NAVIGATING THE TENSION BETWEEN BENEVOLENCE AND HONESTY: ESSAYS ON THE CONSEQUENCES OF PROSOCIAL LIES

Emma Edelman Levine

A DISSERTATION
in
Operations, Information, and Decisions
For the Graduate Group in Managerial Science and Applied Economics
Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

____
Maurice Schweitzer, Associate Professor of Operations, Information, and Decisions
Graduate Group Chairperson

____
Eric Bradlow, K.P. Chao Professor; Professor of Marketing, Statistics, and Education

Dissertation Committee:
Katherine L. Milkman, Associate Professor of Operations, Information and Decisions
Adam Grant, Professor of Management
Andrew Carton, Assistant Professor of Management
Taya R. Cohen, Associate Professor of Organizational Behavior and Theory, Carnegie Mellon University

DEDICATION

This dissertation is dedicated to my better half:

Ethan Levine.

ACKNOWLEDGEMENT

First, I would like to thank my advisor, Maurice Schweitzer. Without his mentorship, I would not have become the scholar I am today and this dissertation would never have been completed. Thank you for your time, encouragement, and inspiration as an advisor, a researcher, and a person.

I am also immensely grateful for the feedback and support of my committee members and my collaborators. In particular, I would like to thank Katherine Milkman, Taya Cohen, Andrew Carton, Adam Grant, and Deborah Small. I have also received invaluable assistance from former and current Wharton doctoral students and post-doctoral fellows. In particular, I would like to thank Alixandra Barasch, Jonathan Berman, T. Bradford Bitterly, Hengchen Dai, Kaitlin Daniels, Berkeley Dietvorst, Celia Gaertig, Jessica Kennedy, Joshua Lewis, Danielle Tussing, and Julianna Pillemer.

I appreciate the financial support from the Operations, Information and Decisions Department, the Wharton Behavioral Lab, the Wharton Risk Center Russell Ackoff Doctoral Student Fellowship, the Russell Sage Foundation, and the Fostering Improvement in End-of-Life Decision Science (FIELDS) Program of the Perelman School of Medicine at the University of Pennsylvania.

Finally, I would like to thank my family and friends for their unwavering love and encouragement. My deepest gratitude goes to my parents, Lucy Berman-Edelman and Eric Edelman, and my brother, Adam Edelman, who have always supported my aspirations and prioritized my happiness. I love you.

ABSTRACT

NAVIGATING THE TENSION BETWEEN BENEVOLENCE AND HONESTY: ESSAYS ON THE CONSEQUENCES OF PROSOCIAL LIES

Emma Levine

Maurice Schweitzer

Many of our most common and difficult ethical dilemmas involve balancing honesty and benevolence. For example, when we deliver unpleasant news, such as negative feedback or terminal prognoses, we face an implicit tradeoff between being completely honest and being completely kind. Using a variety of research methods, in both the laboratory and the field, I study how individuals navigate this tension. Each chapter in this dissertation addresses the tension between honesty and benevolence at a different level. In Chapters One and Two, I examine how honesty and benevolence influence moral judgment. In Chapter Three, I explore how honesty and benevolence influence interpersonal trust. In Chapter Four, I explore how honesty and benevolence influence psychological well-being. Finally, in Chapter Five, I examine how different stakeholders view tradeoffs between honesty and benevolence in an important domain: healthcare. Across these chapters, I identify three key themes. First, for moral judgment and interpersonal trust, benevolence is often more important than honesty. As a result, those who prioritize benevolence over honesty by telling prosocial lies, lies that are intended to help others, are deemed to be moral and trustworthy. Second, despite philosophers' assumption that individuals would rarely consent to deception, I demonstrate that individuals frequently want to be deceived. Individuals want others to deceive them when it protects them from harm. This desire manifests itself in systematic circumstances and during individuals' most fragile moments. Third, honesty and benevolence are associated with interpersonal and intrapersonal tradeoffs. Although benevolence seems to be more central for interpersonal judgments and relationships, honesty seems to be more central for creating personal meaning. Throughout these chapters, I discuss the implications of these findings for the study of ethics, organizational behavior, and interpersonal communication.

TABLE OF CONTENTS

INTRODUCTION

Deception is typically considered to be a vice, and honesty a virtue. For centuries, philosophers have touted the moral inviolability of honesty (Bacon, 1872, Bok, 1978; Kant, 1785; St. Augustine, 421, cited in Gneezy, 2005), and modern psychologists, behavioral scientists, organizational behavior scholars, and practitioners have largely echoed this view (e.g., Mayer, Davis, & Schoorman, 1995; Wojciszke, 2005; Schweitzer, Hershey, & Bradlow, 2006; Tenbrunsel, 1998; Treviño, Weaver, & Reynolds, 2006). Many companies proclaim the importance of honesty in their codes of conduct, assuming that honesty, and honesty alone, is the foundation of ethical practice. For example, Microsoft's number one moral value is "Honesty and integrity" and Dell's key moral claim is "We are honest". Furthermore, many practitioners must take oaths of honesty. For example, physicians are explicitly told they "shall be honest in all professional interaction" (American Medical Association Code of Ethics, 2006). Although honesty is important for interpersonal relationships and organizational conduct, honesty often conflicts with other moral values, such as kindness, compassion, and hope.

Many of our most common and difficult ethical dilemmas involve balancing the tension between honesty and benevolence. People routinely face this conflict in their personal lives, when deciding how to communicate with friends and family members, and in their professional lives, when deciding how to deliver difficult news and critical feedback. Honesty and benevolence also conflict during some of our most demanding and emotional ethical decisions. For example, when healthcare professionals and loved ones communicate information to sick and elderly individuals, they must strike a delicate balance between providing hope and care, and communicating honestly. Or, when employees have to decide whether or not to report the transgression of a close friend, they must decide whether to use honesty or loyalty as a moral guide.

Despite the frequency with which honesty and benevolence collide, little research examines how people navigate this conflict and virtually no research offers prescriptive advice on how to balance this tension in personal or professional relationships. The goal of this dissertation is to fill this gap by a) answering fundamental questions about how people reason through the moral conflict between honesty and benevolence, by b) examining the interpersonal and intrapersonal consequences of this conflict, and by c) examining how communicators and targets judge conflicts between honesty and benevolence in high-stakes organizational settings. This dissertation is composed of five chapters. Each chapter explores the tension between honesty and benevolence through a different lens. Chapters One and Two explore moral judgments of this tension. Chapter Three examines the interpersonal consequences of this tension. Chapter Four examines the intrapersonal consequences of this tension, and Chapter Five examines this tension within the healthcare context.

The moral consequences of honesty and benevolence

In the first chapter of my dissertation, I explore moral judgments of prosocial lies in collaboration with Maurice Schweitzer. Prosocial lies, lies that are intended to help others, reflect a conflict between honesty and benevolence. Across three studies using economic games, we find that individuals who tell prosocial lies are perceived to be *more* moral than individuals who tell the truth.

In Chapter Two, I build on these findings to develop a descriptive moral theory of deception. Through a large inductive study, and a series of experiments (N = 1313) participants, I demonstrate that lay people have a codified set of rules that guide their moral judgments of deception. A basic theory underlies these implicit rules: deception is perceived to be ethical and individuals prefer to be deceived when honesty causes *unnecessary harm*. Perceptions of unnecessary harm are influenced by two dimensions: the degree to which honesty will help or harm an individual at the moment of communication, and the instrumental value of truth. Perceptions of "unnecessary harm" dictate nine implicit rules – pertaining to the targets, topics, and timing of a conversation – that specify the systematic circumstances in which deception is perceived to be ethical. I demonstrate that unnecessary harm is the key driver of moral judgments of deception and I rule out a series of alternative mechanisms that have been proposed in normative and moral psychology (e.g., perceptions of autonomy, self-interest). This research provides insight into how individuals value honesty and deception for making moral judgments, for learning information about themselves, and for communicating with others.

The interpersonal consequences of honesty and benevolence

In Chapter Three, I explore the interpersonal consequences of honesty and benevolence by exploring the relationship between prosocial lying and trust. One of the key claims that philosophers and scholars have made against deception is that it harms trust. For example, philosopher Sir Francis Bacon famously argued that deception deprives, "people of two of the most principal instruments for interpersonal action—trust and belief" (from "On Truth", cited in Tyler & Feldman, 2006). Empirical research in organizational behavior and economics has largely supported this claim, demonstrating that that deception harms relationships (Ford, King, & Hollender, 1988; Lewis & Saarni, 1993; Tyler & Feldman, 2006), elicits negative affect (Planalp, Rutherford, & Honeycutt, 1988), decreases liking, (Tyler, Feldman, & Reichert, 2006) triggers retaliation (Boles et al., 2000; Croson et al., 2003), and does indeed harm trust (Schweitzer et al., 2006).

In this chapter, Maurice Schweitzer and I challenge these claims. In nearly all empirical investigations of the consequences of deception, scholars have confounded deception with self-interest. In this research, we disentangle deception from self-interest, and demonstrate that deception often increases trust. Specifically, we demonstrate that prosocial lies increase benevolence-based trust. Consistent with prior claims, we also find that prosocial lies harm integrity-based trust. We present four studies in which we document the robustness of these results and introduce new paradigms for the study of trust. These findings expand our understanding of the interpersonal consequences of deception and deepen our insight into the mechanics of trust.

The intrapersonal consequences of honesty and benevolence

In Chapter four, in collaboration with Taya Cohen, I explore how honesty and benevolence influence well-being in everyday life. In a large-scale field experiment, we randomly assigned individuals to be honest, kind, or conscious of their communication (our control condition) in every interpersonal interaction for three days. We examine the impact of our interventions on predicted and actual hedonic and eudaimonic well-being and we identify three main results. First, individuals predict that honesty will be far less enjoyable (i.e., less hedonically rewarding) than kind or conscious communication, causing individuals to avoid honesty. Second, this prediction is incorrect: the experience of communicating honestly is more enjoyable than individuals predict. Third, honesty yields greater meaning (i.e., eudaimonia) than kind or conscious communication, and as a result, has greater long-term impact on individuals' lives. This research sheds new light on the relationships among communication, morality, and well-being. Furthermore, this research complements Chapters 1-3 by highlighting the interpersonal and intrapersonal tradeoffs of honesty and benevolence. Although benevolence may be more important for moral judgment and trust, honesty may be more important for promoting personal meaning.

The organizational consequences of honesty and benevolence

Finally, in Chapter Five I examine how individuals navigate the tension between honesty and benevolence in healthcare communication, in collaboration with Joanna Hart, Kendra Moore, Emily Rubin, Kuldeep Yadav, and Scott Halpern. Professional medical organizations (American Medical Association Code of Ethics, 2006; World Medical Association International Code of Ethics, 2016) and ethicists (Apatira et al., 2008; Beste, 2005; Herring & Foster, 2012; Sarafis, Tsounis, Malliarou, Lahana, 2014) often prohibit deception. This prohibition is motivated by normative assumptions regarding the negative consequences of deception, rather than empirical evidence. In this research, we empirically investigate physicians', patients', and healthy adults' moral judgments and preferences for deception and identify three important findings. First, individuals believe that it is more ethical to use deception when discussing future predictions (e.g., prognoses) than discussing present knowledge (e.g., diagnoses). Second, physicians think very differently about lies of omission and commission than patients and healthy adults do. Physicians believe that lies of commission are less ethical than lies of omission, but patients and healthy adults often believe the opposite. We introduce a theoretical framework to explain these findings and we discuss the clinical and psychological implications of this research for medicine, behavioral ethics, and human communication. This work highlights the practical relevance of studying reactions to deceptions and demonstrates how preferences for deception can trigger predictable asymmetries between communicators and targets during challenging conversations.

Conclusion

This dissertation makes fundamental contributions to our understanding of moral judgment, interpersonal relationships, well-being, and practical ethics. In Chapters One, I answer the basic psychological question: Is deception perceived to be ethical? In Chapter Two, I explore this issue further and develop a descriptive moral theory of deception. Just as Kahneman, Knetsch, and Thaler’s (1986) foundational work on community standards of fairness overturned the assumption that individuals universally value self-interest, and demonstrated that concerns about fairness place systematic constraints on market behavior, the first two chapters of this dissertation challenge the assumption that people universally value truth, and demonstrates that concerns about interpersonal harm place systematic constraints on honest communication.

Second, this dissertation provides insight into how honesty and benevolence influence relationships and well-being. Philosophers, theologians, and leaders regularly espouse moral values and make claims about what it means to live a virtuous life. We know very little, however, about what virtues actually improve well-being and relationships. Chapter Three sheds light on this question by exploring the interpersonal consequences of honesty and benevolence, and Chapter Four sheds light on this question by examining how honesty and benevolence influence well-being in everyday life. Although organizational scholars have typically refrained from making normative claims, recently, scholars have called for a greater integration between organizational and normative ethics (e.g., Barry & Rehel, 2014). This dissertation answers this call by providing practical insights on the consequences of distinct moral virtues.

Finally, this dissertation demonstrates the organizational importance of challenging normative assumptions about the consequences of deception, and ethical principles broadly. In medicine, ethical principles are in place to ensure the protection and well-being of patients. And yet, we know very little about the ethical principles that patients care about and how current ethical guidelines affect patient well-being. In Chapter Five of my dissertation, I explore patients' and physicians' attitudes towards deception, and find evidence that physicians' ethical training may not always accommodate patients' desire for (false) hope. This research highlights how individuals' roles shift their preferences for deception, contributing to miscommunication and potential conflict in high-stakes settings.

Taken together, these chapters examine the tension between honesty and benevolence from every angle, thereby contributing fundamental knowledge to the study of moral judgment, trust, and well-being and providing practical advice to those who must manage this tension in their personal and professional relationships.

References

“AMA's Code of Medical Ethics." *AMA's Code of Medical Ethics*. American Medical Association, n.d. Web. 13 Mar. 2016. <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page>.

Apatira, L., Boyd, E. A., Malvar, G., Evans, L. R., Luce, J. M., Lo, B., & White, D. B. (2008). Hope, truth, and preparing for death: perspectives of surrogate decision makers. *Annals of Internal Medicine*, *149*(12), 861-868.

Bacon, F. (1872). *The letters and the life of Francis Bacon* (Vol. 6). City, State: Longmans, Green and Company.

Barry, B., & Rehel, E. M. (2013). Lies, Damn Lies, and Negotiation: An Interdisciplinary Analysis of the Nature and Consequences of Deception at the Bargaining Table. *Handbook of Research in Conflict Management (Elgar, 2014)*.

Beste, J. (2005). Instilling hope and respecting patient autonomy: Reconciling apparently conflicting duties. *Bioethics*, *19*(3), 215-231.

Bok, S. (1978). *Lying: Moral choices in public and private life*. New York, NY: Pantheon.

Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *83*(2), 235-259.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143-159.

Ford, C. V, King, B.H., & Hollender, M.H. (1988). Lies and liars: Psychiatric aspects of prevarication. *American Journal of Psychiatry*, *145*(5), 554-562.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review*, *95*(1), 384-394.

Herring, J., & Foster, C. (2012). Please don't tell me. *Cambridge Quarterly of Healthcare Ethics*, *21*(01), 20-29.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986a). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728-741.

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

Lewis, M. & Saarni, C. (1993). *Lying and deception in everyday life*. New York, NY: The Guilford Press.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734.

Planalp, S., Rutherford, D. K., & Honeycutt, J. M. (1988). Events that increase uncertainty in personal relationships II: Replication and extension. *Human Communication Research*, *14*(4), 516-547.

Sarafis, P., Tsounis, A., Malliarou, M., & Lahana, E. (2014). Disclosing the truth: a dilemma between instilling hope and respecting patient autonomy in everyday clinical practice. *Global Journal of Health Science*, *6*(2), 128.

Schweitzer, M. E., & Hsee, C. K. (2002). Stretching the truth: Elastic justification and motivated communication of uncertain information. *Journal of Risk and Uncertainty*, *25*(2), 185-201.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1-19.

Tenbrunsel, A. E. (1998). Misrepresentation and expectations of misrepresentation in an ethical dilemma: The role of incentives and temptation. *Academy of Management Journal*, *41*(3), 330-339.

Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management*, *32*(6), 951-990.

Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, *42*(1), 69-77.

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology*, *16*(1), 155-188.

CHAPTER 1.

ARE LIARS ETHICAL?

ON THE TENSION BETWEEN BENEVOLENCE AND HONESTY

Emma E. Levine

Maurice E. Schweitzer

Published in *Journal of Experimental Social Psychology* in 2015.

ABSTRACT

We demonstrate that some lies are perceived to be more ethical than honest statements. Across three studies, we find that individuals who tell prosocial lies, lies told with the intention of benefitting others, are perceived to be *more* moral than individuals who tell the truth. In Study 1, we compare altruistic lies to selfish truths. In Study 2, we introduce a stochastic deception game to disentangle the influence of deception, outcomes, and intentions on perceptions of moral character. In Study 3, we demonstrate that moral judgments of lies are sensitive to the consequences of lying for the deceived party, but insensitive to the consequences of lying for the liar. Both honesty and benevolence are essential components of moral character. We find that when these values conflict, benevolence may be more important than honesty. More broadly, our findings suggest that the moral foundation of care may, at times,be more important than the moral foundation of justice.

ARE LIARS ETHICAL? ON THE TENSION BETWEEN BENEVOLENCE AND HONESTY

"To me, however, it seems certain that every lie is a sin…" – St. Augustine (circa 420 A.D.)
"By a lie, a man annihilates his dignity." – Immanuel Kant (circa 1797)
"…deception is unethical." – Chuck Klosterman, The New York Times, "The Ethicist" (2014)

For centuries, philosophers and theologians have characterized lying as unethical (Kant, 1785; for review, see Bok, 1978). Similarly, ethics scholars have argued that honesty is a critical component of moral character (e.g. Wojciszke, 2005; Rosenberg, Nelson, Vivekananthan, 1998) and a fundamental aspect of ethical behavior (e.g. Ruedy, Moore, Gino, & Schweitzer, 2013).

The conceptualization of lying as immoral, however, is difficult to reconcile with its prevalence. Lying is common in everyday life (DePaulo & Kashy, 1998; Kashy & DePaulo, 1996). Not only do people lie to benefit themselves (e.g. lying on one's tax returns), but people also lie to benefit others (e.g. lying about how much one likes a gift) or to serve both self-interested and prosocial motives. This broader conceptualization of lying to include prosocial or mixed-motive deception has been largely ignored in ethical decision-making research.

In studies of ethical decision-making, scholars have routinely confounded deception with self-serving motives and outcomes. This is true of both theoretical and empirical investigations of deception (e.g., Mazar, Amir, & Ariely, 2008; Shalvi, Dana, Handgraaf, & De Dreu, 2011; Shalvi, 2012; Tenbrunsel, 1998; Boles, Croson & Murninghan, 2000; Shu, Mazar, Gino, Ariely, & Bazerman, 2012; Ruedy, Moore, Gino, & Schweitzer, 2013; Mead, Baumeister, Gino, Schweitzer, & Ariely, 2009; Koning, Steinel, Beest, & van Dijk, 2011; Steinel & De Dreu, 2004; Gaspar & Schweitzer, 2013; Schweitzer, DeChurch, & Gibson, 2005). For example, ethics scholars who have conflated lying with self-serving motives have investigated behaviors like cheating on one's taxes (e.g. Shu, et al., 2012), inflating self-reported performance (e.g., Mazar et al., 2008; Ruedy et al., 2013; Mead, et al., 2009), misreporting a random outcome for financial gain (e.g. Shalvi et al., 2011) and lying to a counterpart to exploit them (Koning, et al., 2011; Steinel & De Dreu, 2004).

Related research has studied the interpersonal consequences of deception. This work has found that lying harms interpersonal relationships, induces negative affect, provokes revenge, and decreases trust (Tyler, Feldman, & Reichert, 2006; Boles, Croson & Murnighan, 2000; Schweitzer & Croson, 1999; Schweitzer, Hershey, & Bradlow, 2006; Croson, Boles, & Murnighan, 2003). All of this research, however, has studied lies that are motivated by self-interest, such as the desire for reputational or financial gains. As a result of this narrow conceptualization of deception, what we know about the psychology of deception is limited. Quite possibly, our understanding of deception may simply reflect attitudes towards selfish behavior, rather than deception per se.

In contrast to prior research that has assumed that deception is immoral, we demonstrate that lying is often perceived to be *moral*. In the present research, we disentangle deception from self-interest and explore the moral judgment of different types of lies. Across three studies, we find that lying to help others *increases* perceptions of moral character.

Our research makes two central contributions to our understanding of deception and moral judgment. First, we challenge the universal presumption that deception is immoral and that honesty is moral. We demonstrate that perceptions of honesty and deception are far more complex than prior work has assumed. This qualifies extant research and illustrates the need to explore a broader set of dishonest behaviors when investigating attitudes towards deception. Second, we explore the conflict between two universal moral foundations: justice and care. Justice is a moral foundation that prioritizes fairness, honesty and moral principles and rules; care is a moral foundation that prioritizes the obligation to help and protect other people (Gilligan, 1982; Haidt & Graham, 2007; Walker & Hennig, 2004). Prior studies that have focused on violations of either justice *or* care offer little insight into how individuals resolve dilemmas with competing moral principles. Our investigation has broad practical significance because in many settings, justice and care conflict. Prosocial lies reflect this conflict.

Prosocial lies

In routine interactions, individuals often face opportunities to tell prosocial lies. We may tell a host that their meatloaf was delicious, a child that we love their artwork, or a colleague that his or her work makes an interesting contribution. Consistent with prior research, we define lies as false statements made with the intention of misleading a target (Depaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). We define *prosocial lies* as *false statements made with the intention of misleading and benefitting a target* (Levine & Schweitzer, 2013). We distinguish prosocial lies from altruistic lies and define altruistic lies as a subset of prosocial lies; *altruistic lies* are *false statements that are costly for the liar and are made with the intention of misleading and benefitting a target* (Erat & Gneezy, 2012; Levine & Schweitzer, 2013).

We also distinguish prosocial lies from white lies. White lies involve small stakes and are “of little moral import” (Bok, 1978: 58). White lies can be either self-serving or prosocial. We define white lies as *false statements made with the intention of misleading a target about something trivial*. In contrast, prosocial lies are intended to benefit the target and can have small or substantial consequences. For example, parents may tell prosocial lies about their marriage to protect their children (e.g. Barnes, 2013), government authorities may tell prosocial lies to citizens, hoping to protect them (e.g. Bok, 1978), and doctors may tell prosocial lies about the severity of a prognosis to help a patient (e.g. Park, 2011; Palmieri & Stern, 2009; Iezzoni, Rao, DesRoches, Vogeli& Campbell, 2012). In fact, a recent study found that over 55% of doctors describe prognoses in a more positive manner than warranted, and over 10% of doctors explicitly lie to patients (Iezzoni, et al., 2012).

A few studies have explored the frequency of deception in routine communication. This work found that individuals lie in approximately 20% of their social interactions, and many of these lies are prosocial (DePaulo et al., 1996). Studies have also found that women tell more prosocial lies than men (Erat & Gneezy, 2012; Dreber & Johannesson, 2008) and that prosocial lies are most often told to close family members (DePaulo & Kashy, 1998) and to people who are emotionally invested in the content of the lie (DePaulo & Bell, 1996). Prosocial lies are often told as a form of politeness (Brown & Levinson, 1987; Goffman, 1967).

In the present research, we explore moral judgments of prosocial lies. Prosocial lying is an ethically ambivalent act; prosocial lying signals care for others (a positive moral signal), but also disregard for the moral principle of honesty (a negative moral signal). By pitting the signals of care and honesty against each other, we build our understanding of the relationship between ethical conflicts and moral character judgments.

Judging moral character

To manage and coordinate interpersonal relationships, individuals assess the moral character of those around them (e.g. Reeder, 2009). Research on moral character judgments has largely focused on perceptions of an actor’s motives. When individuals observe an unethical act, they can make either personal or situational attributions for the action (e.g. Knobe, 2004; Yuill & Perner, 1988; Young & Saxe, 2008). In making these attributions, individuals seek to understand the intentionality of the actor’s actions (Alicke, 1992; Darley & Pittman, 2003; Pizarro, Uhlmann, & Bloom, 2003). Individuals make inferences about an actor’s intentionality by using characteristics of the decision-making process as information (see Ditto, Pizzaro, & Tannenbaum, 2009 for review). For example, individuals who make quick moral decisions are perceived to be more moral than individuals who take their time to arrive at a moral decision, because a quick decision signals that an actor was certain about her judgment (Critcher, Inbar, & Pizarro, 2013).

Recent research has expanded our understanding of the different signals, such as decision speed, that influence perceptions of ethicality. However, there is still much to learn about the traits and values that really matter for judgments of moral character (e.g. Leach, Ellemers, & Barreto, 2007; Brambilla, Sacchi, Rusconi, Cherubini, & Yzerbyt, 2012).

Scholars argue that justice and care are two key components of moral character (Walker & Hennig, 2004; Aquino & Reed, 2002; Lapsley & Lasky, 2001). Justice reflects respect for overarching moral rules, such as “do not lie.” Care reflects the obligation to help and protect others (Gilligan, 1982; Haidt & Graham, 2007; Walker & Hennig, 2004). Though many scholars identify these two components as the core foundations of moral reasoning (Kohlberg, 1969; Gilligan, 1982), others have expanded the set of moral foundations to include Purity, Authority, and In-group Loyalty (Haidt & Graham, 2007, Graham, Haidt, & Nosek, 2009). In our investigation, we focus on justice and care.

Extant ethics research has primarily studied acts that violate either justice *or* care (e.g. Tannenbaum, Uhlmann, & Diermeier, 2011). In these cases, the ethical choice is often clear. However, when justice and care conflict, the ethical choice is unclear. Surprisingly, little work has examined the moral judgment of competing moral principles (for an exception, see Uhlmann & Zhu, 2013). In the present research, we explore the tension between justice and care by studying prosocial lies. Prosocial lies represent a justice violation (e.g. “Never tell a lie”) that signals care.

The majority of research in moral psychology argues that, at its core, “morality is about protecting individuals” (Haidt & Graham, 2007: 100). Caring for others is fundamental to the human experience and humans are hardwired to detect harm to others (de Waal, 2008; Graham, et al., 2011; Craig, 2009). For example, individuals often construe immoral acts as causing harm, even when no objective harm has been done (Gray, Schein, & Ward, 2014). Some scholars have even suggested that moral rules of justice evolved to protect people from harm (Gray, Young, & Waytz, 2012). That is, the reason we value justice may have more to do with its role in protecting individuals, than our preference for formal rules (Turiel, 1983; Turiel, Hildebrandt, & Wainryb, 1991; Rai & Fiske, 2011).

Consistent with this notion, we postulate that when justice causes harm to individuals (i.e., when justice and care conflict), concerns for care will supersede concerns for justice. Consequently, we expect observers to judge individuals who tell lies that help others to be more moral than individuals who are honest, but harm others.

The present research

Across three studies, we examine moral judgments of individuals who tell prosocial lies. In Study 1, we find that altruistic lies are perceived to be moral. We compare altruistic lies to selfish truths and find that individuals who lie to help others are perceived to be more moral than individuals who are honest. In Study 2, we disentangle deception, outcomes, and intentions. We find that intentions matter profoundly, but that the outcomes associated with deception do not influence judgments of morality.

In Study 3, we extend our investigation by disentangling the consequences of lying for the liar and the consequences of lying for the deceived party. We find that lies that neither help nor harm others are perceived to be immoral, but lies that help others, regardless of their cost to the liar, are perceived to be moral. Taken together, our studies demonstrate that the perceived ethicality of deception is labile. Intentions matter, and in at least some domains, caring for others is perceived to be more diagnostic of moral character than honesty.

Study 1

In Study 1, we examine moral judgments of altruistic lies and selfish truths. In our first study, participants judged an individual's actions in a deception game. In this experiment, lying benefited the deceived party at a cost to the deceiver. In this study, we find that altruistic lies are perceived to be moral.

Method

Participants. We recruited 215 participants from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee.

Procedure and Materials. We randomly assigned participants to one of two conditions in a between-subjects design. Participants observed and then judged an individual who either told an altruistic lie or was selfishly honest.

We told participants that they would observe the decision another participant had made in a prior exercise, called "The Number Game." The prior participant's decision in The Number Game served as our manipulation of lying.

The Number Game. We modified the deception game (Erat & Gneezy; 2012; Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Gneezy; 2005) to create The Number Game.

In The Number Game, two individuals were paired and randomly assigned to the role of either Sender or Receiver. The payoffs for each pair of participants were determined by the outcome of a random number generator and the choices made by the Sender and the Receiver. We refer to the individual who sent the message (who either lied or was honest) as "the Sender" throughout our studies. We refer to the Sender's partner (the individual who received the message) as "the Receiver." In our studies, participants observed and judged the behavior of one Sender in The Number Game.

The rules of The Number Game were as follows:

1. Senders were told a number supposedly generated by a random number generator (1, 2, 3, 4, or 5). In our study, the number was always 4.
2. The Sender then had to report the outcome of the random number generator to his/her partner, the Receiver. The Sender could send one of five possible messages to the Receiver. The message could read, “The number picked was [1, 2, 3, 4, or 5].”
   - The Sender knew that the number the Receiver chose (1, 2, 3, 4, or 5) determined the payment in the experiment. The Sender also knew that the only information the Receiver would have was the message from the Sender and that most Receivers chose the number indicated in the Sender’s message.
   - The Sender knew there were two possible payment options, A and B. If the Receiver chose the correct number, the Sender and the Receiver would be paid according to Option A. Otherwise, the Sender and the Receiver would be paid according to Option B.
3. In Study 1, the payoffs for Option A were $2 for the Sender and $0 for the Receiver. The payoffs for Option B were $1.75 for the Sender and $1 for the Receiver.
4. After receiving the Sender’s message, the Receiver chose a number: 1, 2, 3, 4 or 5. The Receiver knew that his/her choice determined the payment in the experiment, but the Receiver did not know the payoffs associated

with the choices. The Sender's message was the only piece of information the Receiver had.

Therefore, Senders faced the following options:

A. Send an honest message, e.g. "*The number picked was 4*."

Honesty was most likely to lead to an outcome that was costly for the Receiver, and beneficial for Sender (i.e. selfish).

B. Send a dishonest message, e.g. "*The number picked was [1, 2, 3, or 5]*."

Lying was most likely to lead to an outcome was beneficial for the Receiver, and was costly to the Sender (i.e. altruistic).

Design of the present study. Participants in our study learned the rules of The Number Game and had to pass a comprehension check to continue with the study.

Participants who passed the comprehension check learned about the behavior of a prior Sender. Specifically, participants observed a Sender who either sent an honest, but selfish message (Option A) or sent a deceptive, but altruistic message (Option B). We provide a summary of the payoffs associated with each choice in Table 1.

Dependent variables. After learning about the Sender's choice and the outcome of The Number Game, participants rated the Sender. We used seven-point Likert scales for all ratings.

Participants rated whether the Sender was ethical, moral, and a good person, and the extent to which the Sender's decision was ethical and moral (α = .93). These items were anchored at 1 = "Not at all" and 7 = "Extremely."

Participants also rated the benevolence of the Sender using two items: "This person is kind" and "This person has good intentions," (r(196)=.83), and the honesty of the Sender using two items: “This person is honest” and “This person tells the truth,” ($r(196)=.96$). These items were anchored at 1 = “Strongly disagree” and 7 = “Strongly agree.”

We also asked two multiple-choice recall questions to ensure participants had paid attention to our manipulations: “What message did the Sender send to his or her Receiver?” and “What was the actual number chosen by the random number generator?”[1]

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. We ran this study for the length of one laboratory session and we report all data exclusions and manipulations (Simmons, Nelson & Simonsohn, 2011).

Results

We report results from 196 participants (62.2% female; M_{age}= 20.4 years, SD = 2.38) who passed the comprehension check and completed the entire study; 19 participants failed the comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2. An exploratory factor analysis (Varimax rotation) yielded two factors that accounted for 74.06% of the variance. The first factor (eigenvalue = 5.32) consisted of the five morality items and the two benevolence items (loadings ≥ |.79|), and the second factor (eigenvalue = 1.77) consisted of the two honesty items (loadings ≥ |.86|).

[1] A total of 94.9% of participants correctly answered both recall questions. We report analyses for all participants who completed the entire study, but none of our findings change when we restrict our sample to only those who correctly answered the recall questions.

Although perceived benevolence and moral character are closely linked (e.g. Haidt & Graham, 2007) and loaded onto one factor, benevolence is theoretically distinct from morality (e.g. Haidt & Graham, 2007; Walker & Hennig, 2004; Leach et al., 2007; Brambilla et al., 2012). Consequently, we present analyses of benevolence and moral character separately. However, our results follow the same pattern if we combine these items into one construct. This was the case across all three of our studies.[2]

We conducted a one-way ANOVA to examine the effect of altruistic lying on perceived benevolence, honesty, and moral character. Participants judged altruistic liars to be more moral (*M* = 5.03, *SD* = 1.13) than selfish truth-tellers (*M* = 4.30, *SD* = 1.09), $F(1, 194) = 21.52$, $p < .001$, $\eta_p^2 = .100$ (see Figure 1). Participants also judged altruistic liars to be more benevolent (*M* = 5.36, *SD* = 1.29) than selfish truth-tellers (*M* = 3.98, *SD* = 1.32), $F(1, 194) = 53.90$, $p < .001$, $\eta_p^2 = .217$. However, altruistic liars were judged to be less honest (*M* = 3.50, *SD* = 1.19) than selfish truth-tellers (*M* = 5.06, *SD* = 1.40), $F(1, 194) = 69.98$, $p < .001$, $\eta_p^2 = .265$.

Figure 1 here

Discussion

In contrast to prior research that assumes that dishonesty undermines moral character, we find that, at least in some cases, lying increases moral character. In

[2] We thank an anonymous reviewer for his/her recommendation to explore our factor structure.

particular, we find that individuals perceive those who tell altruistic lies to be more moral than those who tell selfish truths. Study 1 suggests that when benevolence and honesty conflict, benevolence may be more important than honesty.

Study 2

In Study 2, we extend our investigation of deception and judgments of moral character. In this study, we use a deception game similar to the game we used in Study 1. In Study 2, however, we independently manipulate intentions, outcomes, and deception. This design enables us to measure the effect of deception, controlling for (selfish and altruistic) intentions. That is, in this study, we disentangle the effects of honesty and benevolence.

In Study 2, we also introduce informational uncertainty. In many practical contexts, individuals tell lies, but are uncertain of the consequences. For example, we may tell a colleague that his presentation was great with the intention of helping by boosting his confidence. This lie, however, may actually lead to an unintended outcome such as overconfidence and less preparation. We disentangle intentions from outcomes to investigate perceptions of lies that are told with good intentions but lead to negative outcomes.

Method

Participants. We recruited 237 participants from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee.

Procedure and Materials. As in Study 1, participants observed the decisions an individual made in an exercise called “The Number Game.” We randomly assigned participants to one of eight experimental conditions in a 2(Intentions: Altruistic vs.

Selfish) x 2(Lying: Lie vs. Truth) x 2(Outcome: Altruistic vs. Selfish) between-subjects design. Specifically, participants observed a Sender who either lied or sent an honest message, whose intentions were either selfish or altruistic, and whose choice ultimately led to an outcome that was either altruistic or selfish.

The Number Game. The Number Game in Study 2 was similar to the game we used in Study 1, with two notable changes. First, we introduced a stochastic element to the game to disentangle the effects of outcomes and intentions. Specifically, Senders in this game knew that the message that s/he selected was only delivered to the Receiver 75% of the time. Senders learned that 25% of the time, the computer overrode their decision and delivered the opposite message to the Receiver. That is, whether or not the Receiver received a truthful or deceptive message was probabilistically determined. In the actual experiment, the computer overrode the confederate Sender’s decision (i.e. intentions) half of the time so that our cells were evenly balanced.

Second, Senders in this experiment played The Number Game with one of two possible payment structures. These payment structures enabled us to manipulate whether deception or honesty was associated with selfish or altruistic intentions. We provide a summary of the payoffs associated with each choice in Table 1.

The first payment structure was identical to the one we used in Study 1. This payment structure represented the choice between selfish honesty (Option A) and altruistic lying (Option B). The second payment structure represents the choice between altruistic honesty and selfish lying. In the second payment structure, Senders learned that they would receive $1.75 and the Receiver would receive $1 if the Receiver chose the correct number (Option A). Otherwise, the Sender would receive $2 and the Receiver would receive $0 (Option B). (As in Study 1, the correct number was always 4).

Therefore, Senders with the second payment structure faced the following options:

A. Send an honest message, e.g. "*The number picked was 4*."

Honesty was most likely to lead to an outcome that benefitted the Receiver and was costly to the Sender (i.e. altruistic).

B. Send a dishonest message, e.g. "*The number picked was [1, 2, 3, or 5]*."

Lying was most likely to lead to an outcome that was costly to the Receiver and benefitted the Sender (i.e. selfish).

Design of the present study. Participants in our study learned the rules of The Number Game and had to pass a comprehension check to continue with the study.

Participants who passed the comprehension check then learned about the choice the Sender made in The Number Game. Participants observed a Sender who either lied or sent an honest message, who's choice was either intended to be altruistic or selfish, and who's choice led to an outcome (which was probabilistically determined) that was either altruistic or selfish.

For example, in the {*Lying*, *Altruistic Intentions*, *Selfish Outcomes*} condition, participants learned the following: the Sender sent a dishonest message to the Receiver; the Sender intended to help the Receiver earn an extra dollar (at a $0.25 cost to the Sender); the computer overrode the Sender's decision and the Receiver actually received the honest message. Consequently, the Receiver chose the correct number and earned $0 and the Sender earned $2. This selfish outcome, however, was not the Sender's intention.

Dependent variables. After learning about the Sender's choice and the outcome of The Number Game, participants rated the Sender. We collected the same measures in this study as those we used in Study 1 (α = .95; *r*'s > .86).

We also asked three multiple-choice recall questions to ensure participants had paid attention to our manipulations: "What message did the Sender send to his or her Receiver?", "What message did the Receiver receive?" and "What was the actual number chosen by the random number generator?"[3]

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. We ran this study for the length of one laboratory session and we report additional measures we collected in this study in the online supplemental materials.

Results

We report results from 211 participants (63.5% female; M_{age}= 24 years, SD = 7.21) who passed the comprehension check and completed the entire study; 26 participants failed the comprehension check and were automatically eliminated from the study. We present the means and standard deviations of each scale, as well as the inter-scale correlation matrix in Table 2. An exploratory factor analysis (Varimax rotation) yielded one factor that accounted for 77.10% of the variance (eigenvalue = 6.94). Consistent with Study 1, we report the results of our manipulations on moral character,

[3] A total of 75.8% of participants correctly answered all three recall questions. We report analyses for all participants who completed the entire study, but none of our findings change when we restrict our sample to only those who correctly answered all of the recall questions.

benevolence, and honesty separately. However, the pattern of results is the same when we combine all of our items into one measure of moral character.

We conducted a three-way ANOVA on our dependent variables, using *Intentions*, *Lying*, and *Outcomes* as factors. We found no main effects or interaction effects of *Outcomes*, and consequently collapsed across this factor in subsequent analyses. That is, outcomes did not influence moral judgments in this study, and our findings are unchanged when we include *Outcomes* as a factor. In other words, whether or not lying actually led to its intended consequence did not influence perceptions of moral character.

Moral character. A two-way ANOVA revealed a main effect of *Lying*, $F(1, 207) = 34.22$, $p < .001$, $\eta_p^2 = .142$, and a main effect of *Intentions*, $F(1, 207) = 77.26$, $p < .001$, $\eta_p^2 = .272$, on perceptions of the Sender's moral character. Specifically, participants believed that the Sender was more moral when s/he was honest ($M = 4.98$, $SD = 1.34$) than when s/he lied ($M = 3.97$, $SD = 1.46$) and when s/he had altruistic intentions ($M = 5.21$, $SD = 1.26$) than when s/he had selfish intentions ($M = 3.71$, $SD = 1.30$). We did not find a significant *Lying* x *Intentions* interaction, $F(1, 207) = 1.10$, $p = .295$, $\eta_p^2 = .005$.

In order to compare altruistic lying and selfish honesty, we conducted a series of planned contrasts. Consistent with Study 1, a contrast between the *Altruistic Lie* and the *Selfish Truth* conditions revealed that Senders who told altruistic lies were judged to be more moral than Senders who told selfish truths ($M = 4.80$, $SD = 1.30$ vs. $M = 4.31$, $SD = 1.27$), $t(100) = 2.04$, $p = .043$, $d = .38$. We depict these results in Figure 2. Notably, altruistic lies and altruistic truths were rated as moral, (significantly above the midpoint on the scale, $p < .001$). Only selfish lies were rated as immoral (significantly below the midpoint of the scale, $p < .001$).

Figure 2 here

Benevolence. A two-way ANOVA revealed a main effect of *Lying*, $F(1, 207) = 29.52$, $p < .001$, $\eta_p^2 = .125$, and a main effect of *Intentions*, $F(1, 207) = 92.91$, $p < .001$, $\eta_p^2 = .310$, on perceptions of the Sender's benevolence. Specifically, participants believed that the Sender was more benevolent when s/he was honest ($M = 4.92$, $SD = 1.44$) than when s/he lied ($M = 3.91$, $SD = 1.71$) and when s/he had altruistic intentions ($M = 5.32$, $SD = 1.38$) than when s/he had selfish intentions ($M = 3.54$, $SD = 1.42$). We also found a marginally significant *Lying* x *Intentions* interaction, $F(1, 207) = 2.95$, $p = .087$, $\eta_p^2 = .014$, such that selfish intentions, relative to altruistic intentions, were perceived to be less benevolent when they were associated with lying ($M_{\text{altruistic}} = 4.97$, $SD_{\text{altruistic}} = 1.58$ vs. $M_{\text{selfish}} = 2.91$, $SD_{\text{selfish}} = 1.14$), $t(104) = 5.61$, $p < .001$, $d = 1.56$, than when they were associated with honesty ($M_{\text{altruistic}} = 5.64$, $SD_{\text{altruistic}} = 1.10$ vs. $M_{\text{selfish}} = 4.21$, $SD_{\text{selfish}} = 1.40$), $t(105) = 2.64$, $p < .01$, $d = 1.14$. That is, selfishness is perceived to be less benevolent – or more malevolent – when it is associated with deception than when it is associated with honesty.

Planned contrasts between the *Selfish Truth* and the *Altruistic Lie* conditions revealed that participants perceived the Sender to be more benevolent when s/he told an altruistic lie (M = 4.97, SD = 1.58) than when s/he told a selfish truth (M = 4.21, SD = 1.40), $t(100) = 2.91$, $p < .01$, d = .51.

Honesty. A two-way ANOVA also revealed a main effect of *Lying*, $F(1, 207)$ = 167.35, $p < .001$, $\eta_p^2 = .447$, and a main effect of *Intentions*, $F(1, 207) = 35.46$, $p < .001$, $\eta_p^2 = .146$, on perceptions of the Sender's honesty. Specifically, participants believed that the Sender was more honest when s/he told the truth (M = 5.53, SD = 1.31) than when s/he lied (M = 3.14, SD = 1.54) and when s/he had altruistic intentions (M = 4.93, SD = 1.61) than when s/he had selfish intentions (M = 3.75, SD = 1.92).

We also found a significant *Lying* x *Intentions* interaction, $F(1, 207) = 5.18$, p =.024, $\eta_p^2 = .024$, such that the same lie was perceived to be less honest, relative to truth-telling, when it was associated with selfish intentions (M_{truth} = 5.18, SD_{truth} = 1.47 vs. M_{lie} = 2.42, SD_{lie} = 1.20), $t(103) = 10.68$, $p < .001$, d = 2.06, compared to altruistic intentions , (M_{truth} = 5.85, SD_{truth} = 1.07 vs. M_{lie} = 3.91, SD_{lie} = 1.51), $t(106) = 7.59$, $p < .001$, d = 1.48. In other words, an otherwise identical lie is perceived to be less dishonest when it is associated with altruism.

Planned contrasts between the *Selfish Truth* and the *Altruistic Lie* conditions revealed that participants perceived the Sender to be less honest when the Sender told an altruistic lie (M =3.91, SD = 1.51) than when the Sender told a selfish truth (M = 5.18, SD = 1.47), $t(100) = 4.83$, $p < .01$, d = .85.

Discussion

In Study 2, we manipulated intentions, deception, and outcomes independently and found that intentions influenced judgments of moral character more than deception or outcomes. In this study, participants judged Senders who told altruistic lies to be more moral than Senders who told selfish truths. In this study, the only decisions participants judged to be immoral were selfish lies.

We also found that judgments of honesty influenced judgments of benevolence and judgments of benevolence influenced judgments of honesty. Controlling for deceptive behavior, altruistic intentions signaled honest character, and controlling for intentions, honesty signaled benevolent character. That is, a single moral behavior triggered a halo of unrelated moral trait attributions. However, as expected, judgments of benevolence were more sensitive to intentions and judgments of honesty were more sensitive to deception.

Importantly, we also found that outcomes, when disentangled from deception and intentions, had no effect on moral judgments of deception. These findings offer new insight into the psychology of deception. The consequences of deception, and unethical behavior generally, are uncertain. Interestingly, we find that whether or not (dis)honesty actually helped or hurt did not influence judgments of moral character.

Study 3

In Studies 1 and 2, we examined altruistic lies. Altruistic lies are costly for the liar and beneficial for the target. In Study 3, we manipulate the consequences of deception for the Sender and the Receiver independently. This enables us to disentangle attributions of benevolence from attributions of altruism, and to contrast altruistic lies with non-altruistic prosocial lies. In this design, we also include a control condition that directly examines perceptions of lying, free of consequences for the liar and the deceived party.

Method

Participants. We recruited 300 adults to participate in an online survey via Amazon's Mechanical Turk.

Procedure and Materials. As in Studies 1 and 2, participants learned about the decisions an individual made in an exercise, called "The Number Game." In Study 3, we randomly assigned participants to one of eight cells in a 2(Lying: Lie vs. Truth) x 2(Consequences for the Sender: None vs. Cost) x 2(Consequences for the Receiver: None vs. Benefit) between-subjects design. That is, participants learned the following about a Sender: the Sender either lied or was honest; lying was either costly for the Sender or had no effect on the Sender; and lying either benefited the Receiver or had no effect on the Receiver.

The Number Game. The Number Game in Study 3 was similar to the game we used in Study 1. Participants learned about a Sender who either accurately reported or lied about the outcome of a random number generator. We manipulated the payoffs associated with honesty and lying by manipulating the payments associated with decisions in The Number Game.

In Study 3, participants viewed one of four possible payment structures. These payment structures varied the payoffs associated with lying for the Sender and the Receiver. These payment structures, depicted in Table 1, operationalized one of four types of lies:

1. Control Lie: Lying, relative to honesty, had no effect on the Sender or the Receiver.
2. Self-sacrificial Lie: Lying, relative to honesty, hurt the Sender and had no effect on the Receiver.

3. Prosocial Lie: Lying, relative to honesty, had no effect on the Sender and benefited the Receiver.
4. Altruistic Lie: Lying, relative to honesty, hurt the Sender and benefited the Receiver.

Participants learned about a Sender who faced the opportunity to tell one of the four types of lies described above. For example, Senders in the Prosocial Lie conditions had the opportunity to send a dishonest message to the Receiver, which would have no effect on the Sender but would benefit the Receiver. In each condition, participants learned that the Sender either lied or told the truth. Honesty was associated with the same payoffs in all conditions ($2 for the Sender, $0 for the Receiver).

As in Studies 1 and 2, participants had to pass a comprehension check to ensure that they understood The Number Game before they could continue with the experiment. Participants who failed the comprehension check were automatically removed from the study.

Dependent variables. After learning about the Sender's choice and passing the comprehension check, participants rated the Sender. We developed new scales in Study 3 to better distinguish judgments of moral character from judgments of benevolence and honesty.

Moral character. We measured moral character using six items (α = .96) we adapted from Uhlmann, Zhu, & Tannenbaum (2013). Specifically, we asked participants whether the Sender had "good moral character" (1 = "Extremely immoral character", 7 = "Extremely moral character"), was "an ethical person" (1 = "Extremely unethical person," 7 = "Extremely ethical person"), was "a morally good person" (1 = "Extremely morally bad person," 7 = "Extremely morally good person"), "will behave morally in the future" (1= "Extremely likely to behave immorally", 7 = "Extremely likely to behave morally"), "made the morally right decision" (1 = "Extremely immoral decision" 7 = "Extremely moral decision"), and "made the ethical decision" (1 = "Extremely unethical decision", 7 = "Extremely ethical decision.").

Benevolence. Participants rated the Sender's benevolence using four items (α = .89): This person is [benevolent, empathic, caring, selfish (reverse-scored)]. These items were anchored at 1 = "Strongly disagree" and 7 = "Strongly agree." We adapted this scale from Uhlmann et al.'s (2013) perceived empathy scale, but we included additional items to measure benevolence rather than general empathy (e.g. selfish, benevolent).

Honesty. Participants rated the honesty of the Sender using three items (α =.91): This person [is honest, tells the truth, is deceptive (reverse-scored)]; 1 = "Strongly disagree" and 7 = "Strongly agree."

As in Study 1, we also asked two multiple-choice recall questions to ensure participants had paid attention to our manipulations: "What message did the Sender send to his or her Receiver?" and "What was the actual number chosen by the random number generator?"[4]

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. We

[4] A total of 87.0% of participants correctly answered both recall questions. We report analyses for all participants who completed the entire study, but none of our findings change when we restrict our sample to only those who correctly answered the recall questions.

determined our sample size in advance and we report all data exclusions and manipulations.

Results

We report results from 269 participants (45.6% female; M_{age}= 32 years, SD = 11.03) who passed the comprehension check and completed the entire study; 31 participants failed the comprehension check and were automatically eliminated from the study. We present the means and standard deviations of each scale, as well as the inter-scale correlation matrix in Table 2. Although we devised new scales to measure moral character and benevolence, these constructs remained closely related and loaded together on one factor (Exploratory factor analysis, Varimax rotation, loadings ≥ |.65|). Consistent with Studies 1 and 2, we report the results of our manipulations on moral character and benevolence separately, but our findings remain the same when we combine moral character and benevolence into one scale.

We conducted a three-way ANOVA on our dependent variables, using *Lying*, *Consequences for the Sender*, and *Consequences for the Receiver* as factors. We found no main effects or interaction effects of *Consequences for the Sender*. That is, whether or not lying was costly for the Sender did not influence judgments of the Sender's moral character, benevolence, or honesty. Notably, prosocial lies were not judged differently than were altruistic lies. We collapse across *Consequences for the Sender* in our subsequent analyses, but our findings are unchanged when we include *Consequences for the Sender* as a factor.

Moral character. We find no main effects of *Lying*, $F(1, 265) = .02$, $p = .887$, $\eta_p^2 = .000$, or *Consequences for the Receiver*, $F(1, 265) = 2.70$, $p = .100$, $\eta_p^2 = .010$, on perceptions of moral character. Importantly, we did find a significant *Lying* x *Consequences for the Receiver* interaction, $F(1, 265) = 41.20$, $p < .001$, $\eta_p^2 = .135$. When lying helped the Receiver, the Sender was judged to be *more* moral when s/he lied (M = 4.88, SD = 1.36) than when s/he told the truth (M = 3.90, SD = 1.37), $t(132) = 4.41$, $p < .001$, $d = 1.01$. Conversely, when lying had no effect on the Receiver, the Sender was judged to be *less* moral when s/he lied (M = 3.62, SD = 1.25) than when s/he told the truth (M = 4.64, SD = 1.10), $t(135) = 4.66$, $p < .001$, $d = .87$. Consistent with our findings in Studies 1 and 2, prosocial lying increased perceptions of moral character. Lies that neither helped nor harmed the Receiver, however, decreased perceptions of the Sender’s moral character.

Benevolence. A two-way ANOVA revealed main effects of *Lying*, $F(1, 265) = 3.76$, $p = .053$, $\eta_p^2 = .014$, and *Consequences for the Receiver*, $F(1, 265) = 5.61$, $p = .020$, $\eta_p^2 = .021$, on perceived benevolence. Specifically, participants believed that the Sender was more benevolent when s/he lied (M = 4.09, SD = 1.49) than when s/he was honest (M = 3.81, SD = 1.12) and when lying helped the Receiver (M = 4.12, SD = 1.54) than when it had no effect Receiver (M = 3.78, SD = 1.04).

However, these effects were qualified by a significant *Lying* x *Consequences for the Receiver* interaction, $F(1, 265) = 45.98$, $p < .001$, $\eta_p^2 = .148$. When lying helped the Receiver, the Sender was judged to be more benevolent when s/he lied (M = 4.76, SD = 1.50) than when s/he told the truth (M = 3.48, SD = 1.30), $t(132) = 6.13$, $p < .001$, $d = .91$.

Conversely, when lying did not help the Receiver, the Sender was judged to be less benevolent when s/he lied (*M* = 3.41, *SD* = 1.13) than when s/he told the truth (*M* = 4.13, *SD* = 0.79), *t*(135) = 3.44, *p* < .001, *d* = .74. This interaction demonstrates that the main effect of lying on benevolence is driven by judgments of prosocial lies.

Honesty. A two-way ANOVA revealed a significant effect of *Lying*, *F*(1, 265) = 77.76, *p* < .001, η_p^2 = .227, on perceived honesty. Participants rated the Sender as less honest when s/he lied (*M* = 3.41, *SD* = 1.57) than when s/he told the truth (*M* = 4.97, *SD* = 1.39). We find no effect of *Consequences for the Receiver* on perceived honesty, *F*(1, 258) = 1.19, *p* = .276, η_p^2 = .004. The Sender was judged to be similarly honest when lying helped the Receiver (*M* = 4.28, *SD* = 1.51) and when lying had no effect on the Receiver (*M* = 4.11, *SD* = 1.81).

We do find a significant *Lying* x *Consequences for the Receiver* interaction, *F*(1, 265) = 13.11, *p* < .001, η_p^2 = .047. Consistent with our findings in Study 2, the difference in perceived honesty between a truth and a lie was greater when lying had no effect on the Receiver, (M_{truth} = 5.19, SD_{truth} = 1.31 vs. M_{lie} = 2.99, SD_{lie} = 1.57), *t*(135) = 8.84, *p* < .001, *d* = 1.52, than when lying helped the Receiver (M_{truth} = 4.74, SD_{truth} = 1.43 vs. M_{lie} = 3.83, SD_{lie} = 1.46), *t*(132) = 3.65, *p* < .001, *d* = .63. That is, deception was perceived to be more honest when it helped another person.

Judgments of different types of lies. Although *Consequences for the Sender* had no effect on moral judgments, we sought to better understand perceptions of lies with respect to our control condition. We conducted planned contrasts for each type of lie and we depict these results in Figures 3-5. We summarize perceptions of moral character for each type of lie below and in Table 3.

In our control condition, lying was inconsequential. That is, deception and honesty resulted in the same payoffs. In this condition, participants rated the Sender as significantly less moral when s/he lied (*M* = 3.58, *SD* = 1.30) than when s/he told the truth (*M* = 4.52, *SD* = 1.08), *t*(67) = 3.03, *p* < .01, *d* = .79. This contrast documents an aversion to lying.

We find the same pattern of results for self-sacrificial lies: Participants rated the Sender as significantly less moral when s/he told a self-sacrificial lie (*M* = 3.68, *SD* = 1.20) than when s/he told the truth (*M* = 4.75, *SD* = 1.12), *t*(67) = 3.44, *p* = .001, *d* = .92. We find no difference between ratings of self-sacrificial lies and inconsequential lies.

We find the opposite pattern of results for prosocial and altruistic lies. Participants rated the Sender as significantly *more* moral when s/he told a prosocial lie (*M* = 5.03, *SD* = 1.32) than when s/he told the truth (*M* = 3.87, *SD* = 1.45), *t*(62) = 3.58, *p* < .001, *d* = .84. Similarly, participants rated the Sender as significantly *more* moral when s/he told an altruistic lie (*M* = 4.75, *SD* = 1.40) than when s/he told the truth (*M* = 3.93, *SD* = 1.31), *t*(69) = 2.69, *p* < .01, *d* = .60. We find no difference between ratings of prosocial and altruistic lies.

Prosocial lies and altruistic lies were both rated to be more moral than lies that had no consequences (*t*s > 3.92, *p*s < .01, *d*s > .86). Truth-telling was also rated to be more moral in the control condition than truth-telling in the altruistic lie condition (*t* = 2.04, *p* = .042, *d* = .49) and marginally more moral than truth-telling in the prosocial lie condition (*t* = 1.87, *p* = .063, *d* = .51), even though the payoffs for truth-telling were identical across these conditions. Taken together, our results suggest that having the opportunity to lie to help another party causes lying to appear to be more moral *and* causes honesty to appear to be less moral.

Figures 3-5, Table 3

Discussion

In Study 3, we find that individuals who lie to help others, regardless of whether or not the lie is costly for them, are perceived to be more moral than individuals who are honest. Consistent with our findings in Studies 1 and 2, prosocial motives influenced perceptions of moral character more than deception did.

In addition, we find evidence of a direct distaste for lying. Individuals who told lies that had no consequences for either themselves (the liars) or the deceived party were perceived to be less moral than individuals who were honest. Consistent with Study 2, this result suggests that perceptions of deception are not solely determined by the consequences and intentions associated with lying. To our knowledge, this is the first study to examine moral judgments of deception, independent of its consequences.

General discussion

Because extant research has conflated deception with self-serving motives and outcomes, our understanding of deception is limited. We know little about how common forms of deception, and conflicts between honesty and benevolence broadly, influence judgment and behavior.

Across three studies, we explore moral judgments of prosocial lies. In Study 1, we find that altruistic lies are perceived to be *more* moral than selfish truths. In Study 2, we independently manipulate deception, prosocial intentions, and prosocial outcomes. We find that outcomes did not influence judgments of moral character, but, consistent with prior work, intentions mattered profoundly (e.g. Alicke, 1992; Ames & Fiske, 2013). Although deception also had an effect on moral character, we find that the effect of intentions was larger than that of deception. Consequently, individuals with altruistic intentions are perceived to be more moral, more benevolent, and more honest, *even when they lie*.

In our third study, we examine different types of lies. We find that perceptions of prosocial lies do not depend on self-sacrifice; altruistic lies and prosocial lies both increase perceptions of moral character. We also find evidence for a direct aversion to deception; lies that had no consequences for the liar or the deceived party were perceived to be immoral.

Theoretically, our findings make several contributions. First, we demonstrate the importance of a broader conceptualization of deception. Whereas prior studies of ethical decision-making and moral character have conflated deception with selfishness, we distinguish self-serving deception from altruistic, prosocial, and inconsequential deception. We find that individuals who tell lies that help others are perceived to be moral.

Second, our investigation expands the study of ethical decision making to conflicts between honesty and benevolence. Prior work has studied violations of either honesty *or* benevolence in isolation, or acts that violate *both* honesty and benevolence at the same time. To our knowledge, our work is the first to examine character judgments when these values conflict. In our studies, benevolence was more closely related to moral character than honesty. Although we cannot conclude that the principle of benevolence is *always* more important than honesty, we can conclude that, at least in some cases, prosociality has a greater effect on moral character than does deception.

Third, our findings offer insight into lay beliefs about universal moral values. We conceptualize prosocial lying not only as a conflict between honesty and benevolence, but more broadly as a conflict between justice and care. Prosocial lying reflects the violation of an ethical rule in order to care for another person. Providing care and avoiding harm towards others is a fundamental human tendency. Our findings demonstrate that care is, at least sometimes, more important than justice for moral character judgments. Importantly, our work illustrates the importance of studying conflicting moral rules (e.g. Broeders, van den Bos, Müller & Ham, 2011).

Our study of justice and care also extends our understanding of deontological and utilitarian principles. Deontological philosophers argue that lying is immoral because it violates the sacred value of the right to truth (Kant, 1785). Utilitarians argue that the ethicality of lying depends upon its consequences (e.g. Martin Luther, cited in Bok, 1978; Bentham, 1843). Our findings support elements of both schools of thought. When lies are inconsequential, individuals do penalize liars for violating the principle of honesty. However, when lies help others the utilitarian consideration of consequences outweighs the deontological prohibition of deception. These findings reflect the ambivalence that we have for deception and quite possibly, many other moral violations. Perhaps our true moral compass reflects both deontological and utilitarian values.

Our work also contributes to the literature on moral dilemmas. In our investigation, we created a framework to explore a common type of ethical dilemma. Although prior research on ethical dilemmas and moral reasoning has substantially expanded our understanding of ethical decision-making, most of this work has studied extreme circumstances. For example, scholars use paradigms such as the trolley problem to study the dilemma of killing one person to save many (Broeders, van den Bos, Müller & Ham, 2011; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Moore, Clark, & Kane, 2008), and the Heinz dilemma to study the dilemma of stealing an expensive drug to save a life (Kohlberg, 1981). Our investigation extends our understanding of moral judgment by exploring conflicting moral principles in a context that pervades our everyday lives.

Limitations and future directions

Future work is needed to understand judgments of the full range of deceptive behaviors. In our studies, the intentions associated with lying were clear. In many settings, however, a liar's intentions are ambiguous. In addition to benefiting others, many prosocial lies also benefit the deceiver. For example, when a colleague asks if you enjoyed her talk, the prosocial lie ("It was great!") may benefit both the colleague (causing her to feel better) and the deceiver (avoiding a protracted discussion about the fatal flaws in the research). That is, a single act of deception may be both prosocial and self-serving. Future research should examine how individuals judge lies that have mixed or uncertain motives.

Future work should also explore how prosocial lying influences a broader set of perceptions and behaviors. For example, a substantial body of research suggests that deception harms trust (e.g. Boles, Croson & Murninghan, 2000; Schweitzer, Hershey, &

Bradlow, 2006), but trust scholars have primarily investigated the consequences of selfish lies. Recent studies suggest that the relationship between deception and trust depends on the extent to which the liar's motives are believed to be prosocial (Levine & Schweitzer, 2013; Wang & Murnighan, 2013). More research is needed to understand when prosocial lies, and ethical violations broadly, can increase trust and cooperation.

Prosocial lying may also signal negative character traits. For example, prosocial lying may harm perceptions of moral traits other than benevolence and honesty, such as courage (Walter & Hennig, 2004; Uhlmann, Zhu, & Tannenbaum, 2013). If individuals consider prosocial lying to be cowardly, prosocial lying may decrease, rather than increase, perceptions of moral character. Prosocial lying may also have negative effects over time, as the signal value of benevolence weakens and the liar becomes less credible.

More broadly, we call for future research to expand our understanding of conflicts between moral principles. A substantial literature has explored characteristics of ethical decision-making when the ethical choice is clear (e.g., Mazar et al., 2008; Tenbrunsel, 1998; Boles, Croson, & Murninghan, 2000); and a large literature has explored conflicts between deontological and utilitarian principles (e.g. Greene et al., 2004; Moore et al., 2008). However, scholars have largely overlooked behaviors that signal competing moral values (for exceptions, see Gino & Pierce, 2009; Gino & Pierce, 2010).

Ethicists and psychologists have argued that morality reflects a set of values, such as honesty, benevolence, restraint, and loyalty (e.g. Leach et al., 2007; Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Wojciszke, Bazinska, & Jaworski, 1998; Reeder & Spores, 1983; Noddings, 1984; Walker & Hennig, 2004; Blasi, 1984; Aquino & Reed, 2002) and that these values reflect different moral foundations, such as justice, care, purity, and authority (e.g. Haidt & Graham, 2007). We investigate the conflict between justice and care, but important work remains with respect to understanding how individuals resolve—and judge others who resolve—conflicts between other principles, such as fairness and mercy (Kidder, 1995; Wiltermuth & Flynn, 2012; Flynn & Wiltermuth, 2010), and harm versus purity (e.g. Uhlmann & Zhu, 2013). We argue that the study of conflicting moral principles represents a substantial challenge for ethical decision-making scholars.

Conclusion

Scholars, managers, and parents routinely extoll the virtues of honesty and warn of the dire consequences of deception. Deception, however, is not only pervasive but also employed by some of the same people who enjoin others to avoid its' use. In this work, we disentangle deception from intentions and outcomes. We investigate prosocial lies, lies told to benefit others, and find that prosocial lies are judged to be more moral than honesty.

Prosocial lies represent a conflict between two moral foundations: justice and care. Prior work has overlooked how individuals resolve conflicts between moral principles, and we call for future work to develop this important line of investigation.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*(3), 368.

Ames, D. L., & Fiske, S. T. (2013). Intentional Harms Are Worse, Even When They're Not. *Psychological Science*, *24*(9), 1755-1762.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423.

Barnes, MA, Claire. "Should Split Parents Ever Lie to Their Children?." *The Huffington Post*. TheHuffingtonPost.com, 2 Apr. 2013. Web. 11 Dec. 2013. <http://www.huffingtonpost.com/claire-n-barnes-ma/should-split-parents-ever_b_2964221.html>.

Bok, S. (1978). *Lying: Moral choices in public and private life.* New York, NY: Pantheon.

Bentham, J. (1843/1948). *An introduction to the principles of morals and legislation*. Oxford, UK: Basil Blackwell. (Original work published 1843).

Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational behavior and human decision processes, 83*(2), 235-259.

Blasi, A. (1984). Moral identity: Its role in moral functioning. In W. Kurtines & J. Gewirtz (Eds.), Morality, moral behavior and moral development (pp. 128–139). New York: Wiley.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*(2), 135-143.

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., & Yzerbyt, V. Y. (2012). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *51*(1), 149-166.

Broeders, R., van den Bos, K., Müller, P. A., & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology*, *47*(5), 923-934.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge, England: Cambridge University Press.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology*, *45*(6), 1321-1324.

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

Craig, K. D. (2009). The social communication model of pain. *Canadian Psychology/ Psychologie canadienne*, *50*(1), 22.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143-159.

Darley, J. M., & Pittman, T. S. (2003). The psychology of compensatory and retributive justice. *Personality and Social Psychology Review*, *7*(4), 324-336.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships.

Journal of Personality and Social Psychology, *74*(1), 63-79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979-995.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: lies are told to those who care. *Journal of Personality and Social Psychology*, *71*(4), 703.

De Waal, F. B. (2008). Putting the altruism back into altruism: The evolution of empathy. *Annual Review of Psychology*, *59*, 279-300.

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. *Psychology of Learning and Motivation*, *50*, 307-338.

Dreber, A., & Johannesson, M. (2008). Gender differences in deception. *Economics Letters*, *99*(1), 197-199.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

Flynn, F. J. & Wiltermuth, S. S. (2010). Who's with me? False consensus, social networks, and ethical decision making in organizations. *Academy of Management Journal*, 53(5), 1074-1089.

Gaspar, J. P., & Schweitzer, M. E. (2013). The Emotion Deception Model: A Review of Deception in Negotiation and the Role of Emotion in Deception. *Negotiation and Conflict Management Research*, *6*(3), 160-179.

Gilligan, C. (1982). *In a different voice: Psychological theory and women's development* (Vol. 326). Harvard University Press.

Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, *20*(9), 1153-1160.

Gino, F., & Pierce, L. (2010). Lying to level the playing field: Why people may dishonestly help or hurt others to create equity. *Journal of Business Ethics, 95*(1), 89-103.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic Review, 95*(1), 384-394.

Goffman, E. (1967). *Interaction ritual: Essays on face-to-face behavior*. Garden City, NJ: Anchor.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101-124.

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*. Advance online publication. http://dx.doi.org/10.1037/a0036149

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*(1), 98-116.

Iezzoni, L. I., Rao, S. R., DesRoches, C. M., Vogeli, C., & Campbell, E. G. (2012). Survey shows that at least some physicians are not always open or honest with patients. *Health Affairs*, *31*(2), 383-391.

Kant, I. (1785). *Foundation of the metaphysics of morals*. Beck LW, translator. Indianapolis: Bobbs-Merrill; 1959.

Kashy, D. A., & DePaulo, B. M. (1996). Who lies? *Journal of Personality and Social Psychology*, *70*(5), 1037.

Kidder, R.M. (1995). *How good people make tough choices: Resolving the dilemmas of ethical living*. New York: Fireside.

Kohlberg, L. (1969). *Stage and sequence: The cognitive-developmental approach to socialization* (pp. 347-480). New York: Rand McNally.

Knobe, J. (2004). Intention, intentional action and moral considerations. *Analysis*, *64*(282), 181-187.

Koning, L., Steinel, W., Beest, I. V., & van Dijk, E. (2011). Power and deception in ultimatum bargaining. *Organizational Behavior and Human Decision Processes*,*115*(1), 35-42.

Lapsley, D. K., & Lasky, B. (2001). Prototypic moral character. *Identity: An International Journal of Theory and Research*, *1*(4), 345-363.

Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*(2), 234.

Levine, E. E., & Schweitzer, M. E., (2013). Prosocial lies: When deception breeds trust. Working Paper. University of Pennsylvania.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633-644.

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of experimental social psychology*, *45*(3), 594-597.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549-557.

Noddings, N. (1984). Caring: A feminine approach to ethics and moral education. Berkeley, CA: University of California Press.

Nyberg, D. (1993). *The varnished truth: Truth telling and deceiving in ordinary life*. Chicago: University of Chicago Press.

Palmieri, J. J., & Stern, T. A. (2009). Lies in the doctor-patient relationship. *Primary care companion to the Journal of clinical psychiatry*, *11*(4), 163.

Park, Alice. "White Coats, White Lies: How Honest Is Your Doctor." *Time Magazine*, 10 Dec. 2011. Web. 11 Dec. 2013. <http://healthland.time.com/2012/02/09/white-coats-white-lies-how-honest-is-your-doctor/>.

Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39*(6), 653-660.

Reeder, G. D. (2009). Mindreading: Judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, *20*(1), 1-18.

Reeder, G. D., & Spores, J. M. (1983). The attribution of morality. *Journal of Personality and Social Psychology*, *44*(4), 736.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology*, *9*(4), 283.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*(1), 57.

Ruedy, N., Moore, C., Gino, F., & Schweitzer, M. (2013). The Cheater's High: The Unexpected Affective Benefits of Unethical Behavior. *Journal of Personality and Social Psychology*, *105*(4), 531-548.

Schweitzer, M. E., & Croson, R. (1999). Curtailing deception: The impact of direct questions on lies and omissions. *International Journal of Conflict Management*,*10*(3), 225-248.

Schweitzer, M. E., DeChurch, L. A., & Gibson, D. E. (2005). Conflict Frames and the Use of Deception: Are Competitive Negotiators Less Ethical? *Journal of Applied Social Psychology*, *35*(10), 2123-2149.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1-19.

Shalvi, S. (2012). Dishonestly increasing the likelihood of winning. *Judgment and Decision Making*, *7*(3), 292-303.

Shalvi, S., Dana, J., Handgraaf, M. J., & De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, *115*(2), 181-

190.

Shu, L. L., Mazar, N., Gino, F., Ariely, D., & Bazerman, M. H. (2012). Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences*, *109*(38), 15197-15200.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359-1366.

Steinel, W., & De Dreu, C. K. (2004). Social motives and strategic misrepresentation in social decision making. *Journal of Personality and Social Psychology*, *86*(3), 419.

Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, *47*(6), 1249-1254.

Tenbrunsel, A. E. (1998). Misrepresentation and expectations of misrepresentation in an ethical dilemma: The role of incentives and temptation. *Academy of Management Journal*, *41*(3), 330-339.

Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, *42*(1), 69-77.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties, inconsistencies and consistencies: I. *Monographs of the society for research in* *child development*.

Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, 126, 326–334.

Uhlmann, E. L., & Zhu, L. (2013). Acts, persons, and intuitions: Person-centered cues and gut reactions to harmless transgressions. *Social Psychological and Personality Science*.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage*, *40*(4), 1912-1920.

Yuill, N., & Perner, J. (1988). Intentionality and knowledge in children’s judgments of actor's responsibility and recipient's emotional reaction. Developmental Psychology, 24, 358–365.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: Just, brave, and caring. *Journal of Personality and Social Psychology*, 86, 629–647.

Wang, L. & Murninghan, J.K. (2013). Trust, White Lies, and Harsh Truths. Working Paper. City University of Hong Kong.

Wiltermuth, S., & Flynn, F. (2012). Power, Moral Clarity, and Punishment in the Workplace. *Academy of Management Journal*, 56, 1002-1023.

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European Review of Social Psychology*, *16*(1), 155-188.

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251-1263.

TABLES

Table 1
Payoffs used in each study

	Type of Lie		Payoffs associated with Truth	Payoffs associated with Lie
Study 1	Altruistic Lie	Sender	$2.00	$1.75
		Receiver	$0.00	$1.00
Study 2	Altruistic Lie	Sender	$2.00	$1.75
		Receiver	$0.00	$1.00
	Selfish Lie	Sender	$1.75	$2.00
		Receiver	$1.00	$0.00
Study 3	Control Lie	Sender	$2.00	$2.00
		Receiver	$0.00	$0.00
	Self-Sacrificial Lie	Sender	$2.00	$1.75
		Receiver	$0.00	$0.00
	Prosocial Lie	Sender	$2.00	$2.00
		Receiver	$0.00	$1.00
	Altruistic Lie	Sender	$2.00	$1.75
		Receiver	$0.00	$1.00

Note. In Study 2, the values displayed correspond to the intended outcome, but not necessarily the realized outcome, associated with each choice. In Study 2, the computer overrode the Sender’s choice 25% of the time, such that the computer sent an honest message in place of a dishonest message, or a dishonest message in place of an honest message.

Table 2. Scale Statistics in Studies 1, 2, and 3

Study 1

Scale	M(SD)	1	2
1. Moral Character	4.67 (1.16)		
2. Benevolence	4.67 (1.68)	0.77*	
3. Honesty	4.29 (1.51)	0.24*	0.10

Study 2

Scale	M(SD)	1	2
1. Moral Character	4.47 (1.49)		
2. Benevolence	4.44 (1.66)	0.89*	
3. Honesty	4.34 (1.87)	0.72*	0.66*

Study 3

Scale	M(SD)	1	2
1. Moral Character	4.27 (1.37)		
2. Benevolence	3.95 (1.32)	0.83*	
3. Honesty	4.19 (1.67)	0.64*	0.48*

Note. *p < .01.

Table 3. Summary of Results (Study 3)

	Consequences of Lying		Perceptions of
Type of Lie	To Sender	To Receiver	moral character
Prosocial Lie	No consequences	Helps	Increase
Altruistic Lie	Harms	Helps	Increase
Self-Sacrificial Lie	Harms	No consequences	Decrease
Inconsequential Lie (Control)	No consequences	No consequences	Decrease

FIGURES

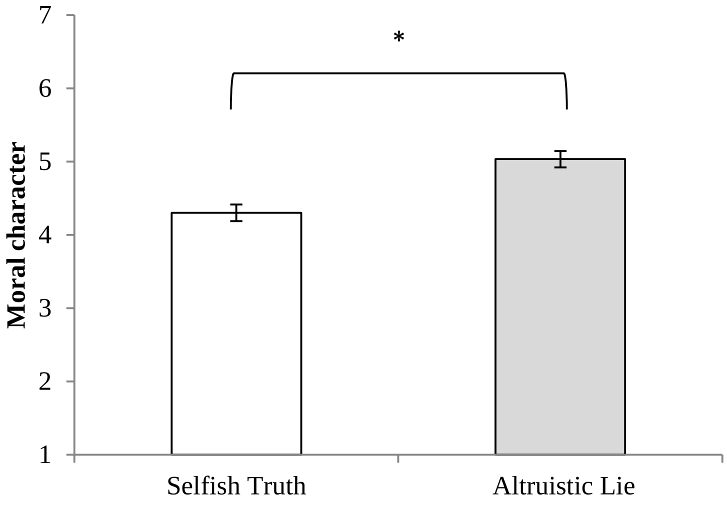


Figure 1: Moral character judgments in Study 1. Error bars represent ±1 SE. * *p* < .05.

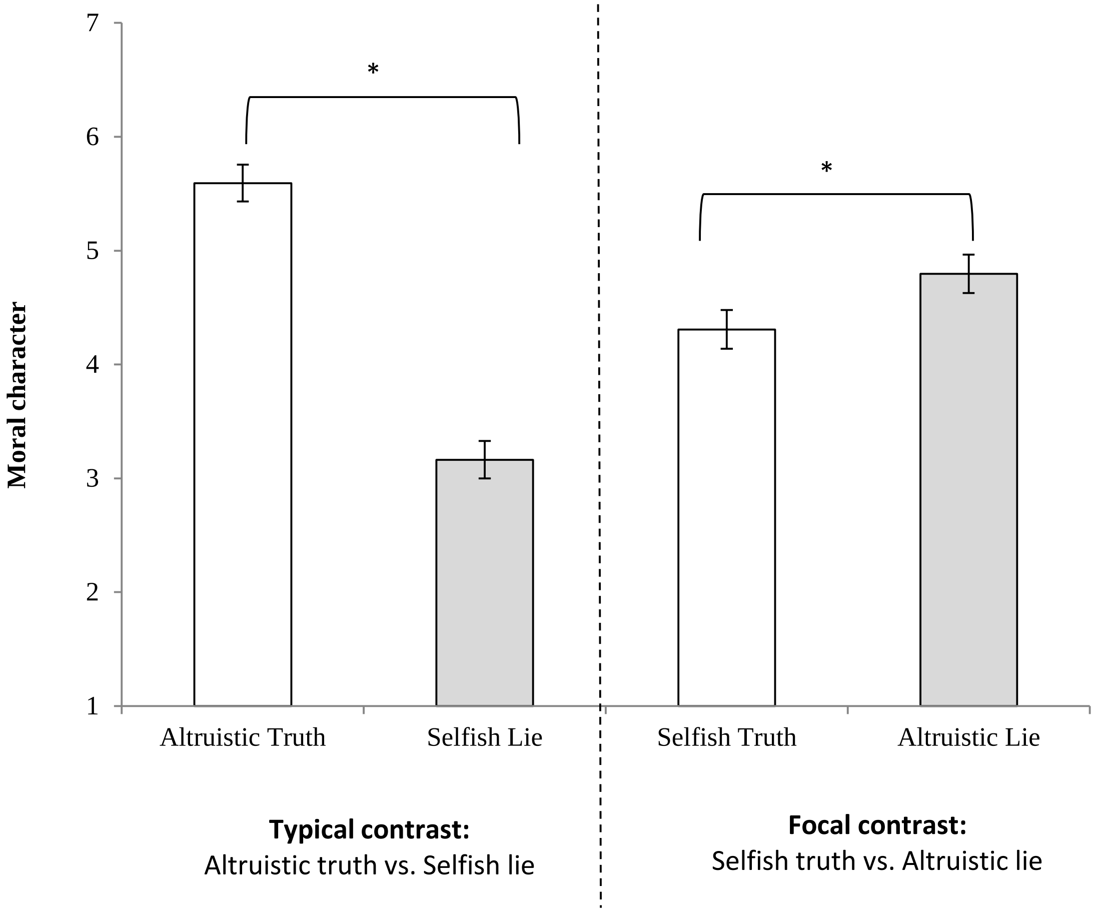


Figure 2: Moral character judgments in Study 2. Error bars represent ±1 SE. * $p < .05$.

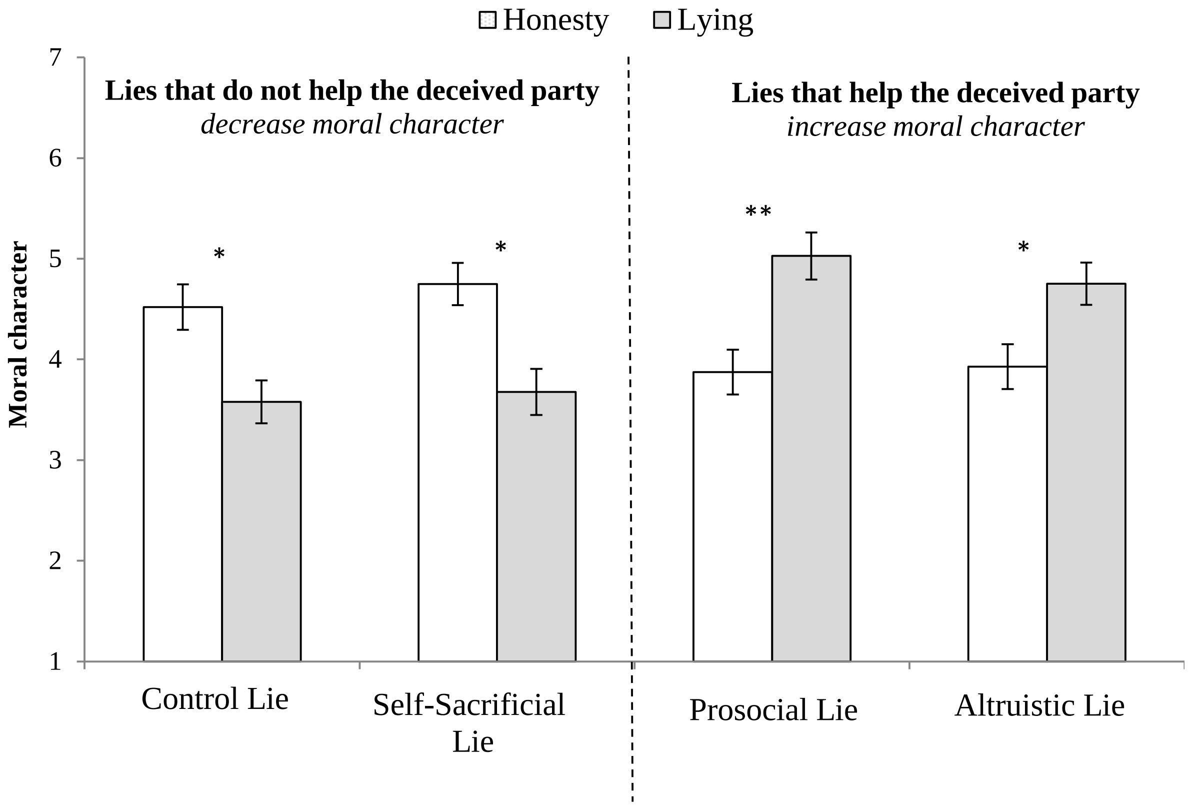


Figure 3: Moral character judgments in Study 3. Error bars represent ±1 SE. * *p* < .01, ** *p* < .001.

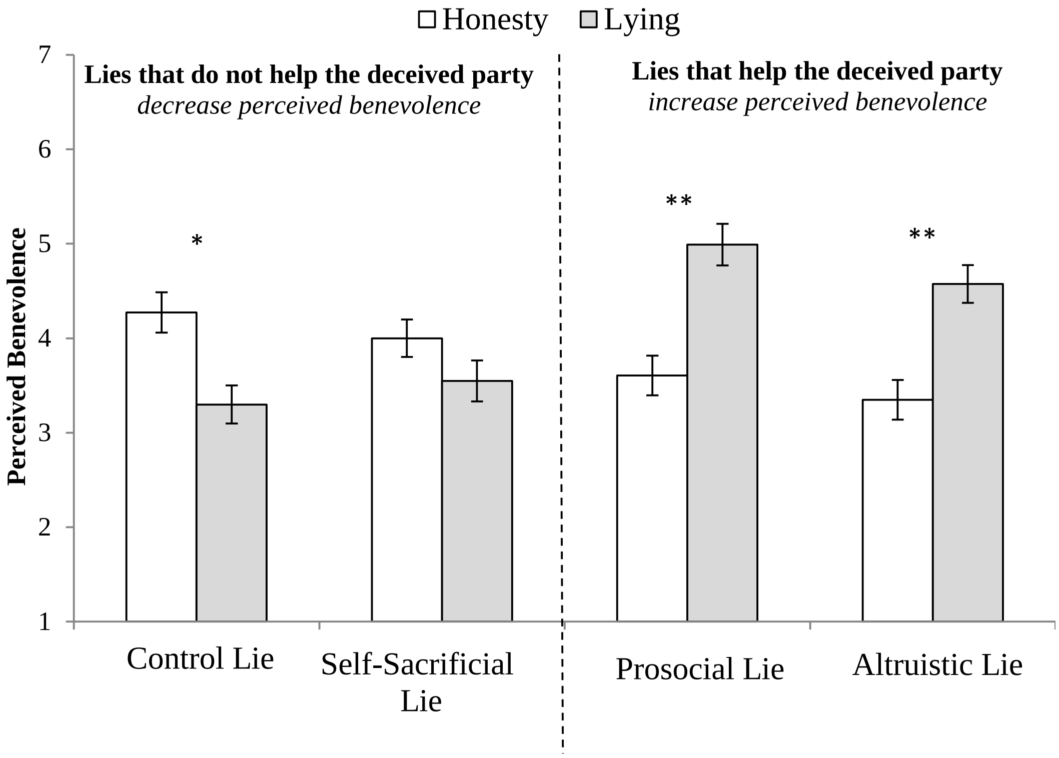


Figure 4: Perceived benevolence in Study 3. Error bars represent ±1 SE. * *p* < .01, ** *p* < .001.

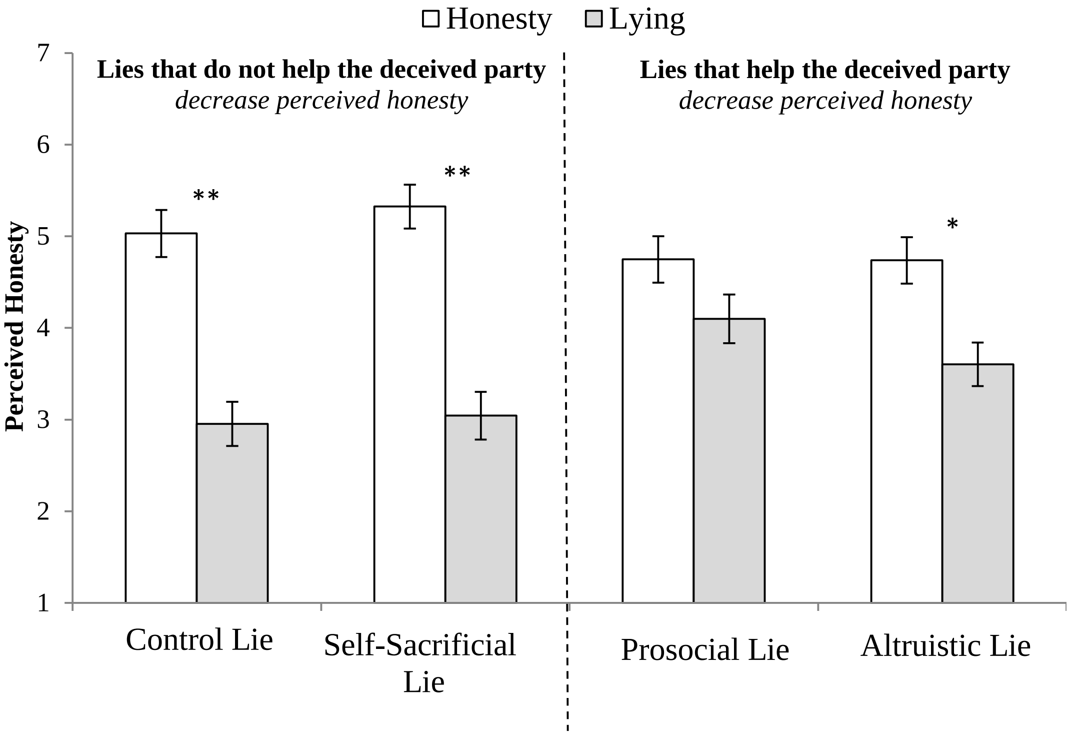


Figure 5: Perceived honesty in Study 3. Error bars represent ±1 SE. * *p* < .01, ** *p* < .001.

CHAPTER 2.

COMMUNITY STANDARDS OF DECEPTION

Emma E. Levine

ABSTRACT

When is lying ethical? Through a large inductive study, and a series of experiments (N = 1313), I develop and test a descriptive moral theory to address this fundamental question. I find that deception is perceived to be ethical when it prevents *unnecessary harm*. There are two key dimensions that influence perceptions of unnecessary harm: the degree to which deception prevents harm to an individual at the moment of communication, and the instrumental value of truth. I identify nine implicit rules – pertaining to the targets of deception and the topic and timing of a conversation – that specify the systematic circumstances in which deception is perceived to be ethical. I document the causal effect of each implicit rule on the endorsement of deception, and I demonstrate that judgments of unnecessary harm explain reactions to these implicit rules better than several other constructs (e.g., self-interest, perceptions of autonomy, moral duty) that have been assumed to motivate the use or avoidance of deception in past philosophical and psychological scholarship. This research provides insight into when and why people value honesty, and paves the way for future research on when and why people embrace deception.

COMMUNITY STANDARDS OF DECEPTION

Moral decency ensures for us the right to be deceived as surely as the right to truth: to extol the latter and deny the former is to misunderstand being human.

– David Nyberg, *The Varnished Truth* (1993)

A central justification for the moral prohibition of deception is the conviction that deception robs individuals of their autonomy and their right to truth (Bacon, 1872; Bok, 1978; Kant, 1959/1785). For example, Sissela Bok, the modern voice on the philosophy of deception, proclaimed that deception is only ethical when it upholds the principle of autonomy: the only lies that are ethical are the ones that can be "openly debated and *consented to* [emphasis added] in advance" (Bok, 1978, p. 181).

This justification for truth telling assumes that people universally value truth and would only consent to deception in rare circumstances. Individuals, however, frequently choose to avoid information and eschew truth (see Sweeny, Melnyk, Miller, & Shepperd, 2010 for a review). In fact, people are often complicit in others' attempts to deceive them. Individuals routinely avoid spoiling surprises and accept false compliments, even when they suspect deceit. Many individuals also avoid learning about negative news that they cannot control (e.g., Yaniv, Benador, & Sagi, 2004). Consider a patient who can learn whether or not he has an incurable disease. He may prefer not to know – or even to be deceived – about the disease precisely because he wishes to maintain his autonomy: the freedom to live as if he were not ill. In this case, the patient may believe that honesty would cause him unnecessary harm and that deception would be ethical.

Existing research on deception has failed to consider when and why people want to be deceived and how this affects the moral judgment and use of deception in interpersonal contexts. In the present investigation, I integrate philosophical and psychological scholarship to unearth community standards of deception, the implicit psychological principles that individuals use to justify deception. Rather than assuming that most people value honesty as a rule and that deception is a rare exception, I assume that people have numerous, *systematic* rules that govern judgments of and preferences for deception.

No prior research has documented these rules. Consequently, basic questions on deception remain unanswered. For example, when specifically do individuals endorse deception? What qualities of a target justify the use of deception? What qualities of true information justify deception? How do individuals' own preferences for information, honesty, and deception influence their moral judgments of deception?

Through a large inductive study, and a series of vignette experiments, I answer these questions. I demonstrate that lay people have a codified set of rules that guide their moral judgments of deception. A basic theory underlies these implicit rules: deception is perceived to be ethical and individuals consent to being deceived when honesty causes *unnecessary harm*. Perceptions of unnecessary harm are driven by two key factors: the degree to which deception will prevent immediate harm to an individual at the moment of communication, and the instrumental value of truth (i.e., the degree to which honest information may yield meaningful learning, growth, or behavioral change). Individuals are particularly likely to endorse deception when honesty causes immediate harm and when honesty has no instrumental value. These two factors are influenced by attributes of the target (i.e., the person being deceived), as well as the timing and topic of conversation. For example, the emotional fragility of the target, the target's capacity to understand truthful information, and the possibility that honest feedback can be implemented in the future all critically influence perceptions of unnecessary harm and consequently, the endorsement of deception.

This research makes important contributions to our understanding of deception, moral judgment, and human communication. In developing a descriptive moral theory of deception, I challenge prior assumptions about individuals' judgments of and preferences for deception. It is important to develop descriptive, rather than normative, moral theories because descriptive theories predict social judgment, moral reasoning, and everyday human behavior (e.g., Knobe & Nichols, 2008; Monin, Pizarro, & Beer, 2007, Haidt, 2001). Just as Kahneman, Knetsch, and Thaler's (1986a, 1986b) foundational work on community standards of fairness overturned the assumption that individuals universally value self-interest, and demonstrated that concerns about fairness place systematic, rather than anomalous, constraints on market behavior, the present research challenges the assumption that people universally value truth, and demonstrates that concerns about unnecessary harm place systematic constraints on honest communication.

Thus, this research highlights the circumstances in which truthful information will not be shared with others and the circumstances in which honesty will be penalized. Integrating community standards of fairness into the study of economic behavior shed light on predictable market failures (Kahneman, Knetsch, & Thaler, 1986a, 1986b). Similarly, integrating community standards of deception into the study of social communication sheds light on predictable communication frictions. This research offers novel insight into the rules that govern how people provide and respond to personal critiques, negative performance feedback, and terminal prognoses.

The Ethics of Deception

Normative Views

Consistent with extant research, I define deception as "*the transmission of information that intentionally misleads others"* (Levine & Schweitzer, 2015, p. 89). For centuries, philosophers and theologians have characterized deception as unethical. Perhaps the most famous condemner of deception is Immanuel Kant, who believed that deception was categorically unethical (Kant, 1959/1785). In Kant's view, deception is unethical because all individuals have a right to truth, and lying undermines that right. Similarly, Sir Francis Bacon (1872) argued that deception is unethical because it deprives people of "trust and belief." Both Kant and Bacon believed that deception is unethical, at least in part, because it destroys trust between individuals and trust in contracts, which ultimately causes societal harm. This deontological view of deception, however, predates Kant and Bacon. For example, Saint Augustine (circa 420 A.D.) argued that "every lie is a sin" (cited in Gneezy, 2005) and the Judeo-Christian Bible positions one's duty not to lie as one of the Ten Commandments (e.g., "Thou shalt not bear false witness," Exodus 20:16).

Some philosophers, however, have proposed alternative credentials for judging deception. In contrast to the deontological prohibition of deception, Utilitarians argue that deception is morally justified when its benefits outweigh its costs, (Bentham, 1843; Martin Luther, cited in Bok, 1978). Importantly, Utilitarians do not consider *who* bears those costs and benefits. For Utilitarians, a small lie that tremendously benefits the liar may be morally indistinguishable from a small lie that tremendously benefits the deceived party.

Despite the prominence of consequentialist thinking on many moral issues, modern rhetoric on deception largely follows the deontological tradition (e.g., Saarni & Lewis, 1993; Klosterman, 2014; Harris, 2013). Economists, for example, who have long positioned consequentialism as a normative standard, notably disparage deception. Experimental economics prohibit deception in laboratory experiments because they believe that deception undermines participants' trust in future experiments (Ariely & Norton, 2007; Jamison, Karlan, & Schechter, 2008; Levitt & List, 2007; Ortmann & Hertwig, 2002). Recently, several public figures and sources of moral guidance have also disparaged deception. For example, in the past two years, three ethics columnists for the *New York Times* wrote articles that warned readers of the dangers of deception (Appiah, Bloom, & Yoshino, 2015a, 2015b; Klosterman, 2014)[5], and best-selling author Sam Harris recently authored a popular philosophy book titled *Lying* in which he asserts that, "lying, even about the smallest matters, needlessly damages personal relationships and public trust" (2014, p. 2).

Often, individuals' discomfort with deception stems not only from the belief that deception undermines trust, but also from the belief that deception undermines the target's autonomy: the ability to make independent and rational decisions. Kant alludes to the importance of autonomy, suggesting that lying violates one's personal right to truth,

[5] Interestingly, while this manuscript was being written, the *New York Times* did feature a column suggesting that deception is often ethical (Dworkin, 2015). The column features many of the same rules and justifications introduced in the present manuscript.

but the modern philosopher Sissela Bok most clearly articulates the importance of autonomy. In her famous “Test of Publicity,” Sissela Bok asks, “which lies, if any, would survive the appeal for justification to reasonable persons” (Bok, 1978, p. 93). To pass this test, a lie must be acceptable to the deceived party. That is, the deceived party must consent in advance to the lie being told (Bok, 1978, p. 181). Although Bok does concede that some lies may pass this test (e.g., lies of trivial importance, or lies in extreme circumstances) she largely assumes that people rarely – if ever – would consent to being deceived. Thus, in Bok’s view, deception is rarely assumed to be ethical.

Descriptive Views

In the present research, I descriptively explore Bok’s Test of Publicity. I build and test a theory that explains when reasonable people justify deception and consent to being deceived. Although individuals are unlikely to consent to being told selfish lies (lies that help the liar and harm the deceived party) recent research suggests that individuals are far more likely to consent to prosocial lies (lies that help the deceived party).

A large body of research documents individuals’ negative reactions to and distaste for selfish lies. Selfish lies can trigger distrust, disliking, negative affect, and retaliation (Boles, Croson, & Murnighan, 2000; Croson, Boles, & Murnighan, 2003; Tyler, Feldman, & Reichert, 2006) and often prompt other forms of fraudulent behavior (Smith-Crowe, Tenbrunsel, Chan-Serafin, Umphress, & Joseph, 2015). Prosocial lies, however, are quite different from selfish lies in both their motivation and their consequences (see Wiltermuth, Newman, & Raj, 2015 for a review). The central motivation for prosocial lying is the desire to help or prevent harm to others. In routine conversations, individuals may tell prosocial lies to make others more confident or to avoid hurting others’ feelings

(DePaulo & Bell, 1996; DePaulo & Kashy, 1998). In economic interactions, individuals may tell prosocial lies to help generate more money for a specific counterpart or to restore equality (Erat & Gneezy, 2012; Gino & Pierce, 2009, 2010; Levine & Schweitzer, 2014, 2015; Wiltermuth, 2011).

Importantly, prosocial lies are often welcomed by targets and can yield interpersonal benefits. For example, Levine and Schweitzer (2014, 2015) found that individuals who told prosocial lies (i.e., lied about the outcome of a coin-flip to earn money for a partner) were perceived to be more ethical and were trusted more than individuals who told the truth, and harmed others.

These results suggest that individuals care more deeply about harm than following moral rules such as “never lie.” Indeed, scholars have suggested that perceptions of harm and care are the core of all moral judgments, and ethical rules only evolved to protect people from harm (Gray & Keeney, 2015; Gray, Schein, & Ward, 2014; Gray, Young, & Waytz, 2012). However, we do not yet know the rules that govern judgments of harm in everyday communication. The present research documents these rules and demonstrates that systematic perceptions of harm influence the justification of deception far more than past normative frameworks have assumed. In the present research, I compare the frequency with which individuals draw upon utilitarian and deontological reasoning to justify deception to the frequency with which individuals draw upon a harm-avoidance framework to justify deception. I also explore the importance of autonomy and consent to justifications of deception, demonstrating that individuals are willing to consent to deception when it prevents harm.

The Present Research

To establish a descriptive moral theory of deception, I begin with an inductive study. The motivation for using an inductive approach is three-fold. First, the goal of this research is to identify an overarching theory that describes moral judgments of deception. Inductive research is well-suited for exploratory theory generation because it does not impose any pre-existing assumptions onto participant responses (Gray, 2013). Second, I wanted to capture the language and context of a wide range of participant responses. This is useful for developing psychological insight and theory, as well as crafting realistic vignettes and experiments to use in the second stage of this research. Third, this approach corresponds with the methods suggested by Bok's Test of Publicity (1978). Bok asks individuals to consider the lies that reasonable people would consent to and justify. The present study empirically addresses Bok's famous thought experiment.

To provide convergent evidence of the implicit rules identified and the theory developed in the inductive study (Study 1), I experimentally manipulate the implicit rules in a series of vignettes (Studies 2 and 3). This empirical approach is informed by Kahneman, Knetsch, and Thaler's (1986a) approach to establishing community standards of fairness. In each vignette, I simply asked participants whether lying or honesty was the ethical decision.

Studies 2 and 3 achieve two main goals. First, they provide causal evidence of the relationship between implicit rules and moral judgments of deception. Whereas Study 1 simply unearths circumstances that are salient when considering the ethicality of deception, Studies 2 and 3 cleanly demonstrate that these circumstances causally influence moral judgments. Second, Studies 2 and 3 explore the underlying mechanisms. Across 12 vignettes, I demonstrate that perceptions of unnecessary harm underlie the effects of implicit rule violations on moral judgments of deception (Study 2) and I rule out a series of alternative mechanisms (Study 3).

Before beginning my investigation, it is important to clearly articulate the scope of the present research. First, this research focuses on understanding moral judgments of deception within a single conversation between a communicator (i.e., a potential liar) and a target (i.e., a potential deceived party). Individuals may view deception as unethical, broadly, because it destroys trust *over time*. The present research does not refute this possibility. In fact, there is interesting research to be done that addresses why people endorse deception in the context of a particular conversation but refuse to endorse it as a general practice. However, as a starting point, the present research examines when and why a lie is seen as ethical within the context of a single conversation.

Second, in line with Bok's Test of Publicity, I focus primarily on the perspective of the target. I unearth the lies that targets would consent to being told and then I explore whether these lies are also perceived to be moral by communicators and impartial third parties. Although communicators and third parties may be guided by additional implicit rules that targets do not see as justified, identifying those rules is beyond the scope of the present research.

Study 1

In Study 1, I used open-ended survey questions to ascertain the circumstances in which individuals would want to be deceived (i.e., consent to deception) and I examined how this converged with moral judgments of deception. I then used a three-stage coding process to establish a common set of implicit rules and develop a descriptive moral theory of deception.

Method

Participants. To ensure that my effects were robust to the characteristics of any particular population, I recruited two separate samples to complete Study 1. The first sample consisted of 117 adults recruited via Amazon Mechanical Turk (50% female; M_{age} = 37 years). The second sample consisted of 187 adults recruited from a U.S. university laboratory pool (59% female; M_{age} = 24 years).[6] I do not find systematic differences across these two samples. Thus, I report results collapsed across samples.

Procedure. All participants completed an online survey in which they answered free-response questions about deception. I randomly assigned participants to one of two conditions in a between-subjects design: *Preferences* or *Ethics*. Participants either answered three questions about their preferences for deception (the *Preferences* condition) or the general ethicality of deception (the *Ethics* condition).

In the *Preferences* condition, I first asked participants to, "T**hink about when you would want someone to lie to you.**" Then participants answered the following three questions, "In what circumstances would you want someone to lie to you?", "In what circumstances would you not want someone to be completely honest with you?", and "Please come up with three concrete examples of instances in which you would want to be lied to." In other words, they indicated the lies that they would consent to being told.

[6] Across all studies, stopping rules for data collection were decided in advance. For every study involving a laboratory sample, I collected data for the length of one laboratory session (3 days), and then stopped data collection. All laboratory participants were paid a $10 show-up fee in exchange for their participation in a 50-minute laboratory session. For MTurk samples, I targeted recruitment to be 100 participants/survey, 250 participants/survey and 150 participants/survey in Studies 1, 2, and 3 respectively. All MTurk participants were paid $.50-$.75/survey.

In the *Ethics* condition, I first asked participants to, “Think about when lying is right and when lying is wrong.” Then participants answered the following three questions, “In what circumstances is lying to someone the right thing to do?”, “In what circumstances is being completely honest with someone the wrong thing to do?”, and “Please come up with three concrete examples of instances in which it is ethical to lie.” In both conditions, participants had to respond to each question for at least one minute, and write at least 500 characters. Then, I collected demographic information for exploratory purposes.

Analytical approach. My goal in this study was to develop a codified set of rules and an underlying theory regarding lay perceptions of deception. Specifically, the goal was to identify the rules and underlying mechanisms that describe when people consent to being deceived and judge deception to be ethical. To do this, I adopted an iterative coding procedure (Strauss & Corbin, 1990). I first read through 50 participants’ responses and developed a preliminary coding scheme, informed by the present data, related research (DePaulo et al., 1996), and pilot data.

To code Study 1, I trained two research assistants to independently code all of the responses from both the *Preferences* and the *Ethics* perspectives using an initial coding scheme. The initial coding scheme required coders to read through each participant’s responses to all three questions and then code each participant’s responses according to the expressed justification for deception. The initial coding scheme included 12 possible justifications. I then met with both research assistants to collectively discuss the coding.

During this conversation, a single construct – (the prevention of) unnecessary harm – emerged as the overarching justification for deception. When discussing unnecessary harm, participants discussed the degree to which deception could prevent harm to the target at the moment of communication and the degree to which honesty could yield instrumental benefits to the target, such as enlightenment and growth. That is, participants generally endorsed deception when it prevented immediate harm to the target *and* when honesty had no potential to benefit the target in the future.

After converging on this overarching justification, we also discussed 20 participant responses in detail and used this discussion to identify new coding categories and to clarify the categorization scheme for the next round of coding. During this discussion, we also realized that participants' responses to the second survey question often repeated content from their response to the first question. Furthermore, some participants misinterpreted the second survey question. Consequently, the final coding procedure focused on analyzing only responses to the first and third questions in the survey ("In what circumstances is lying to someone the right thing to do?/In what circumstances would you want someone to lie to you?" and "Please come up with three concrete examples of instances in which it is ethical to lie./ Please come up with three concrete examples of instances in which you would want to be lied to.").

Whereas the initial coding scheme focused primarily on identifying different reasons that deception is perceived to be ethical (or preferred to the truth), the final coding scheme focused on first categorizing responses along the two proposed components of unnecessary harm, and then categorizing the features of the target, honest information, and context that participants used to explain the existence of unnecessary harm. Specifically, because each participant was asked to broadly identify the circumstances in which lying is ethical [they would like to be lied to], as well as specific examples of these instances, I was able to create a final coding scheme that categorized responses according to the features of specific examples, and the components of unnecessary harm. This approach allowed me to identify specific rules of deception – the contextual circumstances in which honesty would cause unnecessary harm and in which deception would be justified. In the final coding scheme, I also examined the frequency of utilitarian and deontological approaches to deception to explore whether the motivation to prevent harm was more salient than these two justifications that have pervaded rhetoric and scholarship on deception.

I then trained two *new* research assistants to use the final coding scheme (see Table 1). The new research assistants first coded 10 responses together and made revisions to the coding manual as needed. Then, they coded 10 responses individually, met to discuss questions and discrepancies, and then made another set of revisions to the coding manual. We discussed these 20 codes as a group and made one final set of revisions to the coding manual. After this meeting, one research assistant coded the remainder of the data set (304 responses in total). The second research assistant coded 50 randomly selected responses to establish the reliability of the final coding scheme. I report the reliabilities (Kappa) between the two coders in Table 1. These numbers reflect the agreement achieved across the 70 responses that both research assistants coded.

For all subsequent analyses, I use only codes from the single research assistant who coded all responses according to the final coding manual. I only used the second coder's codes to establish reliability.

Final coding scheme. Participants' open-ended responses were classified in four different ways. First, responses were coded according to the participant's framework for justifying deception: Deontology, Utilitarianism, and Harm Avoidance (see Table 1, Panel A for descriptions). All participants that used a Harm Avoidance framework were also coded according to the dimension of unnecessary harm: the immediate harm of honesty and instrumental value of honesty. Additionally, all participants that used a Harm Avoidance framework were coded according to the attributes of the target, the attributes of the honest information (i.e., the topic of conversation), and the context of the conversation that specified the presence of unnecessary harm. Each of these categories had a variety of sub-categories that were not mutually exclusive (see Table 1 for all subcategories, definitions, examples, and reliabilities). There were 9 categories in the final coding scheme. Each participant's responses were coded into as few or as many categories as were relevant. For example, some participants did not mention any attributes of the target that justified deception, whereas other participants stated that deception is ethical when someone is too young to understand the truth *and* when someone is too emotional to handle the truth (coded as "Target cannot understand the truth" and "Target is emotionally fragile," both sub-categories of "Attributes of target").

These nine categories specify nine implicit rules of deception. There were three criteria for maintaining a category in the final coding scheme. First, the category had to be represented in more than one participant's response. This cutoff is intentionally low. Because the inductive study captures the salience of different circumstances in which deception may be justified, rather than the strength of the relationship between any particular circumstance and the justification of deception, I wanted to include rules that may not be particularly salient but very closely map onto the proposed dimensions of unnecessary harm.

Second, the category had to reflect a Harm Avoidance framework. There were a few justifications that appeared with some regularity that I did not include in the final coding scheme because they did not pertain to the prevention of harm: for example, lying to create a surprise or to win a game of poker. It is possible that there are other common justifications for deception that do not pertain to the prevention of harm, but that is not the focus of the present investigation.

Third, the coders needed to come to consensus on the meaning of the category. Several categories were dropped from the final coding scheme because they were too vague and did not lead to strong agreement. For example, the initial coding scheme included a category that read, "When the target is looking for something other than truth." However, this category was too broad and could be more easily categorized into the conditions that would lead the target to avoid truth (e.g., when s/he is fragile). I provide an example response and how it was coded in Appendix A.

--Figure 1 about here--

Results

Below, I report the frequency with which participants rely on three moral frameworks when justifying deception: Deontology, Utilitarianism, and Harm Avoidance. I also review the components of unnecessary harm – the degree to which deception prevents immediate harm to the target, and the degree to which honesty has instrumental value for the target. Then, I review the attributes of the target, topic, and conversation that influence these perceptions, and consequently, specify the implicit rules of deception.

Deontology. A total of 5% of participants took a deontological approach to lying and reported that lying was never acceptable. These participants did not provide any justifications for the use of deception.

Utilitarianism. A total of 36.9% of participants justified deception that helped or prevented harm to parties other than the target (e.g., society, the liar, third parties). I conceptualize these justifications as utilitarian because they involved the calculation of costs and benefits, but were not focused solely on preventing harm to the target. In other words, these participants were not sensitive to who bore the burdens and benefits of deception.

Harm Avoidance (Preventing unnecessary harm). A total of 91% of participants justified deception that prevented unnecessary harm to the target. Participants focused on two types of harm: immediate harm at the moment of communication, and harm that yielded no instrumental benefits.

A total of 70.8% of participants justified deception that prevented immediate harm to the target. For example, individuals justified deception when honesty would immediately hurt a target's feelings or cause embarrassment.

A total of 69.4% of participants justified deception when honesty had no instrumental value to the target. For example, individuals justified deception when honesty would not have any meaningful impact on a target's future thinking or behavior.

---Table 1 about here---

Implicit rules

In addition to revealing the abstract principles that justify the use of deception, participants elucidated the specific circumstances in which those principles apply (i.e., the circumstances in which deception prevents unnecessary harm). These circumstances illustrate a number of implicit rules of deception, which pertain to the attributes of the target, the topic of honest information, and the context of the conversation. These rules can be summarized as:

It is acceptable to lie to targets when they are:

1. *Emotionally fragile*
2. *Unable to understand the truth*
3. *In their final days of life*

It is acceptable to lie about information that is:

4. *Subjective*
5. *Trivial*
6. *Uncontrollable*

It is acceptable to lie when:

7. *Honest information would disrupt a sacred event*
8. *Honest feedback can no longer be implemented*
9. *Honesty would embarrass the target in front of others*

I provide descriptions of these rules and the frequency with which these rules appeared in Table 1 (Panel B). These rules only pertain to honesty that has the potential to be hurtful to the target (e.g., critical feedback or bad news). Each of these rules describes circumstances in which honesty would be particularly harmful at the moment of communication (and thus deception would be particularly beneficial) and/or circumstances in which honesty would not have instrumental value. For example, participants endorsed lying to emotionally compromised targets (Rule 1) because they believed that honesty would cause the greatest immediate harm to fragile targets. Participants also believed that honesty would cause unnecessary harm when a conversation preceded – and had the potential to ruin – an event that was of special significance to the target, like the target's wedding (Rule 7) and when a conversation occurred in public (Rule 9). Honesty causes unnecessary harm in these circumstances because there are *temporary* features of the target or context that increase the intensity of harm and hinder the target's ability to cope. Thus, many participants expressed that communicators *should* lie during these moments, but perhaps reveal the truth at a later time.

The remainder of the rules document circumstances in which honesty is perceived to lack instrumental value. Specifically, participants endorsed lying to targets that could not understand the truth (Rule 2) and targets that were near death (Rule 3). In these circumstances, honesty would not be understood deeply enough to yield instrumental value (Rule 2) or would not alter future learning or behavior because the target's future was limited (Rule 3).

Similarly, the subjectivity (Rule 4) and the triviality (Rule 5) of the honest information influenced the extent to which honesty was perceived to yield meaningful instrumental benefits. Participants did not believe that others were morally obligated to voice their subjective and trivial opinions honestly, nor did participants want to be honestly told all of the subjective and trivial opinions of others. In Table 2, I summarize how each rule relates to the proposed dimensions of unnecessary harm.

---Table 2 about here---

It is important to note that these nine rules may not be an exhaustive list of implicit rules of deception. A key strength of the inductive approach is that it allows researchers to derive theory based on participants' own thoughts and identify overarching constructs, rather than imposing them. However, a weakness of this approach is that it primarily captures the most salient implicit rules. For that reason, I focus my analysis on how each rule describes the presence of unnecessary harm, rather than the percentage of participants that mention each rule (which I report in Table 1). The frequency with which each rule is mentioned may reflect the frequency with which each type of rule is considered in routine conversation, but it does not necessary reflect the predictive power of each rule violation.

Discussion

Deception is perceived to be ethical and individuals consent to being deceived when deception prevents unnecessary harm. Furthermore, individuals are far more likely to focus on avoiding unnecessary harm than they are to engage in purely deontological or utilitarian thinking when considering their preferences for or judgments of deception. Perceptions of unnecessary harm are driven by the degree to which deception will prevent immediate harm to the target and the potential for honesty to yield instrumental benefits.

I depict the relationship between these two factors and the endorsement of deception in Figure 1. When honesty is immediately painful and is not associated with instrument benefits (lower right quadrant), I expect most individuals to endorse deception. In these circumstances, honesty causes *unnecessary* harm. When honesty is immediately painful, but is associated with instrumental benefits (upper right quadrant), I expect individuals to equivocate. In these circumstances, honesty causes *necessary* harm. For this reason, individuals are likely to believe that the honest information should eventually be shared. However, they may advocate for temporary deception or the use of sensitive language to blunt the immediate harm. In other words, individuals are likely to advocate for discretion.

When honesty is not immediately painful and is associated with instrumental benefits (upper left quadrant), I expect most individuals to endorse honesty. When honesty is not immediately painful and is not associated with instrumental benefits (lower left quadrant), I do not expect people to have strong moral preferences. However, they may weakly prefer honesty, consistent with past research demonstrating that individuals prefer honesty when they lack a compelling reason to use deception (Levine, Kim & Hamel, 2010; Levine & Schweitzer, 2014). I empirically explore the validity of this two-dimensional framework in Study 2.

---Figure 1 about here---

This study revealed nine implicit rules that specify the conditions in which honesty causes unnecessary harm. Although these rules were derived inductively, in hindsight many of them could have been hypothesized a priori based on past research. In particular, research on information avoidance provides convergent evidence of many of these implicit rules. For example, past research demonstrates that individuals often avoid painful information about outcomes they cannot control, like incurable diseases (Yaniv et al., 2004; see also Shiloh, Ben-Sinai, & Keinan, 1999). The present research suggests that individuals may actually desire that others deceive them in these same circumstances. In their review of the information avoidance literature, Sweeny et al. (2010) outlined three central causes of information avoidance: 1) the extent to which an individual has control over the consequences of the information, 2) the extent to which an individual can cope with the information, and 3) the ease of interpreting the information. These causes of information avoidance also arise as justifications for deception in the present investigation. Individuals justified deception when they could not control the consequences of the honest information (Rules 6 and 8), when the target would be unable to cope with the honest information (Rule 1), and when the target would have difficulty interpreting the honest information (Rule 2).

Individuals' desire to be deceived about subjective information (Rule 4) also dovetails with research on individuals' desire to avoid uncertainty (Fox & Tversky, 1995; Fox & Weber, 2002; Lazarus & Folkman, 1984). Although participants in Study 1 rarely discussed uncertainty in their free responses, I suspect individuals would likely justify deception about uncertain information for the same reason they justify deception about subjective information: if painful information is not known to be absolute and objective, people believe it is unnecessary to know.

Participants also generated common rules that reflect the importance of personal dignity. For example, individuals are sensitive to the potential for public embarrassment (Rule 9), and they treat the end of life (Rule 3) and sacred events (Rule 7) with special care. Interestingly, many cultural, religious, and practical texts have discussed these circumstances when considering the moral importance of dignity relative to truth. For example, in the Babyloinian Talmud, a book of Jewish teachings compiled between 200 and 500 A.D., two rabbis discuss the ethics of falsely complimenting a bride on her wedding day. After much debate, the rabbis decide that you *should* tell a bride she is beautiful on her wedding day, regardless of the truth. This particular discussion highlights the importance of upholding dignity during sacred events, such as weddings (Telushkin, 1994). Furthermore, the importance of preserving dignity by avoiding public embarrassment is discussed throughout Eastern cultural texts (e.g., Ho, 1976) and the importance of preserving the dignity of those who lack cognitive capacity is discussed in the medical ethics literature (e.g., Beach & Kramer, 1999; Richard, Lajeunesse, & Lussier, 2010). In the present research, I provide a framework for understanding these seemingly disparate ideas.

Study 2

In Study 1, I used an inductive study to ascertain a set of nine implicit rules of deception. Although this approach provides insight into lay theories regarding the justification of deception, it does not allow me to make any causal claims. In Study 2, I use experiments to provide convergent evidence of the rules derived in Study 1 and to causally demonstrate how implicit rule violations influence preferences for and perceptions of deception. In Study 2, I manipulate the nine implicit rules derived in Study 1 across nine vignettes. I document how the violation of each implicit rule influences the two proposed dimensions of unnecessary harm (immediate harm and instrumental value of truth), and consequently, the endorsement of deception.

Participants

As in Study 1, I used multiple samples (participants recruited via Amazon Mechanical Turk and participants recruited by a U.S. university laboratory) to document the robustness of my effects. I conducted three separate surveys at different time points with different samples. Each survey examined a different set of three implicit rules. The choice of sample for each survey reflects the order in which the surveys were conducted and the availability of the sample. Because each survey has the same basic design and ultimately serves the same purpose (to document the causal effect of implicit rule violations on the endorsement of deception), I report the results of these surveys together as a single study.

I collected data from a total of 731 participants across the three surveys. I report the vignettes that appeared in each survey, the sample details, demographics, and any design differences between the three surveys in Table 3 (see note).

---Table 3 about here---

Procedure

Each survey featured three vignettes. Participants responded to multiple vignettes, but never saw more than one version of the same vignette. I randomized the order in which the vignettes were presented within each survey.

Each vignette examined one of the nine implicit rules identified in Study 1. In each vignette, I manipulated whether or not the relevant implicit rule was violated. *Implicit Rule Violation* was a between-subjects factor. Table 3 features the exact vignettes that tested each implicit rule. For example, in the *Presence of Others* vignette, I manipulated whether or not the target had the opportunity to receive negative feedback in public or private. Receiving negative feedback in public reflects the violation of an implicit rule (Rule 9).

The main dependent variable in each vignette was a dichotomous choice: participants chose whether truth-telling or lying was the preferred communication tactic in each vignette. I also manipulated *Perspective*; participants took the perspective of either an observer or the target when judging the preferred communication tactic. The purpose of the *Perspective* manipulation was to examine whether or not targets' preferences for deception converged with moral judgments, as in Study 1. In the first survey, I manipulated *Perspective* between-subjects. In this survey, participants read the vignette from either the perspective of an observer or the perspective of the target. In the remaining two surveys, I manipulated *Perspective* within-subjects and randomized the order of the two perspectives. In these surveys, all participants read the vignette from the perspective of the observer and were asked to imagine they were the target or an observer (order randomized). I manipulated *Perspective* as both a within-subjects and between-subjects factor to examine whether or not perceptions differed in separate versus joint evaluation (Hsee, 1996).

Dependent variables. In the *Observer Perspective*, participants answered the following question, "Which of the following options is the more ethical response?" In the *Target Perspective*, participants answered the question, "Of the following options, how would you prefer that [the communicator] responds?" To answer these questions, participants chose between telling the truth and lying. I include the exact wording of the response options for each vignette in Appendix B.

After participants selected the most ethical [their most preferred] response, participants answered a series of questions intended to examine the proposed mechanisms: immediate harm (e.g., "To what extent would telling the truth in this vignette cause pain to you [the individual]?") and the instrumental value of truth (e.g., "To what extent would telling the truth in this vignette be valuable for your [the individual's] improvement or well-being?"). All items were measured using seven-point rating scales anchored at 1 = "Not at all" and 7 = "Extremely." The items vary slightly in each set of vignettes, as I refined the scales between studies. The scales maintained high reliability in every vignette (all α's > .74). I report all scale items in Appendix C. After participants submitted their responses, I collected demographic information.[7]

Results

The purpose of the vignettes was to demonstrate that judgments of deception are governed by multiple rules, rather than to examine the differences between these rules. Thus, consistent with Kahneman, Knetch, and Thaler (1986a), I analyzed each vignette independently. I also conducted a meta-analysis across all vignettes to test the proposed theory.

Vignette-level results. I found no main effect or interaction effect of *Perspective* on the endorsement of deception in any vignette (all *p*s > .16). In other words, targets and observers did not differ in their endorsement of deception. Thus, I collapsed across *Perspective* for my main analyses.

For my main analyses, I used chi-squared tests to compare the proportion of participants who endorsed deception when an implicit rule was or was not violated. Table 3 includes all proportions and statistical tests. I find a significant main effect of each of

[7] Participants also answered a single-item recall question, which asked about the relevant implicit rule. However, each scenario varied significantly with respect to the percentage of participants that correctly answered the recall question. I report all recall questions and the percentage of participants who answered them correctly in the online supplemental materials. Excluding participants who did not answer the recall questions correctly does not change any main results.

In the first two surveys, participants also provided an open-ended written response, explaining what the communicator should say in each vignette. I do not examine participants' free responses in the present manuscript.

the nine implicit rules violations on the endorsement of deception (all *p*s < .01, see Table 3).

I also mapped each implicit rule on to the proposed theoretical framework (see Figure 2). Specifically, for each vignette, I plot the mean ratings of immediate harm and instrumental value of truth in the *Control condition* and the mean ratings of immediate harm and instrumental value of truth in the *Implicit Rule Violation condition*. The size of each data point is proportional to the percentage of people who endorsed deception in each condition. These graphs demonstrate that violating an implicit rule generally increases perceptions of immediate harm and lowers perceptions of the instrumental value of truth. The four quadrants of the proposed theoretical framework also closely align with my empirical data. Specifically, the majority of participants endorsed deception in vignettes that were judged to be in the high immediate harm-low instrumental value (lower right) quadrant of the theoretical framework. Participants rarely endorsed deception in vignettes that were judged to be in the low immediate harm-high instrumental value (upper left) quadrant. And, participants were torn when reacting to vignettes that were judged to be in the high immediate harm-high instrumental value (upper right) quadrant.

---Figure 2 about here---

Meta-analytic results. I also conducted a meta-analysis to test the proposed theory more precisely. Across the nine vignettes, I expected perceptions of immediate harm and the instrumental value of truth to mediate the effects of implicit rule violations on the endorsement of deception. To conduct this meta-analysis, I combined all the data from the three surveys into one dataset. I ran a series of logistic regressions on the endorsement of deception (1 = lying is endorsed, 0 = truth-telling is endorsed) including *Implicit Rule Violation*, *Perspective,* immediate harm, instrumental value, gender, and age as independent variables (see Table 4). In these regressions, I coded *Implicit Rule Violation* as 1 if the relevant rule was violated (e.g., if the target was not able to understand the information) and 0 otherwise. I coded *Perspective* as 1 in the target condition and 0 otherwise. In all analyses, I included fixed effects for each vignette, and I clustered standard errors at the Vignette and Participant levels.

The logistic regression results demonstrate that implicit rule violations powerfully influence the endorsement of deception (b = 1.70, p < .001, Models 2 and 3), whereas the perspective of the judge matters very little (b =-.10, *ns*, Model 3). The meta-analysis also reveals that perceptions of immediate harm (b =1.19, p < .001) and instrumental value (b =-.73, p < .001) influence the endorsement of deception (Model 5). Interestingly, there is also an interaction between immediate harm and instrumental value (b = .093, p = .03, Model 6), suggesting that these perceptions may have multiplicative, rather than additive, effects on the endorsement of deception.

---Table 4 about here---

In addition to the logistic regressions, I ran mediation analyses. I used the bootstrap procedure with 10,000 samples to test the processes by which implicit rule violations influence judgments of deception (SPSS Process Macro, Model 4, Hayes, 2013; Preacher, Rucker, & Hayes, 2007). The mediation model included *Implicit Rule Violation* as the independent variable, perceptions of immediate harm and instrumental value as simultaneous mediators, and the endorsement of deception as the dependent measure. I find significance evidence of mediation through both perceptions of immediate harm (Indirect effect = 0.82, SE = .06, 95% CI [.71, .95]) and perceptions of the instrumental value of truth (Indirect effect = .75, SE = .05, 95% CI [.65, .86]).

It is important to note, however, that I only have evidence of partial mediation. Models 5 and 6 (Table 4) demonstrate that implicit rule violations have a direct effect on the endorsement of deception, even after controlling for perceptions of immediate harm and instrumental value. In other words, there are likely other features of the vignettes that drive the endorsement of deception.

Discussion

Across nine vignettes, I provide convergent evidence of the implicit rules of deception. Each of the nine implicit rules identified in Study 1 had a significant causal effect on targets' desire for and observers' moral judgments of deception in Study 2. Perceptions of immediate harm and instrumental value underlie these effects. When an implicit rule is violated – for example, when a target does not have time to implement feedback – honesty is perceived to be more painful at the moment of communication, and honesty is perceived to yield less instrumental value. Thus, deception is perceived to be ethical. The meta-analysis and theoretical mappings of each vignette provide strong evidence of the centrality of these two mechanisms in predicting moral judgments of deception.

Study 3

In Study 3, I extend the present investigation in two ways. First, I introduce the perspective of the communicator (i.e., the potential liar). In Studies 1 and 2, I only examined individuals' judgments of deception from the perspective of the target and from the perspective of an impartial moral judge (i.e., observer). Although it is possible that individuals in the *Ethics* condition in Study 1 took the perspective of the liar when discussing the circumstances in which deception is ethical, it is difficult to know whether this influenced their judgments. Thus, in Study 3, I explicitly explore the perspective of liars, compared to targets, and observers.

Second, I rule out alternative mechanisms. Although I propose that perceptions of unnecessary harm are central to moral judgments of deception, philosophical debates have largely focused on three other factors: individuals' moral duty to tell the truth (e.g., Kant, 1959/1785), the societal harm caused by lying (Bacon, 1872; Harris, 2013), and the deleterious effect deception has on individual autonomy (e.g., Bok, 1978). Furthermore, social scientists have largely assumed that individuals only use deception when it is in their self-interest (e.g., when they will benefit from lying and are unlikely to get caught; Shalvi, Gino, Barkan, & Ayal, 2015; Bereby-Meyer & Shalvi, 2015). Thus, I explore whether any of these potential mechanisms underlie the relationship between implicit rule violations and judgments of deception. Specifically, I compare perceptions of unnecessary harm to the following five factors: perceptions of duty, societal harm, autonomy violations, self-interest, and the probability of deception detection.

Methods

Participants. As in Studies 1 and 2, I recruited two separate samples to complete this study. In Study 3, both samples completed the same survey. The first sample consisted of 136 adults recruited via Amazon Mechanical Turk (43% female; M_{age} = 32 years). The second sample consisted of 142 adults from a U.S. university laboratory sample (61% female; M_{age} = 23 years). I find no effects of sample on the endorsement of deception, thus, I report results collapsed across samples.

Procedure. The survey consisted of three vignettes. In each vignette, I manipulated a different implicit rule and I manipulated whether participants judged the ethicality of deception from the perspective of an observer, target, or liar. *Implicit Rule Violation* and *Perspective* were both between-subjects factors. As in Study 2, participants responded to multiple vignettes, but never saw more than one version of the same vignette. I randomized the order in which the vignettes were presented.

Vignettes. I created new vignettes for Study 3 that manipulated three of the nine implicit rules. Although I only explored a subset of the implicit rules in Study 3, I purposefully chose rules that pertain to three different contextual drivers of unnecessary harm. In the first vignette I manipulated an attribute of the target (i.e., his ability to understand the truth). In the second vignette, I manipulated the timing of a conversation (i.e., whether the target had time to implement change). In the third vignette, I manipulated the context of the conversation (i.e., whether others were present).

Ability to understand. In the first vignette, participants had to decide whether or not to inform a target that his daughter had died. I manipulated whether or not the target suffered from dementia. This vignette corresponds with the rule: Lie to targets who cannot understand the truth (Rule 2). This vignette mirrors the medical ethics concept of "therapeutic fibbing" (Beach & Kramer, 1999), suggesting it may be ethical to lie to Alzheimer's and dementia patients to protect them from undue anxiety, suffering, and confusion. This vignette also addresses the limitations of the "Ability to understand" vignette featured in Study 2. In Study 2, I compared an adult to a child, who may differ on qualities other than their ability to understand difficult information. In Study 3, I held age constant and simply manipulated cognitive capacity. The exact vignette appears below.

> *Imagine a caregiver at a nursing home. The caregiver is responsible for Jeff, a 93-year-old man.*
>
> Control condition: *Jeff is in good physical and mental health.*
>
> Violation condition: *Although Jeff is in good physical health, he suffers from severe dementia. This means that he often cannot make sense of his reality and is easily confused.*
>
> *The caregiver recently learned that Jeff's estranged daughter, who he has not heard from for over a decade, died two years ago.*
>
> *One day, out of the blue, Jeff asks his caregiver if she has heard anything about his family.*

Time to implement change. The second vignette depicted an individual who had made an error when writing a manuscript. I manipulated whether the mistake could be corrected. This vignette corresponds with the rule: Lie when honest feedback can no longer be implemented (Rule 8):

> *Imagine a graduate student, Jeff, who is planning to submit a paper for publication. Jeff has poured months into his research and is very proud of the resulting manuscript.*
>
> *Jeff's friend recently read Jeff's manuscript and noticed a few errors.*
>
> Control condition: *Jeff submitted the paper yesterday – meaning he is no longer able to implement changes.*
>
> Violation condition: *Jeff is submitting the final paper tomorrow – after he submits the manuscript he will no longer be able to implement changes.*
>
> *Jeff asks his friend what he thought of the manuscript.*

Presence of others. The third vignette depicted an individual who had delivered a presentation poorly. I manipulated whether the opportunity to give feedback occurred in public or private. This vignette corresponds with the rule: Lie when honesty would embarrass the target in front of others (Rule 9):

> *Imagine a summer intern named Jeff, who just delivered his end-of-internship presentation to his office.*
>
> *Jeff's PowerPoint slides were disorganized and he misspoke several times. Jeff's friend attended the presentation and believed that Jeff's presentation went very poorly. Jeff did not seem to realize that, and it is unclear what other audience members thought.*
>
> Control condition: *Immediately after the presentation, in a private space, Jeff asks his friend what he thought of the presentation.*
>
> Violation condition: *Immediately after the presentation, in front of several remaining audience members, Jeff asks his friend what he thought of the presentation.*

Dependent variables.

Endorsement of deception. After participants read each vignette, I asked participants, "In the course of this conversation, which of the following options is the more ethical response?" Participants chose between: "Tell [the individual] the truth" and "Lie to [the individual]." Unlike Study 2, the main dependent variable and the response options were identical across all perspectives. The response options were followed by short descriptions of the relevant truth or lie for each vignette. I include the exact wording of all response options in Appendix D.

Potential mechanisms. After participants chose to endorse either deception or honesty, participants answered a series of questions intended to examine the proposed mechanisms and rule out alternatives. All items were measured using seven-point rating scales anchored at 1 = “Not at all” and 7 = “Extremely.”

*Immediate harm and Instrumental value***.** Participants responded to four items about the immediate harm of honesty (α = .80): “To what extent would honesty cause pain to [the target] at the moment of communication?”, “To what extent would telling a lie protect [the target’s] feelings at the moment of communication?”, “To what extent would honesty cause harm to [the target] at the moment of communication?”, and “To what extent would lying benefit [the target] at the moment of communication?”

Participants also responded to four items about the instrumental value of truth (α = .83): “To what extent would telling the truth in this vignette have the potential to influence [the target’s] future behavior?”, “To what extent would telling the truth in this vignette be valuable for [the target’s] long-term well-being?”, “To what extent is the honest information necessary for [the target] to know?”, and “To what extent is the honest information useful for [the target’s] learning, growth or enlightenment?” I adapted these items from Study 2.

Moral duty. Participants also responded to a single item about moral duty: “To what extent does [the potential liar] have a moral duty to tell the truth?”

Societal harm. Participants responded to a single item about the degree to which lying could cause societal harm: “To what extent might telling this lie harm society as a whole?”

Autonomy. Participants responded to two items about the degree to which lying violated the target’s autonomy ($r = .46$): “To what extent does this lie infringe upon [the target’s] autonomy?” and “To what extent does telling this lie prevent [the target] from making informed decisions?”

Self-interest. Participants responded to two items about the degree to which lying benefited the liar ($r = .59$): “To what extent is lying the easiest course of action for [the potential liar]?” and “To what extent does lying spare [the potential liar] from conflict?”

Probability of detection. Finally, participants responded to two items about the degree to which lying could ever be discovered ($r = .26$): “To what extent is the honest information verifiable?” and “To what extent is it possible for [the target] to independently uncover the truth?”

After participants submitted their responses, I collected demographic information.[8]

Results

Analytical approach. As in Study 2, I analyzed each vignette independently and then conducted a meta-analysis across all scenarios to test the proposed theory and examine alternative mechanisms. For each vignette, I conducted a set of logistic regressions to examine the effects of *Implicit Rule Violation* and *Perspective* on the endorsement of deception (1 = lying is endorsed, 0 = truth-telling = endorsed). In these regressions, I coded *Implicit Rule Violation* as 1 if the relevant rule was violated (e.g., if the target was not able to understand the information) and 0 otherwise. I created two

[8] Participants also answered a single-item manipulation check, which asked about the relevant implicit rule. I report these items and the corresponding results in the online supplemental materials. I find a significant effect of *Implicit Rule Violation* on the manipulation check in every vignette (*p*s < .01).

dummy variables for the *Perspective* conditions. I created one variable called *Target* that had the value of 1 in the *Target Perspective* condition and 0 otherwise; and I created one variable called *Liar* that had the value of 1 in the *Liar Perspective* and 0 otherwise. The *Observer Perspective* served as the control.

Vignette-level results. The results of the vignette-level logistic regressions appear in Table 5. Figure 3 also depicts the proportion of participants who endorsed deception in each experimental condition in each vignette.

---Table 5 and Figure 3 about here---

In each vignette, I find a significant effect of implicit rule violation (*p*s < .05). In the *Ability to understand* and *Time to implement* vignettes, I find no main or interaction effects of *Perspective* (*p*s > .44). In other words, participants responded to implicit rule violations similarly if they considered the situation from the perspective of a liar, target, or an observer.

In the *Presence of others* vignette, I found significant, but unpredicted, perspective effects (see Table 5, Column 3). Liars believed that lying was more ethical than observers did (b = 1.40, p = .04). There were also significant *Liar* × *Implicit Rule Violation* (b = -1.57, p = .05) and *Target* × *Implicit Rule Violation* (b = -1.65, p = .05) interactions; the implicit rule violation had a stronger effect on liars and targets than observers. These results suggest that observers may fail to fully appreciate the value of deception in public contexts.

Meta-analytic results. I conducted a meta-analysis to test the proposed theory and rule out alternative mechanism. Using the data from all three vignettes, I ran a series of logistic regressions on the endorsement of deception (1 = lying is endorsed, 0 = truth-telling is endorsed) including *Implicit Rule Violation, Perspective,* gender, age, and the seven mechanism measures as independent variables (see Table 6). In these analyses, I included fixed effects for each vignette, and I clustered standard errors at the Vignette and Participant levels.

The logistic regression results demonstrate that implicit rule violations powerfully influence the endorsement of deception (all *b*s > .73, *p*s < .001, Models 2-7), whereas perspective matters much less. As hypothesized, perceptions of immediate harm and instrumental value also significantly influenced the endorsement of deception (all *bs* > .51, *p*s < .05, Models 5-7). As in Study 2, I also found an interaction between immediate harm and instrumental value (*b* = .09, *p* = .06 in Model 6 and *b* = .08, *p* < .001 in Model 7), providing further evidence that immediate harm and instrumental value have multiplicative effects on the endorsement of deception. Of the alternative mechanisms I examined, only perceptions of moral duty significantly impacted the endorsement of deception (*b* = -.66, *p* < .001, Model 7).

---Table 6 about here---

In addition to the logistic regressions, I ran mediation analyses. I used the bootstrap procedure with 10,000 samples to test the processes by which implicit rule violations influence judgments of deception (SPSS Process Macro, Model 4, Hayes, 2013; Preacher, Rucker, & Hayes, 2007). The mediation model included *Implicit Rule Violation* as the independent variable, the seven potential mechanisms as simultaneous mediators, and the endorsement of deception as the dependent measure. I find evidence of mediation through both proposed mechanisms: immediate harm and instrumental value.

I also find significant mediation through perceptions of moral duty. However, the direction of the effect does not echo philosophical assumptions about one's duty to tell the truth (e.g., Kant, 1959/1785). Lay people do not believe that they have a categorical imperative to tell the truth. Rather, they believe that when an implicit rule is violated, they have *less* duty to tell the truth, leading them to endorse deception. Figure 4 depicts the full mediation model and all indirect effects.

As in Study 2, I only have evidence of partial mediation, suggesting there are other features of the vignettes that drive the endorsement of deception.

---Figure 4 about here---

Discussion

Study 3 documents two key results. First, implicit rule violations have largely the same effects on targets', liars', and observers' moral judgments of deception. Second, perceptions of the immediate harm of honesty and the instrumental value of truth, the two hypothesized dimensions of unnecessary harm, underlie the effects of implicit rule violations on the endorsement of deception; perceptions of autonomy, societal harm, self-interest, and the probability of detection do not.

It is important to note, however, that these alternative mechanisms are influenced by implicit rule violations and do influence the endorsement of deception (see Appendices E and F). For example, in the *Ability to understand* vignette, lying to the target was seen as a greater autonomy violation when the target was of healthy mind than when he suffered from dementia. Furthermore, in the *Time to implement change* vignette, lying to the target was seen as more beneficial for the liar when the target could not implement feedback than when he could. Thus, we cannot conclude that autonomy and self-interest do not matter for making judgments of deception. However, we can conclude that autonomy and self-interest (as well as the probability of deception and perceptions of societal harm) do not independently influence the endorsement of deception, once we control for perceptions of harm to the target. Alternatively, perceptions of harm to the target do independently influence the endorsement of deception, above and beyond the effects of all other mechanisms I investigated.

Perceptions of moral duty also independently influence the endorsement of deception. This result reveals novel insights about lay conceptions of duty. Although moral duties are typically conceptualized as immovable obligations (Kant, 1959/1785), lay people seem to conceptualize the moral duty to tell the truth as context-dependent. When the truth causes unnecessary harm, lay people believe that they are freed of their duty to tell the truth.

General Discussion

Across one inductive study and 12 vignettes (*N* = 1313), I unearth community standards of deception, the implicit moral rules individuals use to justify deception. Motivated by Bok's Test of Publicity, I inductively derived these rules in Study 1 by asking participants which lies they would consent to being told and which lies they find to be ethical. I then provide causal evidence for each rule in Studies 2 and 3. Consistent with research on the centrality of harm and care in moral judgment (Gray, Schein, & Ward, 2014; Gray, Young, & Waytz, 2012; Haidt & Graham, 2007), I find that individuals' implicit rules are motivated by an overarching desire to avoid causing unnecessary harm. Each implicit rule describes a circumstance in which honesty would cause unnecessary harm (e.g., when a target is fragile or unable to implement feedback) and thus, in which deception is ethical.

Most of the rules I identify have been explicitly discussed in past philosophical, theological, or psychological scholarship. For example, the medical ethics literature has long discussed the ethics of lying to cognitively impaired patients (e.g., Sokol, 2007) and the information avoidance literature has discussed individuals' desire to avoid information that they cannot control or emotionally handle (Sweeny et al., 2010). Until now, however, these ideas have been siloed in disparate literatures, and we have lacked a parsimonious framework for understanding why individuals endorse deception in various circumstances. Furthermore, rather than carefully consider these systematic circumstances in which deception is seen as ethical, most modern rhetoric focuses on a simpler message: deception is wrong. In the present research, I demonstrate that people often view deception as right, they do so in systematic ways, and their logic is driven by the simple desire to avoid causing unnecessary harm.

The framework I present also clarifies the two ways in which honesty is perceived to cause unnecessary harm. First, there may be features of a particular context that *temporarily* increase the emotional, psychological, or material pain associated with honesty. These features, – such as a target's emotional fragility, the presence of others, or the timing of an important or sacred event – increase the harm associated with honesty at the moment of communication. Second, there may be features of a particular context that limit the instrumental benefits of honesty. Although honesty is often discussed as a moral good in and of itself, the present research suggests that lay people value honesty because of its instrumental benefits, such as enlightenment and growth, rather than its intrinsic value. In circumstances in which honest information does not lead to instrumental benefits – for example, when truthful information is not meaningful, cannot be understood, or cannot be implemented – individuals openly consent to and justify deception.

Importantly, perceptions of unnecessary harm justify the use of both major and minor lies. Minor, or white, lies are often perceived to be trivial and of little moral import (Bok 1978; Brown & Levinson, 1987). The present research demonstrates that white lies are often perceived to be ethical because they can spare the target from emotional harm, and because they do not hinder the target's understanding or growth in a meaningful way. However, the present research also demonstrates that there are a variety of circumstances in which major lies are perceived to be ethical. Lying about significant events, such as infidelity or death, is often perceived to be moral. These judgments, however, hinge on the degree to which the information is useful, rather than meaningful. Meaningfulness and usefulness are two orthogonal qualities that independently influence the perceived instrumental value of truth.

This research also demonstrates that lay theories do not necessarily align with common normative positions on deception. Very few individuals hold a deontological view of deception, believing that deception is categorically wrong. Furthermore, although the proposed framework is notably consequentialist – individuals implicitly weigh the short-term harm against the long-term benefits of truth – it is important to recognize that most individuals focus narrowly on the consequences of lying for the target, rather than the consequences for the liar or society writ large. In other words, lay beliefs reflect the desire to avoid harm to a particular victim rather than a general utilitarian calculus.

Broadly, this research illuminates the moral value of discretion in human communication. Lay people believe that individuals should lie in many situations (e.g., in front of others, or during times of strife), but reveal the truth later. Similarly, participants said they wanted to be protected and deceived during particular moments, but that they would want to uncover the truth at a later point in time. This reflects a pragmatic view of honesty; people believe that the use of both honesty and deception should be constrained by the particular needs of the particular people involved in a particular conversation. In other words, lay people seem to conceptualize honesty and deception as tactics that can and should be used to regulate other virtues and vices, such as enlightenment and harm, rather than conceptualizing honesty and deception as categorical virtues and vices themselves.

Limitations and Future Directions

The present research has several limitations that can be addressed by future research. First, although the inductive study provides a solid foundation for an initial set of implicit rules, future research may be needed to establish a complete set. It may be useful, for example, to more deeply explore the responses of individuals who provided utilitarian justifications for deception; there may be circumstances in which the benefits conferred to liars are great enough that even targets would consent to deception. Asking liars directly when they think lying is ethical may also be a fruitful endeavor. For example, lies that protect the liar’s privacy may be broadly justified, despite not being a salient to targets. Liars may also be more likely to justify lies that surprise or flatter targets. Although the present theory focuses on deception that is motivated by the desire to *prevent* harm, lies that cause pleasure may also be justified broadly.

More research is also needed to understand how moral judgments of deception change across time and perspectives. The two proposed dimensions of unnecessary harm highlight a potential intertemporal tradeoff between immediate harm and long-term instrumental benefits of honesty. At the moment of communication, individuals may overweight the immediate harm caused by honesty. However, when thinking about a potential conversation from a distance, individuals may be more attuned to the potential long-term benefits of honesty. Thus, individuals may intend to be honest (and expect to appreciate honesty) when they consider having an unpleasant conversation, but when the moment to inflict (or experience) pain actually comes, they may prefer deception. Communicators may be particularly likely to overweight the harm of negative information, relative to targets, because they are motivated to avoid inflicting harm. Although I do not find consistent evidence for perspective effects in the present investigation, more research is needed to fully understand when and why communicators' and targets' perceptions of deception differ.

Individuals may also react differently to deception before and after it is used. The present research examines a priori judgments of deception – the circumstances in which individuals *expect* to judge deception as ethical. However, these judgments may diverge from in vivo judgments (individuals' beliefs about deception during a particular conversation) and post hoc judgments (individuals' reactions to telling a lie or being deceived). It is possible, for example, that when individuals are emotionally fragile they do not want to hear the truth, and they appreciate deception in the moment, but upon learning of the deception become furious. Interestingly, the lies that people consent to in advance may not be the same lies that people forgive others for telling.

The proposed implicit rules also highlight a number of circumstances in which deception may be used paternalistically. Although communicators and targets agree that deception is more ethical when the target lacks cognitive capacity, is emotionally fragile, or cannot implement feedback, communicators and targets may not necessarily make identical assessments about the presence or absence of these circumstances. For example, a communicator may be motivated to believe that a target is less competent or more fragile than he or she really is. As a result, a communicator may behave paternalistically and assume that a target cannot handle the truth, rather than soliciting information from the target that would help the communicator make a more informed judgment. Indeed, recent research suggests that individuals resent paternalistic lies. Although individuals embrace deception when there is unambiguous agreement about whether lying benefits the target, individuals resent lies that are motivated by a communicator's assumptions about what benefits the target (Lupoli, Levine, & Greenberg, 2016).

It may be safer to make assumptions about a target's desire for deception in some circumstances than others. In the present research, many individuals discuss extreme events that produce momentary states of fragility or cognitive depletion and call for the use of deception. These events, such as the death of a loved one or a state of drunkenness, could happen to anyone and are likely to produce similar emotional and cognitive consequences across individuals. Thus, these circumstances do not require communicators to make nuanced assumptions about a target's need for deception. In many circumstances, however, communicators make dispositional assessments about a target's fragility, cognition, or need. For example, before delivering feedback, a manager may simply ask himself if a particular employee can emotionally handle negative news. I suspect that these dispositional assessments are far riskier, and more likely to motivate unwelcome deception, than the systematic circumstances described in the present research.

The potential for paternalistic deception may also differ across relationships. In close or hierarchical relationships, individuals may feel a particularly strong need to protect the target. Thus, individuals may be most drawn towards deception in these relationships. Indeed, past work has demonstrated that individuals may have very different standards for morality across different types of relationships (Rai & Fiske, 2011, 2012). Although I find that individuals generally agree on the dimensions that justify the use of deception, this does not mean that communicators will weigh these dimensions rationally, consistently, or selflessly when they actually engage in difficult conversations with different relational partners.

It will also be important to more carefully investigate how different forms of deception are perceived. Across my vignettes, I compare lying and truth-telling in response to direct questions. Although I explicitly force participants to endorse either lying or truth-telling, I do not always specify the language that a communicator will use to lie, or whether the lie is by omission or commission. These nuances likely influence how individuals react to deception. For example, telling an ill-dressed target, "You look fine," may be seen as far more innocuous than saying "You look fantastic." Furthermore, changing the subject or using honest statements to convey a false impression (i.e., paltering, Rogers, Zeckhauser, Gino, Schweitzer, & Norton, 2014) may be seen as more permissible than a blatant lie. Future research should examine how these strategies are perceived, particularly from the perspective of both targets and communicators.

Finally, future research should explore the behavioral consequences of the implicit rules of deception. If honesty is considered immoral when it violates implicit rules, receiving honest information may elicit moral outrage, anger, and contempt in these circumstances. For example, a dying patient who wants his doctor to communicate optimism, despite knowledge of near-certain death, may deeply resent his doctor for crushing his hope. Or, an employee may lose trust in a manager who gives him negative feedback in front of others, instead of using discretion and providing his truthful opinion in private. Exploring the emotional and relational consequences of violating the implicit rules of deception is an important next step for future research.

Conclusion

Deception is typically characterized as unethical, and existing research assumes that individuals would rarely consent to being deceived. In contrast to these views, the present research demonstrates that individuals frequently consent to and morally justify deception, and they do so in systematic ways. Individuals seem to believe, consistent with David Nyberg’s sentiment in the opening quote, that moral decency often demands deception.

References

Appiah, K. A., Bloom, A., & Yoshino, K. (May 13, 2015). Can I change my name to avoid discrimination? *The New York Times Magazine*. Retrieved May 16 2015 from http://www.nytimes.com/2015/05/17/magazine/can-i-change-my-name-to-avoid-discrimination.html?_r=0

Appiah, K. A., Bloom, A., & Yoshino, K. (May 23, 2015). May I lie to my husband to get him to see a doctor? *The New York Times*.

Ariely, D., & Norton, M. I. (2007). Psychology and experimental economics A gap in abstraction. *Current Directions in Psychological Science*, *16*(6), 336-339.

Bacon, F. (1872). *The letters and the life of Francis Bacon* (Vol. 6). City, State: Longmans, Green and Company.

Beach, D. L., & Kramer, B. J. (1999). Communicating with the Alzheimer's resident: Perceptions of care providers in a residential facility. *Journal of Gerontological Social Work*, *32*(3), 5-26.

Bentham, J. (1843/1948). *An introduction to the principles of morals and legislation*. Oxford, UK: Basil Blackwell. (Original work published 1843)

Bereby-Meyer, Y., & Shalvi, S. (2015). Deliberate honesty. *Current Opinion in Psychology*, *6*, 195-198.

Bok, S. (1978). *Lying: Moral choices in public and private life*. New York, NY: Pantheon.

Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *83*(2), 235-259.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). City: Cambridge University Press.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143-159.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology*, *71*(4), 703-716.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63-79.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979-995.

Dworkin, G. (December 14, 2015). Are these 10 lies justified? *New York Times*. Retrieved December 14 2015 from http://opinionator.blogs.nytimes.com/2015/12/14/can-you-justify-these-lies/

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723–733.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, 585-603.

Fox, C. R., & Weber, M. (2002). Ambiguity aversion, comparative ignorance, and decision context. *Organizational Behavior and Human Decision Processes, 88*(1), 476-498.

Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science*, *20*(9), 1153-1160.

Gino, F., & Pierce, L. (2010). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science*, *21*(6), 1176-1194.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 384-394.

Gray, D. E. (2013). Doing research in the real world. City, State: *Sage.*

Gray, K., & Keeney, J. (2015). Impure, or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychology and Personality Science*, 6(8), 859-868.

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.* http://dx.doi.org/10.1037/a0036149.

Gray, K., Young, L., & Waytz, A. (2012) Mind perception is the essence of morality. *Psychological Inquiry*, 23(2), 101–124.

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98-116.

Harris, S. (2013). *Lying*. City, State: Four Elephants Press.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Ho, D. Y. F. (1976). On the concept of face. *American Journal of Sociology*, 867-884.

Howell, J. L., & Shepperd, J. A. (2012). Reducing information avoidance through affirmation. *Psychological Science*, *23*(2), 141-145.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247-257.

Jamison, J., Karlan, D., & Schechter, L. (2008). To deceive or not to deceive: The effect of deception on behavior in future laboratory experiments. *Journal of Economic Behavior & Organization*, *68*(3), 477-488.

Kahneman, D., Knetsch, J. L., & Thaler, R. (1986a). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 728-741.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1986b). Fairness and the assumptions of economics. *Journal of Business*, S285-S300.

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

Klosterman, C. (February 1, 2014). Get out of my subconscious! *The New York Times Magazine*. Retrieved December 1 2015 from http://www.nytimes.com/2014/02/02/magazine/get-out-of-my-subconscious.html

Knobe, J., & Nichols, S. (2008). An experimental philosophy manifesto. *Experimental Philosophy*, 3-14.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive–developmental approach. In T. Lickona (Ed.), *Moral development and behavior* (pp. 31-53). New York, NY: Reehard & Winston.

Lazarus RS, Folkman S. Stress, Appraisal, and Coping. New York: Springer; 1984.

Levine, T. R., Kim, R. K., & Hamel, L. M. (2010). People lie for a reason: Three experiments documenting the principle of veracity. *Communication Research Reports*, *27*(4), 271-285.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107-117.

Levine, E., & Schweitzer, M. 2015. Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, 126, 88-106.

Levitt, S. D., & List, J. A. (2007). What do laboratory experiments measuring social preferences reveal about the real world? *The Journal of Economic Perspectives*, 153-174.

Lupoli, M., Levine, E.E. & Greenberg, A. (2015). Paternalistic Lies. *Working Paper.*

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, *11*(2), 99-111.

Nyberg, D. (1993). *The varnished truth*. Chicago, IL: University of Chicago Press.

Ortmann, A., & Hertwig, R. (2002). The costs of deception: Evidence from psychology. *Experimental Economics,* 5(2), 111—131.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185-227.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118*, 57-75.

Rai, T. S., & Fiske, A. P. (2012). Beyond harm, intention, and dyads: Relationship regulation, virtuous violence, and metarelational morality. *Psychological Inquiry, 23*, 189-193.

Richard, C., Lajeunesse, Y., & Lussier, M. T. (2010). Therapeutic privilege: Between the ethics of lying and the practice of truth. *Journal of Medical Ethics*, *36*(6), 353-357.

Rogers, Todd and Zeckhauser, Richard J. and Gino, Francesca and Schweitzer, Maurice E. and Norton, Michael I., Artful Paltering: The Risks and Rewards of Using Truthful Statements to Mislead Others (September 18, 2014). HKS Working Paper No. RWP14-045. Retrieved from: http://ssrn.com/abstract=2528625

Saarni, C., & Lewis, M. (1993). *Lying and deception in everyday life*. New York, NY: Guilford Press.

Shalvi, S., Gino, F., Barkan, R., & Ayal, S. (2015). Self-serving justifications doing wrong and feeling moral. *Current Directions in Psychological Science*,*24*(2), 125-130.

Shiloh, S., Ben-Sinai, R., & Keinan, G. (1999). Effects of controllability, predictability, and information-seeking style on interest in predictive genetic testing, *Personality and Social Psychology Bulletin*, 25, 1187–1195.

Sokol, D. K. (2007). Can deceiving patients be morally acceptable?. *BMJ: British Medical Journal*, *334*(7601), 984.

Smith-Crowe, K., Tenbrunsel, A. E., Chan-Serafin, S., Brief, A. P., Umphress, E. E., & Joseph, J. (2015). The ethics “fix”: When formal systems make a difference. *Journal of Business Ethics*, *131*(4), 791-801.

Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications, Inc.

Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*, *14*(4), 340-353.

Telushkin, J. (1994). *Jewish wisdom: Ethical, spiritual, and historical lessons from the great works and thinkers*. City, State: W. Morrow.

Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, *17*(2), 223-236.

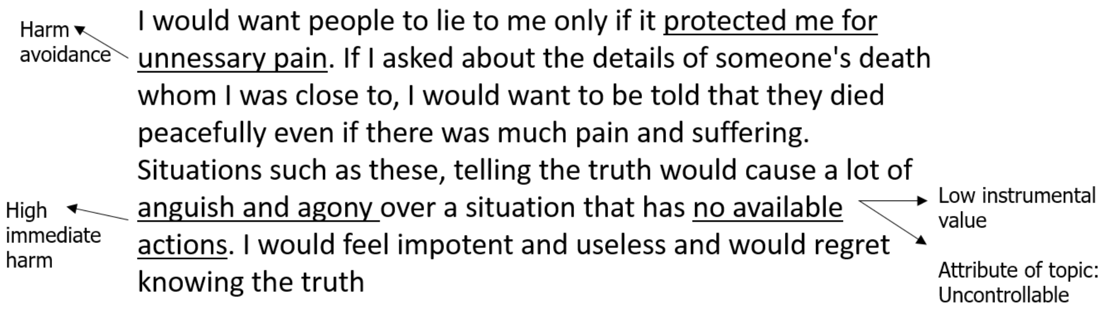
Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, *42*(1), 69-77.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, *115*(2), 157-168.

Wiltermuth, S. S., Newman, D. T., & Raj, M. (2015). The consequences of dishonesty. *Current Opinion in Psychology*, *6*, 20-24.

Yaniv, I., Benador, D., & Sagi, M. (2004). On not wanting to know and not wanting to inform others: Choices regarding predictive genetic testing. *Risk Decision and Policy*, *9*(4), 317-336.

Appendix A. Example Response and Coding



I would want people to lie to me only if it protected me for unnessary pain. If I asked about the details of someone's death whom I was close to, I would want to be told that they died peacefully even if there was much pain and suffering. Situations such as these, telling the truth would cause a lot of anguish and agony over a situation that has no available actions. I would feel impotent and useless and would regret knowing the truth

Note. This is an excerpt from one participant's response to "In what circumstances would you want someone to lie to you?" Key words are underlined and accompanying codes follow arrows.

Appendix B. Choice Options for Each Vignette in Study 2

Emotional fragility vignette

Ethics condition:

- Which of the following options is the more ethical response?
  - Tell the employee the truth – the manager should say that the report was not well done
  - Lie to the employee – the manager should say the report was fine

Preferences condition:

- Of the following options, how would you prefer that your manager respond?
  - Tell you the truth – the manager should say that the report was not particularly well done
  - Lie to you – the manager should say the report was fine

Deathbed vignette

Ethics condition:

- Which of the following options is the more ethical response?
  - Tell the individual the truth – the spouse should admit that s/he cheated on the individual
  - Lie to the individual – the spouse should not mention that s/he cheated on the individual

Preferences condition:

- Of the following options, how would you prefer that your spouse respond?
  - Tell you the truth – the spouse should admit that s/he cheated on you
  - Lie to you – the spouse should not mention that s/he cheated on you

Ability to understand vignette

Ethics condition:

- Which of the following options is the more ethical response?
  - Tell the patient the truth – the doctor should tell the patient s/he is likely to die
  - Lie to the patient – the doctor should not tell the patient that s/he is likely to die

Preferences condition:

- Of the following options, how would you prefer that your doctor respond?

    - Tell you the truth – the doctor should tell you that you are likely to die
    - Lie to you – the doctor should not tell you that you are likely to die

Subjective vignette

Ethics condition:

- Which of the following options is the more ethical response?
    - Tell the employee the truth – the colleague should tell the employee that she thinks the employee looks bad in the scarf
    - Lie to the employee – the colleague should tell the employee that she thinks the employee looks fine (or good) in the scarf

Preferences condition:

- Of the following options, how would you prefer that your colleague respond?
    - Tell you the truth – the colleague should tell you that she thinks you look bad in the scarf
    - Lie to you – the colleague should tell you that she thinks you look fine in the scarf

Trivial vignette

Ethics condition:

- Which of the following options is the more ethical response?
    - Tell the host the truth – the guest should tell the host that the soup is too salty
    - Lie to the host – the guest should tell the host that the soup is good or fine

Preferences condition:

- Of the following options, how would you prefer that your guest respond?
    - Tell you the truth – the guest should tell you that the soup is too salty
    - Lie to you – the guest should tell you that the soup is good or fine

Uncontrollable vignette

Ethics condition:

- Which of the following options is the more ethical response?
    - Tell the intern the truth – the friend should tell the intern that his stutter decreased the quality of his presentation
    - Lie to the intern – the friend should tell the intern that the presentation was fine (or good)

Preferences condition:

- Of the following options, how would you prefer that your friend respond?
  - Tell you the truth – your friend should tell you that your stutter decreased the quality of your presentation
  - Lie to you – your friend should tell you that the presentation was fine (or good)

Disruption to special moments and event vignette

Ethics condition:

- Which of the following options is the more ethical response?
  - Tell the employee the truth – the manager should tell the employee that s/he is getting laid off
  - Lie to the employee – the manager should not tell the employee that s/he is getting laid off

Preferences condition:

- Of the following options, how would you prefer that your manager respond?
  - Tell you the truth – the manager should tell you that you are getting laid off
  - Lie to you – the manager should not tell you that you are getting laid off

Time to implement vignette

Ethics condition:

- Which of the following options is the more ethical response?
  - Tell the employee the truth – the colleague should tell the employee that he thinks the suit is inappropriate
  - Lie to the employee – the colleague should tell the employee that he thinks the suit is fine

Preferences condition:

- Of the following options, how would you prefer that your colleague respond?
  - Tell you the truth – the colleague should tell you that he thinks the suit is inappropriate
  - Lie to you – the colleague should tell you that he thinks the suit is fine

The presence of others vignette

Ethics condition:

- Which of the following options is the more ethical response?

    - Tell the employee the truth – the manager should tell the employee that the report was not particularly well done
    - Lie to the employee – the manager should tell the employee that the report was fine

Preferences condition:

- Of the following options, how would you prefer that your manager respond?
    - Tell you the truth – your manager should tell you that your report was not particularly well done
    - Lie to you – your manager should tell you that the report was fine

Appendix C. Mechanism Questions in Study 2

Immediate harm

- To what extent would telling the truth in this scenario cause unnecessary harm? (Survey Group 1-2)
- To what extent would honesty cause pain to the [individual]? (Survey Group 1-3)
- To what extent would telling a lie protect the [individual]'s feelings? (Survey Group 1-3)
- To what extent would honesty cause harm to the [individual]? (Survey Group 3 only)
- To what extent would lying benefit the [individual]? (Survey Group 3 only)

Instrumental value

- To what extent would telling the truth in this scenario have the potential to influence the [individual]'s behavior? (Survey Group 1-3)
- To what extent would telling the truth in this scenario be valuable to the [individual]'s improvement* overall well-being? (Survey Group 1-3)
- To what extent is the honest information necessary for the [individual] to know? (Survey Group 3 only)
- To what extent is the honest information useful for the [individual]'s growth or enlightenment? (Survey Group 3 only)

Note. *The word improvement was removed from the *Death bed* vignette and from Survey Group 3 to eliminate confusion.

Appendix D. Response Options for Vignettes in Study 3

Ability to understand vignette

In the course of this conversation, which of the following options is the more ethical response?

- Tell Jeff the truth - Jeff's caregiver should tell him about his dead daughter
- Lie to Jeff - Jeff's caregiver should not tell him about his dead daughter

Time to implement vignette

In the course of this conversation, which of the following options is the more ethical response?

- Tell Jeff the truth - Jeff's friend should tell Jeff about the errors
- Lie to Jeff - Jeff's friend should not tell Jeff about the errors

The presence of others vignette

In the course of this conversation, which of the following options is the more ethical response?

- Tell Jeff the truth - Jeff's friend should tell Jeff his presentation went poorly
- Lie to Jeff - Jeff's friend should not tell Jeff his presentation went poorly

Appendix E. Effects of Implicit Rule Violations on Potential Mechanisms, Within Each Vignette (Study 3)

			Immediate harm of truth	**Instrumental value of truth**	**Self-interest (of liar)**	**Autonomy (of target)**	**Probability of detection**	**Societal harm**	**Moral duty**
Ability to understand	Control	M	5.35	4.75	5.20	4.46	5.27	2.51	5.39
		SD	*1.19*	*1.25*	*1.66*	*1.56*	*1.24*	*1.67*	*1.60*
	Violation	M	5.55	3.88	5.47	3.94	4.45	2.35	4.93
		SD	*1.32*	*1.54*	*1.43*	*1.76*	*1.26*	*1.68*	*1.77*
			p = .17	*p* < .01	*p* = .15	*p* = .01	*p* < .01	*p* = .42	*p* = .024
Time to implement	Control	M	3.39	6.05	4.08	4.31	5.50	2.73	5.72
		SD	*1.32*	*0.97*	*1.75*	*1.62*	*1.15*	*1.75*	*1.38*
	Violation	M	4.54	5.03	5.11	3.68	5.36	2.65	4.95
		SD	*1.33*	*1.23*	*1.49*	*1.59*	*1.27*	*1.93*	*1.66*
			p < .01	*p* < .01	*p* < .01	*p* < .01	*p* = .34	*p* = .71	*p* < .01
The presence of others	Control	M	4.73	5.92	5.49	4.20	4.84	2.65	5.14
		SD	*1.07*	*0.97*	*1.42*	*1.47*	*1.27*	*1.73*	*1.43*
	Violation	M	5.04	5.44	5.29	4.12	4.74	2.82	4.78
		SD	*1.27*	*1.17*	*1.39*	*1.33*	*1.12*	*1.85*	*1.61*
			p = .03	*p* < .01	*p* = .24	*p* = .63	*p* = .51	*p* = .44	*p* = .05

Note. The *p*-values reflect the results corresponding with one-way ANOVAs (within each vignette) using *Implicit Rule Violation* (Control vs. Violation) as a factor, and each potential mechanism as the dependent variable.

Appendix F. Correlation Between Potential Mechanisms and Endorsement of Deception from Each Perspective (Study 3)

	Immediate harm of truth	Instrumental value of truth	Self-interest (of liar)	Autonomy	Probability of detection	Societal harm	Moral duty
Liar	.431***	-.434***	.239***	-.242***	-.188**	-.255***	-.496***
Observer	.344***	-.470***	.139*	-.299***	-.201**	-.101$^+$	-.503***
Target	.443***	-.445***	.232***	-.262***	-.254***	-.169**	-.426***

Note. $^+$, *, **, *** denote significance at $p \leq .10$, .05, .01 and .001 respectively

Tables

Table 1. Coding Categories for Justifications and Implicit Rules of Deception

Panel A. Broad Justifications for Deception

		Justification	Description for coders	Examples of participant responses	Kappa	Ethics	Pref	Total
Justifications for deception	Harm to target	Immediate harm of honesty	These justifications include lies that are told to avoid harm to the target at the moment of communication. This type of harm is immediate and not long-lasting.	• From my perspective, lying to someone else is the right thing to do when we can avoid hurting others or make others happy / comfortable • Lying may be the right thing to do when telling that person the truth at that particular moment may be harmful to them.	0.63	78.9%	60.7%	70.8%
		Instrumental value of honesty	These justifications focus on whether or not there are any potential long-term benefits of honesty. Specifically, is the honest information important, actionable, and objective? These responses suggest that lying is ok when honesty does not have the potential to affect future behavior or thinking in a meaningful way or bring about any other benefit.	• As long as it isn't something that's incredibly important for them to know, why bother them with it when you can save them from the truth? • I would want to be lied to under certain circumstances where I cannot change the result.	0.74	65.7%	74.1%	69.4%
		TOTAL	This is a composite category reflecting the presence of either dimension above: Immediate harm or Instrumental value		0.75	92.2%	89.6%	91.0%
	Utilitarian		These justifications incorporate costs and benefits to parties other than the target of the lie. Any responses that mention how a lie will affect the liar, society, or third parties are considered Utilitarian.	• Lying to someone else is the right thing to do when it behooves both you and the other person to have them believe the lie. Lying may prevent conflicts... • It's ok to lie if you're under cover trying to save some prisoners of war. It's ethical if you're trying to capture a killer.	0.65	52.4%	17.8%	36.9%

	Never (Deontological)	"Never" indicates that the participants included a statement expressing that lying is never acceptable. "Never" means that the person does not provide any justifications or examples of when/why lying is right.	• There is no instance where lying to someone else is the right thing to do. • I would never want someone to lie to me.	0.79	2.4%	8.1%	5.0%

Panel B. Implicit Rules of Deception

		Reason to lie	Definition	Examples of participant responses	Kappa	Ethics	Pref	Ttal
Implicit Rules of Deception	Attributes of Target	1. Emotionally fragile	When a person is in an emotionally fragile state (bad day, feeling sad, depressed, drunk, etc.)	• When a person is mentally unstable and his or her emotional well being is at stake.	0.85	4.8%	4.4%	4.7%
		2. Cannot understand truth	When a person cannot cognitively understand the true information (a child, someone with dementia, etc.)	• When children ask quiestions about things that they should not know	0.86	25.3%	3.0%	15.3%
		3. Death Bed	When a person (the target) is at the end of their life	• I would want someone to lie to me about how long I might have to live if I were terminally ill.	0.92	7.8%	6.7%	7.3%
	Attributes of Topic	4. Subjective	When the truth is subjective (a function of different tastes, individual differences, preferences, a specific instance, etc.).	• I find a piece of clothing or accessory that I really like and makes me feel good, I would prefer not to have the person I'm with tell me he or she does not like what I've chosen	0.72	29.5%	29.6%	29.6%
		5. Trivial	When the topic is trivial (does not matter in any meaningful way to the target or others) or honesty is not the purpose of the exchange (e.g., social conventions or politeness are more important than honesty)	• I would rather have someone lie to me in trivial matters than important ones, because the magnitude of the issue at hand is smaller.	0.81	34.9%	22.2%	29.2%
		6. Uncontrollable	When the truth is about something that can never be changed (e.g., someone’s height, a death, a relationship that has ended) or that feels outside of someone’s control (e.g., weight, others' misdeeds)	• If someone knew how my mother really felt about me... I would prefer that the person would lie and tell me said good things about me. My mom is deceased now, so nothing could be changed anyway.	0.83	6.0%	17.8%	11.3%
	Context of Conversation	7. Precedes sacred event	When the truth is hurtful and may upset someone before another unrelated event, such as a wedding, honeymoon, special day.	• Being told there is no bad news before an important event so that the bad news can be postponed.	0.85	4.2%	8.1%	6.0%
		8. Feedback can no longer be implemented	When the conversation occurs after feedback could be implemented (e.g., the person can no longer change their clothing) or the conversation occurs immediately before an event and there is not enough time to implement feedback or changes (e.g., a person is about to go on stage).	• If I were out with my friends at a bar and I asked if I looked okay, I would prefer if my friends said yes because if I did not, there would be nothing I could do about it at the bar.	0.90	6.6%	10.4%	8.3%
		9. In front of others	When the conversation occurs in front of others (and might affect observers’ opinions, or embarrass the target)	• If the truth would embarrass me in front of important people	1.00	0.6%	2.2%	1.3%

Note. The tables above reflect the coding scheme for justifications and implicit rules of deception. Kappa reflects the level of agreement between the two research assistants who coded participant responses, for each coding category. The percentages listed reflect the percentage of participants that listed each justification/implicit rule in the *Ethics* condition, the *Preferences* condition, and in total (respectively).

Table 2. Implicit Rules and Dimensions of Unnecessary Harm

	Source of implicit rule	Moral considerations	Dimension of unnecessary harm
Target	1. Emotional fragility	Emotional reaction (psychological harm)	Increased immediate harm (i.e. harm at the moment of communication)
	2. Ability to understand	Potential to learn from information and implement change Confusion (psychological harm)	Decreased instrumental value Increased immediate harm (i.e. harm at the moment of communication)
	3. Death bed	Potential to implement change Emotional fragility (psychological harm)	Decreased instrumental value Increased immediate harm (i.e. harm at the moment of communication)
Topic	4. Subjective	Importance of information	Decreased instrumental value
	5. Trivial	Importance of information	Decreased instrumental value
	6. Uncontrollable	Potential to implement change	Decreased instrumental value
Context	7. Disruption to special moments and events	Distraction (psychological harm)	Increased immediate harm (i.e. harm at the moment of communication)
	8. Time to implement	Potential to implement change	Decreased instrumental value
	9. The presence of others	Embarrassment (psychological harm)	Increased immediate harm (i.e. harm at the moment of communication)

Table 3. Vignette Details and Results in Study 2

Implicit rule (Vignette name)	Survey group	Scenario Introduction	Control Condition	Implicit Rule Violation Condition
1. Emotional fragility	1	Imagine an employee who just turned in his weekly marketing report to his manager. Although the employee usually delivers good work, the manager - unbeknownst to the employee - does not think this report was done well. The employee has a meeting today with his manager.	The employee asks the manager what he thought of the report. 3% endorse deception χ2 = 18.36, p < .01	The employee's father was unexpectedly hospitalized this morning and the employee is very distressed. The manager knows this information. The employee asks the manager what he thought of the report. 19.5% endorse deception
2. Ability to understand	2	Imagine a doctor who realizes that her patient's cancer is terminal, meaning the cancer is not curable and the patient will likely die.	The doctor's patient is a 44-year-old adult. The patient can tell something is wrong and is very distressed. The patient asks the doctor if s/he is going to die. 7.7% endorse deception χ2 = 36.67, p < .01	The doctor's patient is a 4-year-old child. The patient can tell something is wrong and is very distressed. The patient asks the doctor if s/he is going to die. 33.2% endorse deception
3. Death bed	1	Imagine an individual who is seriously ill. During the individual's illness, his spouse cheated on him. The individual does not know this and still deeply loves his spouse.	Although the individual is still ill, he is very likely to recover. The individual is talking to his spouse about their relationship and asks if the spouse has ever been unfaithful. 31% endorse deception χ2 = 28.67, p < .01	The individual is still ill and is likely to die within the next 24 hours. The individual is talking to his spouse about their relationship and asks if the spouse has ever been unfaithful. 63.8% endorse deception
4. Subjective	3	Imagine an employee who must deliver an important presentation. She plans on wearing her favorite silk scarf during the presentation. She loves the scarf and thinks it brings her good luck. Imagine that the employee's colleague – unbeknownst to the employee - thinks the scarf is hideous.	The colleague also knows that many other employees share this opinion. The day of her presentation, the employee shows up in a suit and her silk scarf and asks how she looks in it. 39.4% endorse deception χ2 = 54.98, p < .01	The colleague also knows, however, that many other employees do not share this opinion. Many colleagues like the scarf. The day of her presentation, the employee shows up in a suit and her silk scarf and asks how she looks in it. 71.2% endorse deception

5. Trivial	2	Imagine an individual who is hosting a dinner party. The host serves soup, which one guest finds to be very salty. The host asks the guest what he thinks of the soup.	This individual, the host, cooks very often. The host is a professional chef and is hosting the party to try out new recipes for his/her restaurant. The host serves soup, which one guest finds to be very salty. The host asks the guest what he thinks of the soup. 18% endorse deception	This individual, the host, does not cook very often. The host has no professional cooking training and is hosting the party for fun. The host serves soup, which one guest finds to be very salty. The host asks the guest what he thinks of the soup. 37.8% endorse deception
			χ2 = 18.82, p < .01	
6. Uncontrollable	3	Imagine a summer intern who just delivered his end-of-internship presentation to his office. The intern stuttered quite a bit during the presentation. The intern's friend attended the presentation and believed that the intern's stutter notably decreased the quality of his presentation, compared to his fellow interns. Aside from the stutter, the presentation was pretty good.	The intern stuttered because he was nervous during this particular presentation. He can likely improve his ability to speak without a stutter. The intern's friend knows this information. The intern asks his friend what he thought of the presentation. 18.8% endorse deception	The intern stuttered because he has a diagnosed speech impediment. The intern cannot improve his ability to speak without a stutter. The intern's friend knows this information. The intern asks his friend what he thought of the presentation. 56.5% endorse deception
			χ2 =80.51, p < .01	

7. Disruption to special moments and events	2	Imagine a manager who must fire 10% of his workforce. It is a Friday afternoon and top management has just given the manager a list of employees to lay off. It is the beginning of December and the manager has until January 1st to inform employees of their work status. After January 1st, employees will have 6 months - at full pay - to search for new jobs and finish their roles. Nothing about their work will change until that time. Imagine an employee who is on the layoff list. This employee has no idea that layoffs are coming, but the employee does know that the company is going through a reorganization.	The employee drops by the manager’s office on his/her way out the door on Friday. The employee asks the manager if there’s any news about the reorganization. 22.9% endorse deception	The employee is getting married this weekend - on Saturday - and s/he drops by the manager’s office on his/her way out the door on Friday. The employee asks the manager if there’s any news about the reorganization. 52% endorse deception
			$\chi2 = 35.16$, $p < .01$	

8. Time to implement	1	Imagine an employee who must deliver an important presentation. He will pitch a new marketing plan to his manager and colleagues. He plans on wearing his favorite black suit during the presentation. Imagine that the employee's colleague – unbeknownst to the employee - thinks this suit is too tight and that the suit is inappropriate for the presentation.	The day before his presentation, the employee tells his colleague that he plans on wearing this suit and he asks the colleague how he looks in it. At this time, the employee has other suits available that he can wear. 7.6% endorse deception	The day of his presentation, the employee shows up in his suit and he asks his colleague how he looks in it. At this time, the employee has no other suits available that he can wear. 64.4% endorse deception
			χ2 = 93.31, p < .01	
9. The presence of others	3	Imagine an employee who just turned in his weekly marketing report to his manager. Although the employee usually delivers good work, the manager - unbeknownst to the employee - does not think this report was well done.	The employee has a one-on-one meeting today with his manager. The employee enters the manager's office. The employee asks the manager what he thought of the report. 1.5% endorse deception	The employee is attending a company-wide networking event today. The employee walks into the event and begins talking to his manager and several other colleagues. In front of a group of colleagues, the employee asks the manager what he thought of the report. 38.3% endorse deception
			χ2 = 115.90, p < .01	

Note.
I ran three separate surveys at different points in time. Each survey (denoted by Survey group) featured three vignettes.
In Survey Group 1: Mturk, N = 267; 46.8% female; Mage = 35, *Perspective* was manipulated between subjects.
In Survey Group 2: U.S. university laboratory, *N* = 195; 52.3% female; M_{age} = 25, *Perspective* was manipulated within subjects.
In Survey Group 3: Mturk, N = 269, 45.4% female; Mage = 38, *Perspective* was manipulated within subjects.

Table 4. Meta-analysis on all Vignettes in Study 2

Dependent variable = Endorsement of deception; 1 = lie, 0 = tell the truth						
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
Intercept	-0.79**	-1.44***	-1.39***	-1.41***	-3.28***	-1.36***
Gender[a]	0.24*					
Age	-0.00					
Implicit Rule Violation[b]		1.70***	1.70***	1.73***	0.94***	.92***
Perspective[c]			-0.10	-0.05		
Perspective x Implicit Rule Violation				-0.08		
Immediate Harm of Truth					1.19***	.81***
Instrumental Value of Truth					-0.73***	-1.17***
Immediate Harm x Instrumental Value						.093*
Vignette Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by Vignette	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by Participant	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3585	3585	3585	3585	3585	3585
R^2	0.07	0.17	0.17	0.17	0.48	0.49

Note. *, **, *** denote significance at $p \leq .05$, < .01 and <.001 respectively

[a] Gender is coded as 1 = female, 0 = male

[b] Violation is coded as 1 = implicit rule violation, 0 = no rule violation

[c] Perspective is coded as 1 = target, 0 = observer

Table 5. Vignette-level Analyses in Study 3

Dependent variable: Endorsement of deception; 1 = lie, 0 = tell the truth

Vignette:	**Ability to understand**	**Time to Implement**	**Presence of Others**
Intercept	-1.21***	-3.11***	-2.61***
Implicit Rule Violation[a]	**.98***	**2.07****	**2.49*****
Liar[b]	-.12	-.72	**1.40***
Target[c]	.35	-.72	.92
Liar x Implicit Rule Violation	.18	.96	**-1.57***
Target x Implicit Rule Violation	-.12	.60	**-1.65***
R^2	0.08	0.22	0.14

Note. *, **, *** denote significance at $p \leq .05$, < .01 and <.001 respectively

[a] Violation is coded as 1 = implicit rule violation, 0 = no rule violation

[b] Liar is coded as 0 = target or observer perspective, 1 = liar perspective

[b] Target is coded as 0 = liar or observer perspective, 1 = target perspective

Table 6. Meta-analysis on all Vignettes in Study 3

Dependent variable = Endorsement of deception; 1 = lie, 0 = tell the truth

	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Intercept	-0.43^{+}	-1.33***	-1.37***	-1.58***	-2.08***	.02	1.64*
Gender[a]	0.49***						
Age	-.02^{+}						
Implicit Rule Violation[b]		1.35***	1.36***	1.67***	.75***	.73***	.91***
Liar[c]			.13	.34			
Target[d]			-.01	.38*			
Liar x Implicit Rule Violation				-.32			
Target x Implicit Rule Violation				-.60			
Immediate Harm of Truth					.90***	.51*	.53***
Instrumental Value of Truth					-.95***	-1.43***	-1.03***
Immediate Harm x Instrumental Value						.09^{+}	.08***
Self-interest							.12
Autonomy							-.07
Probability of Detection							-.05^{+}
Societal Harm							-.01
Moral Duty							-.66***
Vignette Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by Vignette	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cluster by Participant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	834	834	834	834	834	834	834
R^2	0.05	0.10	0.10	0.10	0.34	0.35	0.43

Note. $^{+}$, *, **, *** denote significance at $p \leq .10$, .05, .01 and .001 respectively

[a]Gender is coded as 1 = female, 0 = male; [b]Violation is coded as 1 = implicit rule violation, 0 = no rule violation; [c]Liar is coded as 0 = target or observer perspective, 1 = liar perspective; [d]Target is coded as 0 = liar or observer perspective, 1 = target perspective

Figures

Figure 1. Theoretical Framework

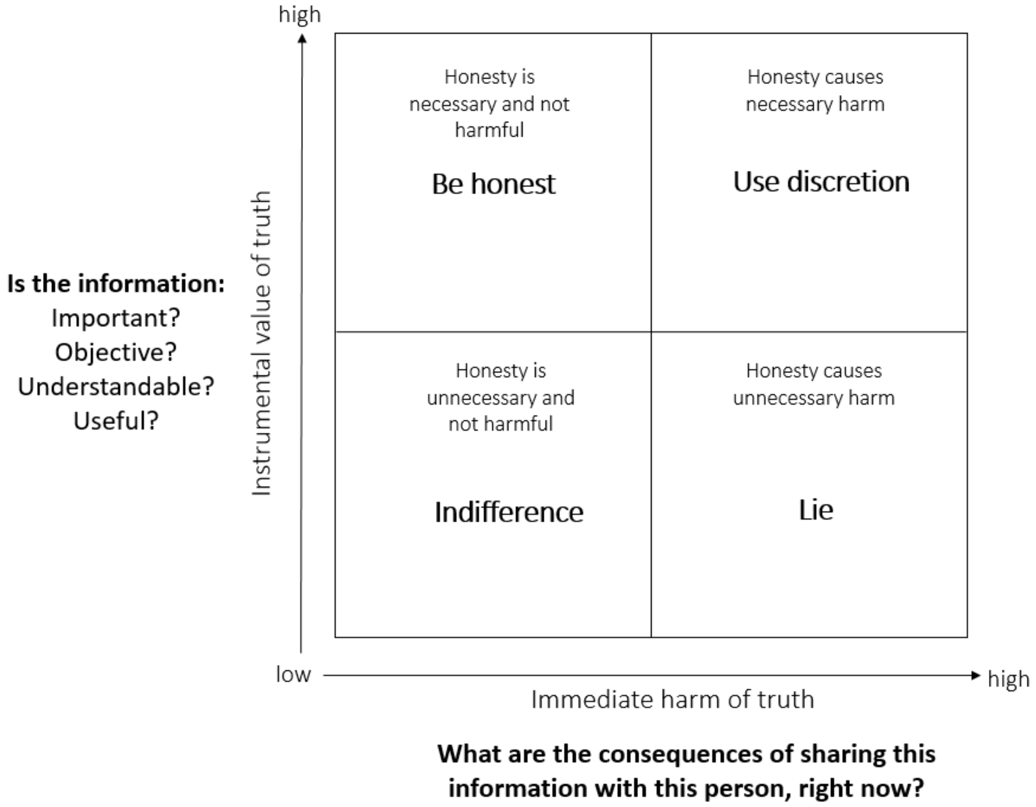
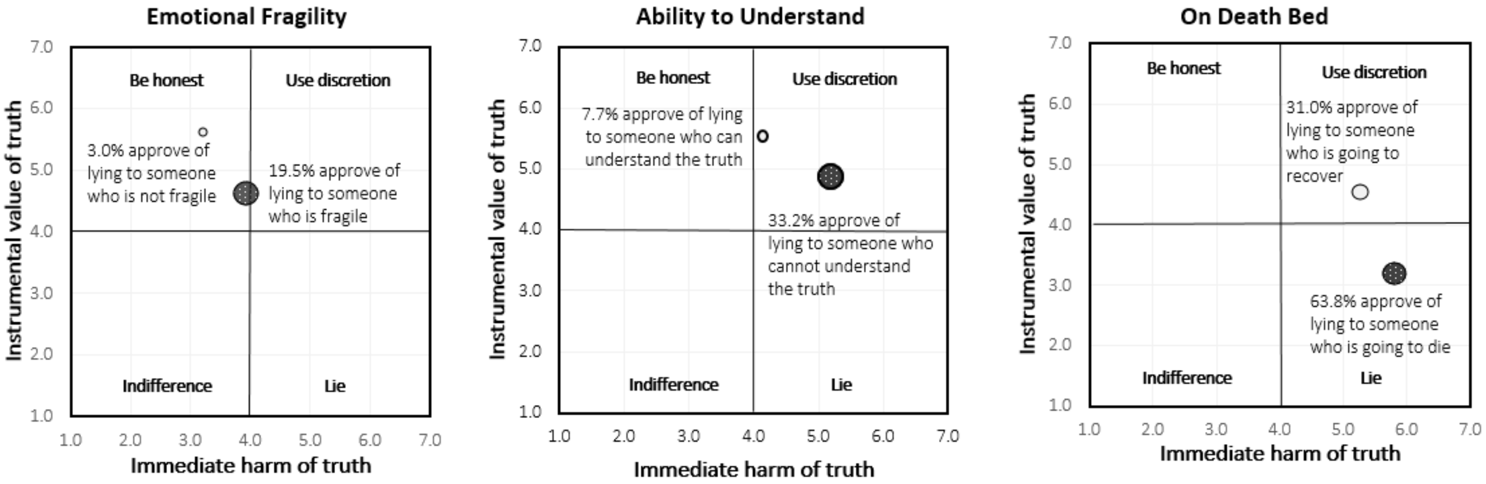
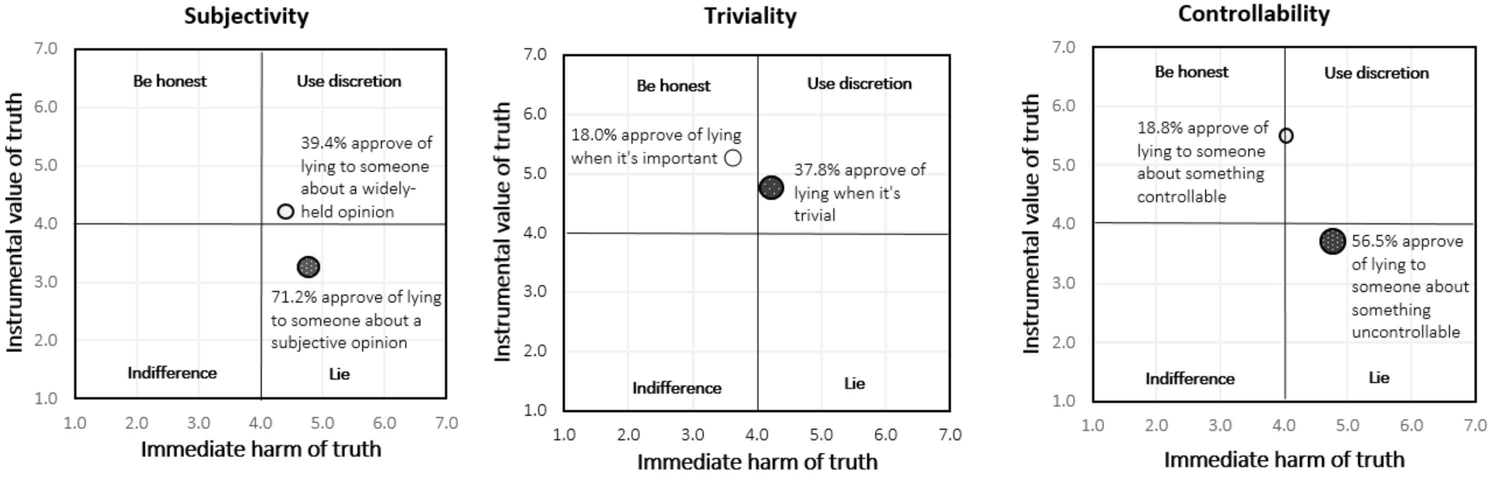

high
Honesty is necessary and not harmful
Be honest
Honesty causes necessary harm
Use discretion
Is the information:
Important?
Objective?
Understandable?
Useful?
Instrumental value of truth
Honesty is unnecessary and not harmful
Indifference
Honesty causes unnecessary harm
Lie
low
high
Immediate harm of truth
What are the consequences of sharing this information with this person, right now?

Figure 2. Mapping the Implicit Rules on to the Theoretical Framework (Study 2)

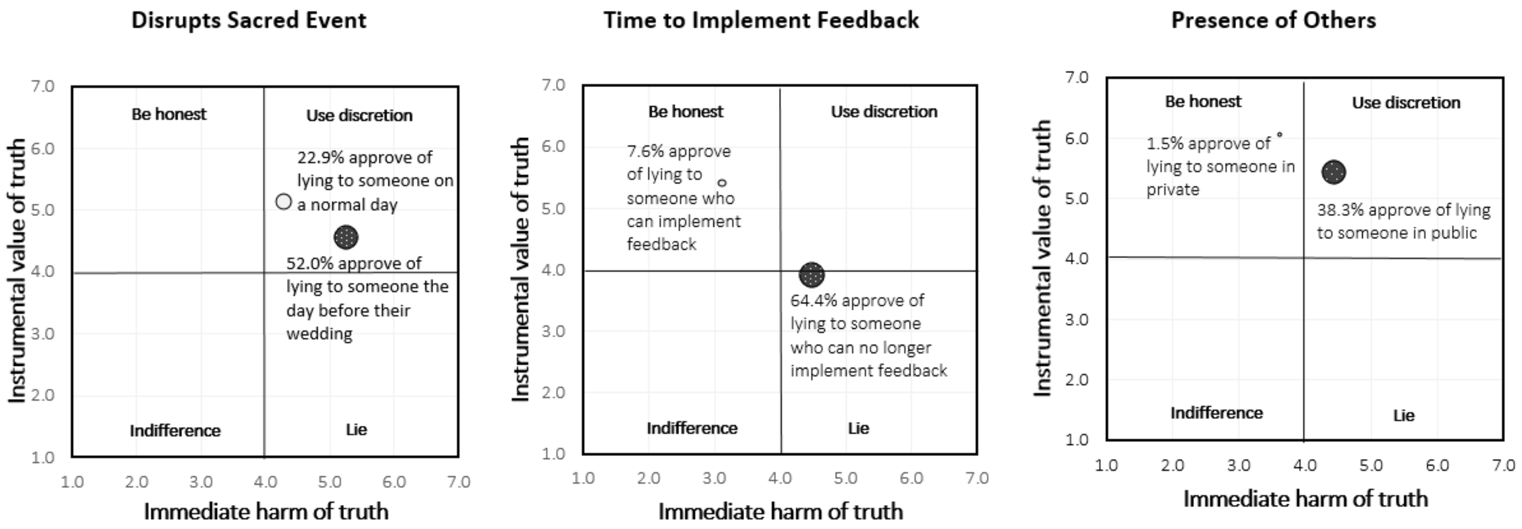
Panel 1: Implicit rules pertaining to attributes of the target



Panel 2: Implicit rules pertaining to attributes of the honest information



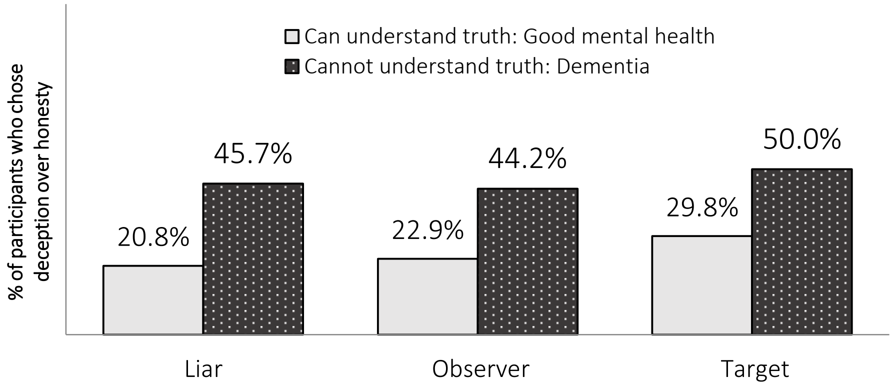
Panel 3: Implicit rules pertaining to attributes of the context



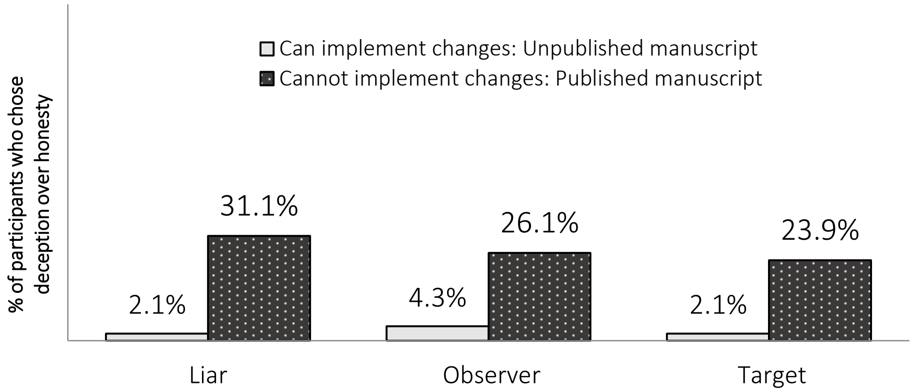
Note. For each vignette, I plot the mean ratings of immediate harm (X axis) and instrumental value (Y axis) in the control condition (white dot) and in the implicit rule violation condition (dark gray, spotted dot). The size of each data point is proportional to the percentage of people who endorsed deception in each condition.

Figure 3. Implicit Rule Violations Across Perspectives (Study 3)

Panel 1. Dementia and a daughter's death (Ability to understand vignette)



Panel 2. Errors in a published versus unpublished manuscript (Time to implement vignette)



Panel 3. Public versus private feedback on a presentation (The presence of others vignette)

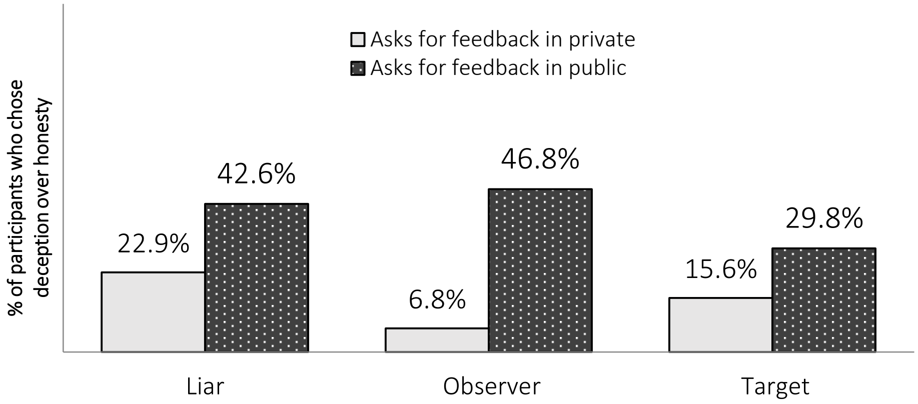
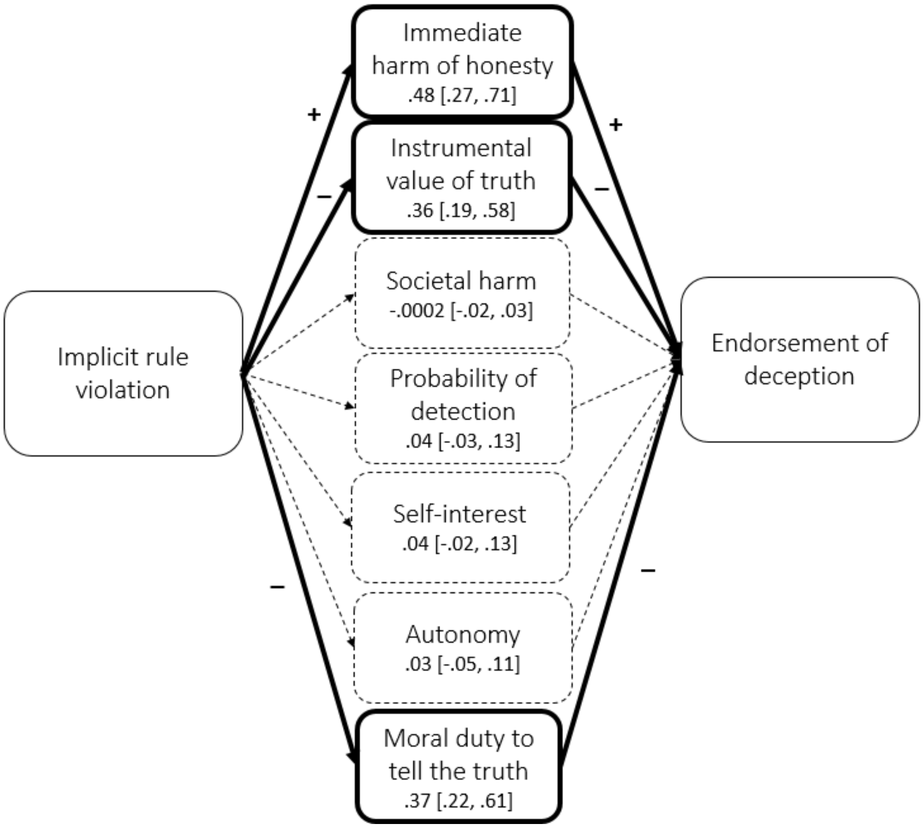

□ Asks for feedback in private
▩ Asks for feedback in public
% of participants who chose deception over honesty
22.9%
42.6%
6.8%
46.8%
15.6%
29.8%
Liar
Observer
Target

Figure 4. Mediation Analysis in Study 3



Note. Numbers reflect the indirect effect and the 95% confidence interval around the indirect effect for each proposed mechanism.

CHAPTER 3.

PROSOCIAL LIES: WHEN DECEPTION BREEDS TRUST

Emma E. Levine

Maurice Schweitzer

Published in *Organizational Behavior and Human Decision Processes* in 2015.

ABSTRACT

Philosophers, psychologists, and economists have long asserted that deception harms trust. We challenge this claim. Across four studies, we demonstrate that deception can increase trust. Specifically, prosocial lies increase the willingness to pass money in the trust game, a behavioral measure of benevolence-based trust. In Studies 1a and 1b, we find that altruistic lies increase trust when deception is directly experienced and when it is merely observed. In Study 2, we demonstrate that mutually beneficial lies also increase trust. In Study 3, we disentangle the effects of intentions and deception; intentions are far more important than deception for building benevolence-based trust. In Study 4, we examine how prosocial lies influence integrity-based trust. We introduce a new economic game, the *Rely-or-Verify* game, to measure integrity-based trust. Prosocial lies increase benevolence-based trust, but harm integrity-based trust. Our findings expand our understanding of deception and deepen our insight into the mechanics of trust.

PROSOCIAL LIES: WHEN DECEPTION BREEDS TRUST

Trust is essential to organizations and interpersonal relationships (e.g., Blau, 1964; Golembiewski & McConkie, 1975; Dirks & Ferrin, 2001; Lewicki, Tomlinson, & Gillespie, 2006; Rempel, Holmes, & Zanna, 1985; Valley, Moag, & Bazerman, 1998). Trust increases leadership effectiveness (Atwater, 1988; Bazerman, 1994; Dirks, 2000), improves the stability of economic and political exchange (Hosmer, 1995), reduces transaction costs (Granovetter, 1985), facilitates cooperation (Valley et al., 1998), and helps firms and individuals manage risk (Sheppard & Sherman, 1998). Golembiewski and McConkie (1975, p. 131) argued that, "There is no single variable which so thoroughly influences interpersonal and group behavior as does trust."

Consistent with prior research, we define trust as, "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another" (Rousseau, Sitkin, Burt, & Camerer, 1998, p. 395). A significant body of research has documented the negative effects of violating trust. For example, trust violations can harm cooperation and bargaining outcomes (Lount, Zhong, Sivanathan, & Murnighan, 2008; Croson, Boles, & Murninghan, 2003), lower organizational commitment (Robinson, 1996), provoke retaliation (Bies & Tripp, 2006), and, in more serious cases, trigger organizational-level failures (Gillepsie & Dietz, 2009).

Although there are many ways to harm trust, existing research identifies one behavior as particularly toxic to trust: deception (e.g., Santoro & Paine, 1993; Boles, Croson, & Murninghan, 2000; Carr, 1968; O'Connor & Carnavale, 1997; Schweitzer & Croson, 1999; Schweitzer, Hershey, & Bradlow, 2006; Croson et al., 2003; Bok, 1978). Prior research suggests that deception is theoretically, philosophically, and empirically antithetical to trust. For example, philosopher Sir Francis Bacon argued that dishonesty deprives, "people of two of the most principal instruments for interpersonal action—trust and belief" (from "On Truth", cited in Tyler & Feldman, 2006). Empirical research has also demonstrated that deception harms relationships (Ford, King, & Hollender, 1988; Lewis & Saarni, 1993; Tyler & Feldman, 2006), elicits negative affect (Planalp, Rutherford, & Honeycutt, 1988), decreases liking, (Tyler, Feldman, & Reichert, 2006) and triggers retaliation (Boles et al., 2000; Croson et al., 2003). Furthermore, trust scholars have found that acts of deception cause enduring harm to trust. Though individuals can often repair trust following a violation (e.g., Kim, Ferrin, Cooper, & Dirks, 2004; Schweitzer et al., 2006), trust violations accompanied by deception irrevocably harm trust (Schweitzer et al., 2006).

We challenge the prevailing assumption that deception harms trust. We argue that most philosophers, psychologists, and economists, have confounded deceptive behavior with selfish intentions and outcomes. As a result, prior research that has documented the harmful effects of deception may really tell us more about the consequences of selfish behavior than deception per se.

We break new ground by demonstrating that some forms of deception increase trust. Across four experiments, we demonstrate that prosocial lying can increase behavioral and attitudinal measures of interpersonal trust. Consistent with prior work, we define deception as *the transmission of information that intentionally misleads others* (see Murnighan, 1991; Boles et al., 2000; Gino & Shea, 2012). We define prosocial deception as a type of deception. Prosocial lies involve the transmission of information that misleads *and benefits* a target (Levine & Schweitzer, 2014).

Our program of research expands our understanding of trust by disentangling the role of benevolence and integrity for building interpersonal trust. In our investigation, we explore distinct forms of both deception and trust. We are the first to demonstrate that some common forms of deception can increase trust.

We report results from a series of experiments. In Studies 1, 2, and 3, participants experienced or observed deception and made decisions in a trust game. Across these studies, we find that prosocial lies increase trust. This is true when deception is directly experienced (Study 1a) and when it is merely observed (Study 1b). This pattern is also true when the prosocial lies are mutually beneficial and help both the target and the deceiver (Study 2).

In Studies 3a and 3b, we disentangle the effects of lying from the effects of prosocial and selfish intentions. When we control for intentions, we find that deception itself has no effect on trusting behavior. In other words, the decision to pass money in the trust game reflects perceptions of benevolence, which is not undermined by deception. Prosocial intentions, regardless of whether they are associated with deception or honesty, significantly increase benevolence-based trust. In Study 3b, we demonstrate that our results do not simply reflect a negative reaction to selfish behavior. Instead, we find that prosocial deception increases trust compared to a neutral control condition.

In our final study, we explore how prosocial deception influences distinct types of trust. The trust game reflects benevolence-based trust; it operationalizes the willingness to be vulnerable to interpersonal exploitation. We introduce a new economic game, the *Rely-or-Verify* game, which reflects integrity-based trust. The *Rely-or-Verify* game operationalizes the willingness to rely on the veracity of another person. Although prosocial lying increases benevolence-based trust, it harms integrity-based trust. We demonstrate that the same action can have divergent effects on different dimensions of trust.

Prosocial lying

Prosocial lying is a common feature of everyday communication. For example, an employee may tell a colleague that they delivered an excellent presentation when they did not, or thank a gift giver for a gift they would have rather not received.

As children, we learn to tell prosocial lies to be polite (Talwar, Murphy, & Lee, 2007; Broomfield, Robinson, & Robinson, 2002). Prosocial deception is also common in adult relationships (Tyler & Feldman, 2004). Adults lie in roughly 20% of their everyday social interactions (DePaulo & Bell, 1996), and most of these lies are prosocial (DePaulo & Kashy, 1998).

Individuals' endorsement of prosocial lies reflects the broader approval of unethical behaviors that help others. For example, individuals are more willing to cheat when cheating restores equity (Gino & Pierce, 2009, 2010a; Schweitzer & Gibson, 2008), helps disadvantaged others (Gino & Pierce, 2010b), and when the spoils of cheating are shared with others (Wiltermuth, 2011; Gino, Ayal, & Ariely, 2013). With respect to deception, prior experimental work has found that individuals are more willing to tell prosocial lies than selfish lies (Erat & Gneezy, 2012) and perceive prosocial lies to be more ethical (Levine & Schweitzer, 2014).

Prosocial lying serves a number of interpersonal aims. While many prosocial lies are motivated by an altruistic desire to protect relational partners (e.g. DePaulo & Kashy, 1998) or provide interpersonal support (Brown & Levinson, 1987; Goffman, 1967), other lies have both prosocial and self-serving motives. For example, prosocial lying can be used to avoid conflict and facilitate uncomfortable social situations. When a wife asks her husband if she looks fat in her dress, the husband may lie not only to protect his wife's feelings, but also to avoid conflict and a lengthy discussion about diet and exercise.

In the present research, we distinguish between lies that are costly for the liar and lies that benefit the liar. We define *altruistic lies* as, "*false statements that are costly for the liar and are made with the intention of misleading and benefitting a target*" (Levine & Schweitzer, 2014: p. 108). We define *mutually beneficial lies* as *false statements that are beneficial for the liar and are made with the intention of misleading and benefitting the target*. We conceptualize altruistic and mutually beneficial lies as a subset of prosocial lies. Consistent with Bok (1978), we also distinguish between prosocial lies and white lies. White lies involve small stakes and can be prosocial or self-serving. Unlike white lies, prosocial lies can have large stakes. For example, some doctors misrepresent prognoses to give their patients comfort in their final weeks of life (e.g., Iezzoni, Rao, DesRoches, Vogeli, & Campbell, 2012).

Prosocial lies and trust

Prosocial lies are particularly relevant to the study of trust because they reflect a conflict between two central antecedents of trust: benevolence and integrity. Trust reflects an individual's expectation about another person's behavior. In contrast with research that conceptualizes trust as a belief about one's ability to carry out organizational duties or effectively perform a particular job (Kim et al., 2004; Kim, Dirks, Cooper, & Ferrin, 2006; Ferrin, Kim, Cooper, & Dirks, 2007), we conceptualize trust as the willingness to be vulnerable to exploitation within an interpersonal interaction (e.g. Lewicki & Bunker, 1995; Rousseau et al., 1998),

Scholars have converged on three qualities of the trustee (the individual who is trusted) that uniquely influence interpersonal trust: benevolence, ability, and integrity (Mayer, Davis, & Schoorman, 1995; Butler, 1991). Benevolence reflects the extent to which an individual has positive intentions or a desire to help the truster (Butler & Cantrell, 1984; Mayer et al., 1995). Ability reflects an individual's technical skills, competence, and expertise in a specific domain (e.g., Mayer et al., 1995; Giffin, 1967; Sitkin & Roth, 1993). Integrity reflects an individual's ethicality and reputation for honesty (Mayer et al., 1995; Butler & Cantrell, 1984). In this work, we investigate the tension between benevolence and integrity.

Existing trust research highlights the importance of benevolence for building interpersonal trust. In dyadic relationships, trust hinges on concerns about exploitation (Barney & Hansen, 1994; Lewicki & Bunker, 1995; Bhattacharya, Devinney & Pillutla, 1998), and perceptions of benevolence can allay these concerns. Individuals who are perceived to have benevolent motives are perceived to be less likely to exploit a potential truster, and consequently, are more likely to be trusted (e.g., Weber, Malhotra, & Murnighan, 2004; Malhotra & Murnighan, 2002; Pillutla, Malhotra, & Murnighan, 2003; Dunn, Ruedy, & Schweitzer, 2012; Lount & Pettit, 2012).

Prior work has also suggested that integrity is a critical antecedent to interpersonal trust. Establishing a direct link between integrity and trust, however, has been difficult. Part of this difficulty stems from the subjective nature of integrity: the belief that "the trustee adheres to a set of principles that the truster finds acceptable" (Mayer et al., 1995, p. 719; Kim et al., 2004). In addition, in nearly every investigation of the link between integrity and trust, integrity has been confounded with benevolence (e.g. Kim et al., 2004; Schweitzer et al., 2006). That is, prior trust research that has studied behaviors that violate ethical principles *and* cause harm to others, reflecting low integrity and low benevolence. For example, prior work has studied lies that exploit others for financial gain (Koning, Steinel, Beest, & van Dijk, 2011; Steinel & De Dreu, 2004; Schweitzer et al., 2006). These lies violate the principle of honesty *and* demonstrate selfishness. Not surprisingly, these lies harm trust. However, an individual may also lie to *benefit* a counterpart. This behavior violates the principle of honesty, but demonstrates benevolence. Existing trust work does not give us insight into how individuals might resolve these competing signals.

Research on corruption and favoritism, however, provides evidence that individuals can place enormous trust in individuals who have demonstrated low integrity. For example, scholars have documented high trust among members of crime rings (Baccara & Bar-Isaac, 2008; Bowles & Gintis, 2004) and among members of communities that have been influenced by organized crime (Meier, Pierce, & Vaccaro, 2013). In these groups, individuals trust in-group members, but distrust out-group members. Individuals within the group are trusted because they care for and protect in-group members, even if they have demonstrated low integrity with respect to their interactions with out-group members.

We conjecture that for interpersonal trust judgments, the concern for benevolence is more deeply rooted than the concern for integrity. The preference individuals have for ethical rules, such as fairness and honesty, may derive from the more fundamental concern for protecting people from harm (Gray, Young, & Waytz, 2012; Turiel, 1983). That is, benevolence may be the primary concern and integrity may be a derivative, secondary concern. Consistent with this proposition, Levine & Schweitzer (2014) found that when honesty harms other people and deception does not, honesty is perceived to be less ethical than deception.

We postulate that individuals who project high benevolence, even if they also project low integrity, will engender trust. We expect this to be particularly true for trust judgments that involve vulnerability to interpersonal exploitation. As a result, we hypothesize that prosocial lies, which demonstrate high benevolence, but low integrity, will build trust.

Overview of current research

Across our studies, we use deception games (adapted from Erat & Gneezy, 2012; Cohen, Gunia, Kim-Jun, & Murnighan, 2009; Gneezy, 2005) and trust games (adapted from Berg, Dickhaut, & McCabe, 1995). We use deception games to operationalize prosocial lies, because these games allow us to cleanly manipulate the intentions associated with deception and, consequently, draw causal inferences about the role of intentions and deception in building trust.

We use the trust game in our first three studies, because it operationalizes the fundamental components of an interpersonal trusting decision: the willingness to be vulnerable based on positive expectations of another (Rousseau et al., 1998; Pillutla et al., 2003). The trust game reflects benevolence-based trust and is the predominant paradigm used to measure trust throughout psychology and economics (e.g., Berg et al., 1995; Schweitzer et al., 2006; McCabe, Rigdon, & Smith, 2003; Glaeser, Laibson, Scheinkman, & Soutter, 2000; Malhotra & Murninghan, 2002; Malhotra, 2004). In the standard trust game, the truster is endowed with money and has the opportunity to keep the money or pass the money to the trustee. The amount of money grows if the truster passes it to the trustee. The trustee then has the opportunity to either return some portion of the money to the truster or keep all of the money for himself. The truster's initial decision to pass money represents trust (Pillutla et al., 2003; McCabe et al., 2003; Glaeser et al., 2000; Malhotra & Murninghan, 2002; Malhotra, 2004). Though trust game decisions may also reflect preferences for equality and risk (Ashraf, Bohnet, & Piankov, 2003), the external validity of trust game decisions has been documented with financial investment decisions (e.g., Karlan, 2005) and prior work has closely linked trust game behavior with attitudinal measures of trust (e.g., Houser, Schunk, & Winter, 2010; Schweitzer et al., 2006).

We begin our investigation by examining the consequences of altruistic lies. In Study 1a, participants were paired with a confederate who either told an altruistic lie to the participant or was selfishly honest. Participants then played a trust game with the confederate. In Study 1a, we find that being deceived increases trust; participants were more trusting of confederates who lied to them than they were of confederates who were honest. In Study 1b, we rule out reciprocity as an alternative explanation. In this study, participants observed, rather than experienced, altruistic deception and then made trust decisions. We find that individuals trust altruistic liars, even when they did not benefit from the prosocial deception.

In Study 2, we extend our investigation by examining different types of lies. In this study, we find that even when prosocial lying helps the liar, deception increases trust; non-altruistic prosocial lies, and mutually beneficial lies increase trust. In Studies 3a and 3b, we isolate the effects of intentions and deception by manipulating them orthogonally. In Study 3a, we find that deception itself has no direct effect on benevolence-based trust, but that intentions matter immensely. Prosocial individuals who told lies or were honest were trusted far more than selfish individuals who lied or were honest. In Study 3b, we include two control conditions and demonstrate that relative to control conditions, prosocial intentions increase trust and selfish intentions decrease trust.

Our first set of studies demonstrate that trust rooted in perceptions of benevolence is not undermined by deception. In our final study, we explore the influence of deception on trust rooted in perceptions of integrity. We introduce a new type of trust game, the *Rely-or-Verify* game, in which trust decisions rely on perceptions of honesty. In this study, we identify a boundary condition of the effect we observe in our initial studies. We find that deception does not harm trust decisions that are rooted in perceptions of integrity.[9]

Study 1

In Studies 1a and 1b, we explore the relationship between altruistic lying and trusting behavior. In Study 1a, participants played a trust game with a counterpart who either told them an altruistic lie or told them a selfish truth. In Study 1b, participants observed an individual who either told an altruistic lie or a selfish truth to a third party. Together, Studies 1a and 1b demonstrate that altruistic deception can *increase* trust and that this result cannot be explained by direct reciprocity.

Study 1a

Method

Participants. We recruited 125 adults to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. In this study, we randomly assigned participants to one of two conditions in a between-subjects design. Participants played a deception game with an individual who either told an altruistic lie or was selfishly honest. Participants then played a trust game with the same partner.

Manipulation of altruistic lies. We used a modified deception game (Erat & Gneezy; 2012; Cohen et al., 2009; Gneezy, 2005; Levine & Schweitzer, 2014) to

[9] Across all of our studies, our sample size or the number of days that the study would run was determined in advance, and no conditions or variables were dropped from any analyses we report.

operationalize altruistic lies. We referred to the deception game as "Exercise 1" in the experiment.

In our version of the deception game, two individuals were paired and randomly assigned to the role of either Sender or Receiver. The payoffs for each pair of participants (one Sender and one Receiver) were determined by the outcome of a computer-simulated coin flip and the choices the participants made. In the deception game, the Sender had the opportunity to lie to the Receiver about the outcome of the coin flip. In the experiment, we refer to the potential liar as "the Sender."

The deception game unfolded in the following steps:

5. Senders were told the outcome of a computer-simulated coin flip. In our study, the coin always landed on heads.
6. The Sender then had to report the outcome of the coin flip to his/her partner, the Receiver. The Sender could send one of two possible messages to the Receiver. The message could read, "The coin landed on heads" or "The coin landed on tails."
    - ➢ The Sender knew that the outcome the Receiver chose (heads or tails) determined the payment in the experiment. The Sender also knew that the only information the Receiver would have was the message from the Sender and that most Receivers chose the outcome indicated in the Sender's message.
    - ➢ The Sender knew there were two possible payment options, A and B. If the Receiver chose the correct outcome, the Sender and the

Receiver would be paid according to Option A. Otherwise, the Sender and the Receiver would be paid according to Option B.

7. In Study 1a, Option A was $2 for the Sender and $0 for the Receiver. Option B was $1.75 for the Sender and $1 for the Receiver. Throughout our studies, we manipulated the payments associated with Option A and Option B to operationalize different types of lies. We summarize the payoffs associated with each choice in Table 1.
8. After receiving the Sender's message, the Receiver had to choose an outcome: heads or tails. The Receiver knew that his/her choice determined the payment in the experiment, but the Receiver did not know the payoffs associated with the choice. The Sender's message was the only piece of information the Receiver had.

Therefore, Senders faced the following options:

C. Send an honest message, e.g. "*The coin landed on heads."*

Honesty was most likely to lead to an outcome that was costly to the Receiver, and benefitted the Sender (i.e. selfish).

D. Send a dishonest message, e.g. "*The coin landed on tails.*"

Lying was most likely to lead to an outcome that benefitted the Receiver, and was costly to the Sender (i.e. altruistic).

In Study 1a, we assigned all participants to the role of Receiver and informed them that their decisions would be matched with the decisions of a previous participant, who had been assigned to the role of Sender. After reading the instructions for the deception game and passing a comprehension check[10], participants received a message from their partner, the Sender. The Sender's message either read "The coin landed on heads" (the *Selfish Honesty* condition) or "The coin landed on tails" (the *Altruistic Lie* condition). Participants then made their prediction by choosing either "Heads" or "Tails."[11] Participants did not know the possible payoffs when they made their choice.

After making their choice, participants learned more information about the deception game. Specifically, we gave them all of the Sender's private information. Participants learned that the Sender knew the coin had landed on heads. Therefore, participants learned that the Sender either lied to them or had been honest. In addition, participants learned the payoffs associated with the Sender's choice. Therefore, participants learned that lying was altruistic and honesty was selfish. This was our manipulation of altruistic lying.

After participants learned about the information their partner knew as the Sender in the deception game, participants played a trust game with the Sender. We referred to the trust game as "Exercise 2" in the experiment. We ran a pilot study with a non-overlapping sample (*N* = 40) in order to generate real decisions with which to match the decisions of participants in our main study.

[10] Participants had to pass two comprehension checks, one for the deception game and one for the trust game, in order to complete the entire study. Participants who failed a comprehension check had the opportunity to reread the instructions for the exercise and retake the comprehension check. If any participant failed a comprehension check twice, they were not allowed to complete the study. We followed this procedure in every study.

[11] A total of 89% of participants actually chose the outcome indicated in their partner's message. Whether or not participants chose the outcome indicated in the message did not influence our results. That is, our results are not influenced by whether or not participants were successfully deceived.

---Table 1 about here---

The trust game. In our trust game, we assigned all participants to the role of Player 1 and told them that they would be paired with the Sender with whom they had just been paired with in Exercise 1 (the deception game). In our version of the trust game, Player 1 was given $1 and could make one of two choices: “Keep $1” or “Pass $1.” Choosing “Keep $1” led to a payout of $1 for Player 1 and $0 for his/her partner, Player 2. If Player 1 passed the $1, the money would grow to $3 and Player 2 could then either choose to “Keep $3” or “Return $1.50.”

Dependent variables.

Trusting behavior. The choice to pass money in the trust game served as our primary dependent variable. In addition, after making a decision, participants rated their partner’s trustworthiness, benevolence, and deception. For all attitudinal measures, we used 7-point Likert scales anchored at 1 = “Strongly disagree” and 7 = “Strongly agree”.

Attitudinal trust. We measured attitudinal trust with two items ($r(121) = .89$): “I trust Player 2 to RETURN money,” and “I am confident that Player 2 will RETURN money.”

Perceived benevolence. We measured perceived benevolence using three items ($\alpha = .80$): “This person is [kind, nice, and selfish (reverse-scored)].”

Perceived deception. We used three items to measure the extent to which our manipulation was recognized as deception ($\alpha = .84$): “This person sent an honest message as a Sender in Exercise 1” (reverse-scored), “This person lied about the outcome of the coin flip in Exercise 1,” and, “This person deceived his/her partner in Exercise 1.”

After participants submitted their responses, we asked two multiple-choice recall questions,[12] collected demographic information, and asked participants what they thought the purpose of the study was. Participants then received a bonus payment based upon their decisions.

Results

We report results from the 121 adults (45% female; M_{age} = 32years, *SD* = 9.77) who passed all comprehension checks and completed the entire study; 4 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

---Table 2 about here---

Trusting behavior. Supporting our thesis, participants were significantly more likely to trust a partner who told them an altruistic lie (56%), than a partner who was honest (32%), χ^2 (1, *N* = 121) = 6.88, *p* < .01. Figure 1 depicts these results.

Attitudinal Trust. Our attitudinal trust measure parallels our behavioral trust results. Participants reported that they trusted their partners more in the *Altruistic Lie* condition (*M* = 3.77, *SD* = 1.91) than in the *Selfish Honesty* condition (*M* = 2.72, *SD* = 1.76), *F*(1, 119) = 9.85, *p* < .01. Our behavioral and attitudinal measures of trust were highly correlated, *r*(121) = .89, *p* < .001, suggesting that passing decisions reflected trust beliefs.

[12] In every study, at least 80% of participants were able to recall the manipulation at the end of the study. For each study, we report analyses for the entire sample, but our results are unchanged when we restrict our sample to only those who answered the recall questions correctly.

Perceived Benevolence. Participants also perceived their partners to be more benevolent in the *Altruistic Lie* condition ($M = 4.19$, $SD = 1.55$) than in the *Selfish Honesty* condition ($M = 3.45$, $SD = 1.32$), $F(1, 119) = 8.12$, $p < .01$.

Perceived Deception. Consistent with our manipulation, participants also perceived their partners to be more deceptive in the *Altruistic Lie* condition ($M = 5.37$, $SD = 1.35$) than in the *Selfish Honesty* condition ($M = 2.88$, $SD = 1.34$), $F(1, 119) = 102.60$, $p < .001$.

Discussion

Consistent with our thesis, individuals trusted altruistic liars more than honest partners. Importantly, participants recognized that they had been deceived, but rated their counterparts as more benevolent and thus, more trustworthy. Study 1a provides initial evidence that deception can increase trust.

---Figure 1 about here---

Study 1b

In Study 1a, participants who were deceived directly benefitted from the deception. Their subsequent trust decisions may have been influenced by reciprocity. In Study 1b, we rule out reciprocity as an alternative explanation. In Study 1b, participants observe, rather than experience, deception. Individuals played a trust game with counterparts who either had or had not told an altruistic lie to a different partner in a previous interaction.

Method

Participants. We recruited 261 participants from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee.

Procedure and Materials. In this study, we randomly assigned participants to one of two conditions in a between-subjects design. Participants observed an individual who either told a prosocial lie or was selfishly honest and then played a trust game with this person.

We seated participants in separate cubicles to complete this study on the computer. The study was titled, "Partner Exercises." We told participants that they would complete two separate exercises with two separate partners. The first exercise, which we called "Exercise 1," was a deception game. Within the experiment, we called the second exercise, the trust game, "Exercise 2." Both games are similar to the games we used in Study 1a. In Study 1b, however, we matched participants with two different partners. Participants first completed the deception game and chose Heads or Tails. We paired participants with a new partner for the trust game. Participants did not learn about their own outcome in the deception game until they completed the entire study.

Manipulation of altruistic lies. We told participants that their partner in the trust game ("Exercise 2") had been matched with a different participant in the deception game ("Exercise 1") and had been assigned to the role of Sender. We then revealed the decision the Sender had made and the information they had prior to making that decision. As in Study 1a, by revealing the Sender's decision and the payments associated with their choice, participants learned that the Sender either told an altruistic lie or was selfishly honest.

The trust game. The trust game in Study 1b was similar to the trust game we used in Study 1a. We assigned every participant to the role of Player 1 and we matched each participant with a Player 2 who was the Sender in the first Exercise. In the trust game in Study 1b, participants started with $2. If Player 1 chose to "Pass $2" the money grew to $5. If Player 1 passed the money, Player 2 had the decision to either "Keep $5" or "Return $2.50." We used larger stakes in this study than those we used in Study 1a because our participants were university students, rather than Mechanical Turk participants.

Dependent variables. As in Study 1a, our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game. All of our other dependent variables were identical to those we collected in Study 1a ($r > .87$; α's > .80).

After participants submitted their responses, we asked two multiple-choice recall questions, collected demographic information, and asked participants what they thought the purpose of the study was. Participants then received bonus payment based on their decisions.

Results

We report the results from 257 participants (60.3% female; M_{age} = 20 years, *SD* = 2.30) who passed all comprehension checks and completed the entire study; 4 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Trusting behavior. Consistent with our prediction, participants were more likely to trust their partner when they learned that their partner had told someone else an altruistic lie (39%), than when they learned that their partner had told someone else the truth (21%), χ^2 (1, *N* = 257) = 9.79, *p* < .01. We depict these results in Figure 1.

Attitudinal Trust. As in Study 1a, our behavioral and attitudinal measures of trust followed the same pattern and were highly correlated, *r*(257) = .70, *p* < .001. Participants reported trusting their partners more in the *Altruistic Lie* condition (*M* = 3.51, *SD* = 1.71) than in the *Selfish Honesty* condition (*M* = 2.66, *SD* = 1.46), *F*(1, 255) = 18.04, *p* < .01.

Perceived Benevolence. Participants also perceived their partners to be more benevolent in the *Altruistic Lie* condition (*M* = 4.14, *SD* = 1.39) than in the *Selfish Honesty* condition (*M* = 3.75, *SD* = 1.07), *F*(1, 255) = 8.12, *p* = .01.

Perceived Deception. Consistent with our manipulation, participants also perceived their partners to be more deceptive in the *Altruistic Lie* condition (*M* = 4.91, *SD* = 1.45) than in the *Selfish Honesty* condition (*M* = 3.40, *SD* = 1.64), *F*(1, 255) = 60.18, *p* < .001.

Discussion

As in Study 1a, our participants trusted altruistic liars more than people who were selfishly honest. In this study, participants observed rather than experienced deception. Results from this study rule out direct reciprocity as an alternative explanation for our findings in Study 1a. Unlike Study 1a, participants in this study did not benefit from the act of deception.

Study 2

In Study 2, we extend our investigation by examining how different types of prosocial lies influence trust. In Studies 1a and 1b, we investigated altruistic lies. Because these lies were costly for the liar, it is possible that our findings reflect a desire to compensate liars for their altruism. We rule out this explanation in Study 2.

In Study 2, we demonstrate that our findings extend to prosocial lies that are not characterized by altruism. We explore how non-altruistic prosocial lies, lies that help the deceived party and have no effect on the liar, and mutually beneficial lies, lies that benefit the deceived party *and* the liar, influence trust.

Method

Participants. We recruited 300 adults to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. As in Study 1b, participants learned about the decisions an individual made as a Sender in a deception game and then played a trust game with that individual. In this study, we randomly assigned participants to one of four experimental conditions in a 2(Deception: Lie vs. Honesty) x 2(Type of lie: Prosocial vs. Mutually beneficial) between-subjects design. That is, participants learned the following about a Sender in the deception game: the Sender either lied or was honest; and lying either had no effect on the Sender and benefited the Receiver (i.e. was prosocial) or benefited both the Sender and the Receiver (i.e. was mutually beneficial).

In this study, participants first learned that they would play a trust game with a partner. We referred to the trust game as "The Choice Game" in the experiment. After participants learned about the trust game, but before they made any decisions, we told them that they would learn more information about their partner. Participants learned that their partner in the trust game had completed the trust game, along with another exercise, "The Coin Flip Game," in a previous study. "The Coin Flip Game" was the same deception game as the one we used in Studies 1a and 1b. Participants in this study, however, observed but did not play the deception game. That is, our participants did not have a chance to earn money before they played the trust game.

Manipulation of prosocial lies. We told participants that their partner in the trust game had been matched with a different participant in the deception game ("The Coin Flip Game") and had been randomly assigned to the role of Sender. We then explained the deception game and revealed the Sender's decision in that game.

In Study 2, we manipulated both the decision to lie and the type of lie that was told. In order to manipulate the type of lie, we manipulated the payments associated with Outcome A (Honesty) and Outcome B (Lying). When lying was prosocial, Outcome A yielded $2 for the Sender, $0 for the Receiver and Outcome B yielded $2 for the Sender, $1 for the Receiver. That is, this lie was prosocial, but not altruistic. When lying was mutually beneficial, Outcome A yielded $2 for the Sender, $0 for the Receiver and Outcome B yielded $2.25 for the Sender, $1 for the Receiver. We summarize the payments associated with each type of lie in Table 1.

Participants learned whether the Sender had been honest or had lied in the deception game, and whether or not lying was prosocial or mutually beneficial. Then, participants played the trust game with the Sender and rated the Sender.

The trust game. We referred to the trust game as “The Choice Game” in the experiment. The trust game we used in Study 2 was similar to the one we used in Study 1a and Study 1b. In this version of the trust game, however, participants played with lottery tickets rather than monetary outcomes. Using lottery tickets allowed us to increase the stakes on Mechanical Turk (a chance to win $25) and prevented participants from directly comparing outcomes in the deception game and the trust game.

In this trust game, we assigned participants to the role of Player 1 and matched them with the confederate Player 2 who had made decisions in “The Coin Flip Game.” In the trust game, Player 1 started with 4 lottery tickets. If Player 1 chose to “Keep 4 lottery tickets,” Player 1 earned 4 lottery tickets and Player 2 earned 0 lottery tickets. If Player 1 chose to “Pass 4 lottery tickets,” the number of tickets tripled to 12 tickets and Player 2 made the decision to either “Keep 12 lottery tickets” or “Return 6 lottery tickets.” Participants knew that the more tickets they had, the more likely they were to win the $25 lottery at the end of the study.

Dependent variables. Our main dependent variable was trusting behavior, measured by Player 1’s decision to pass the lottery tickets in the trust game. Our measures of trusting attitudes and perceived deception were identical to those we collected in Studies 1a and 1b (r > .93; α’s > .82). We modified our measure of perceived benevolence to include new items that were more specific: “This person is benevolent”, “This person would not purposefully hurt others”, “This person has good intentions” (α= .86). We used a 7-point Likert scale anchored at 1 = “Strongly disagree” and 7 = “Strongly agree.”

After participants submitted their responses, we asked two multiple choice recall questions, collected demographic information, and asked participants what they thought the purpose of the study was. We then told participants the number of lottery tickets they received as a result of their decision and their counterpart's decision in the trust game. We conducted the lottery the day the experiment ended.

Results

We report the results from 293 participants (39.9% female; M_{age} = 32 years, *SD* = 11.2) who passed the comprehension checks and completed the entire study; 7 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Trusting behavior. We first conducted a logistic regression on trusting behavior, using *Deception, Type of Lie,* and the *Deception* x *Type of Lie* interaction as independent variables. We found a main effect of *Deception* (*b* = .557, *p* < .01), such that participants were more trusting of individuals who told lies that helped others. Specifically, 63% of participants trusted partners who had lied, whereas only 37% of participants trusted partners who had been honest; χ^2 (1, *N* = 293) = 20.23, *p* < .01.

We found no main effect of *Type of Lie* and we found no significant *Deception* x *Type of Lie* interaction (*p*s > .32). Although lying had a directionally larger effect on trust when the prosocial lie was not mutually beneficial, this difference was not significant. In Figure 2, we display the percentage of participants who passed money in each of our four experimental conditions.

---Figure 2 about here---

Attitudinal Trust. As in Studies 1a and 1b, our behavioral and attitudinal measures of trust were highly correlated, *r*(293) = .73, *p* < .001 and follow the same pattern. A two-way ANOVA revealed a main effect of *Deception* on attitudinal trust, *F*(1,289) = 16.42, *p* < .001. Participants perceived their partner to be more trustworthy when they lied (*M* = 3.83, *SD* = 1.88) than when they had told the truth (*M* = 2.95, *SD* = 1.91). We do not find a main effect of *Type of Lie*, *F*(1,289) = .13, *p* = .71, nor do we find a significant *Deception* x *Type of Lie* interaction, *F*(1,289) = .34, *p* = .56.

Perceived Benevolence. A two-way ANOVA also revealed a main effect of *Deception* on perceived benevolence, *F*(1,289) = 16.42, *p* < .001. Participants perceived their partner to be more benevolent when they lied (*M* = 4.56, *SD* = 1.15) than when they told the truth (*M* = 3.63, *SD* = 1.33).

We also found a marginally significant *Deception* x *Type of Lie* interaction, *F*(1,289) = 3.28, *p* = .07. Lying had a greater effect on perceived benevolence when the lie was prosocial (M_{lie} = 4.73, SD_{lie} = 1.22 vs. M_{honesty} = 3.51, SD_{honesty} = 1.37), *t*(138) = 5.77, *p* < .001; than when the lie was mutually beneficial *(*M_{lie} = 4.44, SD_{lie} = 1.08 vs. M_{thonesty} = 3.75, SD_{honesty} = 1.29), *t*(153) = 3.43, *p* < .001. We do not find a main effect of *Type of Lie, F*(1,289) = .03, *p* = .86.

Perceived Deception. Consistent with our manipulation, participants perceived their partners to be more deceptive when their partner had lied (*M* = 5.39, *SD* = 1.24) than when they told the truth (*M* = 2.83, *SD* = 1.45), *F*(1,289) = 259.69, *p* < .001. We do not find a main effect of *Type of Lie*, *F*(1,289) = .01, *p* = .91, nor do we find a significant *Deception* x *Type of Lie* interaction, *F*(1,289) = 1.29, *p* = .26.

Discussion

In Study 2, we demonstrate that altruism is not a necessary condition for deception to increase trust. Prosocial lies that are not costly for the liar and prosocial lies that benefit the liar both increase trust. These results suggest that trusting behavior does not simply reflect a desire to compensate a liar for altruism. Rather, individuals trust people who help others, even when that help is self-serving and involves deception.

Although mutually beneficial lies are a weaker signal of benevolence than prosocial lies that do not benefit the deceiver, the self-serving nature of these lies did not undermine trust. These results suggest that for trust, judgments of benevolence may be more important than selflessness.

Study 3

Our initial studies demonstrate that prosocial lies can increase trust. In Studies 3a and 3b, we extend our investigation by independently manipulating deception and intentions (Study 3a) and by including two control conditions to disentangle the effects of selfishness from prosociality (Study 3b).

Study 3a

Method

Participants. We recruited 337 participants from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee.

Procedure and Materials. We seated participants in separate cubicles to complete the study on the computer. The study was titled, “Partner Exercise.” As in Study 2, participants learned about the decision a Sender made in a deception game and then played a trust game with that Sender. In Study 3a, we randomly assigned participants to one of four experimental conditions in a 2(Deception: Lie vs. Honesty) x 2(Intentions: Altruistic vs. Selfish) between-subjects design. Specifically, participants observed a Sender who either lied or sent an honest message in a deception game, and whose choice was either altruistic or selfish. Participants then played a trust game with this partner.

Manipulation of lies. The deception game in Study 3a was similar to the one we used in our prior studies. In this game, however, we used a random number generator rather than a coin flip to begin the game. The game was otherwise identical to the game we used in Study 2. That is, the payoffs for each pair of participants (one Sender and one Receiver) were determined by the outcome of a random number generator and the choices made by the Sender and the Receiver. Senders knew the correct number was 4, and could send an honest message (e.g., “The number is 4”) or a dishonest message (e.g., “The number is 5”). We used a random number generator rather than a coin flip so that participants would be less likely to make strategic inferences about the message the Sender sent (e.g., The Sender sent the message: “The coin landed on heads”, hoping their partner would pick “tails”).

Importantly, Senders in this experiment played The Number Game with one of two possible payment structures. These payment structures enabled us to manipulate whether deception or honesty was associated with selfish or altruistic intentions.

The first payment structure was identical to the one we used in Studies 1a and 1b. This payment structure represented the choice between selfish honesty (Option A) and altruistic lying (Option B). In the second payment structure, we reversed the payoffs. This payment structure represented the choice between altruistic honesty and selfish lying.

After learning about the Sender's choice in the deception game, participants played a trust game with the Sender. We ran a pilot study with a non-overlapping sample (*N*=41) to generate decisions with which to match the decisions participants made in Study 3a.

The Trust game. We referred to the trust game as "The Choice Game" in the experiment. "The Choice Game" was identical to the trust game we used in Study 1b. Participants had the choice to either "Keep $2" or trust their partner and "Pass $2."

Dependent variables.

As in Studies 1a, 1b, and 2, our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game. Our measures of attitudinal trust and benevolence were identical to the measures we used in Study 2 (*r*'s > .86, *α* =.91). We made a slight revision to our measure of perceived deception to fit the new version of the deception game. Specifically, we asked participants to indicate their agreement with the following statements: "This person sent an honest message about the number chosen by the random number generator as a Sender in The Number Game," and

"This person lied about the number chosen by the random number generator in The Number Game;" (*r*(312) = .86).

After participants submitted their responses, we asked them two recall questions, collected demographic information and asked participants what they thought the purpose of the study was. At the end of the study, we paid participants a bonus payment based upon their decisions in the trust game.

Results

We report the results from 312 participants (62.8% female; *M*age= 21 years, *SD* = 2.50) who passed all comprehension checks and completed the entire study; 25 participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Passing in the trust game. We first conducted a logistic regression on trusting behavior, using *Deception, Intentions,* and the *Deception* x *Intentions* interaction as independent variables. We found a main effect of *Intentions* (*b* = .498, *p* < .01), such that participants were more trusting of individuals who made altruistic decisions. Specifically, 47% of participants trusted their partners in the *Altruistic* conditions, whereas only 25% of participants trusted their partners in the *Selfish* conditions,χ^2 (1, *N* = 312) = 16.70, *p* < .01. We found no main effect of *Deception* and we found no significant *Deception* x *Intentions* interaction (*p*s > .79). In Figure 3, we display the percentage of participants who passed money in each of the four experimental conditions (*Altruistic Lie, Selfish Lie, Altruistic Honesty,* and *Selfish Honesty*).

---Figure 3 about here---

Attitudinal Trust. As in our previous studies, our behavioral and attitudinal measures of trust were highly correlated, *r*(312) = .72, *p* < .001. A two-way ANOVA revealed a main effect of *Intentions*, *F*(1,308) = 78.74, *p* < .001, such that participants trusted their partners more in the *Altruistic* conditions (*M* = 4.07, *SD* = 1.79) than they did in the *Selfish* conditions (*M* = 2.43, *SD* = 1.49).

Although lying did not significantly influence behavioral trust, it did influence attitudinal trust. We found a main effect of *Deception*, *F*(1,308) = 5.58, *p* = .02 on attitudinal trust, such that participants trusted their partner more in the *Honesty* conditions (*M* = 3.46, *SD* = 1.82) than in the *Lie* conditions (*M* = 3.05, *SD* = 1.85). We find no significant interaction between *Deception x Intentions*, *F*(1,308) = .19, *p* = .66.

Perceived Benevolence. A two-way ANOVA revealed a main effect of *Intentions*, *F*(1,308) = 108.70, *p* < .001, and *Deception*, *F*(1,308) = 18.90, *p* < .01, on perceived benevolence. Participants perceived their partner to be more benevolent in the *Altruistic* conditions (*M* = 4.82, *SD* = 1.22) than in the *Selfish* conditions (*M* = 3.42, *SD* = 1.21) and to be more benevolent in the *Honesty* conditions (*M* = 4.36, *SD* = 1.27) than in the *Lie* conditions (*M* = 3.89, *SD* = 1.49). We find no significant interaction between *Deception x Intentions*, *F*(1,308) = .76, *p* = .36.

Perceived Deception. Consistent with our manipulation, participants also perceived their partner to be more deceptive in the *Lie* conditions (*M* = 6.06, *SD* = 1.30) than in the *Honesty* conditions (*M* = 2.06, *SD* = 1.41), *F*(1,255) = 680.02, *p* < .001. We find no effect of *Intentions*, *F*(1,308) = 1.54, *p* = .22, and we find no significant *Deception* x *Intentions* interaction, *F*(1,308) = .28, *p* = .59.

Mediation Analyses.

We conducted a moderated mediation analysis using the bootstrap procedure (Hayes, 2013; Preacher, Rucker, & Hayes, 2007) to test the process by which lying and intentions influence trusting behavior.

We predicted that altruistic (and selfish) intentions would influence trusting behavior, regardless of whether the target lied, and that this would be mediated by perceived benevolence. Our mediation model included *Intentions* as the independent variable, *Deception* as the moderator variable, Perceived Benevolence and Perceived Deception as the mediator variables, and Trusting Behavior as the dependent measure. Consistent with our hypothesis, we find that Perceived Benevolence mediates in the expected direction in both the *Lie* conditions (Indirect Effect = 1.14, *SE* = .25; 95% CI [0.70, 1.67]), and the *Honesty* conditions (Indirect Effect = .97, *SE* = .23; 95% CI [0.58, 1.44]), and Perceived Deception does not mediate (both confidence intervals for the indirect effect include zero). These results are unchanged when we use Attitudinal Trust, rather than Trusting Behavior, as the dependent measure. Taken together, these results indicate that perceived benevolence, and not perceived deception, influences trust. That is, deception does not harm trust; selfishness does. We present additional regression analyses in Table 3.

---Table 3 about here---

Discussion

In Study 3a, Altruistic individuals were trusted far more than selfish individuals, and this was true whether or not the counterpart's claims were honest or deceptive. Controlling for intentions, we find no direct effect of lying on trusting behavior in either study. This is true even though lying is perceived as deceptive. We use moderated mediation analysis and confirm that perceived benevolence is the primary mechanism linking prosocial lying with increased trust. Interestingly, trust built on perceived benevolence is not diminished by dishonest acts.

Study 3b

In Study 3b, we extend our investigation by including two control conditions in our experiment. By including control conditions, we can disentangle the beneficial effects of altruistic behavior from the harmful effects of selfish behavior. In our control conditions, participants did not learn about the Sender's decision in the deception game.

Method

Participants. For our 12 cell design, we recruited 1000 participants to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. Study 3b was similar to Study 3a, with three notable changes. First, we added two control conditions to disentangle the effects of altruism in increasing trust from the effects of selfishness in decreasing trust. In the control conditions, participants did not learn about the Sender's decision in the deception game.

Second, for simplicity and ease of comprehension we used the Coin Flip game rather than the Number Game for our manipulation of deception. Third, we counterbalanced the order of our behavioral trust measure and our attitudinal trust measure.

In Study 3b, we randomly assigned participants to one of twelve experimental conditions in a 2(Payment Structure: Altruistic Lying-Selfish Honesty vs. Selfish Lying-Altruistic Honesty) x 3(Intentions: Altruistic, Selfish, Control) x 2(Order of measures: behavior first vs. attitudes first) between-subjects design. Participants learned that the Coin Flip Game had one of two possible payment structures. As in Study 3a, these payment structures enabled us to manipulate whether deception or honesty was associated with selfish or altruistic intentions. We used the same payment structures in this study as those we used in Study 3a. The first payment structure reflected the choice between Altruistic Lying and Selfish Honesty, and the second payment structure reflected the choice between Selfish Lying and Altruistic Honesty.

Therefore, participants learned that the Sender either made the *Altruistic* decision (which was associated with *Lying* or *Honesty*), made the *Selfish* decision (which was associated with *Lying* or *Honesty*), or participants did not learn the Sender's decision (the control conditions). Half of the participants in the control condition learned that the Coin Flip Game reflected the choice between altruistic lying and selfish honesty (the first payment structure) and half learned that the Coin Flip Game reflected the choice between selfish lying and altruistic honesty (the second payment structure).

We refer to these six experimental conditions as *Altruistic Lie, Selfish Lie, Altruistic Honesty, Selfish Honesty, Control 1* (learned about the Altruistic Lie-Selfish Honesty payment structure, but did not learn about the Sender's choice), and *Control 2* (learned about the Selfish Lie-Altruistic Honesty payment structure, but did not learn about the Sender's choice).

After participants learned about the Coin Flip Game [and the Sender's decision], participants played a trust game with the Sender.

The Trust Game. We referred to the trust game as "The Choice Game" in this experiment. "The Choice Game" was similar to the trust games we used in our previous studies. Participants had the choice to either "Keep $1" or trust their partner and "Pass $1" in the trust game. If participants passed $1, the amount grew to $2.50 and their partner had the opportunity to keep $2.50 or return half ($1.25).

As in our previous studies, participants had to pass a comprehension check to complete the study.

Dependent variables.

Our primary dependent variable was trusting behavior, measured by the decision to pass money in the trust game. Our measures of attitudinal trust, benevolence, and deception were identical to the measures we used in Study 3a (r = .93, *α's* > .88). However, we did not measure perceived deception in the control conditions because participants did not have any information about whether or not the Sender had deceived their partner.

After participants submitted their responses, we collected demographic information and asked participants what they thought the purpose of the study was. We paid participants a bonus payment based upon their outcome in the trust game before we dismissed them.

Results

We report the results from 974 participants (40.2% female; *M*age= 31 years, *SD* = 10.36) who passed the comprehension checks and completed the entire study; 26 participants failed the comprehension check at the start of the experiment and were automatically eliminated from the study. None of our main results are affected by question order, and we present our analyses collapsed across this factor. We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 2.

Passing in the trust game. We first conducted a logistic regression on trusting behavior, using *Payment Structure, Intentions,* and the *Payment Structure* x *Intentions* interaction as independent variables. In our logistic regression, we coded Intentions such that -1 = Selfish, 0 = Control, 1 = Altruistic. We coded Payment Structure such that Altruistic Lying-Selfish Honesty =1 and Selfish Lying- Altruistic Honesty = -1.

We found a main effect of *Intentions,* (*b* = .938, *p* < .001); participants were significantly more likely to pass money in the trust game in the *Altruistic* conditions (69%) than in the *Control* conditions (47%); χ^2 (1, *N* = 654) = 32.10, *p* < .01, and in the *Selfish* conditions (25%), χ^2 (1, *N* = 650) = 121.43, *p* < .01. Participants were also significantly more likely to trust their partner in the *Control* conditions than they were in the *Selfish* conditions, χ^2 (1, *N* = 644) = 32.53, *p* < .01.

We found no effects of *Payment Structure,* nor did we find a significant *Intentions* x *Payment Structure* interaction (*p*s > .86). In Figure 4, we display the percentage of participants who passed money in each of the six experimental conditions.

---Figure 4 about here---

Attitudinal Trust. As in our previous studies, behavioral and attitudinal measures of trust were highly correlated, *r*(974) = .71, *p* < .001, and followed a similar pattern. A two-way ANOVA revealed a significant main effect of *Intentions, F*(2,968) = 240.74, *p* < .001, such that participants trusted their partners more in the *Altruistic* conditions (*M* = 4.70, *SD* = 1.61) than the *Control* conditions (*M* = 3.22, *SD* = 1.78), *t*(653) = 11.86, *p* < .001; and the *Selfish* conditions (*M* = 1.96, *SD* = 1.37), *t*(649) = 21.94, *p* < .001. Participants were also more trusting of their partner in the *Control* conditions than in the *Selfish* conditions, *t*(643) = 10.00, *p* < .001.

We found no main effect of *Payment Structure*, *F*(1, 968) = 0.25, *p* = .62. There was, however, a significant *Intentions x Payment Structure* interaction, *F*(2, 968) = 4.30, *p* < .05. Participants trusted individuals who told selfish lies (*M* = 1.73, *SD* = 1.09) significantly less than individuals who were selfishly honest (*M* = 2.18, *SD* = 1.56), *t*(319) = 2.54, *p* = .01, but we found no difference in trust between individuals who told altruistic lies (*M* = 4.68, *SD* = 1.64) and individuals who were altruistically honest (*M* = 4.71, *SD* = 1.58), *t*(329) = 0.17, *p* = .87. We also found no difference in trust between the two control conditions (*M* = 3.08, *SD* = 1.66 vs. *M* = 3.35, *SD* = 1.89), *t*(323) = 1.53, *p* = .13. These results suggest that deception in the service of altruism does not undermine trust, but that deception in the service of selfishness does harm trust.

Perceived Benevolence. Perceived benevolence followed the same pattern as attitudinal trust. A two-way ANOVA revealed a significant main effect of *Intentions, F*(2,968) = 377.80, *p* < .001, such that participants perceived their partner to be more benevolent in the *Altruistic* conditions (*M* = 5.20, *SD* = 1.01) than they did in the *Control* conditions (*M* 4.29, *SD* = 0.98), *t*(653) = 11.18, *p* < .001, and the *Selfish* conditions (*M* = 2.98, *SD* = 1.20), *t*(649) = 27.24, *p* < .001. Participants also rated their partners as more benevolent in the *Control* conditions than they did in the *Selfish* conditions, *t*(643) = 16.20, *p* < .001.

We also found a main effect of *Payment Structure*, *F*(1, 968) = 20.01, *p* < .001; partners who faced the opportunity to tell altruistic lies were perceived to be more benevolent (*M* = 4.30, *SD* = 1.32) than were partners who faced the opportunity to tell selfish lies (*M* = 4.04, *SD* = 1.47). This effect was qualified by a significant *Intensions x Payment Structure* interaction, *F*(2, 968) = 17.03, *p* < .001. Participants rated partners who told selfish lies (*M* = 2.54, *SD* = 1.02) to be significantly less benevolent than partners who were selfishly honest (*M* = 3.39, *SD* = 1.22), *t*(319) = 7.28, *p* < .001, but we found no difference in perceived benevolence between partners who told altruistic lies (*M* = 5.25, *SD* = 1.07) and partners who were altruistically honest (*M* = 5.15, *SD* = 0.94), *t*(329) = 0.91, *p* = .36. In other words, selfish deception was perceived to be particularly malevolent. There was no difference in perceived benevolence between the two control conditions (*M* = 4.27, *SD* = 0.92 vs. *M* = 4.32, *SD* = 1.04), *t*(323) = 0.46, *p* = .65.

Perceived Deception. Consistent with our manipulation, a two-way ANOVA revealed a significant *Intentions x Payment Structure* interaction, *F*(1, 645) = 1611.15, *p* < .001, such that altruistic lies were perceived to be more deceptive (*M* = 5.17, *SD* = 1.33) than selfish honesty (*M* = 2.77, *SD* = 1.53), *t*(324) = 18.46, *p* < .001, and selfish lying was perceived to be more deceptive (*M* = 6.47, *SD* = 0.88) than altruistic honesty (*M* = 1.51, *SD* = 0.76), *t*(323) = 38.15, *p* = .001.

We also found a main effect of *Intentions*, *F*(1, 645) = 195.15, *p* < .001, such that selfishness was perceived to be more deceptive (*M* = 4.57, *SD* = 1.07) than altruism (*M* = 3.29, *SD* = 2.13). In other words, the same lie was perceived to be more deceptive when it was associated with selfish, rather than altruistic, intentions. We found no main effect of *Payment Structure*, *F*(1, 645) = 0.08, *p* = .78.

Discussion

In Study 3a, we demonstrate that deception itself has no effect on benevolence-based trust. In Study 3b, we include control conditions and document both a penalty for selfishness and a benefit for altruism. Selfish intentions, whether they were associated with honesty or deception, harmed trust; altruistic intentions, whether they were associated with honesty or deception, increased trust.

Although we find no differences between altruistic lies and altruistic honesty in Study 3b, we do find that selfish lies are penalized relative to selfish honesty. Individuals may perceive honesty as the default decision, whereas lying may reflect a willful departure that is more diagnostic of intentionality. In this case, lying to reap selfish benefits may convey a stronger signal of malevolent intentions than honesty that yields the same outcome.

Study 4

Our studies demonstrate that prosocial lies can increase trust. In Studies 1a, 1b, 2, 3a, and 3b, we measure trust using the trust game, and we conceptualized trust as the willingness to be vulnerable to another person when there is an opportunity for exploitation. In Study 3a we demonstrate that trust behavior and trust attitudes are mediated by perceptions of benevolence and are largely unaffected by deception. Taken together, our studies demonstrate that prosocial deception increases benevolence-based trust.

Benevolence-based trust characterizes some of our most important trust decisions (e.g., Kim et al., 2006). The decision to loan money or property to another person, the decision to rely on someone for emotional support, and the decision to share sensitive information with someone reflect benevolence-based trust (e.g., McAllister, 1995; Currall & Judge, 1995; Glaeser et al., 2000; Levin & Cross, 2004). Some trust decisions, however, reflect perceptions of integrity rather than benevolence.

Integrity-based trust reflects the belief that a trustee adheres to ethical principles, such as honesty and truthfulness (Butler & Cantrell, 1984; Mayer et al., 1995; Kim et al., 2004). Integrity-based trust characterizes trust decisions that reflect perceptions of veracity. For example, the decision to rely upon another person's advice or the information they provide reflects integrity-based trust. In fact, it is exactly this type of trust that Rotter reflects in his definition of trust (1971: p. 444): "a generalized expectancy…that the word, promise, verbal, or written statement of another individual or group can be relied on." For these types of trust decisions, expectations of honesty and integrity may matter more than benevolence. As a result, prosocial lies may decrease integrity-based trust. We explore this proposition in Study 4.

The *Rely-or-Verify* game.

We introduce a new trust game, the *Rely-or-Verify* game, to capture integrity-based trust. We designed the *Rely-or-Verify* game to reflect the decision to trust a counterpart's claim. For example, employers routinely face the decision of whether or not to trust a prospective employee's claim about their prior work experience. An employer could either trust the prospective employee's claim or verify the claim, at a cost. Similarly, negotiators, relational partners, and parents can either trust or verify the claims their counterparts make.

The decision to rely on another person's claim primarily reflects perceptions of integrity. That is, the decision to either rely upon or very another person's claim is fundamentally a judgment about the veracity of the claim: Is the target telling the truth? Perceptions of benevolence may also influence this judgment (e.g., judgments of *why* the target might or might not tell the truth), but perceptions of benevolence are likely to be of secondary import relative to perceptions of integrity.

The following features characterize the *Rely-or-Verify* game: First, the trustee derives a benefit from successful deception (e.g., by over-stating prior work experience). Second, the truster cannot distinguish deception from honesty without verifying a claim. Third, for the truster, relying on the trustee's claim is risky, and fourth, verifying a claim is costly.

In *Rely-or-Verify,* Player 1 (the trustee) makes a claim that is either accurate or inaccurate. Player 2 (the truster) observes the claim and decides to either **Rely** (trust) or **Verify** (not trust) the claim. If Player 1's claim is **inaccurate** and Player 2 **relies** on the claim, Player 1 earns $\boldsymbol{a_1}$ and Player 2 earns $\boldsymbol{a_2}$**.** If Player 1's claim is **inaccurate** and

Player 2 **verifie**s it, Player 1 earns $\boldsymbol{b_1}$ and Player 2 earns $\boldsymbol{b_2}$**.** If Player 1's claim is **accurate** and Player 2 **relies** on it, Player 1 earns $\boldsymbol{c_1}$ and Player 2 earns $\boldsymbol{c_2}$**.** If Player 1's claim is **accurate** and Player 2 **verifies** it, Player 1 earns $\boldsymbol{d_1}$ and Player 2 earns $\boldsymbol{d_2}$**.**

The payoffs for Player 1 are structured such that $\boldsymbol{a_1 > c_1 \geq d_1 > b_1}$. For Player 1, deception is risky; for Player 1, deception yields the highest payoff if Player 2 relies on the deceptive claim, but it yields the lowest payoff if Player 2 verifies the deceptive claim.

The payoffs for Player 2 are structured such that $\boldsymbol{c_2 > d_2 \geq b_2 > a_2}$. In other words, Player 2 earns the highest payoff for relying on accurate information and the lowest payoff for relying on inaccurate information. Verification is costly, but minimizes risk. By verifying information, Player 2 learns the truth. Thus, verification yields the same outcome for Player 2, regardless of whether or not Player 1 told the truth.

In the *Rely-or-Verify* game, Player 2 is always at least weakly better off when Player 1 sends accurate information. That is, sending accurate information is both honest and benevolent. Sending accurate information is also less risky for Player 1.Therefore, Player 1's motive for sending an honest message may include preferences for honesty, benevolence, and risk. We depict the general form of *Rely-or-Verify* in Figure 5.

---Figure 5 about here---

Pilot Study

We report results from a pilot study to demonstrate that trust decisions in *Rely-or-Verif*y reflect perceptions of trustworthiness and integrity. In our study, we term Player 1 the "Red Player" and Player 2 the "Blue Player." The Red Player sends a message to the

Blue Player. In this case, the Red Player reports whether or not the amount of money in a jar of coins is odd or even. The Blue Player (the truster) received this message and can either *Rely* on the message or *Verify* the message. In our study, the payoffs for Player 1 (Red Player) were: $\boldsymbol{a_1}$ = \$1.5, $>$ $\boldsymbol{c_1}$ = \$0.75 $\geq$ $\boldsymbol{d_1}$ = \$0.5 $>$ $\boldsymbol{b_1}$ = \$0; the payoffs for Player 2 (Blue Player) were: $\boldsymbol{c_2}$ = \$1.5 $>$ $\boldsymbol{d_2}$ = \$1 $\geq$ $\boldsymbol{b_2}$ = \$1 $>$ $\boldsymbol{a_2}$ = \$0.

With this payoff structure for the *Rely-or-Verify* game, there is no pure strategy equilibrium. However, there is a mixed strategy equilibrium in which Player 1 (Red Player) provides accurate information with probability 1/3 and Player 2 (Blue Player) relies on that information with probability 2/5. We use this equilibrium as a benchmark in Study 4; if participants are perfectly rational and risk-neutral, they would choose *Rely* 40% of the time. We provide the full instructions and the exact game we used in Appendix A; we include the solution for the game's equilibrium in Appendix B.

Participants. We recruited 198 participants from a city in the northeastern United States to participate in a pilot study of *Rely-or-Verify* in exchange for a \$10 show-up fee.

Method. Participants in the pilot study read the full instructions of the *Rely-or-Verify* game (see Appendix A) and were assigned to the role of the "Blue Player." Participants had to pass a comprehension check in order to complete the entire study. Participants who failed the comprehension check twice were automatically removed from the experiment.

Participants who passed the comprehension check received a message from a confederate "Red Player," informing them that the amount of money in the jar was either odd or even. The decision to *Rely* represents our behavioral measure of integrity-based trust.

After participants made a decision to *Rely* or *Verify*, they rated how much they trusted their partner, and they rated their partner's benevolence and integrity. We measured trusting attitudes using three items (α = .84): "I trust my partner," "I am willing to make myself vulnerable to my partner," and "I am confident that my partner sent me an accurate message;" 1 = "Strongly disagree" and 7 = "Strongly agree." We measured perceived benevolence using the same scale we used in Studies 3a and 3b (α = .78), and we measured perceived integrity using three items (α = .66): "This person has a great deal of integrity," "I can trust this person's word," and "This person cares about honesty and truth;" 1= "Strongly disagree" and 7 = "Strongly agree."

After participants made *Rely-or-Verify* decisions and rated their partner, they answered demographic questions, were paid, and dismissed.

Results. Nearly all of the participants (98%) passed the comprehension check and completed the entire study. A total of 31.3% of participants chose *Rely* and trusted their partner. This result suggests that without knowing any information about their counterpart, participants in the pilot study were relatively distrusting. They chose *Rely* less often than the mixed-strategy equilibrium would predict (40%). We did not identify any gender differences in behavior.

Importantly, the decision to *Rely* was closely related to perceptions of trustworthiness, $r(194) = .71$, $p < .001$. Trusting behavior in *Rely-or-Verify* was correlated with both perceived benevolence, $r(194) = .48$, $p < .001$, and perceived integrity *r*(194) = .52, *p* < .001. In our main study, we demonstrate that integrity is the primary driver of behavior in the *Rely-or-Verify* game.

Main Study

In our main study, participants learned about a counterpart who had either told prosocial lies or who had been honest in a series of prior interactions. After learning this information, participants played either the trust game or the *Rely-or-Verify* game with their counterpart.

Method

Participants. We recruited 500 participants to participate in an online study in exchange for payment via Amazon Mechanical Turk.

Procedure and Materials. Participants in Study 4 learned about a series of decisions a confederate counterpart made as a Sender in the Coin Flip Game. This was the same Coin Flip Game we used in Studies 1a, 1b, 2, and 3b. Participants then played either the trust game or the *Rely-or-Verify* game with this counterpart. We randomly assigned participants to one of four cells from a 2(Deception: Prosocial lie vs. Honesty) x 2(Game: Trust game vs. *Rely-or-Verify*) between-subjects design.

In Study 4, participants learned that the Sender had played the Coin Flip Game four times with four different partners. We altered the payoffs associated with deception in each of the four rounds of the game so that we could include both altruistic and mutually beneficial lies in a single manipulation. By using repeated behavior to manipulate prosocial deception, we strengthened our manipulation. This manipulation made it clear that the Sender was either committed to honesty (telling the truth even when it was costly for themselves) or to benevolence (helping the Receiver even when it was costly for themselves). Specifically, participants learned about four decisions the Sender had made in four rounds of The Coin Flip Game. In rounds 1 and 3, the Sender faced the choice between an altruistic lie and selfish honesty. In rounds 2 and 4, the Sender faced the choice between a mutually beneficial lie and mutually harmful honesty. Participants learned that the Sender made one of the following two sets of decisions: Prosocial Lies {Altruistic lie, mutually beneficial lie, altruistic lie, mutually beneficial lie} or Honesty {Selfish truth, mutually harmful truth, selfish truth, mutually harmful truth}. We include the payoffs associated with each choice in Table 4.

---Table 4 about here---

After participants learned about the Sender's four decisions, participants played either the trust game or the *Rely-or-Verify* game with the Sender. The trust game we used was identical to the version of the trust game we used in Study 3b. The version of the *Rely-or-Verify* game we used was identical to the version we used in the pilot study.

Dependent variables.

Our main dependent variable was trusting behavior, measured by the decision to pass money in the trust game (benevolence-based trust) or *Rely* in the *Rely-or-Verify* game (integrity-based trust). Our measures of attitudinal trust for *Rely-or-Verify* were identical to the measures we used in the pilot study. We adapted the wording of these items to create a parallel measure of attitudinal trust for the trust game (α= *.92*). We provide all of the items and anchors we used in this study in Appendix C.

We measured perceived deception with the same measures we used in our prior studies ($\alpha = .94$). We measured perceived benevolence as we did before, but to be sure to distinguish benevolence from integrity, we eliminated the item, "This person has good intentions;" $r(457) = .72$, $p < .001$.

After participants submitted their responses, we asked a recall question, collected demographic information, and asked participants what they thought the purpose of the study was. The next day, we followed up with participants to pay them a bonus payment based upon their decisions.

Results

We report results from 457 participants (31.6% female; M_{age}= 31 years, *SD* = 9.87) who passed all comprehension checks and completed the entire study; 43 participants failed the comprehension check and were automatically removed from the study.[13] We present the means and standard deviations of each of our scales, as well as the inter-scale correlation matrix in Table 5.

---Table 5 about here---

Trusting behavior. We first conducted a logistic regression on trusting behavior using *Deception, Game,* and the *Deception* x *Game* interaction as independent variables. We found no main effect of *Deception* or *Game* (*p*s > .73).

Importantly, we found a significant *Deception* x *Game* interaction; $b = .37$, $p < .01$, such that prosocial lying increased benevolence-based trust and harmed integrity-

[13] Participants dropped out of the experiment in the *Rely-or-Verify* game at a higher rate, because the comprehension check was more difficult to pass. Although we randomly assigned participants to condition, this resulted in uneven cell sizes.

based trust. Specifically, consistent with our prior studies, participants were more likely to pass money to their partners in the trust game in the *Prosocial Lie* condition (57%) than they were in the *Honesty* condition (40%), χ^2 (1, *N* = 262) = 7.41, *p* < .01. Importantly, we find the opposite pattern of results for behavior in the *Rely-or-Verify* game; participants were *less* likely to rely on their partners in the *Prosocial Lie* condition (37%) than they were in the *Honesty* condition (57%); χ^2 (1, *N* = 195) = 7.75, *p* < .01.

Notably, in the *Rely-or-Verify* game, participants in the *Honesty* condition were significantly more likely to rely on their partners than the equilibrium would predict (57% vs. 40%, one-sample test of proportion: *p* < .001) or than we observed in our pilot study (57% vs. 31%, one-sample test of proportion: *p* < .001). In this case, a history of honest behavior *increased* integrity-based trust. In contrast, behavior in the *Rely-or-Verify* game in the *Prosocial Lie* condition did not differ from the equilibrium prediction (37% vs. 40%, one-sample test of proportion: *p* = .59) or the behavior we observed in our pilot study (37% vs. 31%, one-sample test of proportion: *p* = .17). We depict these results in Figure 6.

Figure 6 about here

Attitudinal Trust. Results from our attitudinal trust measures parallel the results from our behavioral measures. Trusting attitudes were highly correlated with trusting behavior in both games, each $r \geq .80$ (see Table 5).

A two-way ANOVA revealed a significant *Deception* x *Game* interaction, *F*(1,453) = 17.57, *p* < .001, such that prosocial lying increased trusting attitudes in the trust game, but decreased trusting attitudes in the *Rely-or-Verify* game.

Specifically, participants trusted the prosocial liar more than the honest individual in the *Trust game* conditions (*M* = 4.11, *SD* = 2.08 vs. *M* = 3.54, *SD* = 1.86), *t*(261) = 2.48, *p* = .014, but trusted the prosocial liar less than the honest individual in the *Rely-or-Verify* conditions (*M* = 3.57, *SD* = 1.79 vs. *M* = 4.46, *SD* = 1.56), *t*(194) = 3.38, *p* < .01. We did not find a significant main effect of *Deception*, *F*(1,453) = 1.21, *p* = .27, or *Game*, *F*(1,453) = .89, *p* = .34.

Perceived Benevolence. Ratings of perceived benevolence followed a similar pattern. A two-way ANOVA revealed a significant *Deception* x *Game* interaction, *F*(1,453) = 5.93, *p* = .015, but no main effect of *Deception, F*(1,453) = 1.89, *p* = .17, or *Game*, *F*(1,453) = .15, *p* = .70. Specifically, participants judged the prosocial liar to be more benevolent than the honest individual in the *Trust game* conditions (*M* = 4.72, *SD* = 1.74 vs. *M* = 4.16, *SD* = 1.53), *t*(261) = 2.92, *p* < .01, but there was no difference between the prosocial liar condition and the honest condition in the *Rely-or-Verify* game (*M* = 4.30, *SD* = 1.51 vs. *M* = 4.46, *SD* = 1.32), *t*(194) = 0.70, *p* = .48. It is possible that individuals did not rate the prosocial liar as more benevolent in the *Rely-or-Verify* game because of the nature of the game. Decisions in the *Rely-or-Verify* game reflect both benevolence and honesty, and playing the *Rely-or-Verify* game may have caused participants to perceive honest individuals as more benevolent.

Perceived Deception. As expected, individuals who told prosocial lies were perceived to be more deceptive (*M* = 5.81, *SD* = 1.17) than individuals who were honest (*M* = 1.75, *SD* = 1.11), *F*(1, 453) = 1393.2, *p* < .001. We did not find a main effect of

Game, *F*(1,453) = .60, *p* = .44, or a significant *Deception* x *Game* interaction, *F*(1,453) = .04, *p* = .84.

Discussion

Results from this study demonstrate that prosocial lies differentially affect benevolence-based and integrity-based trust. We find that relative to a history of honesty, a history of prosocial deception increases trust rooted in benevolence, but harms trust rooted in integrity.

The prevailing behavioral measure of trust, the trust game, reflects benevolence-based trust. To measure integrity-based trust, we introduce a new tool, the *Rely-or-Verify* game. Although trustworthy behavior in the *Rely-or-Verify* game reflects perceptions of both honesty and benevolence, the trust decisions we observed were significantly more sensitive to signals of honesty than they were to signals of benevolence. We believe that this finding reflects the nature of the trusting decision in the *Rely-or-Verify* game; in this game, the decision to trust reflects beliefs about the veracity of the claim.

It is possible, however, that with different payoffs or different signals of benevolence and integrity, perceptions of benevolence could play a more significant role in trust behavior. Future research should explore how decisions in the *Rely-or-Verify* game change as a function of prior behavior, incentives, and perceptions of benevolence.

General Discussion

Across our studies, we demonstrate that lying can increase trust. In particular, we find that prosocial lies, false statements told with the intention of benefitting others, increase benevolence-based trust. In Study 1a, participants trusted counterparts more when the counterpart told them an altruistic lie than when the counterpart told the truth. In Study 1b, we replicate this result and rule out direct reciprocity as an alternative mechanism. In Study 1b, participants observed, rather than experienced deception.

In Studies 2, 3a, and 3b, we examine different types of lies. We find that participants trusted individuals who told non-altruistic, prosocial lies and mutually beneficial lies more than individuals who told truths that harmed others. Our findings reveal that benevolence, demonstrating concern for others, can be far more important for fostering trust than either honesty or selflessness. In fact, we find that deception per se, does surprisingly little to undermine trust behavior in the trust game.

In Study 4, we investigate how prosocial lying influences distinct types of trust. We introduce a new game, the *Rely-or-Verify* game to capture integrity-based trust. We demonstrate that the same actions can have divergent effects on benevolence-based and integrity-based trust. Specifically, we find that relative to honesty, prosocial lying increases benevolence-based trust, but harms integrity-based trust.

Contributions and Implications

In prior trust research, scholars have singled out deception as particularly harmful for trust. This work, however, has conflated deception with self-serving intentions. We find that although deception can exacerbate the negative inferences associated with selfish actions, deception does not undermine the positive inferences associated with prosocial actions. Our findings demonstrate that the relationship between deception and trust is far more complicated than prior work has assumed. Lying, per se, does not always harm trust.

Our research contributes to the deception and trust literatures in three ways. First, we highlight the importance of studying a broader range of deceptive behaviors. Prosocial lying is pervasive, but we know surprisingly little about the interpersonal consequences of prosocial lies. Although most research assumes that deception is harmful, we document potential benefits of deception. By signaling benevolence, prosocial lies can increase trust and may also afford other inter-personal benefits.

Second, we provide insight into the antecedents of trust. Trust scholars have assumed that both integrity and benevolence are antecedents of trust, yet little research has investigated when each of these values matters. Our research suggests that benevolence may be the primary concern for many—but not all— trust decisions. We are the first to independently manipulate benevolence and honesty and draw causal inferences about how they each impact trust.

Third, we demonstrate that identical actions can have divergent effects on different trust decisions. Scholars have used the term "trust" to refer to a broad range of behaviors. For example, trust has been used to describe the willingness to hire someone (Kim et al., 2004), to give someone responsibility without oversight (Kim et al., 2004; Mayer & Davis, 1999), to rely on someone's word (Johnson-George & Swap, 1982; Rotter, 1971), and to expose oneself to financial risk (Berg et al., 1995; Pillutla et al., 2003; Schweitzer et al., 2006; McCabe et al., 2003; Glaeser et al., 2000; Malhotra & Murninghan, 2002; Malhotra, 2004). Our findings suggest that different types of trust may guide these decisions, and that the same background information may influence these decisions in very different ways.

Our research has both methodological and managerial implications. Methodologically, we introduce a new tool to measure trust. Prior research has relied on the trust game, a tool that measures benevolence-based trust. Although benevolence-based trust underscores many trust decisions, in some trust decisions perceptions of integrity may be more important than benevolence. The *Rely-or-Verify* game provides scholars with a tool to measure integrity-based trust and offers several distinct advantages over the traditional trust game. For example, in contrast with the trust game in which the truster moves first, the truster in the *Rely-or-Verify* game moves second. By moving second, the *Rely-or-Verify* game eliminates alternative motivations for engaging in what might appear to be trusting behavior. For example, by moving first, trusters in the trust game may pass money for strategic reasons, such as to engender reciprocity (Chou, Halevy, & Murnighan, 2011), or for social preferences reasons, such as to promote fairness or altruism (e.g., Ashraf et al., 2006).

Prescriptively, our findings suggest that we should reconsider how we characterize deception. Parents, leaders and politicians often publicly and emphatically denounce lying—even though they often engage in it (Nyberg, 1993; Heyman, Luu, Lee, 2009; Grover, 2005). Acknowledging the benefits of prosocial lies could free individuals of (at least some of) this hypocrisy. In fact, authority figures could explicitly embrace certain types of deception and teach others when and how to lie. This would reflect a stark contrast to the current practice of asserting that lying is universally wrong, while modeling that it is often right.

Managers should also consider if honesty is always the best policy. Honesty, although often considered a virtue, in some cases may be selfish and mean-spirited. In many conversations, individuals make a trade-off between being honest and being kind. In order to engender trust, sometimes benevolence may be far more important than honesty.

Limitations and Future Directions

In our studies, we experimentally manipulated behavior in the deception game, which afforded us experimental control. By altering the monetary payoffs associated with honesty and lies, we were able to send unambiguous signals about the intentions associated with each lie. This enables us to draw causal inferences about how prosocial intentions and deception differentially influence distinct forms of trust. Consistent with prior research (e.g. Bracht & Feltovich, 2009), we find that information about a potential trustee's past behavior dramatically influences trust.

However, many prosocial lies are characterized by features that we did not capture in our experiments. For example, we study lies that generated monetary gains. Although some lies generate monetary outcomes, many lies, and prosocial lies in particular, are motivated by the desire to protect people's feelings (DePaulo, 1992). These lies may be perceived to be more innocuous and be more likely to foster emotional security, an important component of trust in close relationships (Rempel et al., 1985). Furthermore, lies told to avoid losses may be perceived to be more benevolent than lies told to accrue gains. Avoiding a loss is often much more psychologically powerful than generating a gain (Kahneman & Tversky, 1979), and thus, deceived parties may be particularly grateful to be the beneficiaries of these types of lies.

In our studies, the motives and outcomes associated with deception were clear. In practice, however, both motives and the link between acts and outcomes may be difficult to gauge. In some cases, people may even attribute selfish motives to prosocial acts (Critcher & Dunning, 2011; Fein, 1996; Newman & Cain, 2014; Miller, 1999; Lin-Healy & Small, 2013). For example, Wang and Murnighan (2013) found that some lies told to help others, such as a lie told to a medical patient, can be perceived to be low in benevolence and can harm trust, even when the intentions were prosocial.

Our experiments were also free of social context. Although this feature of our investigation enables us to draw clear casual inferences, future work should explore prosocial lies within richer social contexts. It is possible that the effects we observe will be moderated by situational norms, existing relationships, and prior experience. Another critical factor that is likely to influence perceptions of prosocial lies is the target's ability to change and adapt following critical feedback. For example, a husband who tells his wife that she looks great in an unflattering dress may appear benevolent when his wife has no alternative dresses to wear (e.g., out on vacation). However, if the husband is merely impatient and the wife could easy change clothes, this same lie may appear far less benevolent. Importantly, targets, observers, and deceivers may judge the benevolence of the same lie very differently.

The relative importance of benevolence and honesty may also change over time. For example, in early stages of relationship development, emotional security may be a primary concern, and prosocial lying may be particularly beneficial. In late stages of relationships, honesty may be a stronger signal of intimacy than kindness. Perhaps as relationships develop, the role of prosocial lying will change. It is also possible that prosocial lies have detrimental long-term consequences. If an individual develops a reputation for dishonesty, prosocial lies may become less credible. We call for future work to explore the dynamic interplay between trust and prosocial lies.

It is possible that our attitudes towards deception do not reflect intrinsic preferences for honesty and truth, but instead reflect our expectations of different relational partners. We may expect people in some roles to support and help us, but expect others to be objective and provide us with accurate information. Understanding how the nature of prosocial deception and trust differs across relationships is an important next step for trust research.

Gender and power may also influence our preferences for honesty and kindness. For example, women tell more prosocial lies than men (Erat & Gneezy, 2012) and are generally expected to be more polite than men (Brown & Levinson, 1987). Although we identified no gender differences in our studies, there may be circumstances in which women suffer greater backlash for impolite honesty than men. This may also be the case for low-power individuals who are expected to conform to politeness norms (Brown & Levinson, 1987). Sanctions for impolite honesty may have detrimental consequences in organizations by curbing the flow of information and curtailing employee voice.

Conclusion

We challenge the assumption that deception harms trust. Prior studies of deception have confounded lying with selfish intentions. By disentangling the effects of intentions from deception, we demonstrate that the relationship between deception and trust is far more complicated than prior work has assumed. Although prosocial lies harm integrity-based trust, prosocial lies *increase* benevolence-based trust. In many cases, intentions matter far more than veracity.

References

Ashraf, N., Bohnet, I., & Piankov, N. (2006). Decomposing trust and trustworthiness. *Experimental Economics*, *9*(3), 193-208.

Atwater, L. E. (1988). The relative importance of situational and individual variables in predicting leader behavior: The surprising impact of subordinate trust. *Group & Organization Management*, *13*(3), 290-310.

Baccara, M., & Bar-Isaac, H. (2008). How to organize crime. *The Review of Economic Studies*, *75*(4), 1039-1067.

Barney, J. B., & Hansen, M. H. (1994). Trustworthiness as a source of competitive advantage. *Strategic Management Journal*, *15*(S1), 175-190.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, *91,* 3-26.

Bazerman, M. H. 1994. Judgment in managerial decision making. New York: Wiley.

Bracht, J., & Feltovich, N. (2009). Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics*, *93*(9), 1036-1044.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142.

Bhattacharya, R., Devinney, T. M., & Pillutla, M. M. (1998). A formal model of trust based on outcomes. *Academy of Management Review*, *23*(3), 459-472.

Bies, R. J., & Tripp, T. M. (1996). Beyond distrust: “Getting even” and the need for revenge. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations: Frontiers of theory and research:* 246-260. Thousand Oaks, CA: Sage.

Blau, P. M. (1964). *Exchange and power in social life.* Piscataway, NJ: Transaction

Publishers.

Bok, S. (1978). *Lying: Moral choices in public and private life.* New York, NY: Pantheon.

Boles, T. L., Croson, R. T., & Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes*, *83*(2), 235-259.

Bowles, S., & Gintis, H. (2004). Persistent parochialism: trust and exclusion in ethnic networks. *Journal of Economic Behavior & Organization*, *55*(1), 1-23.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage.* New York, NY: Cambridge University Press.

Broomfield, K. A., Robinson, E. J., & Robinson, W. P. (2002). Children's understanding about white lies. *British Journal of Developmental Psychology, 20*(1), 47-65.

Butler Jr, J. K., & Cantrell, R. S. (1984). A behavioral decision theory approach to modeling dyadic trust in superiors and subordinates. *Psychological Reports*, *55*(1), 19-28.

Butler, J. K. (1991). Toward understanding and measuring conditions of trust: Evolution of conditions of trust inventory. *Journal of Management*, *17*(3), 643-663

Carr, A. (1968). Is business bluffing ethical?. *Harvard Business Review, 46*, 143-155.

Chou, E. Y., Halevy, N., & Murnighan, J. K. (2011, June). Trust as a tactic: The calculative induction of reciprocity. In *IACM 24th Annual Conference Paper*.

Cohen, T. R., Gunia, B. C., Kim-Jun, S. Y., & Murnighan, J. K. (2009). Do groups lie more than individuals? Honesty and deception as a function of strategic self-interest. *Journal of Experimental Social Psychology, 45*(6), 1321-1324.

Critcher, C. R., & Dunning, D. (2011). No good deed goes unquestioned: Cynical reconstruals maintain belief in the power of self-interest. *Journal of Experimental Social Psychology, 47,* 1207–1213.

Croson, R., Boles, T., & Murnighan, J. K. (2003). Cheap talk in bargaining experiments: Lying and threats in ultimatum games. *Journal of Economic Behavior & Organization*, *51*(2), 143-159.

Currall, S. C., & Judge, T. A. (1995). Measuring trust between organizational boundary role persons. *Organizational behavior and Human Decision processes*, *64*(2), 151-170.

DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, *111*(2), 203.

DePaulo, B. M., & Bell, K. L. (1996). Truth and investment: Lies are told to those who care. *Journal of Personality and Social Psychology, 71*(4), 703-716.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology, 74*(1), 63-79.

Dirks, K. T. (2000). Trust in leadership and team performance: Evidence from NCAA basketball. *Journal of applied psychology*, *85*(6), 1004-1012.

Dirks, K. T., & Ferrin, D. L. (2001). The role of trust in organizational settings. *Organization Science*, *12*(4), 450-467.

Dunn, J., Ruedy, N. E., & Schweitzer, M. E. (2012). It hurts both ways: How social comparisons harm affective and cognitive trust. *Organizational Behavior and Human Decision Processes*, *117*(1), 2-14.

Erat, S., & Gneezy, U. (2012). White lies. *Management Science*, *58*(4), 723-733.

Fein, S. (1996). Effects of suspicion on attributional thinking and the correspondence bias. *Journal of Personality and Social Psychology, 70*, 1164–1184.

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.

Ford, C. V, King, B.H., & Hollender, M.H. (1988). Lies and liars: Psychiatric aspects of prevarication. *American Journal of Psychiatry, 145*(5), 554-562.

Giffin, K. 1967. The contribution of studies of source credibility to a theory of interpersonal trust in the communication department. *Psychological Bulletin, 68*(2), 104-120.

Gillespie, N., & Dietz, G. (2009). Trust repair after an organization-level failure. *Academy of Management Review*, *34*(1), 127-145.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization*, in press.

Gino, F., & Pierce, L. (2009). Dishonesty in the name of equity. *Psychological Science, 20*(9), 1153-1160.

Gino, F., & Pierce, L. (2010a). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science*, *21*(6), 1176-1194.

Gino, F., & Pierce, L. (2010b). Robin Hood under the hood: Wealth-based discrimination in illicit customer help. *Organization Science*, *21*(6), 1176-1194.

Gino, F., & Shea, C. (2012). Deception in negotiations: The role of emotions. *Handbook of conflict resolution. Oxford University Press, New York*.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., & Soutter, C. L. (2000). Measuring trust. *The Quarterly Journal of Economics*, *115*(3), 811-846.

Gneezy, U. (2005). Deception: The role of consequences. *The American Economic*

Review, *95*(1), 384-394.

Goffman, E. (1967). On face-work. *Interaction ritual*, 5-45.

Golembiewski, R. T., & McConkie, M. (1975). The centrality of interpersonal trust in group processes. In C. L. Cooper (Ed.), *Theories of group processes* (pp. 131-185). London, UK: Wiley.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, *91*(3), 481-510.

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101-124.

Grover, S. L. (2005). The truth, the whole truth, and nothing but the truth: The causes and management of workplace lying. *The Academy of Management Executive*, *19*(2), 148-157.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Heyman, G. D., Luu, D. H., & Lee, K. (2009). Parenting by lying. *Journal of Moral Education*, *38*(3), 353-369.

Hosmer, L. T. (1995). Trust: The connecting link between organizational theory and philosophical ethics. *Academy of Management Review*, *20*(2), 379-403.

Houser, D., Schunk, D., & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, *74*(1), 72-81.

Iezzoni, L. I., Rao, S. R., DesRoches, C. M., Vogeli, C., & Campbell, E. G. (2012). Survey shows that at least some physicians are not always open or honest with patients. *Health Affairs*, *31*(2), 383-391.

Jones, E. E., & Wortman, C. (1973). *Ingratiation: An attributional approach*. Morristown, NJ: General Learning Press.

Johnson-George, C., & Swap, W. C. (1982). Measurement of specific interpersonal trust: Construction and validation of a scale to assess trust in a specific other. *Journal of Personality and Social Psychology, 43*(6), 1306-1317.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Karlan, D. S. (2005). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, *95*(5), 1688-1699.

Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence-vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, *99*(1), 49-65.

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. *Journal of applied psychology*, *89*(1), 104.

Koning, L., Steinel, W., Beest, I. V., & van Dijk, E. (2011). Power and deception in ultimatum bargaining. *Organizational Behavior and Human Decision Processes,115*(1), 35-42.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107*, 34-47.

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science,50*(11), 1477-1490.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology,* doi: 10.1016/j.jesp.2014.03.005

Lewicki, R. J., & Bunker, B. B. (1995). *Trust in relationships: A model of development and decline*. Jossey-Bass.

Lewicki, R. J., Tomlinson, E. C., & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991-1022.

Lewis, M. & Saarni, C. (1993). *Lying and deception in everyday life.* New York, NY: The Guilford Press.

Lin-Healy, F., & Small, D. A. (2013). Nice guys finish last and guys in last are nice: The clash between doing well and doing good. *Social Psychological and Personality Science, 4* (6), 693-699.

Lount Jr, R. & Pettit, N. (2012). The social context of trust: The role of status. *Organizational Behavior and Human Decision Processes*, *117*(1), 15-23.

Lount, R. B., Zhong, C. B., Sivanathan, N., & Murnighan, J. K. (2008). Getting off on the wrong foot: The timing of a breach and the restoration of trust. *Personality and Social Psychology Bulletin*, *34*(12), 1601-1612.

Malhotra, D. (2004). Trust and reciprocity decisions: The differing perspectives of trustors and trusted parties. *Organizational Behavior and Human Decision Processes*, *94*(2), 61-73.

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, 534-559.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*,*84*(1), 123.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.

McAllister, D. J. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24-59.

McCabe, K. A., Rigdon, M. L., & Smith, V. L. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behavior & Organization*, *52*(2), 267-275.

Meier, S., Pierce, L., & Vaccaro, A. (2013). Trust and Parochialism in a Culture of Crime. Washington University of St. Louis: Working Paper.

Miller, D. T. (1999). The norm of self-interest. *American Psychologist*, *54*(12), 1053-1060.

Murnighan, J. K. (1991). *The dynamics of bargaining games*. Englewood Cliffs, NJ: Prentice Hall.

Newman, G., & Cain, D. M. (2014). Tainted Altruism: When doing some good is worse than doing no good at all. *Psychological Science*, *25*(3), 648-655.

Nyberg, D. (1993). *The varnished truth: Truth telling and deceiving in ordinary life*. Chicago: University of Chicago Press.

O'Connor, K. M., & Carnevale, P. J. (1997). A nasty but effective negotiation strategy: Misrepresentation of a common-value issue. *Personality and Social Psychology Bulletin*, *23*(5), 504-515.

Pillutla, M., Malhotra, D., & Murnighan, J. K. (2003). Attributions of trust and the calculus of reciprocity. *Journal of Experimental Social Psychology*, 39, 448-455.

Planalp, S., Rutherford, D. K., & Honeycutt, J. M. (1988). Events that increase uncertainty in personal relationships II: Replication and extension. *Human Communication Research*, *14*(4), 516-547.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, *42*(1), 185-227.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology*, *49*(1), 95-112.

Robinson, S. L. (1996). Trust and breach of the psychological contract. *Administrative Science Quarterly, 41*(4), 574-599.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American psychologist*, *26*(5), 443-452.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393-404.

Santoro, M. A., & Paine, L. S. (1993). Sears auto centers (Harvard Business School case 9-394-010). *Boston: Harvard Business School Publishing*.

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review*, *32*(2), 344-354.

Schweitzer, M. E., & Croson, R. (1999). Curtailing deception: The impact of direct questions on lies and omissions. *International Journal of Conflict Management*, *10*(3), 225-248.

Schweitzer, M. E. & Gibson, D. E. (2008). Fairness, feelings, and ethical decision-making: Consequences of violating community standards of fairness. Journal of

Business Ethics, 77(3), 287-301.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational Behavior and Human Decision Processes, 101*(1), 1-19.

Sheppard, B. H., & Sherman, D. M. (1998). The grammars of trust: A model and general implications. *Academy of Management Review, 23*(3), 422-437.

Sitkin, S. B., & Roth, N. L. (1993). Explaining the limited effectiveness of legalistic “remedies” for trust/distrust. *Organization Science*, *4*(3), 367-392.

Steinel, W., & De Dreu, C. K. (2004). Social motives and strategic misrepresentation in social decision making. *Journal of Personality and Social Psychology*, *86*(3), 419.

Talwar, V., Murphy, S. M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development, 31*(1), 1-11.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451-471.

Tyler, J. M. & Feldman, R. S. (2006). Truth, lies, and self-presentation: How gender and anticipated future interaction relate to deceptive behavior. *Journal of Applied Social Psychology*, *34*(12), 2602-2615.

Tyler, J. M., Feldman, R. S., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Social Psychology*, *42*(1), 69-77.

Tyler, J. M. & Feldman, R. S. (2004). Cognitive demand and self-presentation efforts: The influence of situational importance and interaction goal. *Self and Identity*, *3*(4), 364-377.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

Valley, K. L., Moag, J., & Bazerman, M. H. (1998). A matter of trust: Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior & Organization, 34*(2), 211-238.

Wang, L. & Murnighan, J.K. (2013). Trust, White Lies, and Harsh Truths. Working Paper. City University of Hong Kong.

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior*, *26*, 75-101.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, *115*(2), 157-168.

APPENDIX

Appendix A. Materials for *Rely-or-Verify* (Study 4)

Instructions to participants in Study 4:

As **BLUE Player**, you will report whether the amount of money in the jar of coins below is ODD or EVEN:

If you correctly report whether the amount of money in the jar is ODD or EVEN, you will earn $1.50. If you inaccurately report if the amount of money in the jar is ODD or EVEN, you will earn $0.
Your partner is **RED Player. RED Player's** bonus payment also depends on your choice.
[page break]
You will be asked to report whether the amount of money in this jar of coins is ODD or EVEN.

1. Before you do, you will receive a message from **RED Player.**

- **RED Player** knows the amount of money in the jar.
- **RED Player** will send you one of the following two messages: "The amount of money in the jar is ODD." or "The amount of money in the jar is EVEN."
- The message that **RED Player** sends could be *accurate* or *inaccurate.*
- *You will not know whether or not the message is accurate or inaccurate when you make the choice to **RELY** on the message or **VERIFY** the message.*

2. Once you've received the message from **RED Player,** you can choose to **RELY** on **RED Player's** message, or you can **VERIFY** the message.

3. If you choose **RELY**, you will be paid based on whether or not **RED Player** gave you *accurate* or *inaccurate* information.

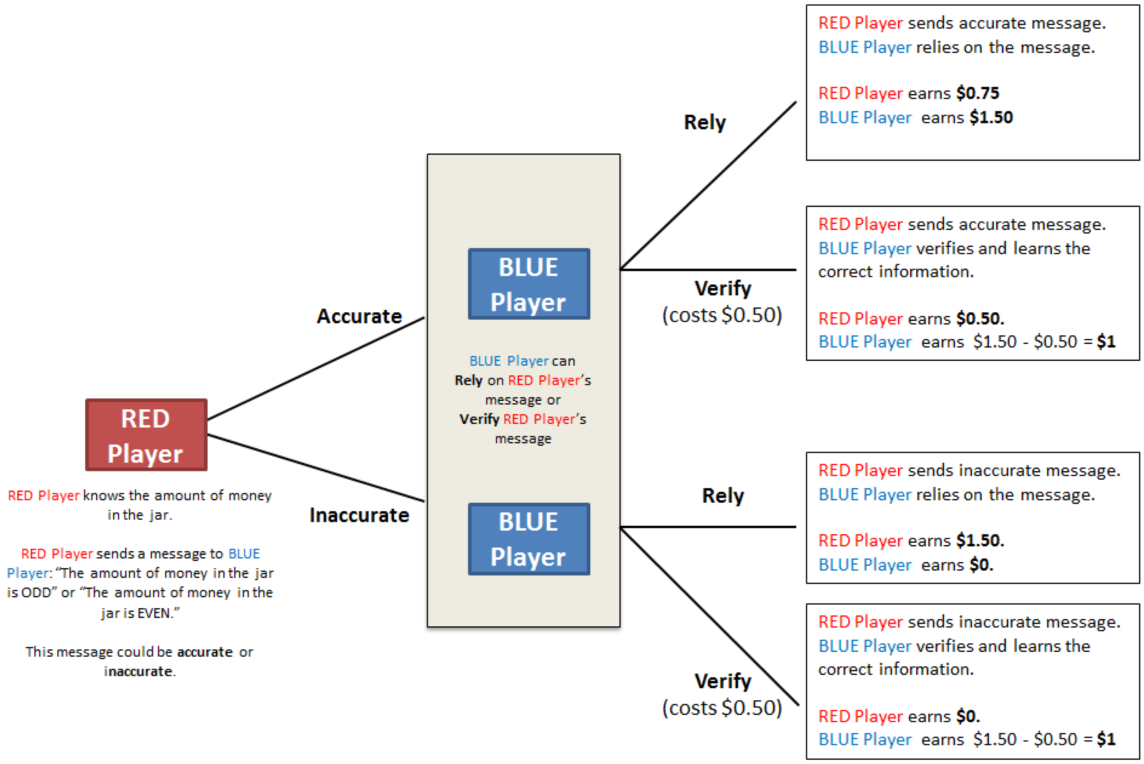
- If **RED Player** gave you *accurate* information and you **RELY** on it, you will correctly report whether the amount of money in the jar is ODD or EVEN, and you will earn $1.50. **RED Player** will earn $0.75.

- If **RED Player** gave you *inaccurate* information and you **RELY** on it, you will incorrectly report whether the amount of money in the jar is ODD or EVEN, and you will earn $0. **RED Player** will earn $1.50.

4. If you choose **VERIFY**, $0.50 will be subtracted from your total earnings and you will learn the correct amount of money in the jar.

- If **RED Player** gave you *accurate* information and you **VERIFY** it, you will earn $1 ($1.50 for the correct answer - $0.50 cost of verification) and **RED Player** will earn $0.50.
- If **RED Player** gave you *inaccurate* information and you **VERIFY** it, you will earn $1 ($1.50 for the correct answer - $0.50 cost of verification) and **RED Player** will earn $0.

Your decisions are represented in the figure below.



RED Player knows the amount of money in the jar.

RED Player sends a message to BLUE Player: "The amount of money in the jar is ODD" or "The amount of money in the jar is EVEN."

This message could be **accurate** or **inaccurate**.

BLUE Player can **Rely** on RED Player's message or **Verify** RED Player's message

RED Player sends accurate message. BLUE Player relies on the message.

RED Player earns **$0.75**
BLUE Player earns **$1.50**

RED Player sends accurate message. BLUE Player verifies and learns the correct information.

RED Player earns **$0.50.**
BLUE Player earns $1.50 - $0.50 = **$1**

RED Player sends inaccurate message. BLUE Player relies on the message.

RED Player earns **$1.50.**
BLUE Player earns **$0.**

RED Player sends inaccurate message. BLUE Player verifies and learns the correct information.

RED Player earns **$0.**
BLUE Player earns $1.50 - $0.50 = **$1**

Comprehension check questions for *Rely-or-Verify*:

1. Suppose **RED Player** sends you an *accurate message*. Will you earn more if you RELY or VERIFY?
2. Suppose **RED Player** sends you an *inaccurate message*. Will you earn more if you RELY or VERIFY?
3. How much does it cost to VERIFY?

4. If you RELY on **RED Player'**s message, would **RED Player** earn more if s/he had sent a message that was *accurate* or *inaccurate?*

Appendix B. Solution to Mixed Strategy Equilibrium for *Rely-or-Verify*

- The *Rely-or-Verify* game took the following form in our studies:

Blue Player
(Participant)

		R	V
Red Player (Confederate)	A	.75, 1.5	.5, 1
	I	1.5, 0	0, 1

- Let p be the probability the Red Player (the confederate) chooses to send an accurate message (A); 1-p is the probability that s/he sends an inaccurate message (I)
- Let q be the probability that the Blue Player (the participant) chooses to rely on the message (R); 1-q is the probability that s/he verifies the message (V)

		q	*1-q*
		R	V
p	A	.75, 1.5	.5, 1
1-p	I	1.5, 0	0, 1

- Solving for mixed strategy equilibrium:

$$p(1.5) + (1 - p)(0) = p(1) + (1 - p)(1)$$
$$p = 2/3$$

$$q(.75) + (1 - q)(.5) = q(1.5) + (1 - q)(0)$$
$$q = 2/5$$

- Red Player will send an Accurate message with probability 2/3 and send an Inaccurate message with probability 1/3

- Blue Player will Rely with probability 2/5 and Verify with probability 3/5

Appendix C. Items used to measure attitudinal trust in Trust game and *Rely-or-Verify* (Study 4)

- I trust my partner. [Rely-or-Verify uses identical measure].
- I am willing to make myself vulnerable to my partner. [Rely-or-Verify uses identical measure].
- I am confident that my partner will return half the money. [I am confident that my partner sent me an accurate message.]

Note. All items were anchored at 1 = "Strongly disagree" and 7 = "Strongly agree."

Tables

Table 1. Payoffs associated with lying and honesty in Studies 1a, 1b, 2, 3a, and 3b

	Experienced or Observed Deception	Deception Game	Type of Lie		Payoffs associated with Truth (Option A)	Payoffs associated with Lie (Option B)
Study 1a	Experienced	Coin Flip Game	Altruistic Lie	Sender	$2.00	$1.75
				Receiver	$0.00	$1.00
Study 1b	Observed	Coin Flip Game	Altruistic Lie	Sender	$2.00	$1.75
				Receiver	$0.00	$1.00
Study 2	Observed	Coin Flip Game	Prosocial Lie	Sender	$2.00	$2.00
				Receiver	$0.00	$1.00
			Mutually beneficial Lie	Sender	$2.00	$2.25
				Receiver	$0.00	$1.00
Studies 3a and 3b[a]	Observed	Number Game (3a)	Altruistic Lie	Sender	$2.00	$1.75
				Receiver	$0.00	$1.00
		Coin Flip Game (3b)	Selfish Lie	Sender	$1.75	$2.00
				Receiver	$1.00	$0.00

Note. [a] Study 3b also included two control conditions. In control condition 1, the Sender faced the Altruistic Lie choice set, and in control condition 2, the Sender faced the Selfish Lie choice set. However, in both control conditions, the Sender's decision was unknown.

Table 2. Descriptive statistics and correlations for measures in Studies 1, 2, and 3

Study 1a

Scale	M(SD)	1	2	3
1. Trusting behavior	43.8%[a]			
2. Attitudinal trust	3.23 (1.91)	0.88**		
3. Benevolence	3.82 (1.48)	0.51**	0.64**	
4. Deception	4.10 (1.84)	0.09	0.08	-0.08

Study 1b

Scale	M(SD)	1	2	3
1. Trusting behavior	29.6%[a]			
2. Attitudinal trust	3.08(1.65)	0.70**		
3. Benevolence	3.95 (1.25)	0.47**	0.61**	
4. Deception	4.15 (1.72)	-0.11[+]	-0.13*	-0.29**

Study 2

Scale	M(SD)	1	2	3
1. Trusting behavior	50.2%[a]			
2. Attitudinal trust	3.41(1.88)	0.73**		
3. Benevolence	4.10 (1.33)	0.49**	0.63**	
4. Deception	4.13 (1.86)	0.08	0.01	0.05

Study 3a

Scale	M(SD)	1	2	3
1. Trusting behavior	36.2%[a]			
2. Attitudinal trust	3.25(1.84)	0.72**		
3. Benevolence	4.12 (1.40)	0.41**	0.67**	
4. Deception	4.09 (2.42)	-0.12*	-0.25**	-0.34**

Study 3b

Scale	M(SD)	1	2	3
1. Trusting behavior	47.2%[a]			
2. Attitudinal trust	3.31(1.95)	0.72**		
3. Benevolence	4.16 (1.40)	0.68**	0.68**	
4. Deception	3.92 (2.27)	-0.26**	-0.26**	-0.38**

Notes. [a] This number represents the percent of participants who chose to pass money in the trust game. ** $p < .001$, *$p < .05$, [+]$p < .10$.

Table 3. Supplemental regressions for Study 3a

Logistic regression on Trusting Behavior

	(1)	(2)	(3)	(4)
	Intentions, Deception, Intentions x Deception	Intentions, Deception, Intentions x Deception, Perceived Benevolence	Intentions, Deception, Intentions x Deception, Perceived Deception	Intentions, Deception, Intentions x Deception, Perceived Benevolence Perceived Deception
Constant	-.601**	-3.887**	0.766^{+}	-2.973***
	(0.122)	(0.601)	(0.411)	(0.839)
Intentions	0.498**	0.019	0.482**	0.042
	(0.122)	(0.151)	(0.125)	(0.153)
Deception	-0.002	0.177	.697**	.506^{+}
	(0.122)	(0.134)	(0.244)	(0.261)
Intentions x Deception	0.032	-0.005	0.025	-0.005
	(0.122)	(0.131)	(0.125)	(0.132)
Perceived Benevolence		0.769**		0.709***
		(0.133)		(0.139)
Perceived Deception			-0.343**	-0.166
			(0.100)	(0.111)
R-Squared	0.054	0.165	0.093	0.181

Notes. ** $p \leq .01$,* p <. 05. $^{+}p < .10$. Standard errors are in parentheses. Independent variables used in each regression are listed in the top row. *Deception* was contrast-coded: -1 = Honest, 1 = Lie. *Intentions* was contrast-coded: -1 = Selfish, 1 = Prosocial.

Table 4. The payoffs associated with prosocial lying in Study 4

Table 4. The payoffs associated with prosocial lying in Study 4

	Type of Lie		Payoffs associated with Truth	Payoffs associated with Lie
Round 1	Altruistic Lie	Sender	$2.00	$1.50
		Receiver	$0.25	$1.00
Round 2	Mutually-beneficial Lie	Sender	$1.50	$2.00
		Receiver	$0.25	$1.00
Round 3	Altruistic Lie	Sender	$1.25	$1.00
		Receiver	$0.25	$1.00
Round 4	Mutually-beneficial Lie	Sender	$1.00	$1.25
		Receiver	$0.25	$1.00

Table 5. Descriptive statistics and correlations for measures in Study 4

Trust game

Scale	M(SD)	1	2	3
1. Trusting behavior	48.50% [a]			
2. Attitudinal trust	3.82 (1.99)	0.84**		
3. Benevolence	4.44 (1.55)	0.49**	0.70**	
4. Deception	3.83 (2.34)	0.07	0.03	0.06

Rely-or-Verify

Scale	M(SD)	1	2	3
1. Trusting behavior	47.20% [b]			
2. Attitudinal trust	4.01 (1.73)	0.80**		
3. Benevolence	4.38 (1.42)	0.41**	0.65**	
4. Deception	3.76 (2.31)	-0.25**	-0.39**	-0.21**

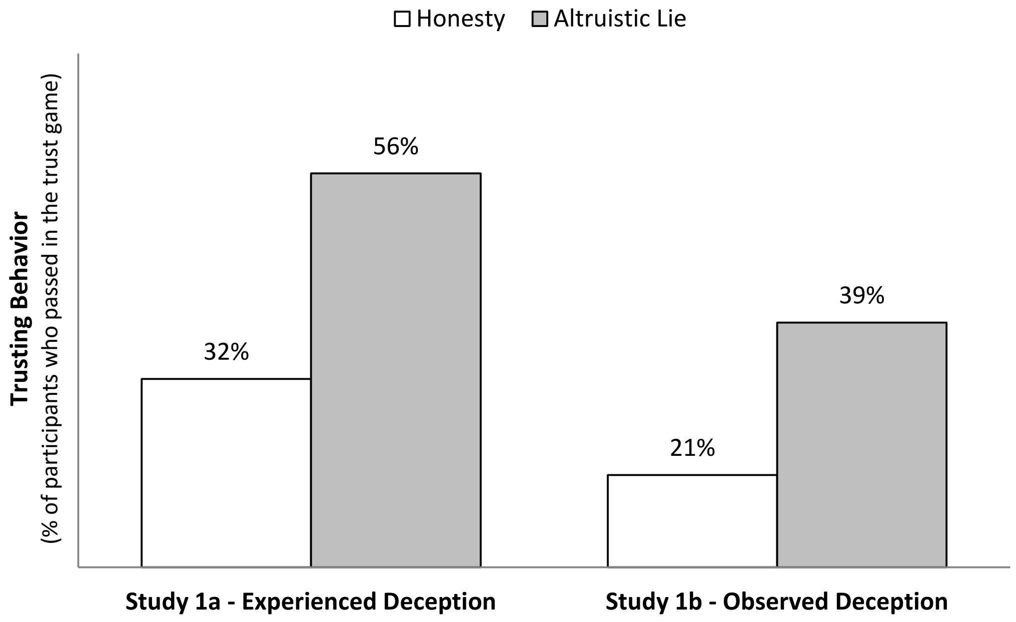
Notes. ***p* < .001.

[a] This number represents the percent of participants who chose to pass money in the trust game.

[b] This number represents the percent of participants who chose *Rely* in *Rely-or-Verify*.

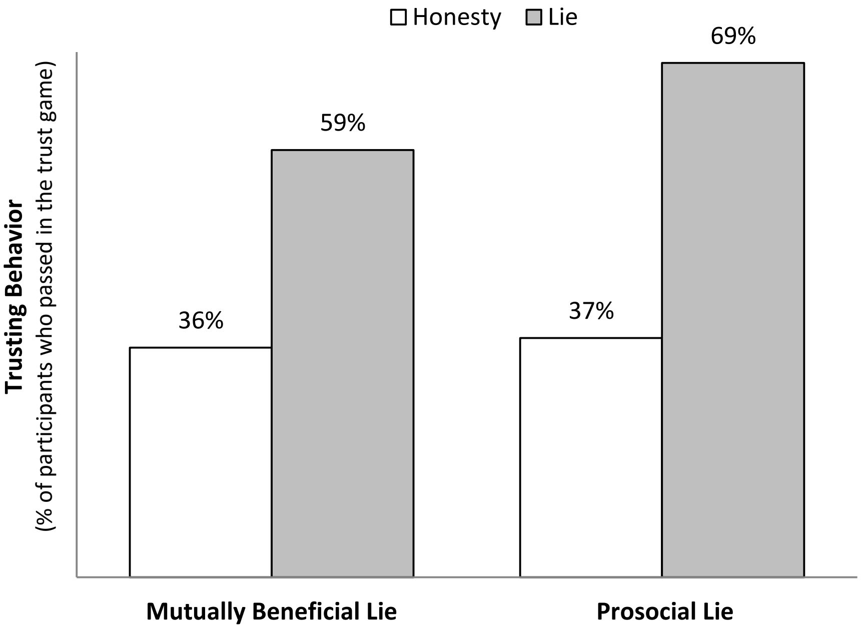
Figures

Figure 1. The effect of altruistic lying on trusting behavior (Studies 1a and 1b).



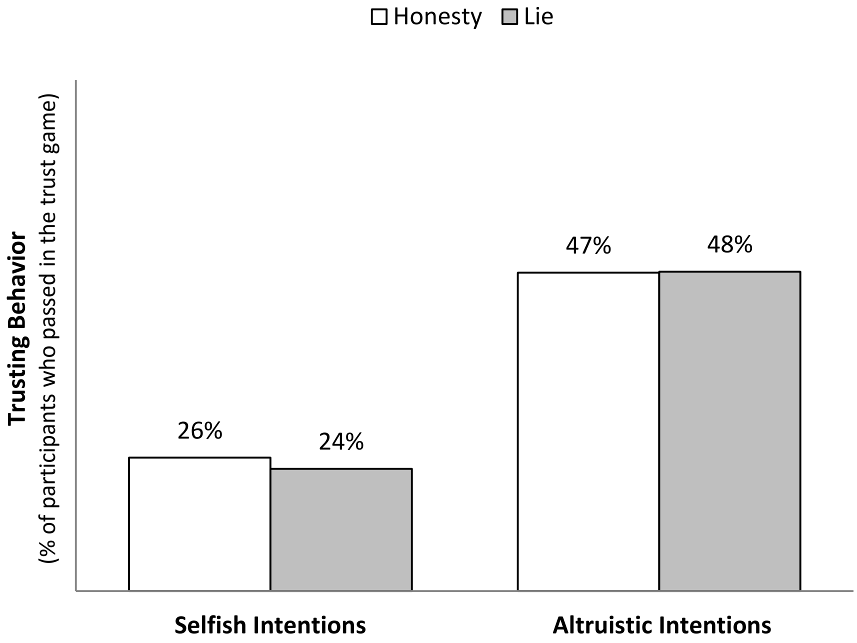
Note. Main effect of altruistic lying in both studies: *p*s < .01.

Figure 2. The effect of prosocial and mutually beneficial lying on trusting behavior (Study 2)



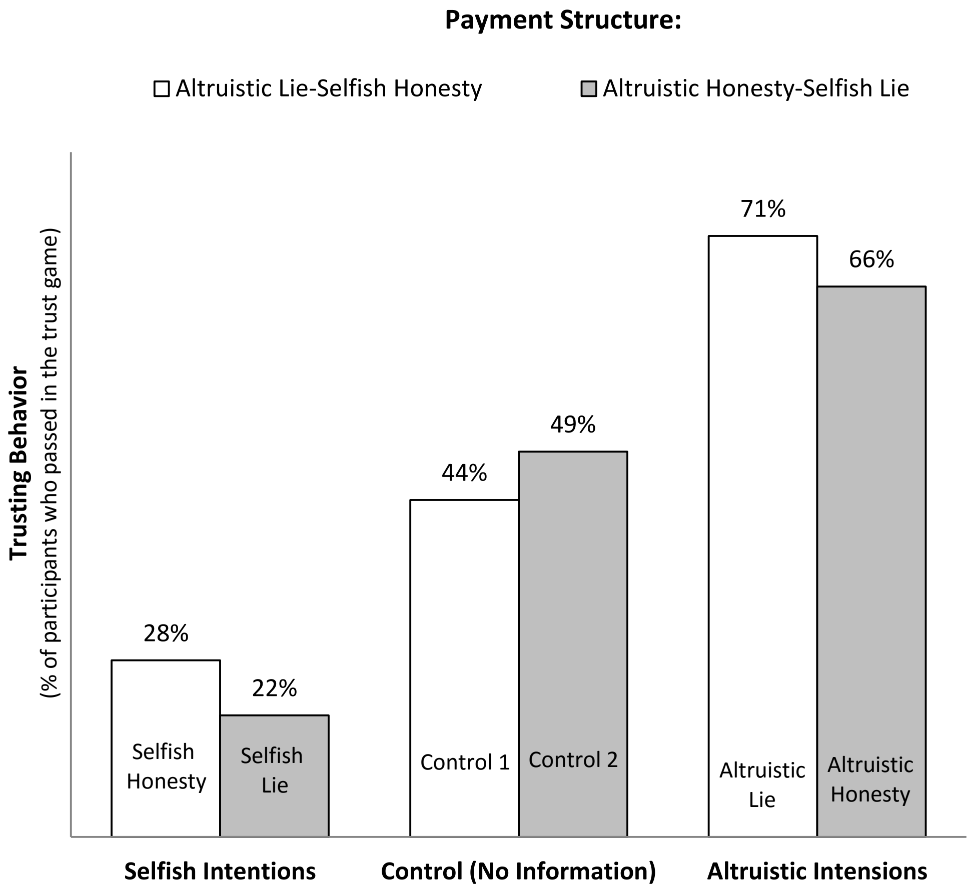
Note. Effect of lying for mutually-beneficial and prosocial lies: each $p < .01$.

Figure 3. Trusting behavior (Study 3a).



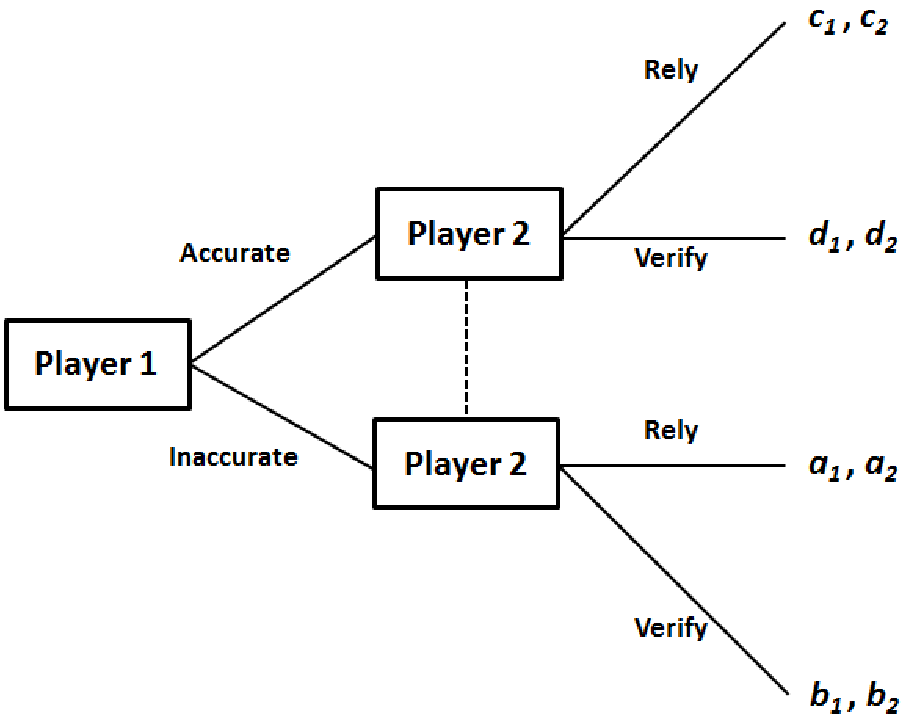
Note. Main effect of intentions: *p* < .01. Main effect of lying: *ns*.

Figure 4. Trusting behavior (Study 3b).



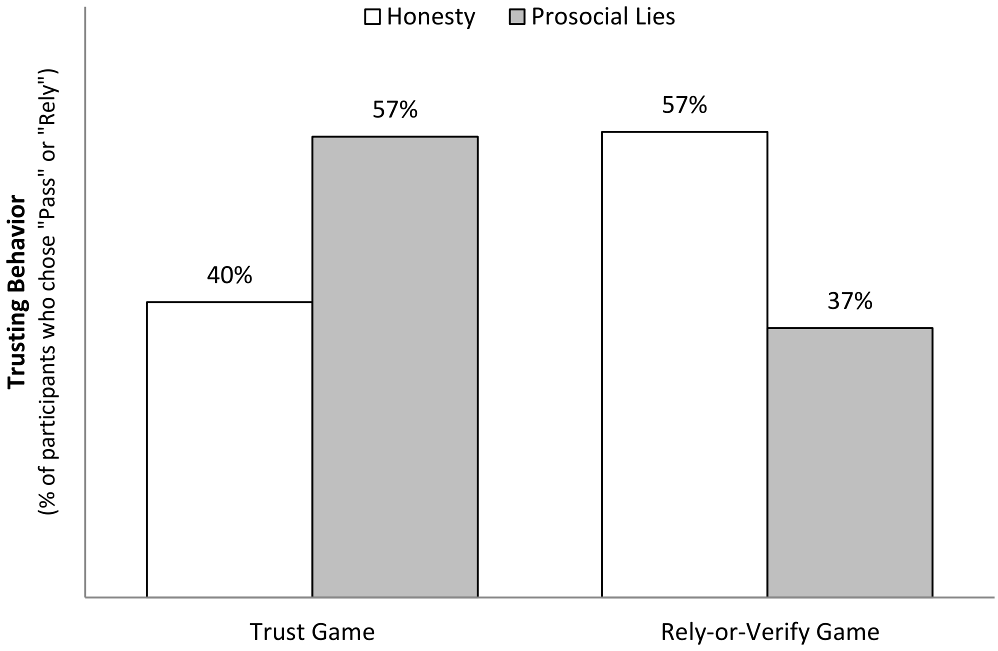
Note. Main effect of decision (Selfish, Control, Altruistic): $p < .01$. Main effect of payment structure: *ns*.

Figure 5. The *Rely-or-Verify* game (Study 4)



Note. This depicts the general form of *Rely-or-Verify*. The exact game we used in Study 4 is depicted in Appendix A. In *Rely-or-Verify*, the payoffs for Player 1 are structured such that $\boldsymbol{a_1} > \boldsymbol{c_1} \geq \boldsymbol{d_1} > \boldsymbol{b_1}$. The payoffs for Player 2 are structured such that $\boldsymbol{c_2} > \boldsymbol{d_2} \geq \boldsymbol{b_2} > \boldsymbol{a_2}$.

Figure 6. Trusting behavior (Study 4).



Note. Deception x Game interaction: p < .01. Main effects of deception and game: *ns.*

CHAPTER 4

YOU CAN HANDLE (SPEAKING) THE TRUTH:

MISPREDICTING THE INTRAPERSONAL CONSEQUENCES OF HONESTY AND KINDNESS

Emma E. Levine

Taya R. Cohen

ABSTRACT

Many of our most difficult conversations involve navigating the tension between honesty and kindness. In the present research, we explore the intrapersonal consequences of communicating honestly and kindly by randomly assigning individuals to be honest, kind, or conscious of their communication (our control condition) in every conversation with every person in their life for three days. We examine the impact of our interventions on predicted and actual hedonic and eudaimonic well-being. We document three main results. First, individuals predict that being honest will be far less enjoyable (i.e., less hedonically rewarding) than being kind, causing individuals to avoid communicating honestly. Second, this prediction is incorrect: the experience of being honest is more enjoyable than individuals expect. Although honesty is less enjoyable than kindness, this difference is significantly smaller than individuals expect. Third, being honest yields greater meaning (i.e., eudaimonic well-being) and has greater long-term impact on individuals' lives than being kind or conscious of one's communication. This research sheds new light on the relationships among communication, morality, and well-being.

YOU CAN HANDLE (SPEAKING) THE TRUTH:

MISPREDICTING THE INTRAPERSONAL CONSEQUENCES OF HONESTY AND KINDNESS

Honesty and kindness are two of the most fundamental moral values in human life. Honesty and kindness are among the most important traits for interpersonal judgment (Anderson, 1968) and they dominate philosophical and religious teaching across time and cultures. For example, the Judeo-Christian Bible contains statements both prohibiting lies (e.g., “Thou shalt not bear false witness,” Exodus 20:16; “Thou shalt not lie to one another,” Leviticus 19:11) and mandating kindness (“Be kind and compassionate to one another,” Ephesians 4:32).

Despite the theoretical importance of these two values, we know very little about the consequences of honesty or kindness in everyday life. This reflects a significant gap between normative and behavioral ethics. For centuries, ethicists have touted the moral significance of different virtuous behaviors, and only recently have psychologists examined the experience and consequences of enacting or violating these virtues (Dunn, Aknin, Norton, 2008; Emmons & McCullough, 2003; Gino, Kouchaki, & Galinsky, 2015; Lyubomirsky, Shelden, & Schkade, 2005). And although the field of behavioral ethics has made enormous contributions to our understanding of human behavior over the past several decades, this research has not been able to offer insight into how individuals *should* balance competing moral values to improve their own well-being (Barry & Rehel, 2014). Instead, the vast majority of behavioral ethics research examines when and why people behave unethically (see Treviño, Weaver, & Reynolds, 2006; Bazerman & Gino, 2012 for reviews).

The present research departs from prior work on behavioral ethics by examining the psychological consequences of enacting distinct moral values. We examine the consequences of honesty and kindness not only because they are two of the most important and salient moral values, but also because they frequently collide in routine human communication (Goffman, 1967; Brown & Levison, 1987). Whenever individuals are faced with opportunities to communicate unpleasant information to others, they implicitly face tradeoffs between being completely honest and being kind. People routinely face this conflict in their personal lives when deciding how to communicate with friends and family members, and in their professional lives when deciding how to deliver negative news and critical feedback. Though this tension is part of everyday life, navigating it can elicit distress and anxiety (e.g., Molinsky & Margolis, 2005). As a result, individuals often avoid engaging in conversations in which honesty and kindness appear to conflict (e.g., Rosen & Tesser, 1970). In this research, we primarily focus on this conflict and compare the consequences of honesty to the consequences of kindness. Although we recognize that these values need not always be in conflict, one goal of this research is to understand whether focusing on either honesty or kindness is more effective for promoting well-being during difficult conversations.

In a large-scale field experiment, we examine the predicted and actual effects of honesty and kindness on psychological well-being. We examine two types of well-being: hedonic well-being and eudaimonic well-being. Hedonic well-being is characterized by pleasure, enjoyment, and happiness. In the hedonic view, well-being consists of the presence of pleasure and the absence of pain (Ryan & Deci, 2001). Eudaimonic well-being is characterized by meaning, fulfillment, and individual autonomy. In the eudaimonic view, well-being consists of the actualization of human potentials, rather than pleasure (Waterman, 1990, 1993). To our knowledge, this is the first research to examine how different moral principles and styles of communication influence these two fundamental forms of well-being.

This research deepens our understanding of communication, morality and well-being. First, we document the psychological forces that (erroneously) push people away from communicating honestly. Second, we demonstrate that different moral proclivities can have very different influences on different forms of well-being. Philosophers, psychologists, public figures, and practitioners have long been motivated to make links between ethical decisions and well-being (e.g., Bentham, 1843/1948; Harris, 2011; Person & Seligman, 2004; Plato, 1976). Some philosophers argue that the entire purpose of morality is to promote well-being (Bentham, 1843/1948; Harris, 2011), and yet, we know very little about the relationship between different - and often competing - moral principles and well-being. This is particularly problematic given the frequency with which practitioners make untested promises about the relationship between honesty and positive life outcomes (e.g., Blanton, 1996; Dalio, 2011; Gaffney, 2002; Newton, 2014). For example, Brad Blanton, a psychotherapist and founder of the cult “Radical Honesty” has promised his thousands of followers that complete honesty is the route to happiness and well-being (Blanton, 1996). The present research explores the validity of these ideas, demonstrating when and why honesty – and kindness – helps and hinders different forms of well-being.

Hypotheses

We make three central predictions regarding the hedonic and eudaimonic consequences of honesty and kindness. First, we hypothesize that individuals expect communicating honestly to be less pleasant (i.e., less hedonically rewarding) than communicating kindly (H1). Consistent with this proposition, past research demonstrates that many individuals choose kindness over honesty when these two values appear to conflict (Lee, 1993; Rosen & Tesser, 1970; Tesser, Rosen, & Rosen, 1971). For example, when delivering difficult news, many individuals naturally focus on being kind and "softening the blow" by using polite and evasive language (Lee, 1993). Many individuals also avoid delivering difficult news altogether (Rosen & Tesser, 1970; Tesser, Rosen, & Rosen, 1971). Individuals avoid honestly sharing unpleasant information or criticisms with others because they worry about others' emotional reactions to the news, and expect the conversation to elicit personal feelings of guilt and distress (e.g., Tesser & Rosen, 1972). Individuals also avoid honestly sharing personal information because they worry about others' judgment and hurt feelings (Rosenfeld, 1979). Individuals' concerns regarding the consequences of honesty pertain primarily to its hedonic, or affective, costs.

We have reason to believe, however, that these concerns are overstated. Although honesty may indeed be unpleasant, we hypothesize that it is less unpleasant than individuals expect (H2). Past research on the experience of performing "necessary evils" such as delivering terminal prognoses or critical performance feedback sheds light on this possibility. Although individuals who candidly communicate unpleasant information do experience psychological duress, many are able to maintain psychological engagement during the process, despite prior assumptions that the discomfort associated with these conversations causes individuals to disengage (Margolis & Molinsky, 2008; Molinsky & Margolis, 2005). This finding suggests that being honest with others may not be as unpleasant as it seems. Furthermore, being honest with oneself by openly sharing one's thoughts and emotions can be quite rewarding. For example, individuals who honestly express their emotions experience lower stress and blood pressure, and develop higher levels of intimacy than individuals who regulate or hide their emotions (Butler et al., 2003; Srivastava et al., 2009).

In other words, we predict an affective forecasting failure (e.g., Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Wilson & Gilbert, 2005) with respect to the hedonic consequences of honesty. Just as individuals overestimate the affective costs of unfortunate events, such as a breakup or the denial of tenure (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998), we expect individuals to overestimate the affective costs of honesty. Individuals are particularly likely to mispredict the affective consequences of honesty because individuals avoid engaging in the conversations that would provide them with accurate feedback about these consequences.

Third, we hypothesize that honesty is more meaningful than kindness (H3). That is, we expect honesty to increases eudaimonic well-being. To communicate honestly, individuals must look inwards and consult their personal feelings and opinions. This process may increase self-actualization and produce feelings of personal control and autonomy, key components of eudaimonia (Deci & Ryan, 2008). Recent research on the experience of inauthenticity is also consistent with this proposition. Behaving inauthentically – by misrepresenting one's emotions or by conforming to social norms that are inconsistent with one's personal beliefs, for example – lowers individual's moral self-regard and sense of moral purity (Gino, Kouchaki, & Galinsky, 2015). Moral identity is closely linked to sense of self (Aquino & Reed, 2002). Thus, decrements in moral identity may undermine meaning and purpose.

In addition to testing these three hypotheses, we also explore the social and long-term consequences of honesty and kindness in the present research.

Overview of Study

We conducted an experiment in which we randomly assigned participants to be completely honest, kind, or conscious of their communication in every interaction for three days. Our study involved two separate samples: Experiencers (Study 1a) and Forecasters (Study 1b). In Study 1a, laboratory participants were randomly assigned to communicate honestly, kindly, or consciously (our control condition).

Although we focus our hypotheses on the differential effects of honesty and kindness, we also include a control condition in our experiment. The control condition serves two purposes. First, it allows us to examine how honesty and kindness each influence well-being above and beyond the experience of the study itself. Second, it allows us to examine the nature of the differences between honesty and kindness. For example, by including a control condition, we can assess whether focusing on kindness increases (predicted) hedonic well-being, or whether focusing on honesty harms (predicted) well-being.

Participants in Study 1a (Experiencers) made forecasts about the three-day experience, provided judgments of the experience every day during the study, and then reflected on their experience during the study two weeks later. We conducted a two-week follow up survey in order to gain greater insight into the long-term impact of honesty and kindness and to examine if individuals' perceptions of the experience changed over time. Participants in Study 1b (Forecasters) did not participate in the main study; they simply learned about the conditions of Study 1a and made forecasts about the experience.

Study 1a: Experiencers

Procedure and Materials. Study 1a consisted of five stages: 1) participants were recruited and took an intake survey, 2) participants were assigned to condition, 3) participants made forecasts of their experience in the study, 4) participants completed the study over three days and completed nightly surveys on their experiences, and 5) participants completed a follow-up survey and reflected on their experiences two weeks later.

Recruitment and intake survey. One-hundred twenty-eight adults (55% female, mean age = 26) agreed to participate in this study. Community members and students were recruited in groups of 10-20 to a United States university laboratory to complete an hour long study in exchange for a $10 show-up fee. For the first thirty minutes of this hour long session, participants completed surveys that were unrelated to the present research.

Thirty minutes into the session, when all participants had completed their surveys, a research assistant who was blind to our hypotheses made an announcement about an optional additional study, called “The Challenging Exercise” study. The experimenter explained that participants could participate in an optional 3-day experiment that would challenge the way they communicate with others. In exchange for their participation, participants would earn $20 and the chance to win an iPad mini. Participants were informed of the time commitment of the study and the potential distress that could be caused by participating. However, they were not provided any information about the experimental conditions at this time. Participants were free to leave if they did not want to participate in the study. We include the exact recruitment announcement in Appendix A.

Participants who chose to participate in “The Challenging Exercise” then completed a link on the computer containing a consent form and an intake survey. The intake survey contained personality measures and other exploratory variables.[14]

Assignment to condition. After participants completed the intake survey, they were assigned to one of three experimental conditions: honesty, kindness, or communication-consciousness (our control condition). We randomized condition at the session-level. That is, each session of participants (i.e., the group of participants that arrived at the lab during the same time) was assigned to the same condition. Participants learned about the experimental condition verbally, and had the opportunity to ask

[14] We also measured satisfaction with life, positive and negative affect, the HEXACO, and general social connection. We report all specific measures and results in our online supplemental materials.

questions. Thus, it was necessary to have each session of participants assigned to a single condition.

The research assistant first provided some basic information about the study. Then, the research assistant instructed participants how to behave for the next three days, according to their experimental condition. Specifically, the research assistant announced:

In this study, you will be asked to reflect upon your social communication. Often, speaking with others requires balancing honesty and kindness. Being completely open and honest about our thoughts, feelings, and opinions, can sometimes upset others and be unkind. Alternatively, being kind, considerate, and helpful towards others sometimes means not being 100% honest.

[Honesty condition]

Throughout the next three days – that means today, tomorrow, and the following day - please strive to be absolutely honest in every conversation you have with every person you talk to. Really try to be completely candid and open when you are sharing your thoughts, feelings, and opinions with others. You should be honest in every conversation you have, in every interaction, with every person in your life. Even though this may be difficult, you should do your absolute best to be honest.

[Kindness condition]

Throughout the next three days – that means today, tomorrow, and the following day - please strive to be kind in every conversation you have with every person you talk to. Really try to be caring and considerate when you are sharing your thoughts, feelings, and opinions. You should be kind in every conversation you have, in every interaction, with every person in your life. Even though this may be difficult, you should do your absolute best to be kind.

[Communication-consciousness– Control condition]

Throughout the next three days – that means today, tomorrow, and the following day - please be conscious of the way you communicate with others. Please act as you normally would throughout the length of this study. You should not change your behavior, but you should be conscious of it.

Note that the research assistant explicitly mentioned the potential conflict between honesty and kindness in every condition. Thus, all participants were primed to consider this difficult tradeoff before engaging in the experiment.

After making this announcement, the research assistant explained the conditions in greater detail and invited questions from participants. Participants were instructed not to tell anyone about the experiment, including their relational and conversational partners. We include the full script for each condition in Appendix B.

Participants were then directed to a link on their computer. They first read the instructions associated with their condition. These instructions were nearly identical to the verbal script read by the research assistant, except they included an additional statement, which said, "Do your best to comply with these instructions**,** but do not do anything you are not comfortable with. When reflecting on your experience, you should answer all surveys accurately and thoughtfully, even if you did not completely comply with the instructions."

Participants then responded to a one-item comprehension check, asking them what their goal in the study was (response-options: "To be honest in all of my communication", "To be kind in all my communication," or "To communicate as I normally do, but be conscious of my communication.") Participants had to answer the comprehension check correctly to proceed with the study.

Next, participants provided their email address to indicate their continued consent, and to allow us to contact them with their nightly surveys. At this point, participants were told that they should let the laboratory staff know if they no longer wished to participate. All participants in our sample continued with the study at this time.

Forecasting the experience. Participants were then directed to the forecasting task, which was on the next page of their survey. Participants rated the extent to which they expected their experience in the study to be: easy, pleasant, meaningful, liberating, fulfilling, and socially connecting. We measured these dimensions using five-point bipolar rating scales with the following anchors: difficult-easy, unpleasant-pleasant, meaningless-meaningful, constraining-liberating, unfulfilling-fulfilling, and socially isolating-socially connecting.

Based on our theoretical assumptions, we combined the first two items into a single measure of Enjoyment (rs > .56), and we combined the middle three items (meaningful, liberating, fulfilling) into a single measure of Meaning (αs > .78). Enjoyment is our measure of hedonic well-being; Meaning is our measure of eudaimonic well-being. We examine Social Connection as a separate construct because social connection could be theoretically conceptualized as either a source of pleasure (hedonic well-being) or meaning (eudaimonic well-being).

Finally, participants were asked to confirm their commitment to the study by typing the following statement into the survey, "*For the next three days, I will [communicate honestly, communicate kindly, be conscious of my communication]*." [15]

[15] We did not include this instruction during the first hour we ran the study. Thus, there are 12 participants (all in the Kindness condition) who did not have to write out their commitment before the study began.

Before leaving the laboratory, participants were reminded of their study condition and instructed to begin the study immediately. Participants were told that they would receive their first nightly survey that evening at 6pm. Participants had to say aloud, “I agree to participate” upon exiting the laboratory.

Nightly surveys. Consistent with past research (e.g., Sonnentag, 2003; Sonnentag, Binnewies, & Mojza, 2008), we tracked behavior over three consecutive days. We emailed participants a nightly survey for three nights at 6pm. We instructed participants to complete the survey as late as possible, but before they were too tired to concentrate.

When completing the nightly survey, participants first completed a communication audit to ensure their commitment to the experiment. We asked participants to recall their longest conversation. Participants reported who they had the conversation with (e.g., friend, spouse, roommate), they described their conversation (free response), and they explained how they [communicated honestly, communicated kindly, or were conscious of their communication] in the conversation (free response). Then, we asked participants if they said anything untrue (yes, no, and explain your answer) and if they said anything unkind (yes, no, and explain your answer) during their conversation.

After participants completed their communication audit, they responded to our focal measures: experiences of Enjoyment (ease, pleasure), Meaning (meaning, liberation, fulfillment), and Social Connection. Participants used the same bipolar scales we administered as a part of the forecasting survey. Next, participants rated their agreement (1 = strongly disagree, 7 = strongly agree) with two manipulation check items:

“I was completely honest and candid in every conversation I had today” and “I was kind and compassionate during every conversation I had today.” [16]

Finally, we asked participants to reflect on their experience that day and to explain how they either did or did not comply with the experiment. We also asked them to write about any challenges they faced and how it felt to focus on [honesty, kindness, their communication]. Participants were given the lead experimenter’s email address and invited to reach out to her with questions or concerns at any time.

Reflection survey. Two weeks after participants completed the third and final day of the experiment, they were emailed a final reflection survey. Participants first responded to several open-ended questions, asking them what they learned, how their behavior and communication had changed, what difficulties they had, any surprises they faced, and how their relationships changed.

Second, participants indicated their agreement with five statements about the degree to which their participation had long-term impact on their lives (1 = strongly disagree, 7 = strongly agree): “As a result of participating in this study [I am more conscious of my communication, I am more thoughtful when speaking to others, I have reconsidered the way I communicate, I have become a better person, I am happier.]” We combined them into a single measure of Long-term Improvement (α = .90).

[16] We also collected measures of general social connection (as in the intake survey), authenticity and self/other focus. We collected these measures during the nightly surveys and during the two-week follow-up. We report the specific measures and results in the online supplementary materials.

Participants also indicated their agreement with four statements about their specific communication (1 = strongly disagree, 7 = strongly agree): “As a result of participating in this study [I am more honest, I communicate more directly, I am more kind, I engage in more conflict].” We combined the first two items into a single measure of Long-term Honesty (r = .72). We examine the latter two items separately and conceptualize them as measures of Long-term Kindness and Long-term Conflict.

Then, participants responded to our hedonic and eudaimonic well-being measures. Participants reflected on their experience and rated the extent to which their experience had provided them with Enjoyment, Meaning, and Social Connection, using the same items we used in the forecasting survey and nightly surveys.

Next, participants answered questions about the degree to which the experiment influenced their relationships. Specifically, we collected a three item measure of Relational Improvement which captured the degree to which participants believed their relationships became better or worse as a result of completing our study (anchored at 1 = much worse and 7 = much better, α = .86): “Do you feel that the people around you know you better or worse than they knew you before this study?”, “Do you feel that the people around you understand you better or worse than they understood you before this study?”, and “Do you feel that the quality of your relationships are better or worse as a result of this study?”[17]

[17] At the two-week follow-up, we also collected two items about whether participants saw themselves as honest [kind] people. We report the specific measures and results in the online supplementary materials.

Before exiting the survey, participants indicated whether they would prefer their $20 payment via paypal or by receiving an amazon.com giftcard. We randomly selected one participant to win the iPad mini and we compensated all participants within one week.

Analytical approach.

First, we created daily average variables by taking the average of all dependent variables that we collected during the nightly surveys. For example, we averaged perceptions of how enjoyable the experience was on Day 1, Day 2, and Day 3 to create a daily average Enjoyment variable.

We conducted four sets of analyses to examine the consequences of honesty and kindness. First, we analyzed our manipulation check measures at the daily average level to examine compliance with the experiment. Second, we conducted our focal analyses: we compare forecasts, experiences, and reflections of Enjoyment, Meaning, and Social Connection. Finally, we conducted a set of analyses to examine the long-term impact of our experiment.

Results. One-hundred twenty-eight adults (55% female, mean age = 26) agreed to participate in the Challenging Exercise and were included in our final data set. We conduct analyses using all participants who responded to each measure. Thus, the degrees of freedom for each analysis may differ slightly. We did not see differential attrition across our experimental conditions throughout the three-day experiment. However, we see slightly greater attrition in the kindness and control conditions, relative to the honesty conditions, at the two-week follow-up. Table 1 depicts the number and percentage of participants who began and completed each stage of the experiment across out conditions.

Manipulation checks. Consistent with the intent of the experiment, a one-way ANOVA revealed a significant effect of *Condition* on participants' average daily honesty, $F(2, 98) = 8.28$, $p < .001$, $\eta_p^2 = .15$. Participants reported being more honest in the *Honesty* condition ($M = 5.66$, $SD = 1.10$), than in the *Kindness* ($M = 4.67$, $SD = 1.14$) or *Control* ($M = 4.80$, $SD = 1.08$) conditions, *p*s < .01. There was no difference between the *Kindness* and *Control* conditions (p =.64).

A one-way ANOVA also revealed a significant effect of *Condition* on participants' average daily kindness, $F(2, 98) = 4.78$, $p = .01$, $\eta_p^2 = .09$. Participants reported being kinder in the *Kindness* condition ($M = 5.45$, $SD = 0.93$), than in the *Honesty* ($M = 4.77$, $SD = 1.21$) or *Control* ($M = 4.71$, $SD = 1.04$) conditions, *p*s $\leq$ .01. There was no difference between the *Honesty* and *Control* conditions (p =.80).

Forecast, Experience, and Reflections. We conducted repeated measure ANOVAs on our measures of enjoyment (i.e., hedonic well-being), meaning (i.e., eudaimonic well-being), and social connection using experimental condition (Honesty, Kindness, Control) as the between-subjects factor, and time-point (Forecast, Experience, Reflection) as the within-subjects factor.

Enjoyment. A repeated measures ANOVA revealed a significant effect of *Condition*, $F(2,87) = 6.94$, $p < .01$, $\eta_p^2 = .14$; participants rated the *Honesty* condition ($M = 3.11$, $SD = 0.73$), as less enjoyable than the *Kindness* ($M = 3.76$, $SD = 0.73$) and

Control conditions (*M* = 3.58, *SD* = 0.73), *ps* $\leq$ .01. There was no difference between the *Kindness* and *Control* conditions (*p* =.37).

There was also a significant effect of *Time-point*, *F*(2,87) = 6.03, *p* < .01, η_p^2 = .07; participants forecasted lower enjoyment (*M* = 3.22, *SD* = 1.08) than they actually experienced over the three-day experiment (*M* = 3.58, *SD* = 0.82) or remembered two-weeks after the experience (*M* = 3.52, *SD* = 0.90), *ps* $\leq$ .01. There was no difference between actual and remembered enjoyment (*p* =.50).

Importantly, these effects were qualified by a significant *Condition* x *Time-point* interaction, *F*(4,87) = 10.07, *p* < .01, η_p^2 = .19. *Honesty* was the only condition in which participants had misforecasted their enjoyment, consistent with hypotheses 1 and 2. Specifically, participants expected the *Honesty* condition (*M* = 2.52, *SD* = 1.05), to yield less enjoyment than the *Kindness* (*M* = 3.72, *SD* = 0.85) and *Control* conditions (*M* = 3.68, *SD* = 0.78), *ps* $\leq$ .01. During the three-day experience, however, *Honesty* (*M* = 3.47, *SD* = 1.00) was only slightly less enjoyable than the *Kindness* condition (*M* = 3.89, *SD* = 0.57), *p* < .05, and was no different from the *Control* condition (*M* = 3.43, *SD* = 0.66), *p* = .83. The nature of this interaction demonstrates that *Honesty* was more enjoyable than individuals expected, but *Kindness* and *Control* did not differ from expectations. There were no differences in remembered enjoyment across any of the three conditions two-weeks later, *p*s > .12. We depict these results in Figure 1.

Meaning. A repeated measures ANOVA revealed a marginal effect of *Condition*, *F*(2,87) = 2.44, *p* = .09, η_p^2 = .05; participants found greater meaning in the *Honesty* condition (*M* = 3.78, *SD* = 0.63), than in the *Kindness* condition (*M* = 3.45, *SD* = 0.63), *p*

< .05, and directionally greater meaning in the *Honesty* condition than in the *Control* condition (M = 3.53, SD = 0.63), p = .10 . There was no difference between the *Kindness* and *Control* conditions (p =.67). We found no main effect of *Time-point*, F(2,87) = 0.85, p = .43, η_p^2 = .01, nor did we find a *Condition* x *Time-point* interaction, F(4,87) = 0.34, p = .85, η_p^2 < .01. We depict these results in Figure 2.

Social connection. A repeated measures ANOVA revealed a significant effect of *Time-point*, F(2,87) = 3.65, p = .03, η_p^2 = .04; participants experienced lower social connection during the 3-day experience itself (M = 3.55, SD = 0.77) than they forecasted before the experience (M = 3.77, SD = 0.90), p = .01, or remembered two-weeks after the experience (M = 3.72, SD = 0.89), p = .03. Although this effect appears to be driven by the *Kindness* condition (directionally, participants expected kindness to be more socially connecting than it was), we found no main effect of *Condition,* F(2,87) = 1.29, p = .28, η_p^2 = .03, nor did we find a significant *Condition* x *Time-point* interaction, F(4,87) = 1.39, p = .24, η_p^2 = .03. We depict these results in Figure 3.

Long-term impact. To assess long-term impact, we conducted one-way ANOVAs on our follow-up measures of Long-term Honesty, Long-term Kindness, Long-term Conflict, Long-term Improvement, and Relational Improvement using experimental condition (Honesty, Kindness, Control) as the between-subjects factor. We display the means and standard deviations of all Long-term impact measures in Table 2.

A one-way ANOVA revealed a significant effect of *Condition* on Long-term Honesty, F(2,97) = 5.93, p < .01, η_p^2 = .11, such that participants became more honest in the *Honesty* condition than in the *Kindness and* Control conditions (ps < .05). There was no difference between the *Kindness* and *Control* conditions (p =.21). We also found a significant effect of *Condition* on Long-term Improvement, F(2,97) = 3.71, p = .03, η_p^2 = .07, such that participants believed that had become better people in the *Honesty* condition, relative to the *Control* condition (p < .01). Participants also believed they became marginally better people in the *Kindness* condition, relative to the *Control* condition (p = .08), but there was not a significant difference between the *Honesty* and *Control* conditions (p = *.45*). We find no effects of our experimental conditions on Long-term Kindness, F(2,97) = .51, p = .60, η_p^2 = .01, Long-term conflict, F(2,97) = 1.17, p = .31, η_p^2 = .03, or Relational Improvement, F(2,97) = 1.82, p = .17, η_p^2 = .03.

These results demonstrate that honesty had a longer-lasting inpact on behavior. Individuals in the *Honesty* condition had become more honest, but no less kind. Individuals in the *Kindness* condition had not changed their levels of honesty or kindness.

--Table 2 here –

To provide greater insight into the impact of our interventions, we examined participants' free responses. We provide example quotes in Table 3 to better illustrate the consequences of honesty, kindness, and communication-consciousness.

---Table 3 about here---

Discussion

The results of Study 1a support our hypotheses. First, consistent with H1 and H2, individuals misforecast the hedonic consequences of honesty; although honesty does yield less pleasure than kindness, individuals expect this gap to be much larger than it actually is. Second, consistent with H3, individuals derive greater meaning from honesty than kindness.

In addition, we find that the experience of being honest with others has longer-lasting consequences than being kind. Although the experience of being honest and the experience of being kind both caused individuals to believe they had become better, happier, and more thoughtful individuals, only individuals who had been honest continued to communicate this way two-weeks later. Interestingly, honesty and kindness did not have differential effects on self-reported social connection or long-term kindness.

We build on these findings in Study 1b by examining forecasts and choices made by individuals who were not involved in Study 1a. This allows us to do a cleaner comparison of forecasters to experiences, and to examine whether individuals' communication choices favor hedonic outcomes (kindness) or eudaimonic outcomes (honesty).

Study 1b: Forecasters

Method. We recruited one-hundred nine adults (50.5% female, mean age = 25) from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee.[18] Participants in Study 1b were drawn from the same subject pool as participants in Study 1a.

Participants learned about an experiment that was taking place, called "the Challenging Exercise" Study. We described the protocol of the Challenging Exercise

[18] We ran Study 1b after running Study 1a. Six participants in Study 1b had previously participated in Study 1a. We removed these participants from the sample before any analyses were performed.

(Study 1a) as closely as possible. Participants learned that individuals who enrolled in the Challenging Exercise would have to make modifications to their communication for three days and complete nightly surveys, and that the experience might cause discomfort. Then, all participants learned about all three conditions of the study – honesty, kindness, and consciousness - and read the exact instructions that participants in the Challenging Exercise (Study 1a) actually received.

Following the procedure of Epley & Schroeder, 2014, we included the same conditions in the forecasting study (1b) as we included in the experience study (1a), but we manipulated the conditions within, rather than between, subjects.

After reading about each of the conditions, participants were asked to imagine participating in the study and to imagine being honest [being kind, being conscious of their communication] for three days. Participants forecasted their level of Enjoyment, Meaning, and Social Connection in each of the experimental conditions using the same items we used in Study 1a.

Then, we asked participants to imagine they actually had to participate in one condition in the study. Participants selected the one condition they would want to participate in. As an exploratory measure, we also asked participants to imagine that they had to participate in the study for an entire year. Participants selected the one condition they would want to participate in for one year. After participants made their choices, they answered demographic questions and were dismissed.

Results

Forecasts. We analyzed the forecasts using a repeated measures ANOVA, in which condition was the within-subjects factor. We find a main effect of *Condition* on expected Enjoyment, $F(1, 108) = 29.53$, $p < .01$, $\eta_p^2 = .21$, such that participants expected the *Honesty* condition ($M = 2.75$, $SD = 1.09$) to be less enjoyable than both the *Kindness* ($M = 3.61$, $SD = 1.06$) and *Control* conditions ($M = 3.43$, $SD = 1.03$), *ps* < .02. We find no difference in expected Enjoyment between the *Kindness* and *Control* conditions ($p = .14$). We depict these results in Figure 1.

There was a main effect of *Condition* on expected Meaning, $F(1, 108) = 6.18$, $p =.02$, $\eta_p^2 = .05$, such that participants expected the *Control* condition ($M = 3.44$, $SD = 0.90$) to be less meaningful than the *Honesty* ($M = 3.66$, $SD = 0.93$) and *Kindness* conditions ($M = 3.68$, $SD = 0.90$), *ps* < .02. We find no difference in expected meaning between the *Honesty* and *Kindness* conditions ($p =.83$). We depict these results in Figure 2.

There was a main effect of *Condition* on expected Social Connection, $F(1, 108) = 18.71$, $p < .01$, $\eta_p^2 = .15$, such that participants expected the *Honesty* condition ($M = 3.04$, $SD = 1.11$) to be less socially connecting than the *Kindness* ($M = 4.03$, $SD = 1.02$) *Control* conditions ($M = 3.51$, $SD = 0.93$), *ps* < .01. Participants also expected the *Control* condition to be less socially connecting than the *Kindness* condition ($p < .01$). We depict these results in Figure 3.

Choice. We conducted a chi-square goodness of fit test against the null hypothesis that there were no differences in preferences across the three conditions (i.e., expected proportion of 33.3% for each of the three conditions). Participants were significantly less likely to choose to participate in the *Honesty* condition (21.1%) compared to the *Kindness* (37.6%) and *Control* conditions (41.3%) for the three-day study, $\chi^2(1) = 7.56$, $p = .02$.

Participants' preferences became more extreme when choosing how to communicate for one year; participants were significantly less likely to choose *Honesty* (9.2%) compared to both *Kindness* (33.9%) and the *Control* condition (56.9%) for a one-year experience, $\chi^2(1) = 37.23$, $p < .01$.

Comparison between Study 1a and 1b

We conducted t-tests between predicted Enjoyment, Meaning, and Social Connection in Study 1b to the daily-average levels of Enjoyment, Meaning, and Social Connection in Study 1a to further test our hypotheses. These results provide convergent evidence that individuals significantly overestimate the hedonic costs of honesty. Specifically, participants in Study 1b expected honesty to be much less pleasant ($M = 2.75$, $SD = 1.09$) than participants in Study 1a actually experienced it to be ($M = 3.47$, $SD = 0.63$), $p < .001$. However, individuals do not seem to mispredict the eudaimonic consequences honesty. Participants in Study 1b did not expect honesty to be more or less meaningful ($M = 3.66$, $SD = 0.93$) than participants in Study 1a experienced it to be ($M = 3.79$, $SD = 0.67$), $p = .44$. Interestingly, Study 1b suggests that individuals may also underestimate the social benefits of honesty. Specifically, participants in Study 1b expected honesty to be less socially connecting ($M = 3.04$, $SD = 1.11$) than participants Study 1a experienced it to be ($M = 3.49$, $SD = 0.80$), $p < .02$. Although we did not find this pattern in Study 1a, it is possible that individuals who were more removed from the experience expected honesty to be more isolating than those who deeply considered what the impending experience would be like. Importantly, Study 1b suggests that individuals' misprediction of the hedonic and social consequences lead them to avoid being honest.

Discussion

In this study, we break new ground by exploring how honesty and kindness, two of the most basic moral principles and facets of human communication, influence psychological well-being. We conducted an intensive three-day field experiment in which individuals had to be honest or kind in all of their social interactions. Our findings make three central contributions to our understanding of human communication, morality, well-being, and affective-forecasting. First, we provide insight into why people avoid being honest with others. Our results suggest that individuals' aversion towards honesty is driven by an affective forecasting failure. Individuals expect honesty to be less pleasant than it is.

Second, we demonstrate that focusing on different moral principles during social communication differentially impacts hedonic and eudaimonic well-being. Focusing on kindness yields greater positive affect and interpersonal engagement, thereby promoting hedonia. Focusing on honesty yields greater self-expression and liberation, thereby promoting eudaimonia. Although the present research focuses exclusively on social communication, these findings likely apply to the broader distinction between justice and care (Levine & Schweitzer, 2014). Individuals who focus on impartial moral principles may experience greater meaning in life, whereas individuals who focus on care towards others may experience greater pleasure. Scholars have long claimed that morality promotes well-being, but to our knowledge, this is the first research to explore how different foundations of morality promote different types of well-being.

Finally, this research provides novel insights into individuals' ability to forecast experiences. Past research has focused solely on *affective* forecasting, concluding that individuals rarely have insight into the affective – or hedonic - consequences of future experiences (e.g., Gilbert et al., 1998; Wilson & Gilbert, 2005). Our findings are consistent with this body of research. However, we also find that individuals do not lack insight into the eudaimonic consequences of future experiences; individuals were much less inaccurate when predicting the meaning associated with our interventions. The forecasting literature has not explored this possibility. Perhaps individuals who experience human suffering – through breakups, death, and defeat (Gilbert, et al., 1998) – do recognize that with hardship comes meaning. Furthermore, perhaps this sense of meaning influences affect over time, which could contribute to adaptation. Our findings pave the way for future research to explore the interplay between forecasted and experienced hedonia and eudaimonia.

Practically, this research also highlights the promise of using short interventions to produce meaningful behavioral and psychological changes. These interventions may be particularly useful in organizations in which employees routinely struggle with the conflict between honesty and kindness. For example, coaches or managers who have to deliver negative feedback may improve their candor and find greater meaning in their work after engaging in short honesty interventions.

Limitations and future directions

Our initial study has a number of limitations that will be addressed in future experiments. First, this study was somewhat exploratory in nature. Although we made a-priori predictions regarding the hedonic and eudaimonic experiences of honesty and kindness, we collected many exploratory measures (see footnotes 14-17). Additionally, we had not conducted power analyses prior to running the experiment because we did not yet have a reasonable estimate of the effect size. As a result, our study was underpowered and some of our key results are of marginal significance. Furthermore, having individuals in Study 1a forecast the experience before engaging in our study may have influenced their reports of the experience. Our next study will address these limitations.

Specifically, in our ongoing research, we are only measuring hedonic and eudaimonic well-being and social connection. We are expanding our scales to be consistent with existing literature (Huta & Ryan, 2010; Hughes, Waite, Hawkley, & Cacioppo, 2004; Steger, Kashdan, & Oishi, 2008, Urry, et al., 2004), and we will not include superfluous measures. We are also increasing our sample size. We performed a sample-size calculation with the goal of achieving 80% power for the critical contrast between honesty and kindness on daily eudaimonic well-being. Estimating the effect size to be d = .62 based on the results of Study 1a, we will require 42 participants per cell. We intend to meet or exceed this sample size in the next study. Finally, we will not have experiencers forecast the experience, consistent with existing literature (e.g., Epley & Schroeder, 2015). In addition to addressing these limitations, we will expand our retrospective measures to better understand the long-term hedonic and eudaimonic consequences of our interventions and individuals’ desire to repeat them.

The purpose of our second study will be to document the robustness of our initial results. However, we have additional studies planned to address several important questions regarding the mechanisms underlying our effects. Although our 3-day intervention allowed us to examine the consequences of communication in everyday life, we do not have the level of control and precision necessary to make claims about the types of conversations that generated pleasure and meaning. It is possible, for example, that one or two significant self-disclosures generated high levels of meaning and that the remainder of the honest conversations were simply uncomfortable. Or it is possible that individuals only focused on self-disclosure, honestly sharing information about the self, rather than other-disclosure, honestly sharing evaluations of others. To explore this more deeply, we intend to manipulate whether individuals are directed to be honest about themselves or others in future studies.

Furthermore, we cannot yet identify the specific processes that led to an affective forecasting failure. It is possible that individuals misforecast others' reaction to their honesty, that individuals misforecast how others' reactions to honesty will impact them, or that individuals misforecast the very experience of self-expression. We may able to disentangle these mechanisms with future studies in which participants and their conversational partners provide judgments about these three processes.

It is also possible that individuals misforecast the experience of honesty because they imagine engaging in conversations that never occur during their three-day experience. For example, perhaps when considering the consequences of honesty, individuals imagine being asked difficult personal questions by threatening relational partners, or they imagine that they have the courage to confront individuals with their most ardent criticisms. But perhaps opportunities to engage in these conversations do not actually arise in the three-day experiment, or individuals actively avoid them. Indeed, some participants in our study did mention avoiding interpersonal interaction that might entail extremely negative honest conversations. To explore differences between participants' predicted and actual behaviors, we plan to ask future participants to generate a list of honest conversations they imagine engaging in, and then track whether or not the actual conversations arise during the study.

We also cannot be sure that every participant fully committed to the intervention and significantly altered their communication for three days. Although we took every effort to ensure commitment to the intervention and participants' free responses suggest that they took the intervention very seriously (see Table 3), it is difficult to confirm this without directly observing behavior. Future studies using controlled or video-taped interactions may be able to provide greater insight into the experience of honesty and kindness. Laboratory experiments may also help us overcome attrition and self-selection issues associated with our current recruitment procedure.

Our results also beg important questions for future research. In particular, it will be important to examine how honesty and kindness influence communicators' relational partners. Although the present research suggests that communicating honestly creates meaning for communicators, it is not clear that the targets of this communication appreciate it. In fact, recent research demonstrates that relational partners often resent painful honesty (Levine & Schweitzer, 2015). If one's goal in considering ethical behavior is to promote overall well-being, as many philosophers have argued it should be (e.g., Bentham, 1843/1948), it is essential to examine the consequences of one's behavior not only for the self, but also for others.

One factor that might influence whether honesty is well-received is whether relational partners are jointly committed to the goal of honesty. Shared honesty may promote intimacy and growth, but simply receiving honesty may be quite unpleasant. The type of relationship may also matter. Close friends may be able to withstand and benefit from difficult honest conversations, but professional or distant relationships may not. Power differences between relational partners may also matter. We may be able to gain initial insight into this question by coding the open-ended reflections during participants' nightly communication audits.

Finally, it is worth noting that honesty and kindness need not be in conflict. Our initial results confirm this, demonstrating that focusing on honesty does not necessarily decrease kindness. We compare and contrast honesty and kindness in the present research as a first step in exploring the relationships among different ethical principles, communication styles, and well-being. However, future research should examine the possibility and consequences of integrating honesty and kindness by instructing individuals to be honest, kindly.

Conclusion

Individuals often shy away from sharing difficult truths, fearing the hedonic costs of honesty. Our findings suggest this may be a mistake. Honesty is not as unpleasant as it seems, and in fact, can promote meaning and long-term growth. In other words, people can handle (speaking) the truth.

References

Anderson, N. H. (1968). Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, *9*(3), 272.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, *83*(6), 1423.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, *8*, 85-104.

Bentham, J. (1948). An introduction to the principles of morals and legislation. Oxford, United Kingdom: Blackwell. (Original work published 1843)

Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.

Emmons, R. A., & McCullough, M. E. (2003). Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. *Journal of Personality and Social Psychology*, *84*(2), 377.

Barry, B., & Rehel, E. M. (2013). Lies, Damn Lies, and Negotiation: An Interdisciplinary Analysis of the Nature and Consequences of Deception at the Bargaining Table. *Handbook of Research in Conflict Management (Elgar, 2014)*.

Blanton, B. (1996). *Radical honesty: How to transform your life by telling the truth*. Dell.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Butler, E. A., Egloff, B., Wlhelm, F. H., Smith, N. C., Erickson, E. A., & Gross, J. J. (2003). The social consequences of expressive suppression. *Emotion*, *3*(1), 48.

Dalio, Ray. "PRINCIPLES." (n.d.): n. pag. 2011. Web. <http://www.bwater.com/Uploads/FileManager/Principles/Bridgewater-Associates-Ray-Dalio-Principles.pdf>.

Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of Happiness Studies*, *9*(1), 1-11.

Dunn, E. W., Aknin, L. B., & Norton, M. I. (2008). Spending money on others promotes happiness. *Science*, *319*(5870), 1687-1688.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social psychology*, *75*(3), 617.

Goffman, E. (1967). On face-work. *Interaction ritual*, 5-45.

Hughes, M. E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys results from two population-based studies. *Research on aging*, *26*(6), 655-672.

Keyes, C. L., Shmotkin, D., & Ryff, C. D. (2002). Optimizing well-being: the empirical encounter of two traditions. *Journal of Personality and Social Psychology*, *82*(6), 1007.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social psychology*, *75*(3), 617.

Gino, F., Kouchaki, M., & Galinsky, A. D. (2015). The Moral Virtue of Authenticity How Inauthenticity Produces Feelings of Immorality and Impurity. *Psychological science*, *26*(7), 983-996.

Harris, S. (2011). *The moral landscape: How science can determine human values*. Simon and Schuster.

Huta, V., & Ryan, R. M. (2010). Pursuing pleasure or virtue: The differential and overlapping well-being benefits of hedonic and eudaimonic motives. *Journal of Happiness Studies*, *11*(6), 735-762.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology, 53*, 107-117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88-106.

Lyubomirsky, S., Sheldon, K. M., & Schkade, D. (2005). Pursuing happiness: the architecture of sustainable change. *Review of general psychology*, *9*(2), 111.

Margolis, J. D., & Molinsky, A. (2008). Navigating the bind of necessary evils: Psychological engagement and the production of interpersonally sensitive behavior. *Academy of Management Journal*, *51*(5), 847-872.

Molinsky, A., & Margolis, J. (2005). Necessary evils and interpersonal sensitivity in organizations. *Academy of Management Review*, *30*(2), 245-268.

Newton, Sheldon D. "Success and Honesty." *The Huffington Post*. TheHuffingtonPost.com, 10 June 2014. Web. 14 Mar. 2016.

<http://www.huffingtonpost.com/sheldon-d-newton/success-and-honesty_b_5482762.html>.

Otake, K., Shimai, S., Tanaka-Matsumi, J., Otsui, K., & Fredrickson, B. L. (2006). Happy people become happier through kindness: A counting kindnesses intervention. *Journal of Happiness Studies*, *7*(3), 361-375.

Plato (1976 [C4 BCE]) *Protagoras*, ed. and trans. C.C.W. Taylor, Oxford: Clarendon Press.

Rosen, S., & Tesser, A. (1970). On reluctance to communicate undesirable information: The MUM effect. *Sociometry*, 253-263.

Rosenfeld, L. B. (1979). Self-disclosure avoidance: Why I am afraid to tell you who I am. *Communications Monographs*, *46*(1), 63-74.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, *52*(1), 141-166.

Steger, M. F., Kashdan, T. B., & Oishi, S. (2008). Being good by doing good: Daily eudaimonic activity and well-being. *Journal of Research in Personality*, *42*(1), 22-42.

Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: a prospective study of the transition to college. *Journal of Personality and Social Psychology*, *96*(4), 883.

Tesser, A., & Rosen, S. (1972). Similarity of objective fate as a determinant of the reluctance to transmit unpleasant information: The MUM effect. *Journal of Personality and Social Psychology*, *23*(1), 46.

Tesser, A., Rosen, S., & Tesser, M. (1971). On the reluctance to communicate undesirable messages (the MUM effect): A field study. *Psychological Reports*, *29*(2), 651-654.

Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management*, *32*(6), 951-990.

Urry, H. L., Nitschke, J. B., Dolski, I., Jackson, D. C., Dalton, K. M., Mueller, C. J., ... & Davidson, R. J. (2004). Making a life worth living neural correlates of well-being. *Psychological Science*, *15*(6), 367-372.

Waterman, A. S. (1990). The relevance of Aristotle's conception of eudaimonia for the psychological study of happiness. *Theoretical and Philosophical Psychology*, 10, 39–44.

Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, 64, 678–691.

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting knowing what to want. *Current Directions in Psychological Science, 14*(3), 131-134.

Appendix A. Verbal Instructions in Study 1a - Recruitment

Please listen carefully.

This next study is optional and will occur outside of this lab session.

The study is about communication in everyday life. In this study, you will be asked to be very conscious of your interpersonal communication. We expect that as a result of participating in this study, you will learn about the way they communicate with and relate to others. However, you may be asked to communicate in ways that could cause discomfort. You should only participate if you are truly willing to be thoughtful about your communication and are open to communicating in different ways.

To participate, you will take an initial survey in this laboratory, which will take roughly 10 minutes. Then, you will learn more about the study. In order to participate, you will have to take nightly online surveys about your emotions, well-being, and relationships, which will each take about five minutes. This will last for three days. You will receive a survey each night, via email. Lastly, you will have to complete a final reflection survey, which will be emailed to you two weeks after the study ends.

This study will take place outside of this lab session and is *in addition* to the session you signed up for. You will be paid the $10 show-up fee for this session regardless of whether or not you enroll in this additional study.

However, in exchange for participating in this additional study, you will earn $20 and the chance to win an iPad mini. You will be paid $20 for your completion of the entire study – that means three nightly surveys, plus the two-week follow-up survey. Your $20 payment will be paid either directly to you by the experimenter, through paypal, or you can choose to receive a $20 amazon e-gift card instead.

In addition to the payment of $20, we will run a lottery for an iPad mini. Thus, you will also have a chance to win an iPad mini in exchange for your participation.

You cannot miss any surveys during this entire study. If you fail to complete a survey, you will not receive payment for this study.

Again, you should only join this study if you are willing to participate in a challenging 3-day study that will require daily surveys and may ask you to communicate with others in certain ways.

If you do not want to join this study, you can check out of the lab at this time.

Please take a moment to think about your decision. You are in no way obligated to participate in this research and you can choose to leave the study at any time. You can head to check out if you do not want to participate.

Appendix A. Verbal Instructions in Study 1a – Assignment to condition

<u>All conditions</u>
In this study, you will be asked to reflect upon your social communication. Often, speaking with others requires balancing honesty and kindness. Being completely open and honest about our thoughts, feelings, and opinions, can sometimes upset others and be unkind. Alternatively, being kind, considerate, and helpful towards others sometimes means not being 100% honest.

Control:
Throughout the next three days – that means today, tomorrow, and the following day - please be conscious of the way you communicate with others. Please act as you normally would throughout the length of this study. You should not change your behavior, but you should be conscious of it.

You should act as you normally would with your closest relational partners. However, you should NOT tell them, or anyone else, any specific information about this study. They can only know that you were asked to pay special attention to your interpersonal communication. After the study has ended, you can share any information you'd like about this study.

Please think about what it means to be conscious of your communication. Feel free to raise your hand if you have questions. [field questions, wait for a moment] Is everyone ready to continue? If so, you can complete the next link on your computer.

Honesty:

Throughout the next three days – that means today, tomorrow, and the following day - be honest in every conversation you have with every person you talk to. Really try to be completely candid and open when you are sharing your thoughts, feelings, and opinions with others. You should be honest in every conversation you have, in every interaction, with every person in your life. Even though this may be difficult, try your best to be honest.

Being authentic, honest, and true to oneself are important virtues. Embrace these virtues every day for the next three days. When someone asks you how you feel, tell them the truth. That means saying you feel happy only when you feel happy and saying you feel sad when you feel sad. When you are giving your opinion, be completely honest. You should provide positive opinions only when you truly feel positive, and you should provide negative opinions when you feel negative.

You should be particularly honest with your closest relational partners. However, you should NOT tell them, or anyone else, any specific information about these instructions. They can only know that you were asked to pay special attention to your interpersonal communication. After the study has ended, you can share any information you'd like about this study.

Please think about what it means to be completely honest. Feel free to raise your hand if you have questions. [field questions, wait for a moment] Is everyone ready to continue? If so, you can complete the next link on your computer.

Kindness:

Throughout the next three days – that means today, tomorrow, and the following day - please strive to be kind in every conversation you have with every person you talk to. Really try to be caring and considerate when you are sharing your thoughts, feelings, and opinions. You should be kind in every conversation you have, in every interaction, with every person in your life. Even though this may be difficult, you should do your absolute best to be kind.

Being kind and helpful, and avoiding harming others are important virtues. Embrace these virtues every day for the next three days. When someone asks you how you feel, give a kind answer. That means taking their feelings and state of mind into consideration. When you are giving your opinion, be kind. You should provide opinions kindly and focus on the needs and feelings of those around you.

You should be particularly honest with your closest relational partners. However, you should NOT tell them, or anyone else, any specific information about these instructions. They can only know that you were asked to pay special attention to your interpersonal communication. After the study has ended, you can share any information you'd like about this study.

Please think about what it means to be kind. Feel free to raise your hand if you have questions. [field questions, wait for a moment] Is everyone ready to continue? If so, you can complete the next link on your computer.

Tables

Table 1. Enrollment and attrition across conditions

	Assignment to condition	**Day 1**	**Day 2**	**Day 3**	**Follow Up**
Honesty	44	42	41	39	40
		95.5%	*93.2%*	*88.6%*	*90.9%*
Kindness	35	30	31	30	27
		85.7%	*88.6%*	*85.7%*	*77.1%*
Control	38	37	33	34	30
		97.4%	*86.8%*	*89.5%*	*78.9%*

Note. Percentages reflect the proportion of individuals assigned to condition that completed surveys at each subsequent time-point.

Table 2. The effects of honesty and kindness on long-term behavioral change

		Long-term honesty	Long-term kindness	Long-term conflict	Long-term improvement	Relational improvement
Honesty	*M*	4.81^a	4.33^a	3.025^a	4.90^a	4.62^a
	SD	*1.16*	*1.31*	*1.42*	*1.04*	*0.81*
Kindness	*M*	3.80^b	4.37^a	2.52^a	4.67^a	4.43^a
	SD	*1.15*	*1.64*	*1.05*	*1.12*	*0.69*
Control	*M*	4.20^b	4.03^a	2.90^a	4.20^b	4.23^a
	SD	*1.34*	*1.33*	*1.47*	*1.08*	*0.58*

Note. Letters within each column indicate significant differences at $p < .05$.

Table 3. Illustrative Quotes about the Experience of Honesty, Kindness, and Communication-Consciousness

Honesty	**Kindness**	**Communication-Consciousness**
• It was difficult but exciting. I felt uncomfortable at first communicating honestly with my coworkers.	• Communicating kindly helped me to think positively, see positively, and feel all around positive.	• I felt that this made me misspeak on fewer occasions and helped me manage expectations when delivering unfavorable news to others.
• It felt weird being so blunt. I'm generally a nice person and always consider others feelings even before my own. This was such a huge change. It took adjusting because I have a passive personality.	• Ordinarily, I am a very sarcastic person with a dry sense of humor. The things I say...are snarky. I tried pretty hard to tone that down today. This was somewhat challenging at times to remember to do.	• I felt like I was more cautious around people. This reflective process also affected my ability to respond to people. I will say that the experience made me feel better.
• It felt good to be honest, though the conversion itself was quite unpleasant. I thought of it as one of those necessary evils	• It was not difficult to be kind to others because I work in a service profession and try to think of others' feelings almost every day	• It was challenging talking to my ex because I was feeling a lot of strong emotions.
• I learned that I previously dedicated a lot of time and energy to stifling my own thoughts and feelings and its a weight off my shoulders now that I've begun trying to stop.	• I found that during the days where I explicitly tried to be only kind, I regretted fewer things I said or how I acted in certain situations.	• Being conscious of my communication has made me more aware of the different levels of friendship I have with people. I realized that I held back a lot to certain friends
• It effected my relationship with my boyfriend. I told him the truth about how i felt sometimes, which lead to our break up. Of course this was bound to happen eventually, I was glad that it happened now rather than later.	• My other struggle was with my boyfriend. ... we often bicker over small things. It was difficult keeping the conversation from turning into an argument and to get my point across while being kind and supportive.	• With my supervisor, I tried to focus on aligning our communication, because I tend to be brusque and get to the point, whereas he's much more old-school, polite, and roundabout, which sometimes makes me frustrated.
• I definitely feel a lot closer to people. Just being honest and opening up to people on a deeper level definitely brings you closer together. I love it!	• Overall I have found that the other appreciates such optimism, thoughtfulness, and kindness. Thus, I am trying to incorporate this positivity and optimism into my everyday behavior	• What also made this difficult was the fact that I went on a first date during this time period which was a bit different than my usually types of conversations with close friends.
• [I learned that] I lie a lot to people because it's easier than telling them the truth. Most people only know snippets of who I really am because I feel that they won't be able to handle everything my life has to offer.	• I feel the same with others as I did before the study, however, if anything I feel as if they view me in a more positive, respectable light.	• It didn't really change the way I interacted with people. However, I would be more conscientious of what I said and how I said it. I would also note how I often slurred my words or mumbled..

• By communicating honestly I found that I had much more meaningful conversations with my friends about anything and everything.	• It felt good to make kindness my ultimate goal, although it was a bit of a challenge to keep from stealing the spotlight of conversation back, which would have been rude and unkind.	• I've always communicated very consciously with my close friends, but I think I've begun to extend it to strangers and acquaintances. So I think my relationship with those individuals has changed for the better

Note. Participants responded to a free-response question in each nightly survey asking them to write about their experience and the extent to which they complied with the experiment. Participants also answered free-response questions about what they learned from the experiment and how it influenced their relationships during the two-week follow up survey. Table 3 presents examples of these responses.

Figures

Figure 1. The anticipated, actual, and retrospective effects of honesty and kindness on hedonic well-being

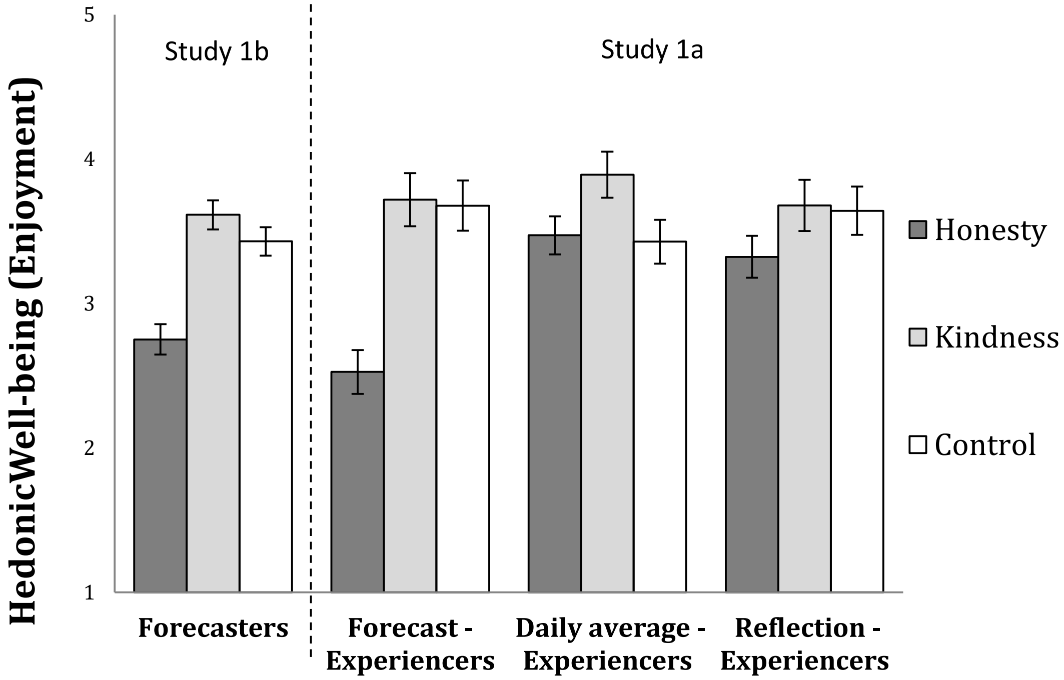


Figure 2. The anticipated, actual, and retrospective effects of honesty and kindness on eudaimonic well-being

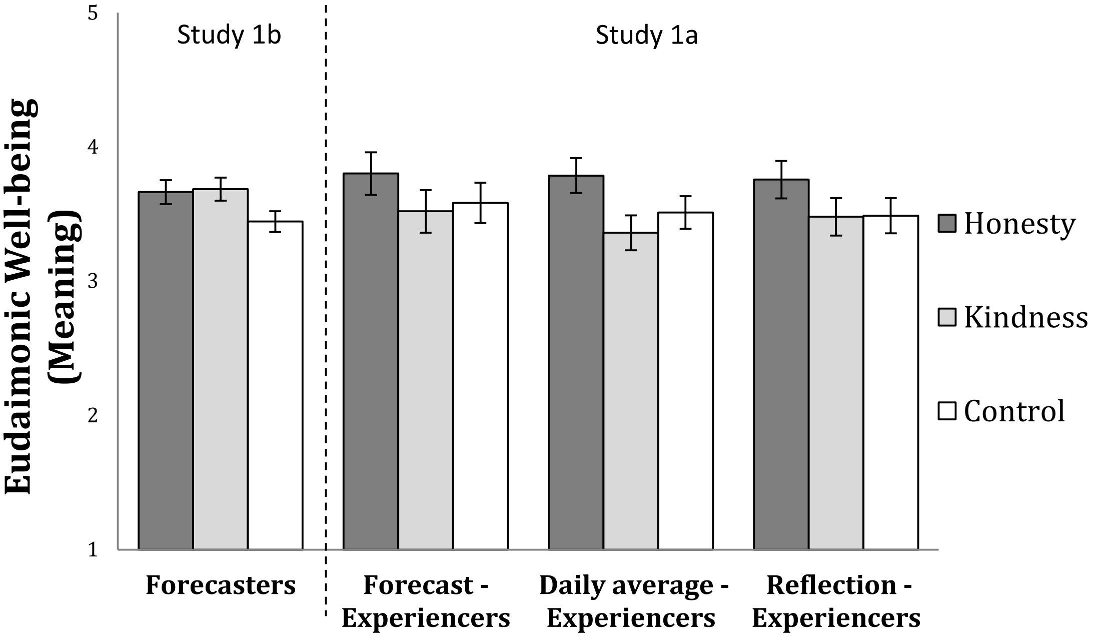
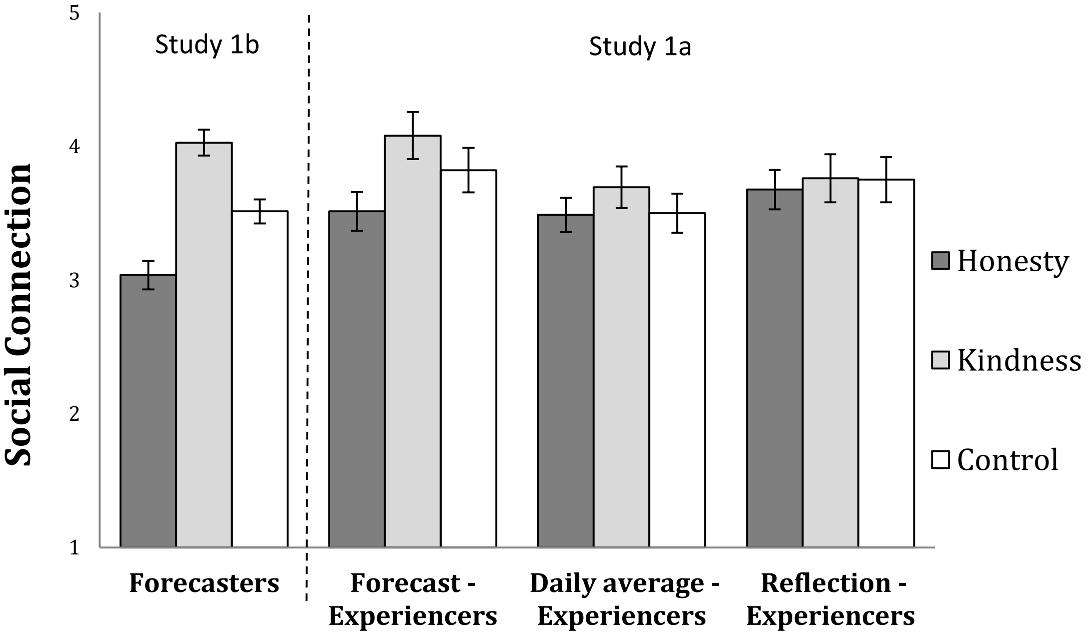


Figure 3. The anticipated, actual, and retrospective effects of honesty and kindness on social connection



CHAPTER FIVE

ON BENEFICENT DECEPTION: ASYMMETRIC PREFERENCES FOR LIES OF OMISSION AND COMMISSION IN HEALTHCARE COMMUNICATION

Emma Levine

Joanna Hart

Kendra Moore

Emily Rubin

Kuldeep Yadav

Scott Halpern

ABSTRACT

The use of deception in doctor-patient communication has long been debated in medical ethics. Although some have advocated for the use of beneficent deception – deception that promotes patient well-being – most scholars and practitioners prohibit it. However, no empirical research has investigated when physicians and their patients engage in and appreciate deception, or when they judge deception to be beneficent. The present research fills this gap. We study physicians', patients', and healthy adults' moral judgments and preferences for deception and we document a robust asymmetry between physicians' and patients' preferences for different forms of deception. Specifically, physicians believe that it is more ethical to lie by omission (i.e., withhold information) than to lie by commission (i.e., provide false hope),

whereas patients often believe the opposite. We document this asymmetry across multiple clinical circumstances with real cancer patients and oncologists and we discuss the psychological and practical implications of this research for medicine, behavioral ethics, and human communication.

ON BENEFICENT DECEPTION:
ASYMMETRIC PREFERENCES FOR LIES OF OMISSION AND COMMISSION IN HEALTHCARE COMMUNICATION

Imagine a patient with terminal cancer. The patient’s cancer is no longer reacting to chemotherapy and the physician knows that the patient is very unlikely to have any other treatment options available to them. The patient has already prepared for the worst, but remains optimistic and wants to pursue any and all options that might prolong their life. The physician must decide what information to share with this patient at this time. Should the physician honestly tell the patient that they have run out of treatment options? Perhaps the physician should say nothing and allow the patient to maintain the illusion of hope. Or perhaps the physician should lie to the patient, saying they too are optimistic about the possibility of future treatment options.

Physicians face these types of ethical dilemmas every day. They must decide how to communicate with vulnerable patients during some of the most challenging and distressing moments in their lives. These decisions are particularly difficult because they reflect a key conflict between two principles of medical ethics: autonomy and beneficence (Beauchamp & Childress, 2003). Autonomy reflects the patient’s right to be fully informed and to make their own decisions. Beneficence reflects the need to promote the patient’s well-being (Gillon, 1994). In the opening example, a physician may choose to be completely honest with the intention of helping the patient make a fully informed decision about how to live the rest of their life. Such honesty, however, often has emotional costs. Thus, a physician may instead choose to engage in some form of deception, by either omitting information or by actively lying to the patient with the intention of preventing emotional distress and promoting psychological well-being during the patient’s final days.

Professional organizations (American Medical Association Code of Ethics, 2006; World Medical Association International Code of Ethics, 2016) and ethicists (Apatira et al., 2008; Beste, 2005; Herring & Foster, 2012; Sarafis, Tsounis, Malliarou, Lahana, 2014) primarily advocate for honesty, suggesting that physicians should prioritize patient autonomy over beneficence. Although some ethicists and practitioners suggest that omission, withholding information until a more appropriate time, is also a reasonable course of action (American Medical Association Code of Ethics Opinion 8.082, 2006), very few advocate for the active use of deception.

These medical guidelines, however, are based on normative assumptions about preferences for and consequences of deception. For example, existing scholarship assumes that patients would rarely, if ever, consent to being deceived (e.g., Bakhurst, 1992; Bok, 1978; Gillon,1994) and that deception will have long-term costs for patient health and eventually erode trust in the doctor-patient relationship (Jackson, 1991). However, empirical data is needed to understand whether these assumptions are correct. Without examining the consequences of different ethical decisions, and patients’ preferences for different ethical principles, we cannot possibly know whether the normative assumptions that guide practice actually promote effective medical practice. In the present research, we fill this gap by examining patients’ and physicians’ judgments of and preferences for deception.

Specifically, we focus on two research questions. First, we explore whether different stakeholders (i.e., doctors, patients, and potential surrogates) have different beliefs about the acceptability and beneficence of deception in healthcare communication. Second, we examine how these stakeholders perceive different types of deception (i.e., lies of omission and commission) within healthcare communication.

Answering these questions deepens our understanding of medical ethics, moral judgment, and human communication and has important practical and theoretical implications. Practically, we document a robust asymmetry between physicians' and patients' preferences for different forms of deception. We demonstrate that these stakeholders have divergent beliefs about the acceptability of lying to provide false hope or manage a patient's anxiety. If physicians and patients have fundamentally different beliefs about the type of communication that is acceptable, this may lead to predictable miscommunication, conflict, and distrust.

Theoretically, this work sheds light on the egocentric biases that guide communicators' and targets' preferences for deception across contexts. We posit that communicators focus on the psychological costs of deception when making judgments of what is right and wrong, whereas targets focus on the benefits of deception to them. In the healthcare context, the costs of deception may include concerns about violating rules of the profession and fear of liability, and the benefits of deception may include patient hope, comfort, and optimism. However, across contexts, communicators may overweigh the guilt of lying, missing the opportunity to provide their conversational partners with emotional support. Thus, this work paves the way for future research on asymmetric evaluations of lies that are intended to help others. These dynamics are likely to be particularly important in conversations that involve balancing honesty with comfort, such as discussions of layoffs, poor performance, or social rejection.

Asymmetric preferences among physicians and patients

In the present research, we conceptualize deception as *any act that intentionally misleads the target.* Thus, deception may include the intentional omission of information, or the intentional provision of false information. We consider omission to be deceptive when it is motivated by a desire to maintain a patient's existing illusion, or to hide new information.

We limit our investigation to circumstances in which honesty is unpleasant. In these circumstances, individuals may use deception with beneficent intentions: to protect the patient from despair and promote the patient's psychological well-being. Indeed, past research has demonstrated that individuals are unwilling to justify selfish deception, but justify and welcome beneficent, or prosocial, deception quite often (Levine & Schweitzer, 2014, 2015; Richard, Lajeunesse, & Lussier, 2010).

We focus the present investigation on understanding perceptions of different types of deception. Existing research on beneficent deception and medical ethics does not always distinguish between lies of omission and lies of commission. However, we expect this distinction to matter. Specifically, we predict that the parties involved in healthcare communication view these two forms of deception very differently. Put formally, we expect the perceived acceptability of lies of commission relative to lies of omission to be moderated by role.

First, we expect physicians to judge lies of commission as less acceptable than lies of omission. We assume that physicians, like most individuals, are motivated to behave ethically (Aquino & Reed, 2002). When individuals engage in unethical behavior, including deception, they experience psychological costs such as guilt and shame (Cohen, Wolf, Panter, & Insko, 2011; Tangney, Stuewig, & Mashek, 2007). Lies of commission may be particularly likely to elicit these negative feelings because they reflect an intentional, active behavior. Indeed, prior work has demonstrated that acts of commission are perceived to be more intentional, harmful, and blameworthy than acts of omission (Alicke, 2000; Cushman, 2008; Spranca, Minsk, & Baron, 1991).

These concerns may be intensified in the medical context because doctors are explicitly advised not to engage in active deception. Physicians may internalize this advice and see lying as inconsistent with their medical duties. They may also be concerned about the potential legal ramifications of actively misleading patients (Herring & Foster, 2012). Thus, we expect that physicians will perceive lies of commission as less acceptable than lies of omission.

We do not expect that patients and potential patients (i.e., healthy adults) will always share this belief. Beneficent lies of commission, despite reflecting a more severe transgression from the perspective of the communicator, may provide greater benefits to the target than lies of omission. Omission itself may cause harm to patients. Specifically, the omission of information is likely to leave patients feeling uncertain. Scholars in many different domains, including economics, cognitive psychology, and medicine have demonstrated that individuals are generally averse to the experience of uncertainty (Dow & da Costa Werlang, 1999; Epstein, 1999; Fox & Tversky, 1995; Fox & Weber, 2002; Politi, Han & Col, 2007). Thus, individuals may resent lies of omission because it prolongs this negative state. Patients are likely to become particularly distressed and confused if their doctor omits information that they had been expecting. Beneficent lies of commission, however, can resolve the aversive experience of uncertainty, at least in the short run, and may improve the patient's psychological experience. It is important to note that honesty also resolves uncertainty, and is likely to be seen as more acceptable than either form of deception. However, perceptions of honesty are not the focus of the current framework.

The proposition that communicators and targets will judge lies of omission and commission differently is consistent with existing research on actors' and recipients' asymmetric evaluations of prosocial behaviors (Zhang & Epley, 2009). Actors focus on costs when evaluating their prosocial actions, whereas recipients focus on the benefits. For example, when exchanging gifts, gift-givers focus on how much they spent on the gift but gift-receivers focus on how much the gift benefited them. We expect the same egocentrism to influence moral judgments of deception. We expect communicators to focus on the potential costs of deception to them, and thus judge lies of commission to be less acceptable than lies of omission. But, we expect targets to focus on how deception benefits them, and thus judge lies of commission to be more acceptable, at least in some cases, than lies of omission.

Overview of study

To test this hypothesis, we examined physicians', patients' and healthy adults' judgments of deception during difficult healthcare conversations. Although our predictions pertain primarily to patients and physicians, we also examine healthy adults to examine whether our effects are unique to the experience of being ill, or whether they generalize to anyone who takes the perspective of the patient.

In this study, we examine judgments of deception by omission and commission across four hypothetical conversations between an oncologist and a cancer patient. We focus on cancer because most individuals have some level of exposure to cancer, and because it is a setting in which the tension between honesty and beneficence is particularly common and intense (Surbone, 2006).

We had participants rate the acceptability and beneficence of omission, commission, and honesty in each conversation. Although our theory focuses on the distinction between omission and commission, we include honesty for completeness. We measure both acceptability and beneficence to distinguish between two possible sources for an aversion towards beneficent deception. On one hand, individuals may not actually see deception as beneficent. In other words, even when deception could presumably provide hope, it may not be seen as improving patient overall well-being, and thus, not consistent with the value of beneficence. On the other hand, individuals may see deception as unacceptable despite believing that deception is sometimes beneficent (i.e., promotes patient well-being). Distinguishing between these two possibilities helps us understand whether individuals see communication as reflecting a tradeoff between different medical obligations, and which of these obligations more heavily influences preferences.

Method

Participants. We recruited 60 participants for this study: 20 healthy adults, 20 oncologists, and 20 cancer patients. All participants were recruited from the same geographic region (a city in the Northeast region of the United States).

Healthy adults (60% female, Mean age = 33) were recruited by a university laboratory. To participate, individuals had to be over 18 years of age, and non-students. We recruited participants to arrive to a laboratory in 30-minute increments. They completed our study one at a time, in a private focus room and received $20 in exchange for their participation.

Oncologists (60% female, Mean age = 43) were recruited by email. We reached out to oncologists that members of the research team knew personally, or that practiced at university-affiliated hospitals. Oncologists received $50 in exchange for their participation. We scheduled appointments with the oncologists and administered the study in their offices.

Patients (45% female, Mean age = 58) were recruited at a university-affiliated hospital. We recruited patients with any cancer at any stage. Patients received $20 in exchange for their participation. When we recruited oncologists, we asked them for permission to approach their patients. If oncologists consented, we approached their patients during their chemotherapy infusions, or while they were waiting for infusion. We administered the study to consenting patients in private infusion suites, while patients were receiving their infusion.

Procedure and materials. All participants judged four clinical scenarios. Data collection included qualitative data and survey responses. Each participant met with a member of our research team in a private space and answered open-ended questions verbally, and answered questions using an iPad with Qualtrics enabled. Each participant's verbal responses were audio-recorded.

This approach allowed the research team to gain insight into participants' thought process, and allowed participants to ask for clarification and discuss the scenarios openly. To ensure this approach did not alter our survey results, we ran a replication of the present study in which the entire survey was completed on the computer and no members of the research team were present.[19]

Clinical scenarios. We presented participants with four clinical scenarios that were designed to reflect difficult clinical conversations that might occur between a physician and a patient over the course of a patient's illness. After reading each scenario, participants were prompted to say, aloud, how they would want their doctor to communicate [physician perspective: how they would communicate]. These responses were audio-recorded. Then, participants judged three communication tactics that a physician might use in each scenario: honesty, omission, or commission. We presented the scenarios in a fixed order to simulate the actual trajectory of conversations that a cancer patient is likely to face (see order below). We randomized the order in which honesty, omission, and commission appeared within each scenario. Patients and healthy adults read scenarios that were written from patient perspective. Oncologists read identical scenarios, except they were written from the physician's perspective.

[19] The replication of Study 1 consisted of identical scenarios and questions as Study 1. However, we administered the survey virtually to a sample of healthy adults on Amazon Mechanical Turk who took perspective of patients (N = 40), and a convenient sample of medical students and trainees (N = 47) who took perspective of physicians. We conceptually replicate all results.

Thus, the full design of the study was a 3(Population: healthy adult, patient, physician) x 3(Communication tactic: Honesty, Omission, Commission) x 4(Scenario). *Population* was a between-subjects factor, and all other factors were manipulated within subjects. The patient perspectives of the four scenarios and their corresponding communication tactics appear below. We add bold text below to indicate the key manipulations.

1. Scenarios 1: Suspicious scan

> *Imagine you go to the emergency room with abdominal pain. It turns out that acid reflux was causing the pain, but while you were there you had a CT scan of your abdomen to help rule out something more serious. Several days later, you follow up with your primary care doctor, who notices that the CT scan shows an abnormal spot on your liver. The spot could be cancerous and the doctor believes that it probably is, but he cannot be certain without additional tests. The spot does not cause any symptoms and it is very unlikely that it will cause any immediate problems. You do not yet know that the doctor has noticed a spot, and you have no reason to suspect that you are ill. The doctor will be more confident in a diagnosis after doing some blood work.*
>
> *[Omission]*
>
> *Imagine the doctor decides* ***not to tell you any information about the spot on your liver*** *at this time. He orders blood work to help confirm the diagnosis, but he tells you the blood work is just "routine."*
>
> [*Commission]*
>
> *Imagine the doctor decides to* ***tell you that everything with your scan*** *is fine at this time. He orders blood work to help confirm the diagnosis, but he tells you the blood work is just "routine."*
>
> *[Honesty]*
>
> *Imagine the doctor decides to* ***tell you that there is an abnormal spot on your liver, that he believes it is likely cancer, but it could also be a benign tumor****. He orders blood work to help confirm the diagnosis.*

2. Scenario 2: Prognosis

> *Imagine you were recently diagnosed with lung cancer. Your doctor has not yet provided you with any prognostic information. However, you've known a*

handful of people in similar medical circumstances who all made full recoveries, so you are optimistic that your chance of recovery is quite high. You have no reason to believe that your prognosis will be unfavorable, and your doctor knows that you are reasonably optimistic.

After looking over all of your test results, the doctor determines that the five-year survival rate for people with your stage of disease (meaning the percentage of patients who are alive five years after diagnosis) is 5%. There is a small chance that your prognosis could actually be much better, however, depending on the results of a test that has not yet come back.

[Omission]

Imagine the doctor decides ***not to provide you with any prognostic information*** *at this moment.*

[*Commission]*

Imagine the doctor decides to ***tell you that there is a good chance you will survive for several years.***

[Honesty]

Imagine the doctor decides to ***tell you that to the best of his knowledge, the five-year survival rate for people with your stage of disease is 5%.***

3. *Scenario 3: Treatment effects*

Imagine you were recently diagnosed with lung cancer that has spread to your liver. Although your prognosis is not optimistic, it is important to you to try all possible courses of action to fight your cancer and prolong your life.
You are visiting a doctor to discuss your treatment options. You are simply discussing your options today – *you will make a decision about your treatment during a future appointment. You discuss chemotherapy, which you believe can eventually eliminate your cancer. The doctor knows that chemotherapy will not be curative. That is, although chemotherapy could potentially prolong your life, it will never cure your cancer.*

[Omission]

Imagine the doctor decides ***not to say anything about the curative nature of chemotherapy at this moment.***

[*Commission]*

Imagine the doctor decides to ***tell you that chemotherapy could cure your cancer.***

[Honesty]

> *Imagine the doctor decides to* ***tell you that chemotherapy will never cure your cancer.***

4. Scenario 4: Future treatment options

> *Imagine you are visiting a doctor to discuss the progress of your lung cancer, and the effectiveness of your chemotherapy and radiation treatment. You were diagnosed with lung cancer seven months ago. The doctor informs you that your treatment has not been effective and that the cancer has now spread to your bones and brain. The doctor knows that there are no existing treatment options that will be effective for prolonging your life and the doctor does not expect any new treatments to be approved within the timeframe that the doctor expects you to live.*
>
> *You still have hope that new options will become available, or that you could qualify for a medical trial in the next few months. It continues to be important to you to try all possible courses of action to prolong your life.*
>
> *[Omission]*
>
> *Imagine the doctor decides* ***not to say anything about whether or not you will have new treatment options available in the future.***
>
> [*Commission]*
>
> *Imagine the doctor decides to* ***tell you might have new treatment options available to you in the future.***
>
> *[Honesty]*
>
> *Imagine the doctor decides to* ***tell you that there are no more options available to stop the spread of the cancer.***

The scenarios were reviewed and revised by pulmonary and critical care faculty as well as researchers without medical backgrounds to ensure understandability and fidelity to real clinical situations. The scenarios were designed to be as realistic as possible and to feature details (e.g., the patient’s desire to prolong his life) that would prompt participants to consider the potential benefits of hope. All scenarios also described circumstances in which there was momentary uncertainty that would be resolved sometime in the future, and thus, described circumstances in which physicians might believe that deception is a reasonable communication tactic.

Dependent variables. Participants answered six questions in response to each communication tactic (omission, commission, honesty), within each of the four scenarios. Participants judged the ethicality of the communication tactic using two items: “How ethical is this behavior?” (1 = completely unethical, 7 = completely ethical) and “This behavior would violate your autonomy [Physician perspective: This behavior would violate the patient’s autonomy]” (1 = strongly disagree, 7 = strongly agree). Participants also judged the desirability of the communication tactic by rating their agreement with one item: “I would want my doctor to behave this way [Physician perspective: I would behave this way]” (1 = strongly disagree, 7 = strongly agree). These three items loaded together on a single factor in an exploratory factor analysis (Principal axis factoring, Varimax rotation). Thus, we combined them into a single measure of acceptability (α = .90).

Participants also rated the beneficence of each communication tactic using three items: “This behavior would spare the patient from anxiety and fear”, “This behavior would promote the patient’s well-being”, “This behavior would improve the patient’s quality of life” (1 = strongly disagree, 7 = strongly agree). These three items also loaded together on a single factor in an exploratory factor analysis (Principal axis factoring, Varimax rotation). Thus we combined them into a single measure of perceived beneficence (α = .89).

At the end of the study, all participants received a signed copy of their consent form and contact information for the research team after completing the study. No personal identifiers were collected in the survey. A professional transcriptionist transcribed participant interviews and all personal identifiers were removed from the interview transcripts.

Results

We focus only on responses to the four clinical vignettes in the present manuscript. We are currently coding the patient-physician conversations. We conducted mixed within-between subject ANOVAs on acceptability and perceived beneficence, using *Communication Tactic* (omission, commission, honesty) as the within-subjects factor and *Population* (healthy adult, patient, physician) as the between-subjects factor. In our main analyses, we include *Scenario* as a covariate. The effects are unchanged if we do not control for *Scenario.*

Acceptability. We found a main effect of *Communication Tactic, F*(2, 236) = 138.39, $p < .001$, $\eta_p^2 = .37$, on perceived acceptability, such that honesty ($M = 5.88$, $SD = 1.24$) was seen as more acceptable than both omission ($M = 2.77$, $SD = 1.54$, $t(159) = 22.37$, $p < .001$) and commission ($M = 3.02$, $SD = 1.96$, $t(159) = 19.72$ $p < .001$). Furthermore, commission was perceived to be marginally more acceptable than omission ($t(159) = 1.93$, $p = .05$).

We also found a main effect of *Population*, $F(2, 236) = 8.41$, $p < .001$, $\eta_p^2 = .07$; such that healthy adults ($M = 4.19$, $SD = .45$) rated the communication tactics as more acceptable than physicians ($M = 3.62$, $SD = .45$, $t(39) = .09$, $p < .001$) and patients ($M = 3.87$, $SD = .45$, $t(39) = 2.29$, $p = .02$) did. Patients also rated the communication tactics as marginally more acceptable than physicians did (t(39) = 1.80, p = .07).

Importantly, these effects were qualified by a significant *Population* x *Communication Tactic* interaction, F(2,236) = 14.42, p < .001, η_p^2 = .11. Consistent with our hypothesis, physicians judged commission to be *less* acceptable than omission (p = .02) but patients judged commission to be *more* acceptable than omission (p < .01). Healthy adults did not judge commission and omission differently (p = .45). All thee populations judged honesty to be more acceptable than either form of deception (ps < .001). We depict this pattern of results in Figure 1.

--Figure 1 and Figure 2 here--

Perceived beneficence. We found a main effect of *Communication Tactic,* F(2, 236) = 13.81, p < .001, η_p^2 = .06, such that honesty (M = 4.14, SD = 1.81) was perceived to be more beneficent than both omission (M = 2.89, SD = 1.50, t(159) = 7.44, p < .001) and commission (M = 3.50, SD = 1.79, t(159) = 3.66, p < .001). Commission was also perceived to be more beneficent than omission (t(159) = 5.42, p < .001).

We also found a main effect of *Population*, F(2, 236) = 7.96, p < .001, η_p^2 = .06, such that healthy adults (M = 3.84, SD = .46) rated the communication tactics as more beneficent than physicians (M = 3.30, SD = .46, t(39) = 3.70, p < .001) and patients (M = 3.38, SD = .46, t(39) = 3.14, p = .002) did. There was no difference between physicians and patients (t(39) = .56, p = .58).

We do not find a significant *Population* x *Communication Tactic* interaction, $F(2,236) = .14$, $p = .87$, $\eta_p^2 = .001$. However, as shown in Figure 2, there were patients and physicians had very different evaluations of lies of commission and omission. Specifically, physicians rated omission and commission as equally beneficent (p = .37), whereas patients rated commission as significantly more beneficent than omission ($p < .001$).

Discussion

In this study, we gain initial insights into stakeholders' beliefs about the acceptability and beneficence of deception in healthcare communication. Importantly, we find that honesty is generally perceived to be more acceptable and beneficent than deception. This finding is consistent with recent physician surveys (e.g., Huang et al., 2015) and suggests that existing guidelines prioritizing honesty may be well-informed. Existing guidelines that suggest omission is more acceptable than commission, however, may be misinformed. Consistent with our hypothesis, we identify an asymmetry between physicians' and patients' judgments of lies of omission and lies of commission. Physicians generally believed that it was less acceptable to lie by commission than omission, but patients believed the opposite. Interestingly, we find that physicians judge lies of commission and omission to be equally beneficent. The discrepancy between physicians' judgments of acceptability and beneficence suggests that physicians may be influenced by their personal concerns about lying, rather than the desire to promote patient well-being. Patients' judgments of acceptability and beneficence largely followed the same pattern.

This research highlights how physicians and patients see ethical dilemmas differently. This is important because these asymmetries may be the source of distrust and miscommunication. For example, a patient may see a physician as immoral and lose trust in him if he fails to provide (even false) hope, which is a consequence physicians are unlikely to anticipate. To overcome this asymmetry, medical communication training should encourage physicians to seek out patient preferences rather than omitting information altogether.

Limitations and future directions

Our initial study was exploratory in nature. Two key strengths of this study were the realism of the vignettes and the use of clinical populations to test our hypotheses. This gives us confidence that these effects do exist within actual healthcare conversations. However, it will be important for future work to more precisely tease apart the mechanisms underlying and boundary conditions surrounding our effects. For example, future work should more carefully control the differences between omission and commission and present knowledge and future outcomes. Small differences in language may significantly alter perceptions of commission. For example, saying “you’re probably fine” is much different than saying “you do not have cancer.”

Future research should also delve deeper into the differences between physicians and patients. We recently ran a study to examine whether the asymmetry between physicians and patients is driven by structural or individual differences (e.g., liability concerns in medicine, medical training, comfort with uncertainty) or whether it is driven by the perspective difference between communicators and targets, as we propose. In this study, we randomly assigned healthy adults to evaluate lies of omission and commission from the perspective of either the physician or the patient. Our initial results reveal that the differences between physicians and patients are driven by perspective differences; we replicate the pattern of results from our first study with a simple perspective manipulation. We also find that individuals in the perspective of the communicator (i.e., the physician) focus more intensely on the guilt associated with lying, whereas individuals in the perspective of the target (i.e., the patient) focus on the benefits of hope. These results suggest that examining perspective differences in preferences for lies of omission and commission across contexts is a valuable endeavor for future research.

The present research also raises important questions that are beyond the scope of the present investigation. For example, will these effects hold across cultures? The United States healthcare system tends to prioritize autonomy above beneficence, but this is not the case across the world (Shahidi, 2010). It will be interesting to examine how physicians and patients respond to these dilemmas in cultures that embrace a more paternalistic model of healthcare.

Finally, future work should examine whether perceptions of beneficence correspond with reality. We believe that it is valuable to study perceptions of beneficence to gain insight into sources of miscommunication and distrust in the doctor-patient relationship. However, to confidently make recommendations to clinicians, we must understand when lies actually promote health and psychological well-being and when they do not. Future research could examine how patients fare, based on the practices exhibited and endorsed by their physicians.

Concluding thoughts

In medicine, ethical principles are in place to ensure the protection and well-being of patients. Surprisingly, we know very little about the ethical principles that patients care about and how current ethical guidelines affect patient well-being. In the present research, we explore patient perceptions of autonomy and beneficence by examining the case of beneficent deception. We document asymmetries between patient and physician preferences for beneficent lies of omission and commission, suggesting that physicians' moral proclivities may not accommodate patients' desire for hope. This research highlights the promise and importance of studying moral judgment in the medical domain and we hope it opens the door for future research on this topic.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, *126*(4), 556.

“AMA's Code of Medical Ethics." *AMA's Code of Medical Ethics*. American Medical Association, n.d. Web. 13 Mar. 2016. <http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics.page>.

Apatira, L., Boyd, E. A., Malvar, G., Evans, L. R., Luce, J. M., Lo, B., & White, D. B. (2008). Hope, truth, and preparing for death: perspectives of surrogate decision makers. *Annals of Internal Medicine*, *149*(12), 861-868.

Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of personality and social psychology*, *83*(6), 1423.

Beauchamp T. & Childress J. (2003) *Principles of Biomedical Ethics*. Oxford University Press, New York, NY, USA.

Beste, J. (2005). Instilling hope and respecting patient autonomy: Reconciling apparently conflicting duties. *Bioethics*, *19*(3), 215-231.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*(2), 353-380.

Dow, J., & da Costa Werlang, S. R. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica: Journal of the Econometric Society*, 197-204

Duffy, T. P. (1987). Agamemnon's fate and the medical profession. *W. New Eng. L. Rev.*, *9*, 21.

Epstein, L. G. (1999). A definition of uncertainty aversion. *The Review of Economic Studies*, *66*(3), 579-608.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics*, 585-603.

Fox, C. R., & Weber, M. (2002). Ambiguity aversion, comparative ignorance, and decision context. *Organizational Behavior and Human Decision Processes*, *88*(1), 476-498.

Gillon, R. (1994). Medical ethics: four principles plus attention to scope. *Bmj*,*309*(6948), 184.

Gillon, R. (2003). Ethics needs principles—four can encompass the rest—and respect for autonomy should be “first among equals”. *Journal of Medical Ethics*,*29*(5), 307-312.

Goodyear-Smith, F., & Buetow, S. (2001). Power issues in the doctor-patient relationship. *Health Care Analysis*, *9*(4), 449-462.

"Health Expenditure, Total (% of GDP)." *Health Expenditure, Total (% of GDP)*. The World Bank Group, 2016. Web: http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS. 13 Mar. 2016.

Herring, J., & Foster, C. (2012). Please don't tell me. *Cambridge Quarterly of Healthcare Ethics*, *21*(01), 20-29.

Huang, H. L., Cheng, S. Y., Yao, C. A., Hu, W. Y., Chen, C. Y., & Chiu, T. Y. (2015). Truth Telling and Treatment Strategies in End-of-Life Care in Physician-Led Accountable Care Organizations: Discrepancies Between Patients' Preferences and Physicians' Perceptions. *Medicine*, *94*(16).

Iezzoni, L. I., Rao, S. R., DesRoches, C. M., Vogeli, C., & Campbell, E. G. (2012). Survey shows that at least some physicians are not always open or honest with patients. *Health Affairs*, *31*(2), 383-391.

Jackson, J. (1991). Telling the truth. *Journal of Medical Ethics*, *17*(1), 5-9.

Kaptchuk, T. J., Friedlander, E., Kelley, J. M., Sanchez, M. N., Kokkotou, E., Singer, J. P., ... & Lembo, A. J. (2010). Placebos without deception: a randomized controlled trial in irritable bowel syndrome. *PloS one*, *5*(12), e15591.

Levine, E.E. Community standards of deception, *Working paper.*

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107-117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88-106.

Liu, P. H., Landrum, M. B., Weeks, J. C., Huskamp, H. A., Kahn, K. L., He, Y., ... & Keating, N. L. (2014). Physicians' propensity to discuss prognosis is associated with patients' awareness of prognosis for metastatic cancers. *Journal of Palliative Medicine*, *17*(6), 673-682.

Lupoli, M. Levine, E.E., & Greenberg, A. Paternalistic lies, *Working paper*.

"Opinion 8.082 - Withholding Information from Patients." *Opinion 8.082 - Withholding Information from Patients*. American Medical Association, Nov. 2006. Web: http://www.ama-assn.org/ama/pub/physician-resources/medical-ethics/code-medical-ethics/opinion8082.page. 13 Mar. 2016.

Peterson, C. (1999). 15 Personal Control and Well-Being. *Well-Being: Foundations of Hedonic Psychology: Foundations of Hedonic Psychology*, 288.

Politi, M. C., Han, P. K., & Col, N. F. (2007). Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making,27*(5), 681-695.

Richard, C., Lajeunesse, Y., & Lussier, M. T. (2010). Therapeutic privilege: between the ethics of lying and the practice of truth. *Journal of Medical Ethics*, *36*(6), 353-357.

Sarafis, P., Tsounis, A., Malliarou, M., & Lahana, E. (2014). Disclosing the truth: a dilemma between instilling hope and respecting patient autonomy in everyday clinical practice. *Global Journal of Health Science*, *6*(2), 128.

Scheier, M. F.; Carver, C.S.; Bridges, M.W., & Chang, E.C. (Ed), (2001). Optimism & pessimism: Implications for theory, research, and practice. , (pp. 189-216). Washington, DC, US: American Psychological Association, xxi, 395 pp.http://dx.doi.org/10.1037/10385-009.

Shahidi, J. (2010). Not telling the truth: circumstances leading to concealment of diagnosis and prognosis from cancer patients. *European Journal of Cancer Care*, *19*(5), 589-593.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of experimental social psychology*, *27*(1), 76-105.

Surbone, A. (2006). Telling the truth to patients with cancer: what is the truth?. *The Lancet Oncology*, *7*(11), 944-950.

Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*,*14*(4), 340.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual review of psychology*, *58*, 345

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin*, *103*(2), 193.

Will, J. F. (2011). A brief historical and theoretical perspective on patient autonomy and medical decision making: part I: the beneficence model. *Chest Journal*, *139*(3), 669-673.

WMA International Code of Medical Ethics. World Medical Association, n.d. Web: http://www.wma.net/en/30publications/10policies/c8/. 13 Mar. 2016.

Zier, L. S., Sottile, P. D., Hong, S. Y., Weissfield, L. A., & White, D. B. (2012). Surrogate decision makers' interpretation of prognostic information: a mixed-methods study. *Annals of Internal Medicine*, *156*(5), 360-366.

Zhang, Y., & Epley, N. (2009). Self-centered social exchange: differential use of costs versus benefits in prosocial reciprocity. *Journal of Personality and Social Psychology*, *97*(5), 796.

Figures

Figure 1. Perceptions of Acceptability by Communication Tactic and Population

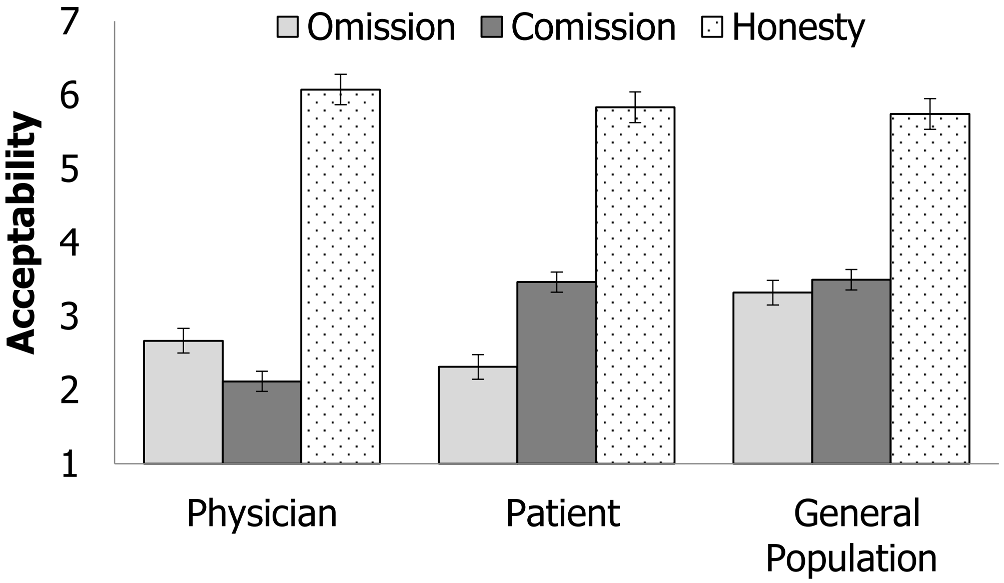
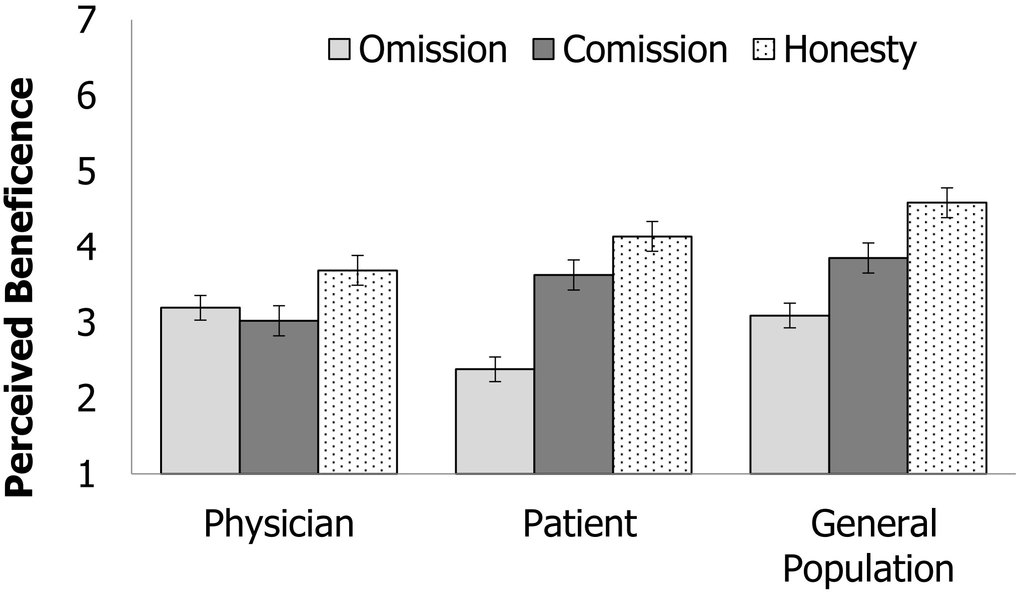


Figure 2. Perceptions of Beneficence by Communication Tactic and Population


Omission
Comission
Honesty
Perceived Beneficence
Physician
Patient
General Population