies on Alcohol*, *64*(3), 331–41. http://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology*, *32*(3), 467–478. http://doi.org/10.1037/0012-1649.32.3.467

Bregant, J., & Robbennolt, J. K. (2013). Neither admit nor deny. *Monitor on Psychology*, *44*(8), 26.

Bregant, J., Shaw, A., & Kinzler, K. D. (2016). Intuitive Jurisprudence: Early Reasoning about the Functions of Punishment. *Journal of Empirical Legal Studies*, *13*(4), 693–717. http://doi.org/10.1017/CBO9781107415324.004

Brodin, M. S. (2016). The Murder of Black Males in a World of Non-Accountability : The Surreal Trial of George Zimmerman for the Killing of Trayvon Martin. *Howard Law Journal*, *59*(3), 765–786.

Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, *60*(5), 930–40. http://doi.org/10.1016/j.neuron.2008.10.016

Bulbulia, J., & Sosis, R. (2011). Signalling theory and the evolution of religious cooperation. *Religion*, *41*(3), 363–388. http://doi.org/10.1080/0048721X.2011.604508

Burg, W. van der. (1991). The Slippery Slope Argument. *Ethics*, *102*(1), 42–65.

Butterfill, S. a., & Apperly, I. a. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. http://doi.org/10.1111/mila.12036

Byrd, B. S. (1989). Kant's Theory of Punishment: Deterrence in its Threat, Retribution in its Execution. *Law and Philosophy*, *8*, 151–200.

Cahill, M. T. (2007). Retributive justice in the real world. *Washington University Law Review*, *85*(4), 815.

Capestany, B. H., & Harris, L. T. (2014). Disgust and biological descriptions bias logical reasoning during legal decision-making. *Social Neuroscience*, *9*(3), 265–277.

Carey, S. (1985). Are Children Fundamentally Different Kinds of Thinkers and Learners than Adults? In S. F. Chipman, J. W. Segal, & R. Glaser (Eds.), *Thinking and Learning Skills, Volume 2: Research and Open Questions* (pp. 485–518). Mahwah, New Jersey: Routledge.

Carey, S., & Spelke, E. (1994). Domain-Specific Knowledge and Coceptual Change. In L. A. Hirschfeld & S. A. Gelman (Eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture* (pp. 169–200). Cambridge: Cambridge University Press.

Carlsmith, K. M. (2006). The Roles of Retribution and Utility in Determining Punishment. *Journal of Experimental Social Psychology*, *42*(4), 437–451. http://doi.org/10.1016/j.jesp.2005.06.007

Carlsmith, K. M. (2008). On Justifying Punishment: The Discrepancy Between Words and Actions. *Social Justice Research*, *21*(2), 119–137. http://doi.org/10.1007/s11211-008-0068-x

Carlsmith, K. M., Darley, J. M., & Robinson, P. H. (2002). Why Do We Punish? Deterrence and Just Deserts as Motives for Punishment. *Journal of Personality and Social Psychology*, *83*(2), 284–299.

Carroll, J. S., Perkowitz, W. T., Lurigio, A. J., & Weaver, F. M. (1987). Sentencing Goals, Causal Attributions, Ideology, and Personality. *Journal of Personality and Social Psychology*, *52*(1), 107–118.

Carter, C. A. (2017). Police Brutality, the Law & Today's Social Justice Movement: How the Lack of Police Accountability Has Fueled #Hashtag Activism. *CUNY L. Rev.*, *20*, 521.

Caruso, E. M. (2010). When the future feels worse than the past: A temporal inconsistency in moral judgment. *Journal of Experimental Psychology: General*, *139*(4), 610–624. http://doi.org/10.1037/a0020757

Casper, J. D., Tyler, T., & Fisher, B. (1988). Procedural Justice in Felony Cases. *Law & Society Review*, *22*(3), 483. http://doi.org/10.2307/3053626

Chakroff, A., Dungan, J., & Young, L. (2013). Harming Ourselves and Defiling Others: What Determines a Moral Domain? *PLoS ONE*, *8*(9), e74434. http://doi.org/10.1371/journal.pone.0074434

Chakroff, A., Russell, P. S., Piazza, J., & Young, L. (2017). From impure to harmful: Asymmetric expectations about immoral agents. *Journal of Experimental Social Psychology*, *69*, 201–209. http://doi.org/10.1016/j.jesp.2016.08.001

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, *72*(2), 263–268. http://doi.org/10.1007/s11336-006-1560-6

Cialdini, R. B., Demaine, L. J., Sagarin, B. J., Barrett, D. W., Rhoads, K., & Winter, P. L. (2006). Managing social norms for persuasive impact. *Social Influence*, *1*(1), 3–15. http://doi.org/10.1080/15534510500181459

Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. *Advances in Experimental Social Psychology*, *24*(C), 201–234. http://doi.org/10.1016/S0065-2601(08)60330-5

Cohen, D., & Nisbett, R. E. (1994). Self-Protection and the Culture of Honor: Explaining Southern Violence. *Personality and Social Psychology Bulletin*, *20*(5), 551–567. http://doi.org/10.1177/0146167294205012

Cohen, J. (1999). Advising clients to apologize. *Southern California Law Review*, *72*, 1009–1069.

Cole, H. L., Mailath, G. J., & Postlewaite, A. (1992). Social Norms, Savings Behavior, and Growth. *Journal of Political Economy*, *100*(6), 1092–1125.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39–67. http://doi.org/10.1177/0149206310388419

Cooter, R. (1998). Expressive Law And Economics. *The Journal of Legal Studies*, *27*(S2), 585–607. http://doi.org/10.1086/468036

Cooter, R. (2000a). Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms. *Virginia Law Review*, *86*(8), 1577–1601.

Cooter, R. (2000b). Three Effects of Social Norms on Law: Expression, Deterrence, and

Internalization. *Oregon Law Review*, *79*, 1–22.

Cornell, N. (2017). *The Possibility of Preemptive Forgiving*. *Philosophical Review* (Vol. 126). http://doi.org/10.1215/00318108-3772018

Corriveau, K. H., Kinzler, K. D., & Harris, P. L. (2013). Accuracy trumps accent in children's endorsement of object labels. *Developmental Psychology*, *49*(3), 470–9. http://doi.org/10.1037/a0030604

Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to Others Outweighs Harm to Self in Moral Decision Making. *Proceedings of the National Academy of Sciences*, *111*(48), 17320–17325. http://doi.org/10.1073/pnas.

Cullen, F. T., Skovron, S. E., Scott, J. E., & Burton, V. S. (1990). Public Support for Correctional Treatment: The Tenacity of Rehabilitative Ideology. *Criminal Justice and Behavior*, *17*(1), 6–18. http://doi.org/10.1177/0093854890017001003

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353–380. http://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in Psychology*, *6*, 97–103. http://doi.org/10.1016/j.copsyc.2015.06.003

Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a "trembling hand" game. *PLoS ONE*, *4*(8). http://doi.org/10.1371/journal.pone.0006699

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, *108*(1), 281–289. http://doi.org/10.1016/j.cognition.2008.02.005

Cushman, F., Sheketoff, R., Wharton, S., & Carey, S. (2013). The development of intent-based moral judgment. *Cognition*, *127*(1), 6–21. http://doi.org/10.1016/j.cognition.2012.11.008

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science*, *17*(12), 1082–9. http://doi.org/10.1111/j.1467-9280.2006.01834.x

Darley, J. M., & Alter, A. L. (2013). Behavioral Issues of Punishment, Retribution, and Deterrence. In E. Shafir (Ed.), *The Behavioral Foundations of Policy*. Princeton University Press.

Darley, J. M., Carlsmith, K. M., & Robinson, P. H. (2000). Incapacitation and Just Deserts as Motives for Punishment. *Law and Human Behavior*, *24*(6), 659–83.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *7*(4), 324–336.

http://doi.org/10.1207/S15327957PSPR0704_05

Darley, J. M., & Shultz, T. R. (1990). Moral Rules: Their Content And Acquisition. *Annual Review of Psychology*, *41*(1), 525–556. http://doi.org/10.1146/annurev.psych.41.1.525

Daubert v. Merrell Dow Pharmaceuticals, 509 U.S. 579 (1993).

De Pelsmacker, P., & Janssens, W. (2007). The effect of norms, attitudes and habits on speeding behavior: Scale development and model building and estimation. *Accident Analysis and Prevention*, *39*(1), 6–15. http://doi.org/10.1016/j.aap.2006.05.011

Demuth, S., & Steffensmeier, D. (2004). Ethnicity effects on sentence outcomes in large urban courts: Comparisons among white, black, and Hispanic defendants. *Social Science Quarterly*, *85*(4), 994–1011. http://doi.org/10.1111/j.0038-4941.2004.00255.x

DeScioli, P., Bruening, R., & Kurzban, R. (2011). The omission effect in moral cognition: Toward a functional explanation. *Evolution and Human Behavior*, *32*(3), 204–215. http://doi.org/10.1016/j.evolhumbehav.2011.01.003

DeScioli, P., Christner, J., & Kurzban, R. (2011). The Omission Strategy. *Psychological Science*, *22*(4), 442–446. http://doi.org/10.1177/0956797611400616

Dhami, M. K. (2011). Offer and Acceptance of Apology in Victim-Offender Mediation. *Critical Criminology*, *20*(1), 45–60. http://doi.org/10.1007/s10612-011-9149-5

Dharmapala, D., & McAdams, R. H. (2003). The Condorcet Jury Theorem and the Expressive Function of Law: A Theory of Informative Law. *American Law and Economics Review*, *5*(217), 1–31. http://doi.org/10.1093/aler/5.1.1

Diamond, S. S., Vidmar, N., Rose, M., Ellis, L., Murphy, B., & Trienens, H. J. (2003). Jury Discussions During Civil Trials: Studying an Arizona Innovation. *Ariz. L. Rev.*, *45*, 1–81.

Dillehay, R. C. (1996). Life under Wainwright v. Witt: Juror dispositions and death qualification. *Law and Human Behavior*, *20*(2), 147–165. http://doi.org/10.1007/BF01499352

Drinking and driving applies to all drinks. (n.d.). Retrieved July 10, 2017, from https://www.iusedtobelieve.com/transport/cars/drink_driving/

Dungan, J. A., Chakroff, A., & Young, L. (2017). The relevance of moral norms in distinct relational contexts: Purity versus harm norms regulate self-directed actions. *PLoS ONE*, *12*(3), 1–17. http://doi.org/10.1371/journal.pone.0173405

Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, *12*(June), 248–253. http://doi.org/10.1016/j.tics.2008.04.006

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing Black: Race, Crime, and Visual Processing. *Journal of Personality and Social Psychology*, *87*(6), 876–893.

http://doi.org/10.1037/0022-3514.87.6.876

Ellsworth, P. C., & Gross, S. R. (1994). Hardening of the Attitudes: Americans' Views on the Death Penalty. *Journal of Social Issues*, *50*(2), 19–52. http://doi.org/10.1111/j.1540-4560.1994.tb02409.x

Enright, R. D., & Kittle, B. A. (1999). Forgiveness in Psychology and Law: The Meeting of Moral Development and Restorative Justice. *Fordham Urb. L.J.*, *27*, 1621–1631.

Everett, R. S., & Wojtkiewicz, R. A. (2002). Difference, Disparity, and Race/Ethnic Bias in Federal Sentencing. *Journal of Quantitative Criminology*, *18*(2), 189–211.

Exline, J. J., Worthington Jr., E. L., Hill, P., & McCullough, M. E. (2003). Forgiveness and Justice: A Research Agenda for Social and Personality Psychology. *Personality and Social Psychology Review*, *7*(4), 337–348. http://doi.org/10.1207/S15327957PSPR0704

Farrington, D. P. (1986). Age and Crime. *Crime and Justice*, *7*(December), 189–250. http://doi.org/10.1086/449114

Fed. R. Evid. 403. (2014). Excluding Relevant Evidence for Prejudice, Confusion, Waste of Time, or Other Reasons.

Fed. R. Evid. 608. (2014). A Witness's Character for Truthfulness or Untruthfulness.

Fed. R. Evid. 701. (2014). Opinion Testimony by Lay Witnesses.

Fed. R. Evid. 803. (2014). Exceptions to the Rule Against Hearsay.

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

Ferrin, D. L., Kim, P. H., Cooper, C. D., & Dirks, K. T. (2007). Silence Speaks Volumes: The Effectiveness of Reticence in Comparison to Apology and Denial for Responding to Integrity- and Competence-Based Trust Violations. *Journal of Applied Psychology*, *92*(4), 893–908. http://doi.org/10.1037/0021-9010.92.4.893

Fincher, K. M., & Tetlock, P. E. (2015). Brutality Under Cover of Ambiguity: Activating, Perpetuating, and Deactivating Covert Retributivism. *Personality and Social Psychology Bulletin*. http://doi.org/10.1177/0146167215571090

Finkel, N. J. (2000). But it's not fair! Commonsense notions of unfairness. *Psychology, Public Policy, and Law*, *6*(4), 898–952. http://doi.org/10.1037//1076-8971.6.4.898

Finkel, N. J., Liss, M. B., & Moran, V. R. (1997). Equal or proportional justice for accessories? Children's pearls of proportionate wisdom. *Journal of Applied Developmental Psychology*, *18*, 229–244.

FitzGibbon, C. D., & Fanshawe, J. H. (1988). Stotting in Thomson's gazelles: an honest signal of

condition. *Behavioral Ecology and Sociobiology*, *23*(2), 69–74. http://doi.org/10.1007/BF00299889

Friedman, O., & Neary, K. R. (2009). First possession beyond the law: Adults' and young children's intuitions about ownership. *Tulane Law Review*, *83*(1991), 1–12.

Friedman, O., Van de Vondervoort, J. W., Defeyter, M. A., & Neary, K. R. (2013). First possession, history, and young children's ownership judgments. *Child Development*, *84*(5), 1519–1525. http://doi.org/10.1111/cdev.12080

Funk, P. (2004). On the effective use of stigma as a crime-deterrent. *European Economic Review*, *48*(4), 715–728. http://doi.org/10.1016/j.euroecorev.2003.11.003

Funk, P. (2007). Is there an expressive function of law? An empirical analysis of voting laws with symbolic fines. *American Law and Economics Review*, *9*(1), 135–159. http://doi.org/10.1093/aler/ahm002

Gaither, S. E. (2015). "Mixed" Results: Multiracial Research and Identity Explorations. *Current Directions in Psychological Science*, *24*(2), 114–119. http://doi.org/10.1177/0963721414558115

Gaither, S. E., Chen, E. E., Corriveau, K. H., Harris, P. L., Ambady, N., & Sommers, S. R. (2014). Monoracial and Biracial Children: Effects of Racial Identity Saliency on Social Learning and Social Preferences. *Child Development*, *85*(6), 2299–2316. http://doi.org/10.1111/cdev.12266

Garcia, S. M., Chen, P., & Gordon, M. T. (2014). The letter versus the spirit of the law : A lay perspective on culpability. *Judgment and Decision Making*, *9*(5), 479–490.

George, W. H., & Martinez, L. J. (2002). Victim Blaming in Rape: Effects of Victim and Perpetrator Race, Type of Rape, and Participant Racism. *Psychology of Women Quarterly*, *26*(2), 110–119. http://doi.org/10.1111/1471-6402.00049

Geraci, A., & Surian, L. (2011). The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Developmental Science*, *14*(5), 1012–1020. http://doi.org/10.1111/j.1467-7687.2011.01048.x

Gerber, A. S., & Rogers, T. (2009). Descriptive Social Norms and Motivation to Vote: Everybody's Voting and so Should You. *The Journal of Politics*, *71*(1), 178–191. http://doi.org/10.1017/S0022381608090117

Gert, H. J., Radzik, L., & Hand, M. (2004). Hampton on the expressive power of punishment. *Journal of Social Philosophy*, *35*(1), 79–90. http://doi.org/10.1111/j.1467-9833.2004.00217.x

Giffin, C., & Lombrozo, T. (2016). Wrong or merely prohibited: Special treatment of strict liability in intuitive moral judgment. *Law and Human Behavior*, *40*(6), 707–720. http://doi.org/10.1037/lhb0000212

Gino, F., Moore, D. A., & Bazerman, M. H. (2009). No harm, no foul: The outcome bias in ethical judgments. *Harvard Business School*, 1–51. http://doi.org/10.2139/ssrn.1099464

Ginther, M. R., Shen, F. X., Bonnie, R. J., Hoffman, M. B., Jones, O. D., Marois, R., & Simons, K. W. (2014). The language of mens rea. *Vanderbilt Law Review*, *67*(5), 1327.

Gneezy, U., & Rustichini, A. (2000). A Fine Is a Price. *The Journal of Legal Studies*, *29*(1), 1–17. http://doi.org/10.1086/468061

Gold, L. J., Darley, J. M., Hilton, J. L., & Zanna, M. P. (1984). Children's Perceptions of Procedural Justice. *Child Development*, *55*(5), 1752–1759.

Goldberg, J. H., Lerner, J. S., & Tetlock, P. E. (1999). Rage and reason: the psychology of the intuitive prosecutor. *European Journal of Social Psychology*, *29*(March 1998), 781–795.

Goodman, G. S., Batterman-Faunce, J. M., Schaaf, J. M., & Kenney, R. (2002). Nearly 4 years after an event: children's eyewitness memory and adults' perceptions of children's accuracy. *Child Abuse & Neglect*, *26*, 849–884.

Gray, K. (2014). Harm concerns predict moral judgments of suicide: Comment on Rottman, Kelemen and Young (2014). *Cognition*, *133*(1), 329–331. http://doi.org/10.1016/j.cognition.2014.06.007

Gray, K., & Keeney, J. E. (2015). Disconfirming Moral Foundations Theory on Its Own Terms: Reply to Graham (2015). *Social Psychological and Personality Science*, *6*(8), 874–877. http://doi.org/10.1177/1948550615592243

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600–1615. http://doi.org/10.1037/a0036149

Green, S. P. G. (2000). Review: The Universal Grammar of Criminal Law. *Michigan Law Review*, *98*(6), 2104–2125.

Greene, L. S. (2015). Before and after Michael Brown - Toward an End to Structural and Actual Violence. *Wash. U. J. L. & Pol'y*, *49*, 1.

Grisso, T. (1996). Society's Retributive Response to Juvenile Violence: A Developmental Perspective. *Law and Human Behavior*, *20*(3), 229–247.

Grisso, T., Steinberg, L., Woolard, J., Cauffman, E., Scott, E., Graham, S., … Schwartz, R. (2003). Juveniles' competence to stand trial: A comparison of adolescents' and adults' capacities as trial defendants. *Law and Human Behavior*, *27*(4), 333–363. http://doi.org/10.1023/A:1024065015717

Gutek, B. a., O'Connor, M. A., Melançon, R., Stockdale, M. S., Geer, T. M., & Done, R. S. (1999). The utility of the reasonable woman legal standard in hostile environment sexual harassment cases: A multimethod, multistudy examination. *Psychology, Public Policy, and*

Law, *5*(3), 596–629. http://doi.org/10.1037/1076-8971.5.3.596

Haidt, J. (2001). The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, *108*(4), 814–834. http://doi.org/10.1037//0033-295X.

Haidt, J. (2007). The new synthesis in moral psychology. *Science (New York, N.Y.)*, *316*(5827), 998–1002. http://doi.org/10.1126/science.1137651

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

Hamlin, J. K. (2013a). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, *128*(3), 451–474. http://doi.org/10.1016/j.cognition.2013.04.004

Hamlin, J. K. (2013b). Moral Judgment and Action in Preverbal Infants and Toddlers: Evidence for an Innate Moral Core. *Current Directions in Psychological Science*, *22*(3), 186–193. http://doi.org/10.1177/0963721412470687

Hamlin, J. K., Mahajan, N., Liberman, Z., & Wynn, K. (2013). Not Like Me = Bad: Infants Prefer Those Who Harm Dissimilar Others. *Psychological Science*, *24*(4), 589–594. http://doi.org/10.1177/0956797612457785

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557–560. http://doi.org/10.1038/nature06288

Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How Infants and Toddlers React to Antisocial Others. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(50), 19931–6. http://doi.org/10.1073/pnas.1110306108

Hampton, J. (1992). Correcting Harms versus Righting Wrongs: The Goal of Retribution. *UCLA Law Review*, *39*, 1659–1702.

Hanna, N. (2008). Say what? A critique of expressive retributivism. *Law and Philosophy*, *27*(2), 123–150. http://doi.org/10.1007/s10982-007-9014-6

Hans, V. P., Blume, J. H., Eisenberg, T., Hritz, A. C., Johnson, S. L., Royer, C. E., & Wells, M. T. (2015). The death penalty: Should the judge or the jury decide who dies? *Journal of Empirical Legal Studies*, *12*(1), 70–99. http://doi.org/10.1111/jels.12065

Hans, V. P., & Jehle, A. (2003). Avoid bald men and people with green socks? Other ways to improve the voir dire process in jury selection. *Chicago-Kent Law Review*, *78*, 1179.

Hans, V. P., & Reyna, V. F. (2011). To Dollars From Sense: Qualitative to Quantitative Translation in Jury Damage Awards. *Journal of Empirical Legal Studies*, (11).

Harcourt, B. E. (2002). Joel Feinberg on Crime and Punishment : Exploring the Relationship

Between The Moral Limits of the Criminal Law and The Expressive Function of Punishment. *Buffalo Criminal Law Review*, *1*(2000), 145.

Harman, J. D. (1983). Consent, harm, and marital rape. *Journal of Family Law*, *22*(3), 423. http://doi.org/10.1525/sp.2007.54.1.23.

Harris, P. L., & Corriveau, K. H. (2011). Young children's selective trust in informants. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1567), 1179–87. http://doi.org/10.1098/rstb.2010.0321

Hastie, C. A. R., Danks, D., Fernbach, P., Lagnado, D., Lempert, R., Lewis, C., … Sloman, S. (2015). Causal Thinking in Judgments (pp. 1–38).

Hauser, M. (2006). *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York, NY: Harper Collins.

Heiphetz, L., Spelke, E. S., & Banaji, M. R. (2013). Patterns of implicit and explicit attitudes in children and adults: tests in the domain of religion. *Journal of Experimental Psychology: General*, *142*(3), 864–79. http://doi.org/10.1037/a0029714

Helwig, C. C., Zelazo, P. D., & Wilson, M. (2001). Children's Judgments of Psychological Harm in Normal and Noncanonical Situations. *Child Development*, *72*(1), 66–81.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies. *American Economic Review*, *91*(2), 73–78.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61-83; discussion 83-135. http://doi.org/10.1017/S0140525X0999152X

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., … Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–70. http://doi.org/10.1126/science.1127333

Heyman, G. D., Sweet, M. A., & Lee, K. (2009). Children's reasoning about lie-telling and truth-telling in politeness contexts. *Social Development*, *18*(3), 728–746. http://doi.org/10.1111/j.1467-9507.2008.00495.x

Ho, B., & Liu, E. (2011). What 's an Apology Worth? Decomposing the Effect of Apologies on Medical Malpractice Payments Using State Apology Laws. *Journal of Empirical Legal Studies*, *8*(December), 179–199.

Horowitz, R., & Pottieger, A. E. (1991). Gender Bias in Juvenile Justice Handling of Seriously Crime-Involved Youths. *Journal of Research in Crime and Delinquency*, *28*(1), 75–100.

Inagaki, K., & Hatano, G. (2006). Young Children ' s Conception of the Biological World. *Psychological Science*, *15*(4), 177–181.

Inbar, Y., & Pizarro, D. (2009). Grime and punishment: How disgust influences moral, social, and legal judgments. *The Jury Expert*, *21*(2), 11–22.

Jaswal, V. K. (2010). Believing What You're Told: Young Children's Trust in Unexpected Testimony about the Physical World. *Cognitive Psychology*, *61*(3), 248–272. http://doi.org/10.1126/scisignal.2001449.Engineering

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young children have a specific, highly robust bias to trust testimony. *Psychological Science*, *21*(10), 1541–1547. http://doi.org/10.1177/0956797610383438

Kahan, D. M. (1996). What Do Alternative Sanctions Mean ? *The University of Chicago Law Review*, *63*(2), 591–653.

Kahan, D. M. (1997). Social Influence, Social Meaning, and Deterrence. *Virginia Law Review*, *83*(2), 349–395.

Kalish, C. (1998). Reasons and Causes: Children's Understanding of Conformity to Social Rules and Physical Laws. *Child Development*, *69*(3), 706–720. http://doi.org/10.1111/j.1467-8624.1998.tb06238.x

Kalish, C., & Sabbagh, M. A. (2007). Conventionality and Cognitive Development: Learning to Think the Right Way. *New Directions for Child and Adolescent Development*, (115), 1–9. http://doi.org/10.1002/cad

Kardas, M., Shaw, A., & Caruso, E. M. (2017). *How to Give Up Your Cake and Eat It, Too (working paper)*.

Kaul, P. (2015). Admit or Deny: A Call for Reform of the SEC's "Neither-Admit-Nor-Deny" Policy. *University of Michigan Journal of Law Reform*, *48*(2), 535.

Kelly, D., Stich, S., Haley, K. J., Eng, S. J., & Fessler, D. M. T. (2007). Harm, Affect, and the Moral/Conventional Distinction. *Mind & Language*, *22*(2), 117–131.

Killen, M., Lynn Mulvey, K., Richardson, C., Jampol, N., & Woodward, A. (2011). The accidental transgressor: Morally-relevant theory of mind. *Cognition*, *119*(2), 197–215. http://doi.org/10.1016/j.cognition.2011.01.006

Kim, P. H., Ferrin, D. L., Cooper, C. D., & Dirks, K. T. (2004). Removing the shadow of suspicion: the effects of apology versus denial for repairing competence- versus integrity-based trust violations. *The Journal of Applied Psychology*, *89*(1), 104–18. http://doi.org/10.1037/0021-9010.89.1.104

Kinzler, K. D., Corriveau, K. H., & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106–11. http://doi.org/10.1111/j.1467-7687.2010.00965.x

Kinzler, K. D., & Dautel, J. B. (2012). Children's essentialist reasoning about language and race.

Developmental Science, *15*(1), 131–8. http://doi.org/10.1111/j.1467-7687.2011.01101.x

Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*(30), 12577–12580. http://doi.org/10.1073/pnas.0705345104

Kinzler, K. D., Shutts, K., Dejesus, J., & Spelke, E. S. (2011). Accent trumps race in guiding children's social preferences, *27*(4), 1–11. http://doi.org/10.1521/soco.2009.27.4.623.Accent

Kinzler, K. D., & Spelke, E. S. (2007). Core systems in human cognition, *164*, 257–264. http://doi.org/10.1016/S0079-6123(07)64014-X

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race? *Cognition*, *119*(1), 1–9. http://doi.org/10.1016/j.cognition.2010.10.019

Kirchler, E., Hoelzl, E., & Wahl, I. (2008). Enforced versus voluntary tax compliance: The "slippery slope" framework. *Journal of Economic Psychology*, *29*(2), 210–225. http://doi.org/10.1016/j.joep.2007.05.004

Kleider, H. M., Cavrak, S. E., & Knuycky, L. R. (2012). Looking like a criminal: Stereotypical black facial features promote face source memory error. *Memory & Cognition*, 1200–1213. http://doi.org/10.3758/s13421-012-0229-x

Knobe, J. (2009). Folk judgments of causation. *Studies in History and Philosophy of Science Part A*, *40*(2), 238–242. http://doi.org/10.1016/j.shpsa.2009.03.009

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, *12*(2), 93–120. http://doi.org/10.1159/000270857

Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, *28*(2), 75–84. http://doi.org/10.1016/j.evolhumbehav.2006.06.001

Lagnado, D. a, & Channon, S. (2008). Judgments of cause and blame: the effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–70. http://doi.org/10.1016/j.cognition.2008.06.009

Landrum, A. R., Mills, C. M., & Johnston, A. M. (2013). When do children trust the expert? Benevolence information influences children's trust more than expertise. *Developmental Science*, *16*(4), 622–638. http://doi.org/10.1111/desc.12059

Lane, J. D., Wellman, H. M., & Gelman, S. A. (2013). Informants' traits weigh heavily in young children's trust in testimony and in their epistemic inferences. *Child Development*, *84*(4), 1253–1268. http://doi.org/10.1111/cdev.12029

Lerner, M. J. (1980). The belief in a just world. In *The Belief in a just World* (pp. 9–30). Springer.

Levine, S., & Leslie, A. (2015). Preschoolers use the “means principle” to make deontic moral judgments.

Levinson, J. D., Cai, H., & Young, D. (2010). Guilty by implicit racial bias: The guilty/not guilty Implicit Association Test. *Ohio State Journal of Criminal Law*, *8*(187), 187–208.

Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1–10. http://doi.org/10.3758/s13428-016-0727-z

Loftus, E. F. (1974). Reconstructing Memory: The Incredible Eyewitness. *Jurimetrics*, *15*, 188–193.

London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, and Law*, *11*(1), 194–226. http://doi.org/10.1037/1076-8971.11.1.194

MacCoun, R. J., & Tyler, T. R. (1988). The basis of citizen’s perceptions of the criminal jury - Procedural fairness, accuracy, and efficiency. *Law and Human Behavior*, *12*(3), 333–352. http://doi.org/10.1007/BF01044389

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*(2), 101–121. http://doi.org/10.1006/jesp.1996.1314

Martin, J. W., & Cushman, F. (2016). The Adaptive Logic of Moral Luck. *A Companion to Experimental Philosophy*, 190–202.

Matsuda, M. J. (1991). Voices of America: Accent, Antidiscrimination Law, and a Jurisprudence for the Last Reconstruction. *Yale Law Journal*, *100*(5), 1329. http://doi.org/10.2307/796694

Mazzone, J. (1999). When Courts Speak: Social Capital and Law’s Expressive Function. *Syracuse Law Review*, *49*, 1039–1066. http://doi.org/10.1525/sp.2007.54.1.23.

McAdams, R. H., & Nadler, J. (2005). Testing the Focal Point Theory of Legal Compliance: The Effect of Third-Party Expression in an Experimental Hawk/Dove Game. *J. Emp. L. Stud.*, *2*(1), 87–123. http://doi.org/10.1111/j.1740-1461.2005.00032.x

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, *134*(JANUARY), 1–10. http://doi.org/10.1016/j.cognition.2014.08.013

McCloskey, M. (1983). Naive Theories of Motion. In D. Gentner & A. L. Stevens (Eds.), *Mental Models*. Hillsdale, NJ: Erlbaum.

McCorkle, R. C. (1993). Research Note: Punish and Rehabilitate? Public Attitudes Toward Six Common Crimes. *Crime & Delinquency*, *39*(2), 240–252. http://doi.org/10.1177/0011128793039002008

McCullough, M. E. (2001). Forgiveness: Who Does It and How Do They Do It? *Current*

Directions in Psychological Science, *10*(6), 194–197. http://doi.org/10.1111/1467-8721.00147

McKee, I. R., & Feather, N. T. (2008). Revenge, Retribution, and Values: Social Attitudes and Punitive Sentencing. *Social Justice Research*, *21*(2), 138–163. http://doi.org/10.1007/s11211-008-0066-z

Mikhail, J. (2007). Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152. http://doi.org/10.1016/j.tics.2006.12.007

Mikhail, J. (2009). Moral Grammar and Intuitive Jurisprudence: A Formal Model of Unconscious Moral and Legal Knowledge. *Psychology of Learning and Motivation*, *50*, 27–100. http://doi.org/10.1016/S0079-7421(08)00402-7

Milgram, S., Bickman, L., & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology*, *13*(2), 79–82. http://doi.org/10.1037/h0028070

Model Penal Code § 2.02 (1962).

Molho, C., Tybur, J. M., Güler, E., Balliet, D., & Hofmann, W. (2017). Disgust and Anger Relate to Different Aggressive Responses to Moral Violations. *Psychological Science*, *28*(5), 609–619. http://doi.org/10.1177/0956797617692000

Moll, J., de Oliveira-Souza, R., Moll, F. T., Ignácio, F. A., Bramati, I. E., Caparelli-Dáquer, E. M., & Eslinger, P. J. (2005). The moral affiliations of disgust: A functional MRI study. *Cognitive and Behavioral Neurology*, *18*(1), 68–78.

Morris, R. D. (1986). Signalling and Agency Theories and Accounting Policy Choice. *Accounting and Business Research*, *18*(69), 47–69.

Mull, M. S., & Evans, E. M. (2010). Did she mean to do it? Acquiring a folk theory of intentionality. *Journal of Experimental Child Psychology*, *107*(3), 207–28. http://doi.org/10.1016/j.jecp.2010.04.001

Murphy, J. G., & Hampton, J. (1988). Introduction (Forgiveness and Mercy). In *Forgiveness and Mercy* (pp. 1–13). Cambridge, UK: Cambridge University Press.

Murphy, J. G., & Hampton, J. (1990). *Forgiveness and Mercy*. Cambridge University Press.

Murphy, R. (1988). The Significance of Victim Harms: Booth v. Maryland and the Philosophy of Punishment in the Supreme Court. *The University of Chicago Law Review*, *55*(4), 1303–1333.

Mustard, D. B. (2001). Racial, Ethnic, and Gender Disparities in Sentencing: Evidence from the U.S. Federal Courts. *The Journal of Law & Economics*, *44*(1), 285–314.

Neary, K. R., & Friedman, O. (2014). Young Children Give Priority to Ownership When

Judging Who Should Use an Object. *Child Development*, *85*(1), 326–337. http://doi.org/10.1111/cdev.12120

Neff, K. D., & Helwig, C. C. (2002). A constructivist approach to understanding the development of reasoning about rights and authority within cultural contexts. *Cognitive Development*, *17*(3–4), 1429–1450. http://doi.org/10.1016/S0885-2014(02)00126-0

Nobes, G., Panagiotaki, G., & Pawson, C. (2009). The influence of negligence, intention, and outcome on children's moral judgments. *Journal of Experimental Child Psychology*, *104*(4), 382–397. http://doi.org/10.1016/j.jecp.2009.08.001

North Carolina v. Alford (1970). Supreme Court.

Nucci, L. (1981). Conceptions of personal issues: A domain distinct from moral or societal concepts. *Child Development*, *52*(1), 114–121. http://doi.org/10.2307/1129220

Nucci, L., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*(2), 400–407. http://doi.org/10.2307/1128704

Nussbaum, M. C. (2013). *Anger: Downranking, Weakness, Payback*.

Olatunji, B. O., David, B., & Ciesielski, B. G. (2012). Who am I to judge? Self-disgust predicts less punishment of severe transgressions. *Emotion*, *12*(1), 169.

Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*(1), 222–231. http://doi.org/10.1016/j.cognition.2007.12.003

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. http://doi.org/10.1126/science.1107621

Orth, U. (2003). Punishment Goals of Crime Victims. *Law and Human Behavior*, *27*(2), 173–86.

Packin, N. G. (2014). Breaking Bad: Too-Big-to-Fail Banks Not Guilty as Not Charged. *Wash. U. L. Rev.*, *91*, 1089. http://doi.org/10.3366/ajicl.2011.0005

Palmquist, C. M., & Jaswal, V. K. (2012). Preschoolers expect pointers (even ignorant ones) to be knowledgeable. *Psychological Science*, *23*(3), 230–231. http://doi.org/10.1177/0956797611427043

Patil, I., Calò, M., Fornasier, F., Cushman, F., & Silani, G. (2017). The behavioral and neural basis of empathic blame. *Scientific Reports*, *7*(1), 1–14. http://doi.org/10.1038/s41598-017-05299-9

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the Story Model for juror decision making. *Journal of Personality and Social Psychology*, *62*(2), 189–206. http://doi.org/10.1037/0022-3514.62.2.189

Perkins, H. W. (2002). Social norms and the prevention of alcohol misuse in collegiate contexts.

Journal of Studies on Alcohol. Supplement, (14), 164–72. http://doi.org/http://dx.doi.org/10.15288/jsas.2002.s14.164

Perry, N. W., McAuliff, B. D., Tam, P., Claycomb, L., Dostal, C., & Flanagan, C. (1995). When lawyers question children. Is justice served? *Law and Human Behavior*, *19*(6), 609–629. http://doi.org/10.1007/BF01499377

Piaget, J. (1997). *The Moral Judgment of the Child*. New York, NY: Free Press Paperbacks. (Original work published 1932).

Piazza, J., Sousa, P., & Holbrook, C. (2013). Authority dependence and judgments of utilitarian harm. *Cognition*, *128*(3), 261–70. http://doi.org/10.1016/j.cognition.2013.05.001

Pimentel, P. S., Arndorfer, A., & Malloy, L. C. (2015). Taking the blame for someone else's wrongdoing: The effects of age and reciprocity. *Law and Human Behavior*, *39*(3), 219–231. http://doi.org/10.1037/lhb0000132

Pizarro, D., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review*, *3*(3), 267–268. http://doi.org/10.1177/1754073911402394

Pontell, H. N., Black, W. K., & Geis, G. (2014). Too big to fail, too powerful to jail? On the absence of criminal prosecutions after the 2008 financial meltdown. *Crime, Law and Social Change*, *61*(1), 1–13. http://doi.org/10.1007/s10611-013-9476-4

Posner, R. A. (2016). On hearsay. *Fordham L. Rev.*, *84*, 1465–1471.

Price, M. E., Cosmides, L., & Tooby, J. (2002). Punitive Sentiment as an Anti-Free Rider Psychological Device. *Evolution and Human Behavior*, *23*(3), 203–231. http://doi.org/10.1016/S1090-5138(01)00093-9

Quas, J. a., Qin, J., Schaaf, J. M., & Goodman, G. S. (1997). Individual differences in children's and adults' suggestibility and false event memory. *Learning and Individual Differences*, *9*(4), 359–390. http://doi.org/10.1016/S1041-6080(97)90014-5

Rabin, R. L. (2006). Pain and Suffering and Beyond: Some Thoughts on Recovery for Intangible Loss. *DePaul Law Review*, *55*.

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does Unconscious Racial Bias Affect Trial Judges? *Notre Dame Law Review*, *84*(3), 1195–1246. http://doi.org/10.1525/sp.2007.54.1.23.

Rakoczy, H., Warneken, F., & Tomasello, M. (2008). The sources of normativity: young children's awareness of the normative structure of games. *Developmental Psychology*, *44*(3), 875–881. http://doi.org/10.1037/0012-1649.44.3.875

Rakoczy, H., Warneken, F., & Tomasello, M. (2009). Young children's selective learning of rule games from reliable and unreliable models. *Cognitive Development*, *24*(2009), 61–69. http://doi.org/10.1016/j.cogdev.2008.07.004

Redding, R. E. (1998). How Common-Sense Psychology Can Inform Law and Psycholegal Research. *U. Chi. L. Sch. Roundtable*, *1*.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences*, *9*(2), 95–123. http://doi.org/10.1016/S1041-6080(97)90002-9

Rhodes, M. (2013). How Two Intuitive Theories Shape the Development of Social Categorization. *Child Development Perspectives*, *7*(1), 12–16. http://doi.org/10.1111/cdep.12007

Rhodes, M., & Chalik, L. (2013). Social categories as markers of intrinsic interpersonal obligations. *Psychological Science*, *24*(6), 999–1006. http://doi.org/10.1177/0956797612466267

Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology*, *59*(3), 244–274. http://doi.org/10.1016/j.cogpsych.2009.05.001

Richardson, C. B., Mulvey, K. L., & Killen, M. (2012). Extending social domain theory with a process-based account of moral judgments. *Human Development*, *55*(1), 4–25. http://doi.org/10.1159/000335362

Riger, S., Foster-Fishman, P., Nelson-Kuna, J., & Curran, B. (1995). Gender bias in courtroom dynamics. *Law and Human Behavior*, *19*(5), 465–480. http://doi.org/10.1007/BF01499338

Riggs, A. E., & Kalish, C. W. (2016). Children's evaluations of rule violators. *Cognitive Development*, *40*, 132–143. http://doi.org/10.1016/j.cogdev.2016.09.001

Robbennolt, J. K. (2006). Apologies and Settlement Levers. *Journal of Empirical Legal Studies*, *3*(08).

Robbennolt, J. K. (2010). Apologies and Reasonableness: Some Implications of Psychology for Torts. *DePaul Law Review*, *59*, 489.

Roberts, S. O., & Gelman, S. A. (2015). Do Children See in Black and White? Children's and Adults' Categorizations of Multiracial Individuals. *Child Development*, *86*(6), 1830–1847. http://doi.org/10.1111/cdev.12410

Roberts, S. O., & Gelman, S. A. (2016). Can White children grow up to be Black? Children's reasoning about the stability of emotion and race. *Developmental Psychology*. Roberts, Steven O.: Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI, US, 48109, sothello@umich.edu: American Psychological Association. http://doi.org/10.1037/dev0000132

Robinson, P. H. (2008). Competing Conceptions of Modern Desert: Vengeful, Deontological, and Empirical. *Cambridge Law Journal*, *67*(1), 145–175.

Robinson, P. H., & Darley, J. M. (1997). The Utility of Desert. *Northwestern University Law*

Review, *91*(2), 453–499. http://doi.org/10.2139/ssrn.10195

Rochat, P., Robbins, E., Passos-Ferreira, C., Donato Oliva, A., Dias, M. D. G., & Guo, L. (2014). Ownership reasoning in children across cultures. *Cognition*, *132*(3), 471–484. http://doi.org/10.1016/j.cognition.2014.04.014

Rottman, J., & Kelemen, D. (2012). Aliens behaving badly: Children's acquisition of novel purity-based morals. *Cognition*, *124*(3), 356–360. http://doi.org/10.1016/j.cognition.2012.06.001

Rozin, P., & Fallon, A. E. (1987). A Perspective on Disgust. *Psychological Review*, *94*(1), 23–41. http://doi.org/10.1037/0033-295X.94.1.23

Rozin, P., Haidt, J., & Fincher, K. M. (2009). From Oral to Moral. *Science*, *323*(February), 1179–1180.

Salerno, J. M., & Peter-Hagene, L. C. (2013). The interactive effect of anger and disgust on moral outrage and judgments. *Psychological Science*, *24*(10), 2069–2078.

Saltzburg, S. A. (2016). Rethinking the Rationale(s) for hearsay exceptions. *Fordham L. Rev.*, *84*, 1485–1497.

Santos, L. R., & Rosati, A. G. (2015). The Evolutionary Roots of Human Decision Making. *Annual Review of Psychology*, *66*, 321–347. http://doi.org/10.1038/nbt.3121.ChIP-nexus

Schein, C., Ritter, R. S., & Gray, K. (2016). Harm mediates the disgust-immorality link. *Emotion*, *16*(6), 862.

Scher, S. J., & Darley, J. M. (1997). How Effective Are the Things People Say to Apologize? Effects of the Realization of the Apology Speech Act. *Journal of Psycholinguistic Research*, *26*(1), 127–140.

Schmidt, M. F. H., Rakoczy, H., & Tomasello, M. (2012). Young children enforce social norms selectively depending on the violator's group affiliation. *Cognition*, *124*(3), 325–333. http://doi.org/10.1016/j.cognition.2012.06.004

Schmidt, M. F. H., & Tomasello, M. (2012). Young Children Enforce Social Norms. *Current Directions in Psychological Science*, *21*(4), 232–236. http://doi.org/10.1177/0963721412448659

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1096–1109. http://doi.org/10.1177/0146167208317771.Disgust

Schraub, D. (2013). Sticky Slopes. *California Law Review*, *101*(5), 1249–1313.

Schulhofer, S. J. (1974). Harm and Punishment: A Critique of Emphasis on the Results of Conduct in Criminal Law. *U. Pa. L. Rev.*, *122*(6), 1–80.

Sevier, J. (2014). The truth-justice tradeoff: Perceptions of decisional accuracy and procedural justice in adversarial and inquisitorial legal systems. *Psychology, Public Policy, and Law*, *20*(2), 212–224. http://doi.org/10.1037/law0000009

Shaw, A., DeScioli, P., & Olson, K. R. (2012). Fairness versus favoritism in children. *Evolution and Human Behavior*, *33*(6), 736–745. http://doi.org/10.1016/j.evolhumbehav.2012.06.001

Shaw, A., & Olson, K. (2014). Fairness as partiality aversion: The development of procedural justice. *Journal of Experimental Child Psychology*, *119*, 40–53. http://doi.org/10.1016/j.jecp.2013.10.007

Shaw, A., & Olson, K. R. (2012). Children discard a resource to avoid inequity. *Journal of Experimental Psychology: General*, *141*(2), 382–95. http://doi.org/10.1037/a0025907

Shen, F. X., Hoffman, M. B., Jones, O. D., Greene, J. D., & Marios, R. (2011). Sorting guilty minds. *New York University Law Review*, 1306–1360.

Shultz, T. R., Schleifer, M., & Altman, I. (1981). Judgments of causation, responsibility, and punishment in cases of harm-doing. *Canadian Journal of Behavioural Science*. Canada: University of Toronto Press. http://doi.org/10.1037/h0081183

Shutts, K., Pemberton Roben, C. K., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*(1), 35–62. http://doi.org/10.1080/15248372.2011.638686

Shweder, R. A., Turiel, E., & Much, N. C. (1981a). The moral intuitions of the child. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 288–305). CUP Archive.

Shweder, R. A., Turiel, E., & Much, N. C. (1981b). The Moral Intuitions of the Child. In J. H. Flavell & L. Ross (Eds.), *Social Cognitive Development: Frontiers and Possible Futures* (pp. 288–305). Cambridge: Cambridge University PressC.

Smetana, J. G. (1981a). Preschool Children's Conceptions of Moral and Social Rules. *Child Development*, *52*(4), 1333–1336.

Smetana, J. G. (1981b). Reasoning in the personal and moral domains: Adolescent and young adult women's decision-making regarding abortion. *Journal of Applied Developmental Psychology*, *2*(3), 211–226. http://doi.org/10.1016/0193-3973(81)90002-2

Smetana, J. G. (1985). Children's impressions of moral and conventional transgressors. *Developmental Psychology*, *21*(4), 715–724. http://doi.org/10.1037/0012-1649.21.4.715

Smetana, J. G., Jambon, M., Conry-Murray, C., & Sturge-Apple, M. L. (2012). Reciprocal Associations Between Young Children's Developing Moral Judgments and Theory of Mind. *Developmental Psychology*, *48*(4), 1144–55. http://doi.org/10.1037/a0025891

Smetana, J. G., Killen, M., & Turiel, E. (1991). Children's Reasoning about Interpersonal and

Moral Conflicts. *Child Development*, *62*(3), 629–644. http://doi.org/10.2307/1131136

Smetana, J. G., Schlagman, N., & Adams, P. W. (1993). Preschool Children's Judgments about Hypothetical and Actual Transgressions. *Child Development*, *64*(1), 202–214.

Sommers, S. R. (2006). On racial diversity and group decision making: Identifying multiple effects of racial composition on jury deliberations. *Journal of Personality and Social Psychology*, *90*(4), 597–612. http://doi.org/10.1037/0022-3514.90.4.597

Sommers, S. R., & Ellsworth, P. C. (2000). Race in the Courtroom: Perceptions of Guilt and Dispositional Attributions. *Personality and Social Psychology Bulletin*, *26*(11), 1367–1379. http://doi.org/10.1177/0146167200263005

Sperber, D., & Hirschfeld, L. (2001). Culture, Cognition, and Evolution. In R. A. Wilson & F. C. Keil (Eds.), *The MIT Encyclopedia of the Cognitive Sciences* (pp. cxi–cxxxii). Cambridge: The MIT Press.

Steffensmeier, D., & Demuth, S. (2000). Ethnicity and Sentencing Outcomes in U.S. Federal Courts: Who is Punished More Harshly? *American Sociological Review*, *65*(5), 705–729.

Steinman, K. (2002). Sex tourism and the child: Latin America's and the United States' failure to prosecute sex tourists. *Hastings Women's Law Journal*, *13*(1), 53–76.

Stern, L. (1970). Deserved punishment, deserved harm, deserved blame. *Philosophy*, *45*(174), 317–329. http://doi.org/10.1017/S0031819100047690

Stevenson, B. A., & Friedman, R. E. (1994). Deliberate Indifference : Judicial Tolerance of Racial Bias in Criminal Justice Deliberate Indifference : Judicial Tolerance of Racial Bias in Criminal Justice, *51*(2).

Stok, F. M., de Ridder, D. T. D., de Vet, E., & de Wit, J. B. F. (2014). Don't Tell Me What I Should Do, But What Others Do: The Influence of Descriptive and Injunctive Peer Norms on Fruit Consumption in Adolescents. *British Journal of Health Psychology*, *19*(1), 52–64. http://doi.org/10.1111/bjhp.12030

Sunstein, C. R. (1995). On the Expressive Function of Law. *U. Pa. L. Rev.*, *144*(5), 2021–2053.

Sunstein, C. R. (2005). Moral heuristics. *The Behavioral and Brain Sciences*, *28*(4), 531-42; discussion 542-73. http://doi.org/10.1017/S0140525X05000099

Thompson, R. A., & Newton, E. K. (2013). Baby Altruists? Examining the Complexity of Prosocial Motivation in Young Children. *Infancy*, *18*(1), 120–133. http://doi.org/10.1111/j.1532-7078.2012.00139.x

Tisak, M. S. (1993). Preschool Children's Judgments of Moral and Personal Events Involving Physical Harm and Property Damage. *Merrill-Palmer Quarterly*, *39*(3), 375–390.

Tisak, M. S., & Turiel, E. (1984). Children's Conceptions of Moral and Prudential Rules. *Child*

Development, *55*(3), 1030–1039.

Turiel, E. (1978). Social regulations and domains of social concepts. *New Directions for Child and Adolescent Development*, *1978*(1), 45–74. http://doi.org/10.1002/cd.23219780105

Turiel, E. (2008). Thought about actions in social domains: Morality, social conventions, and social interactions. *Cognitive Development*, *23*(1), 136–154. http://doi.org/10.1016/j.cogdev.2007.04.001

Tyler, T. R. (1987). Conditions Leading to Value-Expressive Effects in Judgments of Procedural Justice: A Test of Four Models. *Journal of Personality and Social Psychology*, *52*(2), 333–344. http://doi.org/10.1037/0022-3514.52.2.333

Tyler, T. R. (2005). Policing in black and white: Ethnic group differences in trust and confidence in the police. *Police Quarterly*, *8*(3), 322–342. http://doi.org/10.1177/1098611104271105

Tyler, T. R. (2006). *Why People Obey the Law*. Princeton University Press.

Tyler, T. R., & Boeckmann, R. J. (1997). Three Stikes and You Are Out, but Why? The Psychology of Public Support for Punishing Rule Breakers. *Law and Society Review*, *31*(2), 237–266.

Tyler, T. R., & Rasinski, K. (1991). Procedural Justice, Institutional Legitimacy, and the Acceptance of unpopular U.S. Supreme Court Decisions: A Reply to Gibson. *Law & Society Review*, *25*(3), 621–630. http://doi.org/10.2307/3053729

Tyler, T. R., & Sevier, J. (2013). How do the Courts Create Popular Legitimacy?: The Role of Establishing the Truth, Punishing Justly, and/or Acting Through Just Procedures. *Albany Law Review*, *77*, 1095–1137.

United States General Accounting Office. (1990). Death Penalty Sentencing: Research Indicates Pattern of Racial Disparities. http://doi.org/10.1037/e543292006-001

Vaish, A., Carpenter, M., & Tomasello, M. (2010). Young Children Selectively Avoid Helping People With Harmful Intentions. *Child Development*, *81*(6), 1661–1669.

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *The British Journal of Developmental Psychology*, *29*(Pt 1), 124–30. http://doi.org/10.1348/026151010X532888

Van de Vondervoort, J. W., & Friedman, O. (2015). Parallels in preschoolers' and adults' judgments about ownership rights and bodily rights. *Cognitive Science*, *39*(1), 184–198. http://doi.org/10.1111/cogs.12154

Van Hoecke, M., & Warrington, M. (1998). Legal Cultures, Legal Paradigms and Legal Doctrine: Towards a New Model for Comparative Law. *International & Comparative Law Quarterly*, *47*(1998), 1–48. http://doi.org/10.1017/S0020589300062163

van Prooijen, J.-W., Coffeng, J., & Vermeer, M. (2014). Power and retributive justice: How trait information influences the fairness of punishment among power holders. *Journal of Experimental Social Psychology*, *50*, 190–201. http://doi.org/10.1016/j.jesp.2013.10.004

Vidmar, N., & Miller, D. T. (1980). Socialpsychological processes underlying attitudes toward legal punishment. *Law and Society Review*, *14*(3), 565–602. http://doi.org/10.2307/3053193

Volokh, E. (2003). The mechanisms of the slippery slope. *Harvard Law Review*, *116*(4), 1026–1137. http://doi.org/10.2307/1342743

Waldman, E. A. (2003). Healing Hearts or Righting Wrongs?: A Meditation on the Goals of "Restorative Justice." *Journal of Public Law & Policy*, *25*, 355–374.

Ward, T., & Durrant, R. (2011). Evolutionary psychology and the rehabilitation of offenders: Constraints and consequences. *Aggression and Violent Behavior*, *16*(5), 444–452. http://doi.org/10.1016/j.avb.2011.02.011

Ward, T., & Langlands, R. (2009). Repairing the rupture: Restorative justice and the rehabilitation of offenders. *Aggression and Violent Behavior*, *14*(3), 205–214. http://doi.org/10.1016/j.avb.2009.03.001

Warneken, F. (2015). Precocious prosociality: Why do young children help? *Child Development Perspectives*, *9*(1), 1–6. http://doi.org/10.1111/cdep.12101

Warneken, F., & Tomasello, M. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology*, *44*(6), 1785–1788. http://doi.org/10.1037/a0013860

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth About False Belief. *Child Development*, *72*(3), 655–84.

Wells, G. L., & Olson, E. A. (2003). Eyewitness Testimony. *Annual Review of Psychology*, *54*(1), 277–295. http://doi.org/10.1146/annurev.psych.54.101601.145028

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*(6), 603–647.

Wetzel, A. C. (2006). Georgia v. Randolph: A Jealously Guarded Exception-- Consent and the Fourth Amendment (Comment). *Valparaiso University Law Review*, *41*, 499–515. http://doi.org/10.1525/sp.2007.54.1.23.

Winship, V., & Robbennolt, J. K. (2018a). Admissions of Guilt in Civil Enforcement. *Minnesota Law Review*, *102*, 1077–1146.

Winship, V., & Robbennolt, J. K. (2018b). An Empirical Study of Admissions in SEC Settlements. *Ariz. L. Rev.*, *60*, 1–66.

Wittlin, M. (2011). Buckling Under Pressure: An Empirical Test of the Expressive Effects of Law. *Yale Journal on Regulation*, *28*, 419–471.

Witvliet, C. V. O., Worthington, E. L., Root, L. M., Sato, A. F., Ludwig, T. E., & Exline, J. J. (2008). Retributive justice, restorative justice, and forgiveness: An experimental psychophysiology analysis. *Journal of Experimental Social Psychology*, *44*(1), 10–25. http://doi.org/10.1016/j.jesp.2007.01.009

Wood, E. J., & Toppelberg, N. (2017). The persistence of sexual assault within the US military. *Journal of Peace Research*, *54*(5), 620–633. http://doi.org/10.1177/0022343317720487

Woodward, A. (2013). Infant Foundations of Intentional Understanding. In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (pp. 75–80). Oxford University Press.

Woolard, J. L., Reppucci, N. D., & Redding, R. E. (1996). Theoretical and methodological issues in studying children's capacities in legal contexts. *Law and Human Behavior*, *20*(3), 219–228. http://doi.org/10.1007/BF01499021

Yang, F., Shaw, A., Garduno, E., & R. Olson, K. (2014). No one likes a copycat: A cross-cultural investigation of children's response to plagiarism. *Journal of Experimental Child Psychology*, *121*(1), 111–119. http://doi.org/10.1016/j.jecp.2013.11.008

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. http://doi.org/10.1016/j.cognition.2011.04.005

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children's judgments of actor's responsibility and recipient's emotional reaction. *Developmental Psychology*, *24*(3), 358–365. http://doi.org/10.1037/0012-1649.24.3.358

Zatz, M. S. (1987). The Changing Forms of Racial/Ethnic Biases in Sentencing. *Journal of Research in Crime and Delinquency*, *24*(1), 69–92.

Zelazo, P. D., Helwig, C. C., & Lau, A. (1996). Intention, Act, and Outcome in Behavioral Prediction and Moral Judgment. *Child Development*, *67*(5), 2478–2492.

Zhao, X., & Kushnir, T. (2017). Young children consider individual authority and collective agreement when deciding who can change rules. *Journal of Experimental Child Psychology, (in press)*. http://doi.org/10.1016/j.jecp.2017.04.004