Organizational Behavior and Human Decision Processes 146 (2018) 31–50

Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

journal homepage: www.elsevier.com/locate/obhdp

Paternalistic lies

Matthew J. Lupoli[a,b,*], Emma E. Levine[c], Adam Eric Greenberg[d,e]

[a] Deakin University, Australia
[b] University of California, San Diego, United States
[c] University of Chicago, United States
[d] University of California, Los Angeles, United States
[e] Bocconi University, Italy

ARTICLE INFO

Keywords:
Deception
Paternalism
Prosocial lies
Moral judgments

ABSTRACT

Many lies that are intended to help others require the deceiver to make assumptions about whether lying serves others' best interests. In other words, lying often involves a paternalistic motive. Across seven studies (N = 2,260), we show that although targets appreciate lies that yield unequivocal benefits relative to honesty, they penalize paternalistic lies. We identify three mechanisms behind the harmful effects of paternalistic lies, finding that targets believe that paternalistic liars (a) do not have benevolent intentions, (b) are violating their autonomy by lying, and (c) are inaccurately predicting their preferences. Importantly, targets' aversion towards paternalistic lies persists even when targets receive their preferred outcome as a result of a lie. Additionally, deceivers can mitigate some, but not all, of the harmful effects of paternalistic lies by directly communicating their good intentions. These results contribute to our understanding of deception and paternalistic policies.

1. Introduction

People often lie with the intention of benefitting others (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). In many cases, however, it is not immediately obvious whether lying will, in fact, benefit the recipient of the lie (henceforth "target"). For example, an employee may inflate impressions of a colleague's performance on a presentation because he believes honesty will cause emotional harm and demotivate the colleague. Yet this belief may not necessarily be correct. A truthful statement might be seen as more beneficial in the eyes of the colleague, and could actually motivate the colleague to learn from his shortcomings and improve his performance in the future. If this colleague were to find out that the employee lied about his performance, how might he react?

In this research, we investigate how targets respond to lie-tellers (henceforth "deceivers" or "liars") whose lies require them to make subjective judgments about the target's best interests. We label these lies as paternalistic lies. Paternalistic lies are ubiquitous and have important consequences in a variety of contexts. For example, government officials might tell paternalistic lies to citizens by concealing facts about potential security threats to avoid inciting national panic; doctors might tell paternalistic lies to patients by giving them overly optimistic prognoses in order to provide hope; and friends and romantic partners might tell paternalistic lies to each other by delivering false praise with the intention of preventing emotional harm. In all of these cases, deceivers might lie out of genuine concern for the well-being of the targets, but targets may not appreciate these lies because judgments about whether the lie is ultimately more beneficial than the truth are inherently subjective. Thus, well-intended paternalistic lies may backfire. Because paternalistic lies are prevalent and can have important effects on people's lives, it is crucial to understand how they influence interpersonal judgment and behavior.

Here, we provide the first investigation of paternalistic lies. In addition to providing practical advice to those who might be tempted to tell paternalistic lies, we fill an important gap in existing deception research by introducing the construct of paternalistic lies, distinguishing this construct from related forms of deception, and documenting a strong distaste towards paternalistic lies and those who tell them across several dependent variables. This research also deepens our understanding of the primacy of perceived intent in moral judgment; we find that the perceived intentions of paternalistic liars play a critical role in responses to these lies.

1.1. Prosocial and paternalistic lies

Research investigating the consequences of deception has linked lying with a number of harmful effects. Lies have been shown to increase negative affect, damage trust, provoke revenge, harm

* Corresponding author: Rady School of Management, University of California, Wells Fargo Hall #3N130, San Diego, CA 92093-0553, United States.
E-mail address: Matthew.lupoli@rady.ucsd.edu (M.J. Lupoli).

https://doi.org/10.1016/j.obhdp.2018.01.001
Received 19 August 2016; Received in revised form 6 August 2017; Accepted 2 January 2018
Available online 08 March 2018

Table 1
Definitions of terms, with examples.

Prosocial lies False statements made with the intention of misleading and benefitting a target (Levine & Schweitzer, 2014, 2015)	
Unequivocal prosocial lies False statements made with the intention of misleading a target, *and are known to both the deceiver and the target to be in the target's best interests*	**Paternalistic lies** False statements made with the intention of misleading and benefitting a target, *and require the deceiver to make assumptions about the target's best interests*
Example: Your spouse has terminal cancer. **You and your spouse told your doctor in the past that you both would prefer to remain hopeful about the prognosis rather than receive complete candor**. Your doctor falsely tells you that your spouse may be eligible for a new experimental treatment soon.	**Example:** Your spouse has terminal cancer. **You and your spouse had never discussed with your doctor whether you both would prefer to remain hopeful about the prognosis or receive complete candor**. Your doctor falsely tells you that your spouse may be eligible for a new experimental treatment soon.

relationships, and promote further dishonesty (Boles, Croson, & Murnighan, 2000; Croson, Boles, & Murnighan, 2003; Greenberg, 2016; Greenberg & Wagner, 2016; Schweitzer & Croson, 1999; Schweitzer, Hershey, & Bradlow, 2006; Tyler, Feldman, & Reichert, 2006). However, the majority of this work has studied the effects of *selfish lies*, or lies that benefit the deceiver, potentially at a cost to the target. Given the conflation of deception with self-interested motivations in much of the existing literature, it has been difficult to conclude whether interpersonal penalties towards deception reflect an opposition to selfish behavior or deception per se.

To shed light on this issue, scholars have recently examined the consequences of prosocial lies. People tell *prosocial lies*, or false statements made with the intention of misleading and benefitting a target (Levine & Schweitzer, 2014, 2015; Lupoli, Jampol, & Oveis, 2017), on a regular basis (DePaulo et al., 1996). Given that individuals not only consider actions, but also the intentions behind and the consequences of those actions when making moral judgments of themselves (Shalvi, Dana, Handgraaf, & De Dreu, 2011; Shalvi, Gino, Barkan, & Ayal, 2015) and others (Cushman, 2008, 2013; Gino, Shu, & Bazerman, 2010; Greene et al., 2009; Miller, Hannikainen, & Cushman, 2014; Shu, Gino, & Bazerman, 2011), it is likely that prosocial lies are perceived differently than selfish lies.

Indeed, recent work provides evidence for this assertion. Individuals who tell prosocial lies that yield monetary benefits to the target are viewed as more ethical than those who tell the truth, regardless of whether the deceiver benefitted from lying (Levine & Schweitzer, 2014). Importantly, this research demonstrates that positive moral judgments of prosocial liars are driven by the perceived benevolence, rather than honesty, of the deceiver. In addition, prosocial liars are sometimes perceived to be more trustworthy: Levine and Schweitzer (2015) found that individuals were more likely to pass money in a trust game to those who told a prosocial lie than those who told harmful truths. Although prosocial lies increased benevolence-based trust (the willingness to make oneself vulnerable based on beliefs about another person's good intentions, which is captured by the trust game), the authors also found that prosocial lies harmed integrity-based trust—that is, the willingness to make oneself vulnerable based on beliefs about another person's adherence to moral principles, such as honesty and truthfulness. Thus, reactions towards prosocial lies are not universally positive.

While this research has advanced our understanding of prosocial lies, it has focused on one specific type of prosocial lie: lies with objective monetary benefits. Specifically, the majority of research on prosocial lies has utilized economic games to study the decisions to lie (Erat & Gneezy, 2012), as well as reactions to lying (Levine & Schweitzer, 2014, 2015). In these studies, lying is unambiguously beneficial for the target relative to the truth because a dishonest statement from a deceiver results in a monetary gain for the target, the magnitude of which exceeds the payoff resulting from honesty. Other work has investigated prosocial lying that helps a third party, whereby individuals cheat on a task for the monetary benefit of another individual (Gino, Ayal, & Ariely, 2013; Gino & Pierce, 2009; Wiltermuth, 2011). We conceptualize all of these lies as *unequivocal prosocial lies* because lying is known to both the target and the deceiver to be in the best interest of the target or third party. When people tell unequivocal prosocial lies, targets perceive the liars' benevolent intentions to be sincere, and thus, targets react favorably to deception (Levine & Schweitzer, 2014).

However, in many cases, both the consequences and true intentions associated with prosocial lies are unclear. For example, imagine that an employee (Bob) asks a colleague (Joe) for feedback on a presentation. When Bob asks Joe how he performed, what should Joe say? One option is to provide an honest opinion, believing that Bob would prefer to hear the truth and that knowing his presentation was unsatisfactory might help him improve in the future. Alternatively, Joe could lie to Bob, believing that Bob is looking for positive reinforcement and that hearing his performance was poor would devastate him. Without knowing how the truth or a lie would affect Bob emotionally or help him in the future, Joe must rely on his assumptions about Bob's best interests when deciding whether to be truthful. This scenario illustrates that when given the opportunity to tell a prosocial lie, individuals often lack insight into others' preferences for truthfulness, as well as the negative consequences that lying might have on them. Thus, this type of lie can be considered a *paternalistic lie*.

We define paternalistic lies as *lies that are intended to benefit the target, but require the deceiver to make assumptions about targets' best interests.* As such, paternalistic lies are a subset of prosocial lies (see Table 1). When individuals tell paternalistic lies, they are motivated by the assumption that targets are better off being lied to, even though this assumption cannot be objectively verified. Thus, the targets themselves might not agree with this assessment. In short, while unequivocal prosocial lies are known to help the target, paternalistic lies help the target only according to the beliefs of the deceiver. By studying paternalistic lies, we build knowledge of how different types of lies influence interpersonal judgment and behavior, and gain insight into the circumstances in which targets believe versus discredit the prosocial intentions of liars.

It is important to note that although we dichotomize the distinction between unequivocal prosocial lies and paternalistic lies for the ease of investigation, the degree to which deceivers have insight into targets' best interests—and thus the degree to which a lie is paternalistic—falls along a continuum. We use the terms "paternalistic lies" and "unequivocal prosocial lies" as endpoints on this continuum. We do not claim that there are lies that are unequivocally prosocial to all people in all settings. However, we do claim that there are cases in which a deceiver can be more or less confident about what benefits the target. For instance, consider the aforementioned example of Joe, who is asked to give feedback on Bob's poor presentation. If the two have an existing relationship and have already discussed how Bob responds to blunt critiques and words of encouragement, Joe's assumptions about whether honesty or deception are in his colleague's best interests may be fairly accurate. However, if the two have no existing relationship, then his assumptions will be less informed. Without explicit knowledge about how a lie will affect the target and the target's preferences for

lying itself, lying with prosocial intent always requires some assumption regarding the target's interests. That said, if a deceiver is able to gain insight into the target's preferences (e.g., through discussion or past experience), then the deceiver is no longer required to rely on as many assumptions. Thus, lies are distinguishable with respect to how paternalistic they are.

The distinction between paternalistic lies and unequivocal prosocial lies is not merely theoretical, but one that lay people recognize as well. In a pilot study ($N = 90$), we asked participants to generate an example of one circumstance in which someone lied with the intention of helping or protecting someone else. We then asked participants to categorize their example as either a paternalistic lie or an unequivocal prosocial lie. A total of 36% of participants indicated that "the liar assumed that lying was in the person's best interests without knowing for certain" (i.e., told a paternalistic lie), rather than "the liar knew for certain that lying was in the person's best interests" (i.e., told an unequivocal prosocial lie). For instance, one participant gave the example of a person giving overly positive feedback of another's appearance, and offered the following explanation: "It might actually be in the other person's best interest to tell them they don't look good, if this would cause them to change something about their appearance that would lead to better treatment and higher self-esteem." We also asked participants to rate how often they have been the target of both types of lies and found that participants believe they are told unequivocal prosocial lies and paternalistic lies with equal frequency.[1] Together, these results suggest that (a) people recognize that some lies are paternalistic, according to our definition, (b) people perceive being the target of paternalistic lies as often as being the target of unequivocal prosocial lies, and (c) people distinguish between paternalistic and unequivocally prosocial lies. Given that paternalistic lies are common, consequential, and viewed as distinct from unequivocal prosocial lies, it is important to understand their consequences.

1.2. Perceptions of paternalistic lies

Our central thesis is that those who tell paternalistic lies are judged to be less moral than those who are honest. To explain this prediction, we draw on three streams of research: research on procedural justice (e.g., Brockner et al., 1994; Tyler, Degoey, & Smith, 1996), research on the primacy of perceived intentions in moral judgments (e.g., Cushman, 2008; Greene et al., 2009), and research on reactance and the importance of individual autonomy (e.g., Brehm, 1966).

The procedural justice literature suggests that paternalistic lies, unlike unequivocal prosocial lies, will be viewed harshly. A robust finding in this literature is that the desirability of outcomes and the perceived fairness by which those outcomes are obtained interact to influence responses to outcomes (for a review, see Brockner & Wiesenfeld, 1996). Specifically, if people view an outcome as desirable, they will respond favorably regardless of the fairness of the process that yielded the outcome. However, if the outcome is undesirable, their response hinges on the perceived fairness of the process that yielded the outcome; people will respond more favorably if the process seemed fair and less favorably if the process seemed unfair. For example, one study found that organizational commitment was relatively unaffected by the perceived fairness of procedures when satisfaction with job outcomes (e.g., compensation) was high; however, when satisfaction with outcomes was low, organizational commitment was strongly influenced by procedural fairness (McFarlin & Sweeney, 1992). This pattern of results has been observed across a wide range of dependent variables, including job performance, job satisfaction, and trust in management, in both organizational and laboratory contexts (Brockner & Wiesenfeld, 1996).

Surprisingly, no existing work has applied this lens to the study of deception. We build on procedural justice research to explain why individuals have positive reactions towards unequivocal prosocial lies, but may have negative reactions towards paternalistic lies. By definition, unequivocal prosocial lies result in outcomes that are objectively desirable (compared to the outcomes associated with honesty). Thus, in line with the procedural fairness/outcome desirability interaction, individuals are likely to respond favorably to these lies despite potentially objecting to the process (i.e., deception) in general. Indeed, this notion is consistent with past findings on positive perceptions of prosocial lies (Levine & Schweitzer, 2014, 2015). Paternalistic lies, however, result in outcomes that are not objectively desirable compared to those associated with honesty. Thus, when people are targets of paternalistic lies, they are likely to shift their focus towards the process by which outcomes are obtained (i.e., deception or honesty). Because honesty is generally perceived to be more moral than deception (Graham, Meindl, Koleva, Iyer, & Johnson, 2015)—particularly in the absence of clear benevolent motives for deception (Levine & Schweitzer, 2014)—we expect that those who tell paternalistic lies will be judged as less moral than those who tell the truth.

What specific inferences about those who tell paternalistic lies might underlie a potential decrease in perceived moral character? We hypothesize that the perceived intent of deceivers plays a key role in moral judgments, and in particular, that targets will view paternalistic deceivers as not acting with benevolent intent. Moral judgments of actions often hinge on the perceived motives of the actor (e.g., Cushman, 2008). Consistent with this notion, past work on unequivocal prosocial lies suggests that these lies are seen as moral precisely because they credibly signal benevolent intent (Levine & Schweitzer, 2014, 2015). Lies that do not signal benevolent intent, in contrast, are deemed to be less moral than the truth (Levine & Schweitzer, 2014). We propose that paternalistic lies signal a lack of benevolent intent for two reasons. First, the subjective nature of the benefits afforded by paternalistic lies may obscure the good intentions of deceivers. People's ability to take the perspective of others and understand the emotions, beliefs, and motivations that drive them is notably limited (e.g., Epley, Keysar, Van Boven, & Gilovich, 2004; Gilbert & Malone, 1995; Van Boven & Loewenstein, 2003). Thus, if a target thinks that he may have been better off or equally well off receiving the truth, he may incorrectly think that the deceiver with good intentions was also aware of this belief. Furthermore, personal experience with the harmful effects of selfish lies (Boles et al., 2000; Croson et al., 2003; Greenberg, Smeets, & Zhurakhovska, 2015; Greenberg & Wagner, 2016; Schweitzer & Croson, 1999; Schweitzer et al., 2006; Tyler et al., 2006) may have spillover effects on responses to prosocially motivated lies. As a result, individuals might generally be skeptical of deceivers' prosocial intentions, unless the benefits of lying over honesty are clear and unequivocal.

In addition to hypothesizing that paternalistic lies lead targets to doubt deceivers' benevolent motivation, we predict that paternalistic lies are perceived to violate targets' autonomy. Autonomy has been defined as the perceived internal locus of causality (deCharms, 1968; Ryan & Deci, 2000), or a sense that one's actions "emanate from oneself and are one's own" (Deci & Ryan, 1987). Autonomy has been found to thrive when individuals experience choice, when others acknowledge their feelings, and when individuals have the ability to take self-directed actions (Deci & Ryan, 1985). In contrast, autonomy can be diminished by deadlines, directives, pressured evaluations, and imposed goals (Ryan & Deci, 2000).

One reason why paternalistic lies might be seen as a violation of one's autonomy is that people feel they have a right to know the truth, and that acts of dishonesty impinge upon this right. Similarly, lying

[1] In addition, participants read three vignettes depicting paternalistic lies (adapted from the vignettes used in Study 7) and were asked for each vignette, "To what extent is this lie paternalistic? By paternalistic, we mean limiting the freedom or autonomy of the person who has been lied to, in the presumed best interest of that person" (1 = *not at all*, 7 = *very much so*). Collapsing across vignettes, participants rated the lies as significantly paternalistic ($M = 5.20$, $SD = 1.31$; $t(89) = 8.73$, $p < .001$, one sample t-test against the midpoint). We provide additional details on this pilot study in our online supplementary materials.

might be seen as an attempt to control someone else's view of the world, imposing a framework on targets that deceivers deem superior to the reality shaped by the truth. Indeed, philosophers ranging from Kant (1785) to Bok (1978) have opposed deception on these grounds. Paternalistic lies might also threaten targets' autonomy because these lies, by definition, result in an outcome that the target may not have chosen for himself. Thus, paternalistic lying is likely to be perceived as an attempt to influence or coerce the target. Unequivocal prosocial lies, in contrast, generate an outcome that is known to be in the target's best interest. In other words, unequivocal prosocial lying is the course of action that the target would have chosen for himself. Thus, it is less likely that unequivocal prosocial lies would be perceived as autonomy violations. Given the importance of autonomy to moral judgment (Rozin, Lowery, Imada, & Haidt, 1999; Shweder, Much, Mahapatra, & Park, 1997), we predict that the perception that paternalistic lies violates one's autonomy will further contribute to judgments of deceivers' immorality.

If perceived autonomy violations underlie moral judgments of paternalistic liars, then paternalistic lies could elicit reactance. Reactance is a psychological state that arises when individuals feel that their freedom or autonomy is being eliminated or threatened by another (Brehm, 1966). This state can manifest as the derogation of the agent restricting the freedom (Miron & Brehm, 2006). Judging those who tell paternalistic lies to be less moral than those who are honest is one way in which targets might derogate deceivers who are perceived to be violating their autonomy. However, another indicator of reactance that could result from paternalistic lies is a decrease in the attractiveness the outcome resulting from the lie (Brehm, Stires, Sensenig, & Shaban, 1966). For example, recommendations by experts that contradict consumers' initial impressions cause consumers to oppose the recommendations more intensely because they experience a state of reactance (Fitzsimons & Lehmann, 2004). Similarly, if paternalistic lies elicit reactance, targets' preferences may shift as a result of being lied to. Specifically, targets may dislike outcomes associated with lying, even if they would have liked the same outcome had it been associated with honesty. As a result, targets may feel that paternalistic liars are incorrectly predicting their preferences. If targets believe that deceivers made a wrong decision on their behalf—a decision that is potentially seen as immoral—it is possible that this could result in a halo effect (e.g., Nisbett & Wilson, 1977) whereby the deceiver is also viewed as immoral. Thus, it is possible that perceptions that deceivers inaccurately predicted one's preferences may influence moral judgments of paternalistic lies.

In summary, we consider three potential processes that may underlie moral judgments of those who tell paternalistic lies: perceptions that (a) paternalistic liars are not motivated by benevolent intent, (b) paternalistic lies violate one's autonomy, and (c) paternalistic liars are inaccurately predicting targets' preferences. We expect that these processes can operate in tandem, but that each independently influences moral judgments.

1.3. Overview of studies

In seven experiments, we provide the first investigation of paternalistic lies by examining how individuals judge paternalistic lies and those who tell them. We focused primarily on moral judgments of paternalistic deceivers (Studies 1–3, 5–7). We also measured positive affect (Studies 1–3, 5–6) to assess psychological responses to paternalistic lies, in addition to social judgments of deceivers. In Studies 1–5, we examined judgments of paternalistic lies in a well-controlled economic game in which the consequences of lying (relative to truth-telling) for the target were directly manipulated. In Study 1, we explored how both paternalistic lies and unequivocal prosocial lies influence moral judgments and emotional responses. In Study 2, we conceptually replicated Study 1 with a larger sample size and eliminated a potential confound. In Study 3, we examined the mechanisms underlying the effect of paternalistic lies on moral judgments. In Study 4, we moved beyond moral judgments and affect by exploring how paternalistic lies alter preferences for the outcomes associated with lying and honesty. In Study 5, we explored the robustness of our results by (a) using a behavioral measure to capture targets' distaste for paternalistic lies—that is, the degree to which targets punish their deceivers—and (b) testing whether the distaste for paternalistic deception persists even when deceivers communicate their benevolent intentions. We also provided further evidence for the underlying mechanisms identified in Study 3. In Study 6, we assessed external validity of these results by examining judgments of paternalistic and unequivocal prosocial lies in several realistic vignettes. In these vignettes, we again manipulated whether the deceiver communicated benevolent intentions to the target. In Study 7, we used a vignette design similar to that of Study 6 to directly manipulate the deceiver's benevolent intent, rather than the deceiver's *claimed* benevolent intent, to obtain causal evidence for a hypothesized mechanism underlying moral judgments of paternalistic liars.

1.4. Deception game

A large body of research demonstrates the capacity for economic games to teach us about decision-making in real-world dilemmas and social interactions (e.g., Fehr & Fischbacher, 2003; Halevy & Chou, 2014; Halevy & Halali, 2015; Murnighan & Wang, 2016; Zhong, 2011). Games have several advantages, including clean experimental control over endogenous and exogenous factors, ease of comparison across experimental designs and results (Ostrom, Gardner, & Walker, 1994), and unambiguously defined actions and consequences for players (Rapoport, 1973). Given these advantages, deception has often been studied using variations of an economic game called the sender-receiver game (Erat & Gneezy, 2012; Gneezy, 2005; Gneezy, Rockenback, & Serra-Garcia, 2013; Gunia, Wang, Huang, Wang, & Murnighan, 2012; Levine & Schweitzer, 2014, 2015; Zhong, 2011). Although different types of lies have been operationalized using this game (e.g., altruistic lies that benefit others at a cost to oneself; Erat & Gneezy, 2012), no work that we are aware of has used the game to explore paternalistic lies. Thus, in Studies 1–5, we adapted a version of the sender-receiver game to study paternalistic lies, hereafter referred to as the Deception Game.

In this game, all participants learned that they had been assigned to the role of "Receiver," and that they were paired with an anonymous "Sender." In actuality, there was no Sender; the Sender's role was simulated by a set of pre-programmed responses. Participants were told that the computer had simulated a fair coin flip, and that only the Sender knew the actual outcome of the coin flip. They were informed that after learning the outcome of the flip, the Sender sent one of two messages to the Receiver (participants): "The coin landed on HEADS" or "The coin landed on TAILS." Participants were then told that after receiving the Sender's message, they would choose "heads" or "tails," and what they earned would be based on whether their choice corresponded to the actual outcome of the coin flip. Importantly, both the Sender and the Receiver knew that only the Sender was informed about the potential payoffs associated with the Receiver's choice.[2]

Next, participants were randomly assigned to receive one of the two possible messages ostensibly from the Sender (i.e., "The coin landed on HEADS/TAILS"). After viewing the message, participants were asked to choose either "heads" or "tails."

Once they made their choice, participants were told that they would learn about the private information that was available to the

[2] After reading the instructions, participants completed a comprehension check to ensure that they understood the instructions. If they answered either of the comprehension check questions incorrectly, they were given the exercise instructions again, followed by a second comprehension check. If they failed the second comprehension check, they were unable to continue the experiment.

Sender—that is, the possible payoffs and full instructions the Sender received. We then revealed the Sender's information to participants. Specifically, participants learned three new pieces of information. First, they learned that the outcome of the coin flip was heads. Thus, the Sender's message about the outcome of the coin flip constituted our between-subjects manipulation of (dis)honesty. Those who were told by the Sender that the coin landed on heads received the truth, while those who were told that the coin landed on tails were deceived.

Second, participants were informed that the Sender was told, "previous studies have found that almost all Receivers choose the outcome that the Sender indicates in his/her message." We included this statement because we wanted participants to believe the Sender would expect them to follow the message. This was intended to reduce noise in participants' perceptions of why the Sender lied.

Third, participants learned about the payment structure that the Sender faced. According to the Sender's instructions, if the Receiver chose correctly (i.e., her choice corresponded to the actual outcome of the coin flip), the Receiver would be paid according to Option A. If the Receiver chose incorrectly (i.e., her choice did not correspond to the actual coin flip outcome), the Receiver would be paid according to Option B. As such, Senders had faced the choice of sending an honest message, which would likely result in the Receiver getting Option A, or sending a dishonest message, which would likely result in the Receiver obtaining Option B. Importantly, the Sender's own incentives were not tied to either Option A or Option B. Thus, the Sender was simply making a decision that would affect the Receiver, not herself.

In all studies employing the deception game (Studies 1–5), Option A and Option B were pretested to be equally desirable in the aggregate, but involved some tradeoff that could be perceived differently at the individual level. For example, in Study 1, one Option was a low-risk, low-reward gamble, while the other Option was a higher-risk, higher-reward gamble. Structuring the game such that the outcomes were equally desirable on average simulates conditions under which a paternalistic lie might be told; from the Sender's perspective, there is necessarily uncertainty about which outcome is in the Receiver's best interest. Lying to ensure that an individual received a low-risk, low-reward gamble may be well-intended, given that it protects the target from some risk. Yet, this lie is necessarily paternalistic because the deceiver does not know what the target's risk preferences are, and thus, must make assumptions about what the target would want. Indeed, a pretest revealed that Senders who lied in the Deception Game did so because they believed it was in the best interest of Receivers.[3] However, from the Receiver's perspective, the Sender's motivations are intentionally ambiguous because in the real world, targets often are not fully aware of deceivers' motives. Table 2 includes a summary of the outcomes associated with Options A and B for each study, along with an example of the type of paternalistic lies these options model.

In each study, we counterbalanced the outcomes associated with Options A and B between-subjects to ensure that our results were robust across any particular tradeoff. For example, in Study 1, half of the participants saw that Option A was the low-risk, low-reward gamble, and that Option B was the high-risk, high-reward gamble; and the other half of participants saw that Option A was the high-risk, high-reward gamble, and Option B was the low-risk, low-reward gamble. Furthermore, in all studies, the potential payoffs in the game were incentive-compatible, as one participant was randomly selected to receive the Option obtained in the game.

[3] We ran a pilot study in which all participants were assigned to the role of Sender ($N = 148$). After making the decision to send an honest or dishonest message, Senders were asked to indicate their agreement with the statement, "I chose the message I believed was in the best interest of the Receiver" (1 = *strongly disagree*, 7 = *strongly agree*). A *t*-test against the midpoint indicated that Senders who lied ($N = 44$) significantly agreed with this statement ($M = 5.32$, $SD = 2.18$), $t(43) = 4.00$, $p < .001$, suggesting that their deception was motivated by their assumptions about what benefitted the Receiver (i.e., their deception was paternalistic).

In all studies, we did not conduct statistical analyses prior to the completion of data collection. Given that we did not have sufficient precedent to make precise estimates of effect sizes, we decided on sample size using the following heuristics: For laboratory studies (Studies 1 and 4), we aimed to obtain as many participants as possible within the lab time allotted; for online studies (Studies 2–3, 5–7), we aimed to obtain 100 participants per cell (collapsed across choice set for studies using the Deception Game; see Study 1 Procedure and Materials). We report all measures, manipulations, and data exclusions.

2. Study 1

In Study 1, we investigated individuals' moral judgments of those who tell paternalistic lies. In this experiment, both honesty and dishonesty resulted in participants being entered into one of two gambles that were equally desirable on average. Using gambles with different levels of risk as outcomes captures the uncertainty often associated with paternalistic lies. For example, a mentor might lie to an employee if she thinks a low-risk, low-reward career is better for the candidate than a high-risk, high-reward career. Similarly, a doctor might lie to a patient to lead her to choose a low-risk (or high-reward) treatment.

In order to disentangle whether reactions to the Sender were due to the subjective nature of the message's consequences or reactions to deception in general, we also included conditions in which the Sender's message resulted in participants being given one or two lottery tickets for entry into the same gamble. In other words, we compared paternalistic lies (lies that require the deceiver to make assumptions about the best interests of the target) to unequivocal prosocial lies (lies that are known to the target and the deceiver to be in the best interest of the target).

2.1. Procedure and materials

We recruited 200 adults from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee. Eight participants failed a comprehension check at the start of the experiment and were automatically eliminated from the study. We thus report the results from 192 participants (59.9% female; $M_{age} = 20$) who passed the comprehension checks and completed the entire study.

Participants were randomly assigned to one of eight experimental conditions in a 2(Deception: honesty vs. lying) × 2(Lie type: paternalistic lie vs. unequivocal prosocial lie) × 2(Choice Set: choice set 1 vs. choice set 2) between-subjects design. In the paternalistic lie conditions, the benefit associated with lying was subjective. In the unequivocal prosocial lie conditions, deception was unambiguously prosocial (i.e., it made the target strictly better off).

Participants engaged in the Deception Game as previously described. After learning the rules of the game and receiving the randomly assigned message from the Sender, we revealed the Sender's private information. Participants learned that the Sender had either been honest or dishonest. We also revealed the payoffs associated with Options A and B, which were lotteries into which Receivers would be entered. Here, we manipulated whether or not the Sender's lie could yield an objectively beneficial payout. In the paternalistic lie conditions, Options A and B were associated with a 50% chance of winning $1 and a 50% chance of winning $0, versus a 25% chance of winning $2.25 and a 75% chance of winning $0. These gambles were rated as equally preferable ($p > .40$) in a pilot study with a non-overlapping sample ($N = 46$). As mentioned, we counterbalanced the outcomes associated with Options A and B between-subjects so that there were two different choice sets. That is, in the paternalistic lie condition of Choice Set 1, Option A resulted in the Receiver getting 1 lottery ticket for the 50% chance of $1/50% chance of $0 lottery, while Option B yielded 1 lottery ticket for the 25% chance of $2.25/75% chance of $0 lottery. In the paternalistic lie condition of Choice Set 2, the lotteries associated with Options A and B were reversed.

Table 2
Summary of the Deception Game across Studies 1–5. In each study, the payoffs associated with Options A and B were counterbalanced between-subjects.

Study	Type of payoffs	Outcomes associated with Options A and B	Real-world example
1	Gambles	50% chance of $1, 50% chance of $0/25% chance of $2.25, 75% chance of $0	Lying to a patient to ensure s/he chooses a low-risk medical procedure
2	Gift cards	$25 McDonald's gift card/$25 Whole foods gift card	Lying to a friend to ensure s/he chooses a healthy snack
3, 4, 5	Intertemporal choice	$10 today/$30 3 months from now (Studies 3, 4) $17.50 today/$30 3 months from now (Study 5)	Lying to a client to ensure s/he saves money for the future

In the unequivocal prosocial lie conditions, Options A and B were associated with one lottery ticket or two lottery tickets for identical gambles, respectively. As in the paternalistic lie conditions, we counterbalanced the types of gambles associated with Options A and B so that there were two different choice sets. In the unequivocal prosocial lie condition of Choice Set 1, participants saw that Option A resulted in the Receiver receiving 1 lottery ticket for the 50% chance of $1/50% chance of $0 gamble, and that Option B resulted in 2 lottery tickets for this same gamble. The other half of participants (those in the choice set 2/unequivocal prosocial lie condition) saw that Option A yielded 1 ticket and Option B yielded 2 tickets for the 25% chance of $2.25/75% chance of $0 gamble. Because Option B dominates Option A in the both choice sets of the unequivocal prosocial lie conditions, sending a dishonest message would result in an outcome that was objectively better for the Receiver.

2.1.1. Dependent variables

After viewing the Sender's private information and the potential outcomes associated with Options A and B, participants provided ratings of the Sender's moral character by indicating their agreement (1 = *strongly disagree*, 7 = *strongly agree*) with the following statements: "I trust the Sender"; "The Sender had good intentions"; "The Sender wanted to help me"; "The Sender is a good person"; "The Sender is unethical" (reverse-scored); and "The Sender made the wrong decision for me" (reverse-scored). This last item was not included in analysis of moral judgments, as it conflates perceptions of the Sender with personal preferences. However, inclusion of this item does not alter results. The remaining items were highly reliable ($\alpha = 0.89$).

We also measured participants' emotional responses to the Sender's message. Participants were asked to indicate the degree to which they felt the following emotions "in response to the Sender's behavior" (1 = *not at all*, 7 = *extremely*): grateful, excited, happy, and content ($\alpha = 0.88$).[4]

Finally, we included a three-item manipulation check to ensure that participants recognized the act of deception. Participants rated their agreement (1 = *strongly disagree*, 7 = *strongly agree*) with the following items: "The Sender sent an honest message" (reverse-scored); "The Sender lied about the outcome of the coin flip;" and "The Sender was deceptive" ($\alpha = 0.86$). Participants concluded the study by providing demographic information and answering three attention checks.

2.2. Results

For all studies, we report results collapsing across choice set for the sake of brevity. Models with choice set included as a factor are included in the Supplementary Materials.

[4] In addition, we measured negative affect using the following four items: angry, disappointed, sad, and anxious ($\alpha = 0.82$). Positive and negative affect loaded on separate factors. We measured both positive and negative affect in Studies 1–3, 5, and 6. However, for the sake of brevity, we only report positive affect. Positive and negative affect followed the inverse pattern in every study and the results for negative affect are reported in the Supplementary Materials.

2.2.1. Manipulation check

A *t*-test revealed that the deception manipulation was successful. Collapsing across lie type and choice set, participants in the lie condition ($M = 5.56$, $SD = 1.39$) rated the Sender as more dishonest than those in the truth condition ($M = 2.33$, $SD = 1.28$), $t(190) = 16.74$, $p < .001$, $d = 2.42$.

2.2.2. Moral character

A two-way ANOVA revealed a significant Deception × Lie Type interaction, $F(1, 188) = 26.15$, $p < .001$, $\eta_p^2 = 0.12$. Consistent with Levine and Schweitzer (2014), Senders were seen as more moral when they told an unequivocal prosocial lie ($M = 4.54$, $SD = 1.70$) than when they told the truth ($M = 3.99$, $SD = 1.10$), $t(92) = 1.88$, $p = .06$, $d = 0.39$. Importantly, however, this effect reversed for paternalistic lies. When lying was associated with subjective benefits, Senders were seen as more moral when they told the truth ($M = 4.95$, $SD = 1.02$) than when they lied ($M = 3.62$, $SD = 1.19$), $t(96) = 5.93$, $p < .001$, $d = 1.20$. This pattern of results is depicted in Fig. 1. In addition, those who told paternalistic lies ($M = 3.62$, $SD = 1.19$) were seen as less moral than those who told unequivocal prosocial lies ($M = 4.54$, $SD = 1.70$), $t(94) = 3.12$, $p < .01$, $d = 0.64$.

We also found a main effect of deception, $F(1, 188) = 4.94$, $p = .03$, $\eta_p^2 = 0.03$, such that participants generally believed that Senders were more moral when they told the truth ($M_{\text{Honesty}} = 4.47$, $SD_{\text{Honesty}} = 1.16$ vs. $M_{\text{Lying}} = 4.06$, $SD_{\text{Lying}} = 1.52$). There was no main effect of lie type, $p > .90$.

2.2.3. Positive affect

A two-way ANOVA revealed a significant Deception × Lie Type interaction, $F(1, 188) = 12.23$, $p < .001$, $\eta_p^2 = 0.06$. Targets experienced more positive affect in response to unequivocal prosocial lies ($M = 3.32$, $SD = 1.90$) compared to honesty ($M = 2.67$, $SD = 1.36$), $t(92) = 1.92$, $p = .06$, $d = 0.40$. However, targets experienced less positive affect in response to paternalistic lies ($M = 2.80$, $SD = 1.39$) than to honesty ($M = 3.71$, $SD = 1.50$), $t(96) = 3.11$, $p < .01$, $d = 0.63$. There were no main effects of deception or lie type (*ps* > .20).

2.2.4. Robustness check: perceived deception

One potential alternative account for our results is that paternalistic lies are perceived as more deceptive than unequivocal prosocial lies. To test this, we ran a two-way ANOVA with deception and lie type included as factors, using perceived deception (our manipulation check) as the dependent variable. In addition to a main effect of deception, $F(1, 188) = 325.57$, $p < .001$, $\eta_p^2 = 0.63$, we also found a significant Deception × Lie Type interaction, $F(1, 188) = 32.86$, $p < .001$, $\eta_p^2 = 0.15$. In the unequivocal prosocial lie condition, lying had a smaller effect on perceived deception ($M_{\text{Lying}} = 5.03$, $SD_{\text{Lying}} = 1.64$ vs. $M_{\text{Honesty}} = 2.85$, $SD_{\text{Honesty}} = 1.30$), $t(92) = 7.17$, $p < .001$, $d = 1.48$, relative to the paternalistic lie condition ($M_{\text{Lying}} = 6.05$, $SD_{\text{Lying}} = 0.87$ vs. $M_{\text{Honesty}} = 1.82$, $SD_{\text{Honesty}} = 1.05$), $t(66) = 21.77$, $p < .001$, $d = 4.40$. We found no main effect of lie type ($p > .90$). These findings suggest that unequivocal prosocial lies were seen as less deceptive than paternalistic lies.

To rule out the possibility that moral judgments of unequivocal prosocial lies and paternalistic lies were driven by this difference in

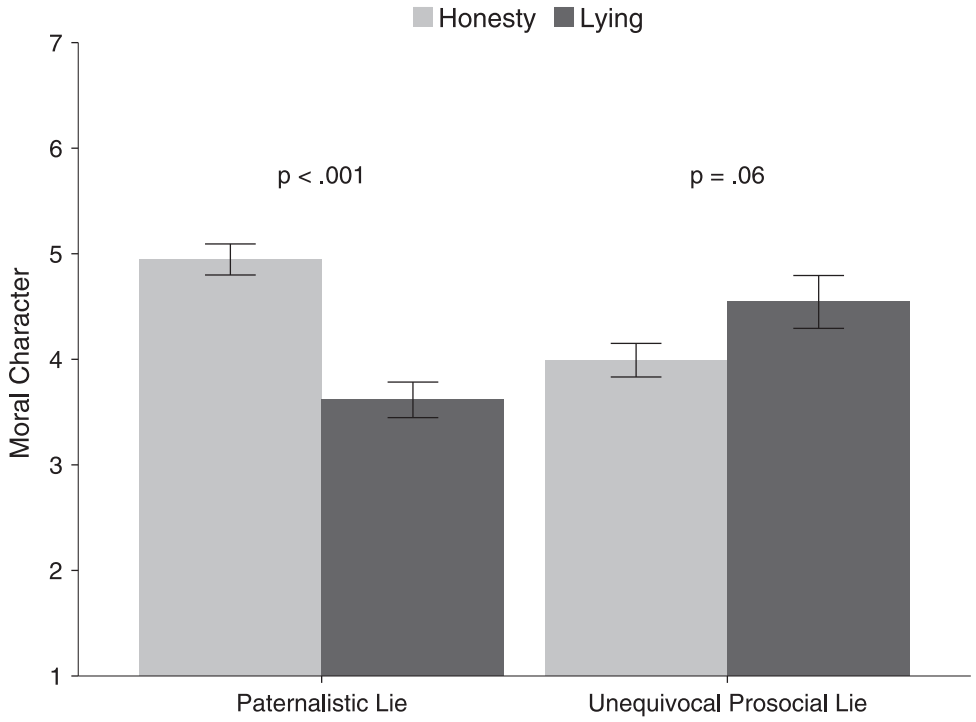

Fig. 1. The effects of paternalistic lies and unequivocal prosocial lies on perceived moral character in Study 1. In the paternalistic lie conditions, lying and honesty were each associated with 1 lottery ticket to either a high-risk/high-reward gamble or a low-risk/low reward gamble. In the unequivocal prosocial lie conditions, lying and honesty were associated with 2 vs. 1 lottery tickets to the same gamble, respectively. Error bars reflect ± 1 SE.

perceived deception, we ran a model to examine the Deception × Lie Type interaction, controlling for perceived deception. The Deception × Lie Type interaction remained significant in this model, $B = 0.78$, $p = .02$. Moral judgments were also significantly predicted by perceived deception, $B = -0.54$, $p < .001$, such that higher perceived deception was associated with lower moral judgments of the Sender. A full regression table for these analyses is available in the Supplementary Materials.

2.3. Discussion

Study 1 documents three main results. First, individuals who told paternalistic lies were seen as less moral than those who made honest statements. Second, receiving a paternalistic lie decreased targets' positive affect. Finally, paternalistic lies were judged differently than unequivocally prosocial lies; whereas unequivocal prosocial lies boosted positive affect and improved moral judgments relative to truth-telling, paternalistic lies had the opposite consequences.

3. Study 2

Study 2 builds upon Study 1's results in three ways. First, in Study 2, we aimed to conceptually replicate Study 1's finding that those who tell paternalistic lies are viewed as less moral and elicit less positive affect than those who are honest. Here, we investigated lies with different types of outcomes. Whereas Options A and B in Study 1 were gambles with different risk profiles, Options A and B in Study 2 were gift cards for healthy or unhealthy food. This setup mirrors another type of setting in which paternalistic lies may be told. For instance, a mother might falsely exaggerate the negative consequences of eating candy for breakfast in order to coerce her child into making healthier choices. In Study 2, participants learned that the Sender was faced with the choice of whether to tell the truth or lie to endow the target with either of two gift cards for food, both of which he/she may like, but differ in healthiness. Importantly, this design involves a decision in which the Sender must make assumptions about the best interests of the Receiver.

Second, in Study 2, participants received an explicit statement in the game instructions that the Sender had no stake in the game—that is, that the Sender would not receive a bonus regardless of the Receiver's choice of heads or tails. This is an important detail because it removes any lingering doubt about whether the Sender has selfish motivations for lying. Because it is clear that the Sender had no monetary incentive to lie, participants may be more apt to recognize benevolent motives for lying.

A final difference from Study 1 is that in Study 2, we just examined paternalistic lies (rather than comparing paternalistic lies to unequivocal prosocial lies) and used a larger sample size in order to increase statistical power and confidence in our results.

3.1. Procedure and materials

We received 198 complete responses on Amazon Mechanical Turk (Mturk). Two hundred five participants began the experiment, but seven participants were automatically excluded from the experiment for failing the comprehension check. We also excluded nine participants who failed an attention check at the beginning of the survey, leaving a final sample of 189 participants (46.0% female; $M_{age} = 33$).

Participants were in the role of Receiver and were given the same instructions as those in Study 1, with one exception. At the end of the instructions, participants were told that the Sender would earn no bonus, regardless of the Receiver's choice of heads or tails. This statement was included to minimize heterogeneity in inferences about the Sender's prosocial intentions, as well as in expectations about the Sender's payoffs, which were not specified in Study 1.

In this 2(Deception: honesty vs. paternalistic lying) × 2(Choice Set: choice set 1 vs. choice set 2) between-subjects design, participants were randomly assigned to receive an honest or dishonest message from the Sender. We followed the same procedure outlined in Study 1, except that the outcomes associated with Options A and B (i.e., the choice sets) were now either 1 lottery ticket for a $25 McDonalds gift card or 1 lottery ticket for a $25 Whole Foods gift card (see Table 2). A pilot study with a sample drawn from the same population ($N = 96$) revealed that participants would be equally satisfied receiving either of these gift cards ($p > .20$). We thus used these two gift cards for Study 2.

3.1.1. Dependent variables

After learning the veracity of the Sender's message and seeing the Sender's private information, participants answered a series of questions aimed to assess their judgments of the Sender. Participants indicated their agreement (1 = *strongly disagree*, 7 = *strongly agree*) with nine questions about the Sender's morality ($\alpha = 0.96$): "I trust the Sender"; "The Sender is caring"; "The Sender is benevolent"; "The Sender is selfish" (reverse-scored); "The Sender is empathic"; "The Sender is trustworthy"; "The Sender is ethical"; "The Sender is immoral" (reverse-scored); and "The Sender is a good person."[5]

We also measured participants' positive affect in response to the Sender. Participants received the same prompt as in Study 1, which asked them to indicate the extent to which they felt happy and grateful "in response to the Sender's behavior" (1 = *not at all*, 7 = *very much*; $r = 0.89$).

On the same page in which the dependent variables were assessed, participants saw a summary of the actions taken in the game. All subsequent experiments contained the same summary at the time the dependent variables were assessed.

3.2. Results

3.2.1. Moral character

Participants viewed Senders as less moral when they told a paternalistic lie ($M = 3.50$, $SD = 1.34$) than when they told the truth ($M = 5.26$, $SD = 0.90$), $t(187) = 10.63$, $p < .001$, $d = 1.55$.

3.2.2. Positive affect

Paternalistic deception also had a significant effect on positive affect. Participants who received a paternalistic lie ($M = 2.84$, $SD = 1.84$) reported less positive affect than those who were told the truth ($M = 4.81$, $SD = 1.68$), $t(187) = 7.68$, $p < .001$, $d = 1.12$.

3.3. Discussion

Study 2 provides further evidence for an aversion to paternalistic deception. In this study, we extended our investigation to lies that promote or inhibit specific consumption habits and found that paternalistic lies again harmed moral judgments and decreased positive affect. A paternalist may believe a person ought to choose healthy food or unhealthy food, the basis of which depends on the paternalist's own ideas about what is best for the target. Our results suggest that these types of lies would not be well-received if uncovered.

4. Study 3

In Study 3, we expanded our investigation in two ways. First, we tested potential mechanisms underlying targets' moral judgments of paternalistic lies. Specifically, we measured the degree to which targets question the motivations of deceivers, the degree to which targets perceive the deceiver as violating their autonomy, and the degree to which deceivers are perceived as inaccurately predicting targets' preferences.

In addition, we examined lies with another type of tradeoff: intertemporal monetary payoffs. Many acts of paternalistic deception involve making an intertemporal choice on behalf of others. For example, when deciding whether to give overly positive feedback to a colleague on a poor performance, one faces the choice of whether to provide a short-term gain (i.e., inflate the positive feedback to avoid causing emotional harm) or long-term gain for the other (i.e., give honest feedback in hopes of improving their future performance).

4.1. Procedure and materials

Five hundred seventy participants began our experiment on Mturk, but 36 participants failed the comprehension check and were automatically eliminated from the study. Zero participants failed the attention check, so we used all 534 complete responses in our analyses (46.9% female; $M_{age} = 33$).

As in Study 1, we randomly assigned participants to one of eight experimental conditions in a 2(Deception: honesty vs. lying) × 2(Lie Type: paternalistic lie vs. unequivocal prosocial lie) × 2(Choice Set: choice set 1 vs. choice set 2) between-subjects design. The description of the Sender's information and the procedure for revealing the Sender's deception was identical to that given in Study 2. The main change we made was in the outcomes associated with Options A and B. In the paternalistic lie conditions, Options A and B resulted in the Receiver getting "1 lottery ticket for the chance to win $10 TODAY," or "1 lottery ticket for the chance to win $30 3 MONTHS FROM NOW." We ran two separate pretests on Mturk ($N = 59$, $N = 155$) using the matching method to elicit time preferences (Hardisty, Thompson, Krantz, & Weber, 2013). Both pretests revealed that $30 was the median amount that would make participants indifferent between receiving $10 today and that amount in 3 months. In the unequivocal prosocial lie conditions, Options A and B resulted in 1 or 2 tickets for one of these two lotteries, respectively (counterbalanced across participants).

4.1.1. Dependent variables

After learning the veracity of the Sender's message and the Sender's information, participants provided their moral judgments of the Sender ($\alpha = 0.96$). This scale was identical to that used in Study 2, except that the item "I trust the Sender" was not included, given its redundancy with the item "The Sender is trustworthy."

On the next survey page, participants evaluated their positive affect in response to the Sender's behavior using the same items we used in Study 2 (happy, grateful; $r = 0.79$).

Finally, participants answered questions designed to assess our proposed mechanisms: perceived benevolent intent, perceived autonomy violations, and inaccurate prediction of preferences. To measure perceived benevolent intent, participants indicated their agreement with the statement, "The Sender was trying to do what he/she thought was best for me" (reverse-scored). We assessed perceived autonomy violations directly by asking participants to rate their agreement with the following statement: "The Sender violated my autonomy." We also measured whether participants believed the Sender inaccurately predicted their preferences with the item, "The outcome I wanted was not the one the Sender thought I wanted." All items were displayed in a randomized order and were on a 1–7 scale (1 = *strongly disagree*, 7 = *strongly agree*).[6]

[5] In addition, we included the following exploratory items to assess mechanism (1 = *strongly disagree*, 7 = *strongly agree*): "The Sender wanted to help me"; "The Sender didn't care about what was best for me" (reverse-scored); "The Sender was making assumptions about my preferences"; "The Sender made the wrong decision for me" (reverse-scored); and "The Sender did what was right." However, the mediation analysis with these items is included in the Supplementary Materials because (a) the item "The Sender did what was right" is conceptually similar to items assessing moral judgments, and (b) the items used to assess mechanism here are different from those in Studies 3–5, where we implemented a consistent set of mediation items that were more conceptually distinct from the dependent variables measured.

[6] In addition, we included several exploratory items to assess potential alternative explanations. We measured the extent to which the Sender made assumptions about the preferences of the target ("The Sender was making assumptions about my preferences"); the extent to which they acted based on their own preferences ("The Sender made his/her decision based on his/her own preferences"); and perceived that the Sender attempted to exert influence over them ("The Sender was trying to influence me"). In this study as well as in Studies 4 and 5, there were no significant indirect effects of the three latter items, and inclusion of these items in mediation models did not alter results. Based on the guidance of the review team, we focus our mediation analyses on the first three items (perceived benevolent intentions, perceived autonomy violation, inaccurate prediction of preferences) in the manuscript. We report mediation results with all items in the Supplementary Materials.

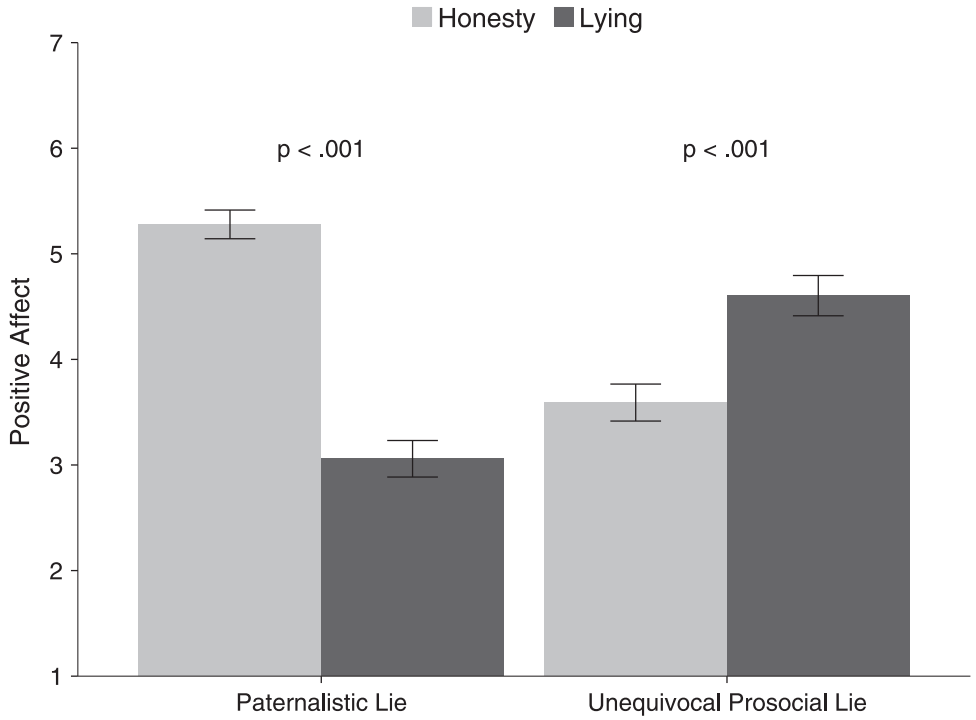

Fig. 2. The effects of paternalistic lies and unequivocal prosocial lies on positive affect in Study 3. In the paternalistic lie conditions, lying and honesty were each associated with either less money today or more money in the future (i.e., different intertemporal choices). In the unequivocal prosocial lie conditions, lying and honesty were associated with 2 vs. 1 lottery tickets, respectively, for the same monetary outcome at the same point in time. Error bars reflect ± 1 SE.

4.2. Results

4.2.1. Moral character

A two-way ANOVA revealed a significant Deception × Lie Type interaction, $F(1, 530) = 65.95$, $p < .001$, $\eta_p^2 = 0.11$. Participants judged Senders who told paternalistic lies ($M = 3.67$, $SD = 1.45$) as significantly less moral than those who were honest ($M = 5.40$, $SD = 0.94$), $t(265) = 11.45$, $p < .001$, $d = 1.40$. In contrast, when the benefits of lying were unequivocal, honesty did not have a significant effect on participants' moral judgments of Senders ($p > .20$). Although unequivocal prosocial lies ($M = 4.70$, $SD = 1.66$) were not perceived to be significantly more moral than truth-telling ($M = 4.50$, $SD = 1.30$) in this study, the results directionally support our hypotheses and past research (Levine & Schweitzer, 2014). In addition, those who told paternalistic lies ($M = 3.67$, $SD = 1.45$) were seen as less moral than those who told unequivocal prosocial lies ($M = 4.70$, $SD = 1.66$), $t(264) = 5.36$, $p < .001$, $d = 0.66$.

There was also a main effect of deception, $F(1, 530) = 41.73$, $p < .001$, $\eta_p^2 = 0.07$, such that participants who received a dishonest message ($M = 4.17$, $SD = 1.64$) judged Senders as less moral than those who received an honest message ($M = 4.94$, $SD = 1.22$), $t(532) = 6.10$, $p < .001$, $d = 0.53$. There was no main effect of lie type ($p > .60$).

4.2.2. Positive affect

A two-way ANOVA revealed a significant Deception × Lie Type interaction, $F(1, 530) = 90.28$, $p < .001$, $\eta_p^2 = 0.17$. Participants reported experiencing significantly less positive affect in response to paternalistic lies ($M = 3.06$, $SD = 2.02$) than honesty ($M = 5.28$, $SD = 1.56$), $t(265) = 10.04$, $p < .001$, $d = 1.23$. In contrast, participants reported experiencing significantly more positive affect in response to unequivocal prosocial lies ($M = 4.60$, $SD = 2.17$) than honesty ($M = 3.59$, $SD = 2.05$), $t(265) = 3.92$, $p < .001$, $d = 0.48$. These results are shown in Fig. 2.

There was also a significant main effect of deception, $F(1, 530) = 12.54$, $p < .001$, $\eta_p^2 = 0.02$. Participants reported more positive affect overall when they received an honest message ($M = 3.81$, $SD = 2.23$) rather than a dishonest one ($M = 4.42$, $SD = 2.01$), $t(532) = 3.26$, $p < .01$, $d = 0.28$. There was no main effect of lie type ($p > .60$).

4.2.3. Mediation

We entered the three focal mechanism items (perceived benevolent intentions: "The Sender was trying to do what he/she thought was best for me"; perceived autonomy violation: "The Sender violated my autonomy"; inaccurate prediction of preferences: "The outcome I wanted was not the one the Sender thought I wanted") simultaneously into a multiple-mediation model using bootstrapping with bias-corrected confidence estimates (Preacher & Hayes, 2004, 2008). We ran a moderated mediation model with 10,000 resamples using deception as the independent variable, lie type as the moderator, and moral character as the dependent variable (PROCESS Macro for SPSS, Model 7, Hayes, 2016).

Table 3
Mediation analyses results from Study 3. Each set of numbers signifies the lower-level and upper-level 95% confidence intervals around the indirect effect for the corresponding item in the first column. The model that was tested included all items in the first column as simultaneous mediators, deception as the IV, moral character as the DV, and lie type as the moderator. We used Hayes' (2016) PROCESS Macro for SPSS, Model 7. Bold numbers indicate confidence intervals that do not contain zero.

	Paternalistic Lies	Unequivocal Prosocial Lies	Index of moderated mediation
1. The Sender was trying to do what he/she thought was best for me	**−1.06, −0.62**	**0.37, 0.82**	**−1.77, −1.09**
2. The Sender violated my autonomy	**−0.28, −0.10**	−0.05, 0.08	**−0.33, −0.09**
3. The outcome I wanted was not the one the Sender thought I wanted	**−0.21, −0.05**	**0.03, 0.16**	**−0.35, −0.08**

Results of the mediation analyses are presented in Table 3. These results suggest that at least three specific processes underlie moral judgments of paternalistic lies. First, targets believed that paternalistic deceivers did not have benevolent intentions. Specifically, paternalistic lies decreased participants' beliefs that the Sender was trying to do what was best for them, $B = -1.84$, $p < .001$. Second, targets believed that the Sender violated their autonomy, $B = 1.00$, $p < .001$. Finally, targets did not believe that the Sender accurately predicted their preferences, $B = 1.33$, $p < .001$. All three of these judgments in turn were significantly associated with moral judgments of the Sender (perceived benevolent intentions: $B = 0.58$, $p < .001$; perceived autonomy violation: $B = -0.49$, $p < .001$; inaccurate prediction of preferences: $B = -0.35$, $p < .001$), and there was a significant indirect effect of paternalistic lies on moral judgments through each of the three mediators (perceived benevolent intentions: 95% CI [−1.06, −0.62]; perceived autonomy violation: 95% CI [−0.28, −0.10]; inaccurate prediction of preferences: 95% CI [−0.21, −0.05]).

Importantly, we found significant evidence of moderated mediation for each of these three processes. As mentioned, a decrease in the belief that the Sender had benevolent intentions (i.e., was trying to do what was best for the target) partially mediated the decrease in perceived moral character resulting from paternalistic lies. In contrast, unequivocal prosocial lies *increased* the belief in benevolent intentions of the Sender, $B = 1.32$, $p < .001$, and this belief partially mediated the positive effect of unequivocal prosocial lying on perceived moral character (95% CI [0.37, 0.82]). We found the same pattern for beliefs about whether the Sender correctly anticipated the outcome the target wanted: while paternalistic lies led targets to believe that Senders were not accurately predicting their preferences, unequivocal prosocial lies increased the belief that Senders *were* accurately predicting their preferences, $B = -0.99$, $p < .001$, which in turn led to more favorable moral judgments (95% CI [0.03, 0.16]). Finally, we found that while targets viewed paternalistic lies as autonomy violations, they did not view unequivocal prosocial lies as such ($p > .70$, 95% CI [−0.05, 0.08]).

4.3. Discussion

Study 3 provides further evidence for the results of Studies 1 and 2 and also lends support for three mechanisms underlying the effect of paternalistic lies on moral judgments. First, beliefs that the Sender did not have benevolent intentions partially explained the effect of paternalistic lies on moral judgments. When the benefits of lying were subjective—that is, in the paternalistic lie conditions—individuals perceived that deceivers did not have targets' interests in mind. When the benefits of lying were obvious, as was the case in the unequivocal prosocial lie conditions, participants perceived that deceivers *did* have their best interests in mind. Because it was reasonable to expect that all targets would prefer two lottery tickets over one lottery ticket for the same outcome, individuals did not doubt the motives of Senders who lied to obtain this outcome for targets.

We also found evidence that perceptions of autonomy violation partially explained the effect of paternalistic lies on moral judgments. These results suggest that individuals believe that paternalistic lies send a coercive signal about the desire to control the deceived party. A related interpretation is that paternalistic lies represent a restriction of the "freedom" to have an undistorted view of the world—a view that is afforded by the truth. Interestingly, when a lie provides individuals with clear benefits over the truth, this freedom is no longer a priority, as unequivocal prosocial lies were not seen as autonomy violations.

Moreover, we obtained evidence for a third mechanism: those who told paternalistic lies were perceived as inaccurately predicting targets' preferences. This finding is particularly striking given that we counterbalanced choice set, or the outcomes that were paired with honesty and dishonesty. Participants thought senders chose incorrectly for them when they lied, regardless of which outcome was associated with the lie. This suggests that receiving an outcome via paternalistic lying may have decreased the attractiveness of the outcome itself, consistent with reactance theory (Brehm et al., 1966; Miron & Brehm, 2006). We explored this possibility further in Study 4.

5. Study 4

Thus far, we have shown (a) that paternalistic lies lead to harsher moral judgments of deceivers; (b) that these lies decrease positive affect amongst targets; and (c) that these effects are driven by doubts about the benevolent motivations of deceivers, the perception that paternalistic liars violated the targets' autonomy, and the perception that paternalistic deceivers inaccurately predicted targets' preferences (Studies 1–3). We also showed that these results are unique to paternalistic lies and do not extend to judgments of unequivocal prosocial lies.

In Study 4, we explored whether individuals' preferences for outcomes change as a result of being the target of a paternalistic lie. In Study 3, we found that targets of paternalistic lies did not believe that the deceiver had correctly anticipated their preferences. As mentioned, one explanation for this result is that the experience of being lied to influenced targets' preferences. To test this notion, we examined whether targets are less satisfied with an outcome resulting from a paternalistic lie than they are when that same outcome is obtained via honesty. Shedding light on this issue has important implications for understanding responses to paternalistic lies from policymakers. Sometimes policies are put in place via dishonest means. For instance, a government might monitor its citizens' personal data under the guise of preventing a terrorist threat, but might also plan to use that data to target other crimes. Examining outcome satisfaction allows us to make claims not only about how targets might respond to these policymakers, but also how they feel about the policies themselves.

This experiment also investigated the moderating effect of individual preferences. Although Study 3 demonstrated that participants believed Senders incorrectly predicted their preferences, it remains unclear whether this effect was driven by a shift in preferences as a result of paternalistic lies, or whether participants happened to receive their less preferred outcome when they were deceived. It is possible that targets who actually received their preferred outcome may reward rather than penalize paternalistic deception. To test this, we conducted a two-part study in which we first measured individual preferences for the outcomes that would be used in the Deception Game. Then, after a period of time had elapsed, participants played the Deception Game. This procedure allowed us to match targets' ex-ante preferences for outcomes to be used in the game with the outcome they actually obtained in the game.

5.1. Procedure and materials

We recruited adult participants from a city in the northeastern United States to participate in a study in exchange for a $10 show-up fee. Two hundred sixty-six participants began the study, but 11 failed the comprehension check and were automatically excluded from the experimenting, yielding 255 complete responses. Two participants failed an attention check prior to the Deception Game, and 30 participants failed a second comprehension check after the Deception Game. Excluding these participants left us with a final sample of 223 participants (73.1% female, $M_{age} = 20$).

Before showing up to the laboratory, participants were required to fill out a short online questionnaire in which we measured individuals' preferences for the outcomes associated with Options A and B in the Deception Game. Specifically, we asked participants whether they would prefer to receive $10 immediately or $30 3 months from now (dichotomous choice). We switched the more immediate option to "$10 immediately" from "$10 today" to strengthen the plausibility of the cover story to laboratory participants. Participants were told that the Sender with whom they were paired in the game had previously completed the study; if participants were scheduled for the first experimental session of the day, it would seem implausible that the Sender could have already participated in a survey that could result in the

participant receiving money that same day.

After arriving at the laboratory, participants were randomly assigned to one of four experimental conditions in a 2(Deception: honesty vs. paternalistic lying) × 2(Choice Set: choice set 1 vs. choice set 2) between-subjects design. The instructions for the Deception Game were the same as those used in Study 3, except that this time the component of the game in which participants chose heads or tails after viewing the Sender's message was eliminated. Instead, participants were told that their payment would be determined by the message chosen by the Sender, rather than their choice as the Receiver (as was the case in Studies 1–3). This change was implemented to ensure that participants could not arrive at an outcome by going against the Sender's message, which could introduce noise in the data (Sutter, 2009). That is, whether one arrives at an outcome by adhering to or going against the Sender's message might moderate outcome satisfaction. By eliminating this possibility, we ensured that outcome satisfaction could only be influenced by (a) preferences for the outcome received and (b) the Sender's honest or dishonest message. In the Sender's information, we described the message as follows: "If you send a message that does (does not) correspond to the actual coin flip outcome, the Receiver can win a $10 bonus IMMEDIATELY ($30 bonus 3 MONTHS FROM NOW)."

5.1.1. Dependent variables

After the Deception Game, participants answered questions to assess their satisfaction with the outcomes by indicating their agreement (1 = *strongly disagree*, 7 = *strongly agree*) with the following statements: "I am satisfied with the outcome I received," and "I am unhappy about the outcome I received" (reverse-scored; $r = 0.66$). We also measured one item regarding satisfaction with the process: "I am satisfied with the process the Sender used to arrive at my outcome." We did not include this in our outcome satisfaction measure because it does not address outcomes per se. However, it follows a similar pattern to that of the other items, and our results do not change if we include it in our outcome satisfaction measure. Following our measures of outcome satisfaction, we also assessed mechanism with the same items used in Study 3.[7]

5.2. Results

5.2.1. Summary statistics

In Part 1 of the study, 28.7% of participants reported preferring $10 immediately and 71.3% preferred $30 3 months from now.

5.2.2. Outcome satisfaction

We conducted a two-way ANOVA entering deception and preferred outcome as factors. Preferred outcome was a binary variable indicating whether participants received their preferred outcome or not. One hundred nineteen participants received the outcome they preferred (53.3%); 104 participants did not (46.7%).

This analysis revealed a main effect of deception, $F(1, 219) = 20.18$, $p < .001$, $\eta_p^2 = 0.08$, such that participants who received an honest message were more satisfied with the outcome they received ($M = 5.49$, $SD = 1.43$) than those who received a paternalistic lie ($M = 4.65$, $SD = 1.63$), $t(221) = 4.04$, $p < .001$, $d = 0.54$. Unsurprisingly, there was also main effect of preferred outcome, $F(1, 219) = 53.80$, $p < .001$, $\eta_p^2 = 0.20$. Those who received their preferred outcome ($M = 5.70$, $SD = 1.31$) were more satisfied than those who did not ($M = 4.33$, $SD = 1.57$), $t(221) = 7.13$, $p < .001$, $d = 0.96$. Interestingly, however, there was no Deception × Preferred Outcome interaction ($p > .90$). Thus, the effect of honesty on outcome satisfaction did not differ depending on whether one's preferred outcome was received. These results are depicted in Fig. 3.

In addition, we examined whether the same outcome was less satisfying when it was obtained through a paternalistic lie than when it was received through honesty. Indeed, participants who received the $30 in 3 months option were significantly less satisfied than when they had obtained that option via a paternalistic lie ($M = 4.73$, $SD = 1.59$) versus honesty ($M = 5.71$, $SD = 1.30$), $t(109) = 3.50$, $p < .001$, $d = 0.67$. Similarly, those who received $10 immediately via a paternalistic lie ($M = 4.56$, $SD = 1.69$) were significantly less satisfied than those who had received that outcome through honesty ($M = 5.28$, $SD = 1.52$), $t(110) = 2.37$, $p = .02$, $d = 0.45$.

Finally, we examined whether individuals were more satisfied when they received their less-preferred outcome via truth-telling or their more-preferred outcome via paternalistic lying. Those who received their preferred outcome via lying ($M = 5.30$, $SD = 1.47$) were marginally more satisfied than those who received their non-preferred outcome via the truth ($M = 4.75$, $SD = 1.53$), $t(108) = 1.92$, $p = .06$, $d = 0.37$.

5.3. Discussion

Study 4 demonstrates that paternalistic lies result in reduced satisfaction with outcomes obtained via those lies.[8] These results suggest that the findings in Studies 1–3 were indeed driven by distaste for paternalistic lies, rather than by dissatisfaction with the outcome obtained in the Deception Game. In addition, across all studies we found that participants' distaste for paternalistic lies (relative to honesty) held regardless of the actual outcomes participants received within the choice sets (see detailed analyses in Supplementary Materials), thus providing further evidence that our results were not driven by initial preferences for outcomes not received. While receiving one's preferred outcome was a stronger predictor of outcome satisfaction than honesty in Study 4, even those who received their preferred outcome were less satisfied when it followed dishonesty rather than honesty.

These findings also suggest that individuals experience reactance towards paternalistic lies and those who tell them. In Study 3, targets believed that paternalistic liars inaccurately predicted their preferences. This result is consistent with reactance theory, whereby the attractiveness of an imposed option is decreased by its imposition (Brehm et al., 1966; Miron & Brehm, 2006). In Study 4, we obtained additional evidence of the belief that deceivers inaccurately predicted targets' preferences by observing that paternalistic lies decreased satisfaction with outcomes. Thus, Study 4 further implicates the role of reactance in responses to paternalistic lies.

Moreover, these results highlight the importance of both honesty and the perceived desirability of outcomes on satisfaction. While individual preferences for a policy, decision, or product may largely influence their satisfaction with these outcomes, the perceived honesty with which these outcomes are obtained likely plays a key role in the extent to which people are satisfied with these outcomes.

6. Study 5

In all studies reported thus far, we employed a version of the Deception Game in which no communication between the Sender and Receiver was permitted, except for the honest or dishonest message from the Sender. This design allows us to isolate the impact of paternalistic lies on participant responses, and also simulates targets' uncertainty about deceivers' motivation for lying. However, sometimes when an individual discovers that she has been the target of a lie, she may confront the deceiver. Given that the deceiver has acted in what she believes is the best interest of the target, the former is likely to

[7] We focus on mechanisms underlying moral judgments in the main text, consistent with Studies 3 and 5, and include mediation with outcome satisfaction as the dependent variable in the Supplementary Materials.

[8] These findings were replicated in an additional experiment with a sample more than twice the size of the current sample (reported in the Supplementary Materials).

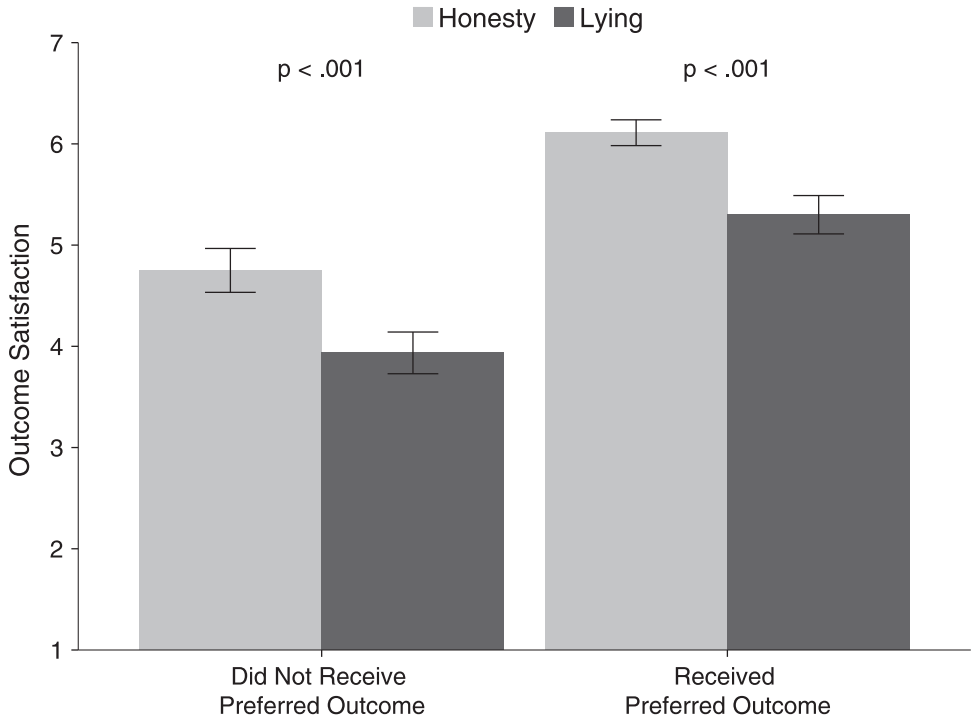

Fig. 3. The effects of receiving one's preferred outcome and paternalistic lies on outcome satisfaction in Study 4. Receiving one's preferred outcome was dummy-coded based on participants' reported preference for either "$10 immediately" or "$30 3 months from now" in Part 1 of the Study. Error bars reflect ± 1 SE.

directly express these good intentions in her defense. But how effective would this defense be at mitigating the target's unfavorable responses to the deceiver? To answer this question, we introduced a new component of the Deception Game in which the Sender could include a personalized message to the Receiver. We explored whether a message conveying the Sender's good intentions would moderate targets' responses to paternalistic lies.

In addition, all studies reported thus far have focused primarily on targets' perceptions of and reactions to paternalistic lies and those who tell them. Here, we introduced a behavioral measure of punishment to document the strength of targets' distaste for paternalistic deception.

6.1. Procedure and materials

Five hundred forty-eight Mturk participants began our study, but 19 participants were automatically excluded from the experiment because they failed the comprehension check. We also excluded one participant who failed an attention check at the beginning of the survey, yielding a final sample of 528 participants (47.2% female; M_{age} = 34).

We randomly assigned participants to one of eight experimental conditions in a 2(Deception: honesty vs. lying) × 2(Communication: communication vs. no communication) × 2(Choice Set: choice set 1 vs. choice set 2) between-subjects design. Those in the no communication conditions engaged in the Deception Game as described in Study 4 (Part 2). Those in the communication conditions received identical procedures, except with additional information about the Sender's ability to send a "personal communication" to the Receiver. These participants were told that the personal communication would be delivered along with the message about the outcome of the coin flip. Participants were told that this personal communication would not affect the bonus participants could earn. On the same screen that displayed the Sender's honest or dishonest message about the coin flip, participants in the communication condition received the Sender's personal communication. This communication read: "Just trying to get you the outcome I thought you'd want." As in the previous experiments, all participants viewed the Sender's private information (i.e., the Sender's deception or honesty, and the payoffs associated with these choices) before we collected our dependent variables.

6.1.1. Dependent variables

After viewing the Sender's information, which included intertemporal payoffs as the choice sets,[9] participants learned about the punishment decision. They were told that the Sender would be entered into a lottery for a $10 bonus, and that they could take away any integer amount (between $0 and $10) from the Sender, though any money they took away would not be added to their own payment. Participants indicated the amount they chose to take away from the Sender, if any.

Next, we measured participants' moral judgments of the Sender (α = 0.94), as well as their positive affect (r = 0.79), using the same items from Study 3. Rather than asking participants to indicate their emotions in response to the Sender, we asked them to indicate the extent to which they felt happy and grateful "right now." Items to measure moral judgments and emotions were displayed on separate and counterbalanced survey pages. Finally, we assessed mechanisms using the same items as in Study 3.

6.2. Results

6.2.1. Manipulation check

To ensure that participants read the Sender's communication and understood that the expressed good intentions were genuine, we examined the effect of communication on the mechanism item, "The Sender was trying to do what he/she thought was best for me," collapsing across whether they received a deceptive message. A t-test revealed a significant difference, such that those who received the communication (M = 5.34, SD = 1.69) expressed greater belief in this statement than those who received no communication (M = 4.93, SD = 1.55), $t(526) = 2.87$, $p < .01$, d = 0.25.

[9] For this study, we ran another pretest (N = 54) with a different method to elicit time preferences. The results of this pretest suggested participants were roughly indifferent between receiving $30 in 3 months and $17.50 today. Thus, we used these options in the choice sets.

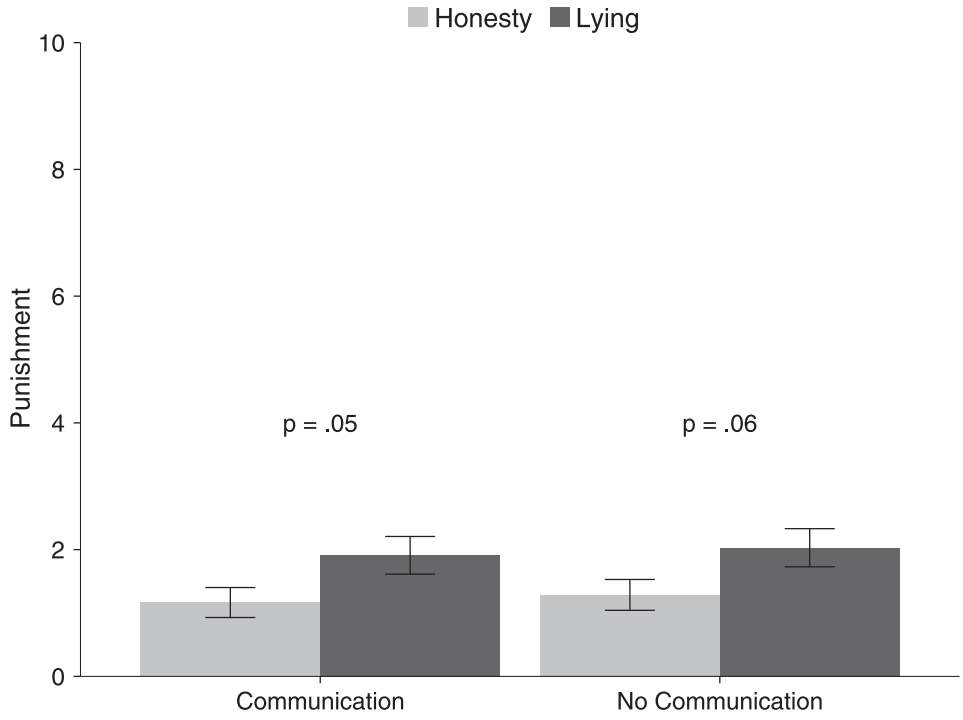

Fig. 4. The effects of communication and paternalistic lies on punishment in Study 5. Participants in the communication condition received a personal communication from the Sender signaling benevolent intent. Those in the no communication condition received no additional communication. Error bars reflect ± 1 SE.

6.2.2. Punishment

A two-way ANOVA with deception and communication included as factors revealed a main effect of deception, $F(1, 524) = 7.46$, $p < .01$, $\eta_p^2 = 0.01$. Those who were told a paternalistic lie ($M = 1.97$, $SD = 3.46$) punished Senders more than those who received an honest message ($M = 1.23$, $SD = 2.73$), $t(526) = 2.74$, $p < .01$, $d = 0.24$. Interestingly, there was no main effect of communication ($p > .60$) and no interaction ($p > .90$). We also looked at the difference in punishment between those who had been lied to with and without communication; the difference was not significant ($p > .70$). These results are depicted in Fig. 4.

6.2.3. Moral character

A two-way ANOVA with moral character as the dependent variable also revealed a main effect of deception, $F(1, 524) = 60.10$, $p < .001$, $\eta_p^2 = 0.10$. Senders who told paternalistic lies ($M = 4.53$, $SD = 1.30$) were judged as less moral than Senders who had been honest ($M = 5.29$, $SD = 0.93$), $t(526) = 7.69$, $p < .001$, $d = 0.67$. Unlike our results for punishment, there was a significant main effect of communication, $F(1, 524) = 8.18$, $p < .01$, $\eta_p^2 = .02$, such that those who communicated benevolent intent ($M = 5.04$, $SD = 1.18$) were perceived as more moral than those who did not ($M = 4.77$, $SD = 1.20$), $t(526) = 2.62$, $p < .01$, $d = 0.23$. This effect held when comparing those who received a lie with communication ($M = 4.74$, $SD = 1.32$) to those who received a lie with no communication ($M = 4.31$, $SD = 1.25$), $t(266) = 2.78$, $p < .01$, $d = 0.34$. There was no Deception × Communication interaction ($p > .10$).

6.2.4. Positive affect

A two-way ANOVA examining the effects of deception and communication on affect yielded results analogous to those for punishment: There was a significant main effect of deception, $F(1, 524) = 32.90$, $p < .001$, $\eta_p^2 = 0.06$, such that dishonesty ($M = 4.09$, $SD = 1.73$) resulted in less positive affect than honesty ($M = 4.91$, $SD = 1.58$), $t(526) = 5.74$, $p < .001$, $d = 0.50$. There was neither a main effect of communication nor a Deception × Communication interaction ($ps > .50$). Furthermore, there was no effect of communication on positive affect among those who had received a dishonest message ($p > .90$).

6.2.5. Mediation

As in Study 3, we assessed the mechanisms underlying moral judgments of paternalistic lies. We ran a moderated mediation model, with deception as the independent variable, communication as the moderator, and moral character as the dependent variable (PROCESS Macro for SPSS, Model 7, Hayes, 2016). We entered all mechanism items (perceived benevolent intentions: "The Sender was trying to do what he/she thought was best for me"; perceived autonomy violation: "The Sender violated my autonomy"; inaccurate prediction of preferences: "The outcome I wanted was not the one the Sender thought I wanted") simultaneously into a multiple-mediation model using bootstrapping with bias-corrected confidence estimates (Preacher & Hayes, 2004, 2008).

Results of the mediation analyses are presented in Table 4. We found significant evidence for mediation for the same three mechanisms identified in Study 3. Furthermore, we found no evidence for moderated mediation. The same mechanisms drove perceptions of moral character in both the communication and the no communication conditions.

Specifically, in both conditions, paternalistic lies resulted in decreased beliefs that the Sender had prosocial intentions (communication: $B = -0.52$, $p = .01$; no communication: $B = -0.66$, $p < .001$). Perceptions of the Sender's prosocial intentions were significantly associated with moral judgments (communication: $B = 0.52$, $p < .001$; no communication: $B = 0.54$, $p < .001$), and there was a significant indirect effect of paternalistic lies on moral judgments through this mechanism (communication: 95% CI[−0.41, −0.06]; no communication: 95% CI[−0.45, −0.12]). This result is a testament to the robustness of the skepticism about deceivers' benevolent intentions resulting from paternalistic lies. Although communicating benevolent intent improved moral judgments of the Sender relative to no communication, lying increased the belief that the Sender was not acting in target's best interest even when benevolent intentions were communicated. This belief in turn led to lower judgments of moral character towards paternalistic lies than honesty.

Table 4
Results of mediation analyses from Study 5. Each set of numbers signifies the lower-level and upper-level 95% confidence intervals around the indirect effect for the corresponding item in the first column. The model that was tested included all items in the first column as simultaneous mediators, deception as the IV, moral character as the DV, and communication as the moderator. We used Hayes' (2016) PROCESS Macro for SPSS, Model 7. Bold numbers indicate confidence intervals that do not contain zero.

	Communication	No communication	Index of moderated mediation
1. The Sender was trying to do what he/she thought was best for me	**−0.41, −0.06**	**−0.45, −0.12**	−0.28, 0.18
2. The Sender violated my autonomy	**−0.11, −0.01**	**−0.12, −0.01**	−0.07, 0.05
3. The outcome I wanted was not the one the Sender thought I wanted	**−0.17, −0.05**	**−0.14, −0.02**	−0.04, 0.10

Furthermore, in both conditions, targets believed that the Sender violated their autonomy (communication: $B = 0.38$, $p = .02$; no communication: $B = 0.42$, $p = .01$) and thought that the Sender did not accurately predict their preferences (communication: $B = 0.98$, $p < .001$; no communication: $B = 0.71$, $p < .001$). Moral judgments were significantly predicted by both perceived autonomy violation (communication: $B = -0.44$, $p < .001$; no communication: $B = -0.38$, $p < .001$) and perceived inaccurate predictions of preferences (communication: $B = -0.26$, $p < .001$; no communication: $B = -0.31$, $p < .001$), and there was a significant indirect effect of paternalistic lies on moral judgments through each of these mechanisms for both those in the communication and no communication conditions (autonomy violation, communication: 95% CI[−0.11, −0.01]; no communication: 95% CI[−0.12, −0.01]; inaccurate prediction of preferences, communication: 95% CI [−0.17, −0.05]; no communication: 95% CI[−0.14, −0.02]).

6.3. Discussion

When deception is uncovered, a possible response of the deceiver is to defend her actions, explaining that she lied because she believed it was in the target's best interest. In Study 5, we tested the effectiveness of this type of defense. While communication of benign intentions did improve judgments of the Sender's moral character, it had no effect on punishment of deceivers or on targets' emotional responses to deception. Moreover, targets believed deceivers were not prosocially motivated, viewed lying as an autonomy violation, and thought deceivers inaccurately predicted their preferences, even when the deceiver tried to communicate good intentions.

7. Study 6

In Studies 1–5, we provided consistent evidence of an aversion to paternalistic lies. However, one potential criticism of these studies is that the Deception Game, while well-controlled, does not fully capture the essence of paternalistic lies. Though the subjective nature of the benefits of paternalistic lies in the game are analogous to the uncertainty associated with real-world outcomes of paternalistic lies, the abstract framing of the game is quite dissimilar to the real-world contexts in which paternalistic lies are told. Furthermore, interactions in the game were between strangers, whereas paternalistic lies in everyday life often occur between friends, colleagues, romantic partners, and other relationships in which both parties are at least acquainted with one another.

Considering these issues, in Study 6, we implemented a different methodology that allowed us to measure judgments of paternalistic lies in a more externally valid setting. Here, participants read several vignettes in which they were asked to imagine that they discovered they had been deceived. In each vignette, we manipulated (a) whether the interests of the target were known or unknown to the deceiver (i.e., whether the lie was paternalistic or unequivocally prosocial) and (b) whether the deceiver communicated his/her benevolent intent to the target. We included both paternalistic and unequivocal prosocial lies in this study to determine whether individuals indeed respond to these lies differently in more realistic contexts. Specifically, we sought to provide further evidence of Study 1 and 3's findings that those who tell paternalistic lies (i.e., when the interests of the target are unknown) are viewed as less moral than those who tell unequivocal prosocial lies (i.e., when the interests of the target are known). Additionally, we extended Study 5's investigation of the effects of communication on judgments of paternalistic deceivers in order to determine whether communicating benign intent has differential effects when there is an existing relationship between the target and deceiver.

7.1. Procedure and materials

We received 394 complete responses from Mturk. Three participants failed an attention check at the beginning of the study and were thus excluded. We also excluded three responses from participants who had already taken the survey (though the original responses of these participants were retained). This left a final sample of 388 (42.8% female, $M_{age} = 38$).

Participants were randomly assigned to one of four conditions in a 2(Lie Type: paternalistic lie vs. unequivocal prosocial lie) × 2(Communication: communication vs. no communication) between-subjects design. Within each condition, all participants read three vignettes that were displayed in a randomized order, with page breaks separating each vignette. In these vignettes, participants were asked to imagine that they had been the target of a paternalistic or unequivocal prosocial lie (depending on condition). Whereas in Studies 1–5 we manipulated paternalistic deception by altering whether the benefits of lying were subjective or objective, in Study 6 we directly manipulated the degree to which the deceiver was aware of the target's preferences, while holding constant the amount of time the deceiver and target knew each other in each vignette. We also manipulated whether the deceiver communicated his/her intentions to help the target by lying (depending on condition). For example, one scenario read as follows:

> *You and your friend Jill are out to dinner at Jill's favorite restaurant. You are trying to lose weight and eat healthy. You ask what Jill recommends. She says that the signature salad is her favorite item on the menu. Weeks later you learn that Jill lied and that her favorite menu item is actually the double cheeseburger.*
> Paternalistic lie: *You and Jill have been friends for about 6 months. You two had never discussed whether you desired to lose weight and avoid temptation or to indulge in tasty but unhealthy foods.*
> Unequivocal prosocial lie: *You and Jill have been friends for about 6 months. You two had discussed your desire to lose weight and avoid temptation or to indulge in tasty but unhealthy foods.*
> Communication: *Jill tells you that she lied because she wanted you to eat healthy.*
> No communication: [*No additional information*]

The other two vignettes, which depict lies from a coworker and a doctor, are reprinted in the Appendix.

After each vignette, participants provided moral judgments of the deceiver and rated the positive affect they expected to experience in response to the deceiver's behavior. The items and scales used to measure moral judgments were the same as those in Studies 3 and 5. For moral judgments, the prompt read, "Please indicate the extent to which the following words characterize [Jill] from the scenario above. [Jill] is…" For affect, the prompt read, "If you were actually the person

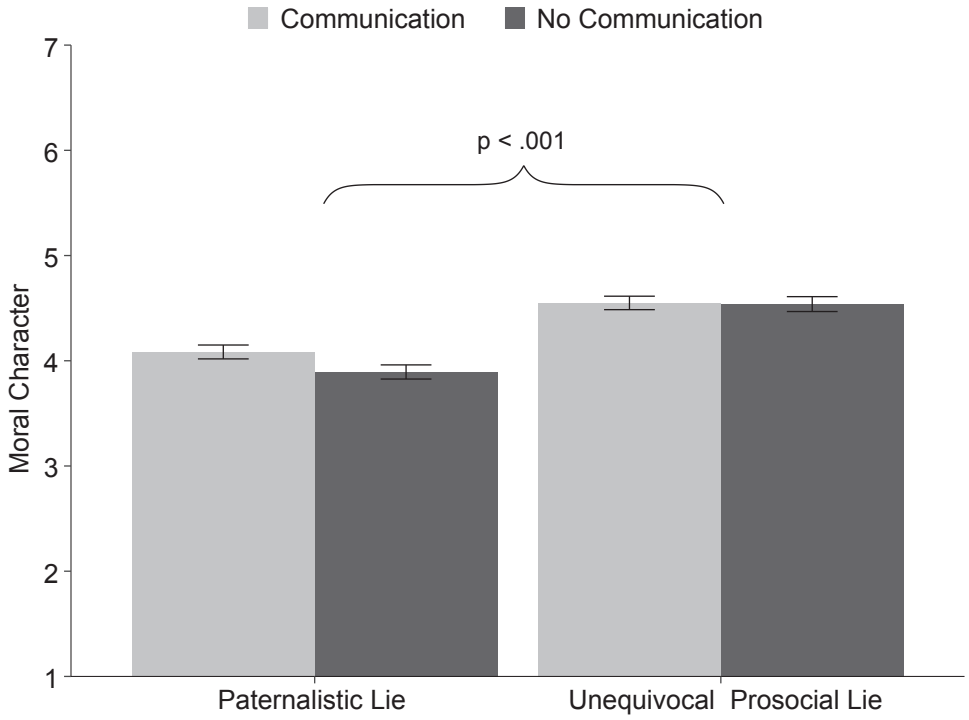

Fig. 5. The effects of communication on perceived moral character for those who received a paternalistic lie or an unequivocal prosocial lie in Study 6. Participants in the communication condition viewed a statement from the deceivers depicted in the vignettes that signaled their benevolent intent. Those in the no communication condition saw no additional communication. Error bars reflect ± 1 SE.

in the above scenario, please indicate the extent to which you would experience the following emotions in response to [Jill's] behavior." Each vignette was displayed to participants as they made their ratings.

7.2. Results

In Study 6, we were interested in examining the effect of lie type, communication, and their interaction on judgments of moral character and positive affect. We therefore report the results of 2(Lie Type: unequivocal prosocial lie vs. paternalistic lie) × 2(Communication: communication vs. no communication) ANOVAs on judgments of moral character and affect collapsed across vignettes. Mixed-model ANOVAs that include the effects of vignette are included in the Supplementary Materials. However, inclusion of vignette in the models does not moderate our results.

7.2.1. Moral character

There was a significant effect of lie type on judgments of moral character, $F(1, 384) = 67.56$, $p < .001$, $\eta_p^2 = 0.15$. Those who imagined they were targets of unequivocal prosocial lies ($M = 4.54$, $SD = 0.66$) judged deceivers as more moral than those who were targets of paternalistic lies ($M = 3.99$, $SD = 0.66$), $t(386) = 8.20$, $p < .001$, $d = 0.83$. There was no main effect of communication ($p > .10$) and no Lie Type × Communication interaction ($p > .10$). We also tested whether communication had an effect within each lie type; there was a marginally significant effect of communication for those in the paternalistic lie conditions, $t(192) = 2.02$, $p = .05$, $d = 0.29$. Communication marginally improved moral judgments of those who told paternalistic lies ($M_{\text{Communication}} = 4.08$, $SD_{\text{Communication}} = 0.67$ vs. $M_{\text{No communication}} = 3.89$, $SD_{\text{No communication}} = 0.64$). There was no effect of communication for those in the unequivocal prosocial lie conditions ($p > .90$). These results are displayed in Fig. 5.

7.2.2. Positive affect

Similar results were obtained for positive affect. There was a significant effect of lie type, $F(1, 384) = 73.62$, $p < .001$, $\eta_p^2 = 0.16$, such that those who imagined they were targets of unequivocal prosocial lies ($M = 3.81$, $SD = 1.12$) reported more positive affect than those who were targets of paternalistic lies ($M = 2.83$, $SD = 1.11$), $t(386) = 8.57$, $p < .001$, $d = 0.87$. There was no effect of communication ($p > .30$), and no interaction ($p > .20$). We also examined the effect of communication within each lie type; the effect of communication was not significant for either those who received a paternalistic lie or those who received an unequivocal prosocial lie ($ps > .10$).

7.3. Discussion

Study 6 provides evidence for the external validity of individuals' aversion to paternalistic lies. Using a design that depicted realistic contexts and relationships in which paternalistic lies are told, we replicated Studies 1 and 3's findings that paternalistic lies result in harsher moral judgments than unequivocal prosocial lies.

In addition, this study offers evidence of the limitations of communication on mitigating the negative effects of paternalistic lies. In both Studies 5 and 6, communication had no effect on affect resulting from being the target of a paternalistic lie. Unlike in Study 5, however, where communication improved moral judgments of paternalistic liars, in Study 6, communication had only a marginally significant effect on moral judgments of those who tell paternalistic lies. Given the use of a more realistic context for studying paternalistic lies in Study 6, these results suggest that the ability of the communication of benevolent intent to reduce the harmful effects of paternalistic liars may be limited.

8. Study 7

In Studies 3 and 5, we assessed the mechanism behind paternalistic lies' effects on moral judgments using mediation analysis. While consistent mediation results across these studies provide evidence for the underlying process, this type of analysis is limited by its correlational nature (Spencer, Zanna, & Fong, 2005). In Study 7, we offer stronger causal evidence for one of the mechanisms uncovered in Studies 3 and

5—perceived lack of benevolent intent—by directly manipulating this construct.

In this experiment, we employed a vignette design similar to that of Study 6, where participants read three different vignettes depicting the telling of paternalistic lies. Here, we varied between-subjects whether the benevolent intent of the deceiver was ambiguous or made clear to participants through a statement that provided an omniscient third-person perspective into the inner state of the deceiver. Although we found that personally communicating one's good intent was not seen as credibly signaling benevolence in the context of paternalistic lies, this study directly examines the role of perceived benevolent intent by manipulating it directly. We predicted that if targets knew for certain that deceivers lied with good intentions, this would improve moral judgments relative to when the motivations of the deceiver are more ambiguous. In Study 7, we also included new vignettes that depicted paternalistic lies told by individuals in leadership positions, thus allowing us make inferences about the effects of paternalistic lies in relatively higher-stakes contexts.

8.1. Procedure and materials

We received 214 complete responses from Mturk. Eight participants were excluded from analyses for failing an attention check. This left a final sample of 206 (53.9% female, M_{age} = 38).

Participants were randomly assigned to one of two conditions: ambiguous motivation or benevolent motivation of deceivers. Within each condition, participants read three vignettes in a randomized order. As in Study 6, for each vignette participants were asked to imagine that they were the target of a paternalistic lie. In Study 6, we manipulated whether the deceiver verbally communicated benevolent intent to the target. In Study 7, we manipulated whether the deceivers' benevolent intent was clear to participants by including a statement from a third-person omniscient perspective that described the deceiver's private thoughts and motivation. For example, one vignette depicted a government official who lied to constituents:

> *Imagine that you are at a community board meeting listening to a local government official speak. There have been rumors about a possible security threat in your city, and the government official is addressing those concerns.*
> *The official insists that the rumors are unsubstantiated, and that there is no security threat.*
> *Weeks later, however, news emerges that there was in fact substantial evidence of a security threat, and the government official knew about this evidence at the time of the community board meeting.*
> *This government official had been in his/her position for around 6 months, and was unaware of your preferences and other constituents' preferences to be fully informed in the event of a threat, or to be uninformed in order to not worry.*
> Benevolent Motivation: *In actuality, the government official lied about the security threat because s/he believed there was nothing the public could do about the threat and that everyone would be better off not worrying. S/he was sincerely trying to do what s/he thought was best for you and the public.*
> Ambiguous Motivation: [*No additional information*]

The second vignette depicted a lie from a doctor (adapted from Study 6), and the third depicted a lie from a financial advisor. These vignettes are reprinted in the Appendix.

After each vignette, participants provided moral judgments of the deceiver, using the same items and prompt as in Study 6. Each vignette was displayed to participants as they made their ratings. In addition, we included items to measure each of the three mechanisms identified in Studies 3 and 5: perceived benevolent intent, autonomy violation, and inaccurate prediction of preferences. The items used to measure these constructs were the same as those in Studies 3 and 5, except "the Sender" was replaced with the deceiver depicted in the vignette (i.e., the doctor, the financial advisor, the government official). The item, "The [deceiver] was trying to do what s/he thought was best for me" served as a manipulation check of perceived benevolent intent. The items "The outcome I wanted was not the outcome the [deceiver] thought I wanted" and "The [deceiver] violated my autonomy" served as tests of discriminant validity—that is, if our experimental treatment indeed manipulated benevolent intent only, the manipulation should not produce changes in these items measuring other constructs.

8.2. Results

In Study 7, we sought to test the impact of benevolent intentions on moral judgments of paternalistic lies. Because there were only two between-subjects treatments in this experiment, we report results of *t*-tests to compare moral judgments of deceivers across conditions, collapsing across vignettes. Mixed-model ANOVAs that include the effects of vignette are reported in the Supplementary Materials.

8.2.1. Manipulation check

A *t*-test indicated that our benevolent intent manipulation worked as planned. Those in the benevolent motivation condition exhibited higher scores (M = 4.68, SD = 1.19) on the item, "The [deceiver] was trying to do what s/he thought was best for me," than those in the ambiguous motivation condition (M = 4.18, SD = 1.07), $t(204) = 3.15$, $p < .01$, $d = 0.44$.

8.2.2. Moral character

There was a significant effect of the motivation manipulation on moral judgment of deceivers. Those in the benevolent motivation condition (M = 3.76, SD = 0.69) rated deceivers as more moral than those in the ambiguous motivation condition (M = 3.53, SD = 0.65), $t(204) = 2.48$, $p = .01$, $d = 0.35$.[10]

8.2.3. Discriminant validity

As mentioned, our manipulation successfully increased perceived benevolent intent. In order to assess the discriminant validity of this manipulation, we examined whether this manipulation also affected perceived autonomy violation or the perception that the deceiver inaccurately predicted one's preferences. There were no differences across conditions for the item "The [deceiver] violated my autonomy" ($p > .10$), or for the item "The outcome I wanted was not the one the [deceiver] thought I wanted" ($p > .60$).

8.3. Discussion

In Study 7, we provide causal evidence that perceived benevolent motivation is a mechanism underlying the effects of paternalistic lies on moral judgments. An experimental manipulation that made explicit deceivers' internal desire to benefit the target via lying improved moral judgments, relative to when deceivers' motivations were not specified. Moreover, the manipulation of benevolent intent did not influence perceived autonomy violation or perceived inaccurate prediction of preferences, thereby highlighting the discriminant validity of this manipulation, as well as offering evidence that these three mechanisms are indeed unique constructs. These results bolster evidence from mediation analyses in Studies 3 and 5, which illustrate the importance of perceived motivation in determining responses to paternalistic lies. In

[10] As described in the Supplementary Materials, there was also a significant Motivation × Vignette interaction, $F(2, 408) = 4.60$, $p = .01$, $\eta_p^2 = 0.02$. The effect of motivation was significant for the government ($M_{Benevolent}$ = 3.65, $SD_{Benevolent}$ = 0.83 vs. $M_{Ambiguous}$ = 3.27, $SD_{Ambiguous}$ = 0.94; $F(1, 204) = 9.34$, $p < .01$, $\eta_p^2 = 0.04$) and finance ($M_{Benevolent}$ = 3.64, $SD_{Benevolent}$ = 0.92 vs. $M_{Ambiguous}$ = 3.29, $SD_{Ambiguous}$ = 0.88; $F(1, 204) = 8.19$, $p < .01$, $\eta_p^2 = 0.04$) vignettes, but not for the healthcare vignette ($p > .25$).

addition, this study expands the contexts in which we investigate paternalistic lies. Compared to the vignettes in Study 6, those in Study 7 depict lies from individuals in leadership positions in relatively higher-sakes situations, thereby highlighting the potentially detrimental effects of paternalistic lies.

9. General discussion

This work adds to our understanding of deception, highlighting how responses to lies hinge on the perceived benefits afforded by lying, as well as the perceived motives of deceivers. Although targets may reward lies that yield unequivocal benefits, they penalize lies that involve others making subjective judgments about their best interests. We identify a robust distaste towards paternalistic deception across moral judgments, affect, punishment, and satisfaction with the outcomes associated with lying.

Our research makes several contributions to theory on deception. First, we broaden the taxonomy of lies by introducing the construct of paternalistic lies. Although paternalistic lies are ubiquitous and have important consequences for both targets and deceivers, no prior research has examined these lies. We distinguish paternalistic lies from unequivocal prosocial lies, another class of lies that are intended to benefit others that have been studied in past work, and demonstrate how responses to paternalistic lies differ from responses to unequivocal prosocial lies.

This research also extends the growing body of research on prosocial lying. Our results identify a boundary condition of the positive effects of prosocial lying (Levine & Schweitzer, 2014, 2015), showing that paternalistic lies and unequivocal prosocial lies can yield divergent moral judgments and affective responses. In Levine and Schweitzer's (2014, 2015) work, unequivocal prosocial lies were perceived to be benevolent. Similarly, in the present research, unequivocal prosocial lies elicited the judgment that the deceiver was truly trying to do what they thought was best for the target (see mediation results in Studies 3 and 5), which is also indicative of perceived benevolent intent. This credible signal of benevolence led to positive judgments of moral character. In contrast, for paternalistic lies, the signal of benevolence is less credible. We find that targets do not believe that deceivers who tell a paternalistic lie were truly trying to do what they thought was best for the target, and that this diminished belief in deceivers' benevolent intent in turn drove the decrease in perceived moral character. Thus, this research highlights the theoretical and practical importance of perceived benevolence in shaping moral judgments.

In addition to identifying perceived benevolent intent as a mechanism behind negative responses to paternalistic lies, we uncover two additional mechanisms underlying these responses: the perception that paternalistic lies violate targets' autonomy, and the perception that paternalistic liars inaccurately predicted targets' preferences. Not only do these findings shed further light on the processes that drive responses to paternalistic lies, but they also suggest that paternalistic lies can elicit reactance amongst targets. According to Miron and Brehm (2006), behavioral indicators of reactance include derogation of the agent restricting one's freedom, as well as a decrease in attractiveness of the imposed option or an increase in the attractiveness of the restricted option. In our experiments, we see evidence for both of these phenomena. Participants derogated deceivers via moral judgments (Studies 1–3, 5–7), and punishment (Study 5). Furthermore, perceptions that the deceiver had inaccurate predicted the target's preferences drove decreases in moral judgments (Studies 3 and 5), and paternalistic lies actually decreased satisfaction with outcomes that were received as a result of these lies (Study 4). We also found that paternalistic lies harm affective responses—another sign of reactance (Miron & Brehm, 2006). Taken together, these findings provide the first evidence to our knowledge that deception can produce reactance.

Our results also present a novel application of theory on procedural justice. A widespread finding in the justice literature is the interaction between procedural fairness and outcome desirability, such that the relationship between procedural fairness and individuals' reactions is stronger when outcome desirability is low (Brockner & Wiesenfeld, 1996). This finding has not yet been applied to judgments of deception, yet our results are consistent with this theory: when individuals are the target of an unequivocal prosocial lie (i.e., a lie with an objectively desirable outcome), they respond favorably, despite the arguably unfair or immoral action that was taken to produce that outcome. When they are the target of a paternalistic lie, however, (i.e., a lie with an outcome that is not objectively desirable), they become more sensitive to the fact that they were deceived, and thus, respond harshly. Similarly to how perceptions of outcomes and procedures interact to produce individuals' reactions in an organizational context, the degree to which individuals react negatively or positively to lies depends on the relative desirability of the outcomes associated with those lies.

Apart from its theoretical contributions, this work also has practical implications for interpersonal interactions, management, and policy-making. Leaders and policy-makers often withhold or distort the truth in the perceived best interests of their stakeholders. Although targets may respond positively when the lie is clearly favorable to them, individuals often lack full insight into others' preferences (e.g., Hsee & Weber, 1997), and there is often uncertainty about the ultimate consequences of deception. Our results indicate that well-intended lies may backfire if deceivers lack sufficient knowledge about what is actually in targets' best interests. Targets are likely to penalize paternalistic lies, as well as the policies, people, and products associated with them.

Relatedly, our work indicates that paternalistic lies have detrimental effects not only on interpersonal perceptions, but also perceptions of outcomes resulting from these lies. Sometimes individuals need to make decisions on behalf of stakeholders that require a choice between two alternatives that have different assets and tradeoffs. For example, a government organization may be faced with the decision of whether to protect citizens' privacy, or obtain personal data to screen for a terrorist threat (e.g., Nakashima, 2016). The decision-maker may act in what she truly believes is the stakeholder's best interest, and the stakeholders' preferences for each of these options are clearly important in determining their satisfaction with the decision. However, our work suggests that the stakeholders may respond more favorably to the outcome that is delivered with transparency than to the outcome that is delivered via deception.

These results open up several potential avenues for further research. One important area of future study would be to investigate moderators of responses to paternalistic lies to determine how opposition to these lies might be reduced. We obtained mixed evidence regarding whether communicating benign intent can soften the blow of paternalistic lies: Communication did improve moral judgments of paternalistic liars in Study 5, but only marginally so in Study 6. Communication also did not decrease punishment of paternalistic liars (Study 5). However, in Study 7, knowledge of deceivers' good intentions via insight into their internal thoughts did improve moral judgments. This suggests that communication in Studies 5 and 6 may not have effectively convinced participants of deceivers' benign intentions. It is possible that communication that *does* successfully convey deceivers' good intent would allay the negative effects of paternalistic lies. Given the limited effectiveness of communication in our work, future research should examine other ways in which liars can successfully convey their benevolent intentions in order to mitigate the harmful effect of paternalistic lies.

Conversely, there are likely other factors that can exacerbate negative responses to paternalistic lies that are worthy of further investigation. In our research, we purposely structured the Deception Game such that the deceiver had no stake in the game so that we could cleanly study paternalistic lies (relative to the truth and unequivocal prosocial lies), without confounding paternalism with self-interest. Likewise, in the vignettes used in Studies 6 and 7, no ulterior motives of deceivers are mentioned. In the real world, however, deceivers may have mixed motives. For example, one tasked with delivering feedback

about a poor performance may upwardly inflate this feedback to prevent causing emotional harm, but also to avoid the discomfort of an awkward situation. In this work, we find that perceived intentions of deceivers play a key role in the divergent effects of paternalistic lies and unequivocal prosocial lies. We would expect, then, that if the liar were known or believed to have ulterior motives, paternalistic lies would be penalized to an even greater extent. Future research should explore this notion.

It will also be important for future work to examine the situations in which lies are more likely to be perceived as paternalistic versus unequivocally prosocial. In certain circumstances, there may be broad consensus that lying serves a target's best interests. In these cases, lies are likely to elicit positive reactions. For example, most people would probably agree that telling a bride she is beautiful on her wedding day is in the bride's best interest, regardless of the truth. Thus, an individual who tells a lie to this effect may be rewarded. However, in other circumstances, there may be little consensus on whether lying is beneficial. In these circumstances, the lie will likely be perceived as paternalistic, and elicit negative reactions. For example, there may be considerable disagreement about whether falsely telling a woman she looks beautiful on an ordinary day is in the woman's best interest. Thus, an individual who tells such a lie may be penalized. Recent research suggests that there are systematic circumstances in which lies are generally perceived to benefit targets (Levine, 2017). It will be interesting for future research to examine if judgments of paternalism are reduced in these contexts.

Another possibility for future work would be to investigate how the relationship between the deceiver and the target influences perceptions of paternalistic lies. In close interpersonal relationships, targets may trust communicators to accurately predict their preferences and may be less skeptical of their motives. In these settings, individuals may experience less hostility towards paternalistic lies. Consistent with this proposition, recent research suggests that perceptions of paternalistic policies hinge on trust in the policy-maker (Tannenbaum & Ditto, 2016; Tannenbaum, Fox, & Rogers, 2016). While we investigate paternalistic lies between strangers in Studies 1–5 and a variety of closer relationships in Studies 6 and 7, more research is necessary to isolate how paternalistic lies are viewed in close versus distant relationships, and how other specific features of a deceiver-target relationship may moderate responses to these lies.

A final potential avenue for future research would be to explore how the method of deception influences perceptions of paternalistic lies and those who tell them. In our research, we explore paternalistic lies in the form of a false statement from deceivers. However, there are other forms of deception that can be considered paternalistic. For example, when faced with the opportunity to tell a paternalistic lie, one can omit information in order to deceive someone for their purported benefit (i.e., lies of omission; Levine et al., 2018). One can also choose to change the subject of conversation (e.g., palter; Rogers, Zeckhauser, Gino, Norton, & Schweitzer, 2017), or actively choose to not disclose any information (e.g., pleading the Fifth Amendment). Recent work suggests that opting to not disclose negative information can result in worse judgments than honest disclosure (John, Barasz, & Norton, 2016). It would be interesting for future work to test how paternalistic lies fare against these alternate modes of communication in terms of influencing social judgments of the communicator.

People are frequently faced with opportunities to engage in paternalistic deception. Though individuals might be tempted to lie with the intent to help others, the uncertainty laden in how the lie will affect the targets should give the potential deceivers pause about the decision. When the consequences of dishonesty are not unequivocally preferable to those of honesty, these parties may be better off telling the truth.

Acknowledgments

We are grateful for financial support from the Wharton Behavioral Lab; the Wharton Operations, Information, and Decisions Department; the Russell Ackoff Fellowship of the Wharton Risk Management and Decision Processes Center; as well as the Center for Decision Research and the Charles E. Merrill Faculty Research Fund at the University of Chicago.

Appendix A

A.1. Vignettes in Study 6

A.1.1. Healthcare vignette

Imagine that your spouse has a fatal cancer. You and your spouse met with the doctor, who informed you that your spouse's existing treatment has not been effective and that the cancer has spread to your spouse's bones and brain. You know your spouse may pass away soon and you have already prepared for the worst. However, the doctor says that there is always hope and that your spouse may qualify for a new experimental treatment soon. A few weeks later, your spouse passes away. You subsequently find out that the doctor knew that your spouse was too sick to receive any experimental treatments in the future.

Unequivocal prosocial lie: *The doctor had known you and your spouse for around 6 months. He had discussed your and your spouse's preferences for negative information. He knew that you both wanted to remain hopeful and optimistic rather than receive complete candor in such dire circumstances.*

Paternalistic lie: *The doctor had known you and your spouse for around 6 months. He had never discussed your and your spouse's preferences for negative information. He did not know whether you and your spouse wanted to remain hopeful and optimistic, or whether you and your spouse wanted complete candor in such dire circumstances.*

No communication: [*No additional information*]

Communication: *The doctor tells you that he lied about the experimental treatment options because he wanted to preserve your and your spouse's hope.*

A.1.2. Feedback vignette

Imagine that you are an employee of a large consumer packaged-goods company. You have been chosen to deliver a speech to thousands of your fellow coworkers at this year's annual sales meeting. The day before the meeting, you practice your speech in front of your coworker, Nick. Nick tells you that the speech is wonderful. At the sales meeting, your speech went fine. However, several weeks later, you find out that Nick actually did not think the speech was particularly interesting or engaging when he first heard it.

Unequivocal prosocial lie: *Nick has been your coworker for about 6 months. You had told him in the past that you benefit from encouragement and reassurance rather than criticism before giving speeches.*

Paternalistic lie: *Nick has been your coworker for about 6 months. You had not discussed in the past whether you would benefit from encouragement and reassurance or criticism before giving speeches.*

No Communication: [*No additional information*]

Communication: *Nick tells you that he lied about his opinion of the speech because he thought it would help you feel and perform better.*

A.2. Vignettes in Study 7

A.2.1. Healthcare vignette

Imagine that your spouse has a fatal cancer. You and your spouse met with the doctor, who informed you that your spouse's existing treatment has not been effective and that the cancer has spread to your spouse's bones and brain. You know your spouse may pass away soon and you have already prepared for the worst. However, the doctor says that there is always hope and that your spouse may qualify for a new experimental treatment soon. A few weeks later, your spouse passes away. You subsequently find out that the doctor knew that your spouse was too sick to receive any experimental treatments in the future. The doctor had known you and your spouse for around 6 months. He had never discussed your and your spouse's

preferences for negative information. He did not know whether you and your spouse wanted to remain hopeful and optimistic, or whether you and your spouse wanted complete candor in such dire circumstances.

Ambiguous motivation: [*No additional information*]

Benevolent motivation: *In reality, the doctor lied about the experimental treatment options because he wanted to preserve your and your spouse's hope. He was sincerely trying to do what he thought was best for you and your spouse.*

A.2.2. Financial advisor vignette

Imagine that you are meeting with your financial advisor about potentially investing in a new fund. Investing in this fund would bring significant financial risk to you, but could also yield high rewards. You tell your financial advisor that you would like to invest in this fund. However, your advisor tells you that you do not meet the minimum criteria to invest. Several weeks later, you find out that you do in fact meet the criteria to invest in this fund, and that your financial advisor knew this. You and your financial advisor have known each other for around 6 months. You two had never discussed your desire to invest in high-risk/high-reward funds, or to stick with low-risk, low-reward funds.

Ambiguous motivation: [*No additional information*]

Benevolent motivation: *In reality, your advisor lied about you not meeting the criteria because s/he thought it would make you financially better off. S/he was sincerely doing what s/he thought was best for you.*

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.obhdp.2018.01.001.

References

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

Brehm, J. W. (1966). *A theory of psychological reactance.* New York: Academic Press.

Brehm, J. W., Stires, L. K., Sensenig, J., & Shaban, J. (1966). The attractiveness of an eliminated choice alternative. *Journal of Experimental Social Psychology, 2*(3), 301–313.

Brockner, J., Konovsky, M., Cooper-Schneider, R., Folger, R., Martin, C., & Bies, R. J. (1994). Interactive effects of procedural justice and outcome negativity on victims and survivors of job loss. *Academy of Management Journal, 37*(2), 397–409.

Brockner, J., & Wiesenfeld, B. M. (1996). An integrative framework for explaining reactions to decisions: Interactive effects of outcomes and procedures. *Psychological Bulletin, 120*(2), 189–208.

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.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353–380.

Cushman, F. (2013). Action, outcome, and value a dual-system framework for morality. *Personality and Social Psychology Review, 17*(3), 273–292.

deCharms, R. (1968). *Personal causation.* New York: Academic Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Deci, E. L., & Ryan, R. M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology, 53*(6), 1024–1037.

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.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327–339.

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

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791.

Fitzsimons, G. J., & Lehmann, D. R. (2004). Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science, 23*(1), 82–94.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21–38.

Gino, F., Ayal, S., & Ariely, D. (2013). Self-serving altruism? The lure of unethical actions that benefit others. *Journal of Economic Behavior and Organization, 93*, 285–292.

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

Gino, F., Shu, L. L., & Bazerman, M. H. (2010). Nameless + harmless = blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior. *Organizational Behavior and Human Decision Processes, 111*(2), 93–101.

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

Gneezy, U., Rockenback, B., & Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization, 93*(C), 293–300.

Graham, J., Meindl, P., Koleva, S., Iyer, R., & Johnson, K. M. (2015). When values and behavior conflict: Moral pluralism and intrapersonal moral hypocrisy. *Social and Personality Psychology Compass, 9*(3), 158–170.

Greenberg, A. E. (2016). Essays in behavioral economics.

Greenberg, A. E., & Wagner, A. F. (2016). The ripple effects of deceptive reporting. Available at SSRN: < http://ssrn.com/abstract=2649392 > .

Greenberg, A. E., Smeets, P., & Zhurakhovska, L. (2015). Promoting truthful communication through ex-post disclosure. Available at SSRN: < http://ssrn.com/abstract=2544349 > .

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364–371.

Gunia, B. C., Wang, L., Huang, L., Wang, J., & Murnighan, J. K. (2012). Contemplation and conversation: Subtle influences on moral decision making. *Academy of Management, 55*(1), 13–33.

Halevy, N., & Chou, E. Y. (2014). How decisions happen: Focal points and blind spots in interdependent decision making. *Journal of Personality and Social Psychology, 106*(3), 398–417.

Halevy, N., & Halali, E. (2015). Selfish third parties act as peacemakers by transforming conflicts and promoting cooperation. *Proceedings of the National Academy of Sciences, 112*(22), 6937–6942.

Hardisty, D. J., Thompson, K. F., Krantz, D. H., & Weber, E. U. (2013). How to measure time preferences: An experimental comparison of three methods. *Judgment and Decision Making, 8*(3), 236–249.

Hayes, A. (2016). PROCESS macro for SPSS and SAS. Retrieved February 1, 2016 from < http://www.processmacro.org/index.html > .

Hsee, C. K., & Weber, E. U. (1997). A fundamental prediction error: Self–others discrepancies in risk preference. *Journal of Experimental Psychology: General, 126*(1), 45–53.

John, L. K., Barasz, K., & Norton, M. I. (2016). Hiding personal information reveals the worst. *Proceedings of the National Academy of Sciences, 113*(4), 954–959.

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

Levine, E. E. (2017). 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.

Levine, E., Hart, J., Moore, K., Rubin, E., Yadav, K., & Halpern, S. (2018). The surprising costs of silence: Asymmetric preferences for prosocial lies of commission and omission. *Journal of Personality and Social Psychology, 114*(1), 29–51.

Lupoli, M. J., Jampol, L. E., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General, 146*(7), 1026–1042.

McFarlin, D. B., & Sweeney, P. D. (1992). Distributive and procedural justice as predictors of satisfaction with personal and organizational outcomes. *Academy of Management Journal, 35*(3), 626–637.

Miller, R. M., Hannikainen, I. A., & Cushman, F. A. (2014). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Emotion, 14*(3), 573–587.

Miron, A. M., & Brehm, J. W. (2006). Reactance theory-40 years later. *Zeitschrift für Sozialpsychologie, 37*(1), 9–18.

Murnighan, J. K., & Wang, L. (2016). The social world as an experimental game. *Organizational Behavior and Human Decision Processes, 136*(C), 89–94.

Nakashima, E. (2016). Apple vows to resist FBI demand to crack iPhone linked to San Bernardino attacks. *The Washington Post.* Retrieved March 16, 2015 from < https://www.washingtonpost.com/world/national-security/us-wants-apple-to-help-unlock-iphone-used-by-san-bernardino-shooter/2016/02/16/69b903ee-d4d9-11e5-9823-02b905009f99_story.html > .

Nisbett, R., & Wilson, T. (1977). Telling more than we know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Ostrom, E., Gardner, R., & Walker, J. (1994). *Rules, games, and common-pool resources.* Ann Arbor: The University of Michigan Press.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36*(4), 717–731.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891.

Rapoport, A. (1973). *Two-person game theory.* Courier Corporation.

Rogers, T., Zeckhauser, R., Gino, F., Norton, M. I., & Schweitzer, M. E. (2017). Artful paltering: The risks and rewards of using truthful statements to mislead others. *Journal of Personality and Social Psychology, 112*(3), 456–473.

Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (Contempt, Anger, Disgust) and three moral codes (Community, Autonomy, Divinity). *Journal of Personality and Social Psychology, 76*(4), 574–586.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

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., 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., 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.
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.
Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin, 37*(3), 330–349.
Shweder, R., Much, N., Mahapatra, M., & Park, L. (1997). Divinity and the "big three" explanations of suffering. *Morality and Health, 119*, 119–169.
Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*(6), 845–851.
Sutter, M. (2009). Deception through telling the truth?! Experimental evidence from individuals and teams. *Economic Journal, 119*(534), 47–60.
Tannenbaum, D., & Ditto, P. H. (2016). Information asymmetries in default options. *Working paper*.
Tannenbaum, D., Fox, C. R., & Rogers, T. (2016). On the misplaced politics of behavioral policy interventions. *Working paper*.
Tyler, T. R., Degoey, P., & Smith, H. J. (1996). Understanding why the justice of group procedures matters: A test of the psychological dynamics of the group-value model. *Journal of Personality and Social Psychology, 70*, 913–930.
Tyler, J., Feldman, F., & Reichert, A. (2006). The price of deceptive behavior: Disliking and lying to people who lie to us. *Journal of Experimental Psychology, 42*(1), 69–77.
Van Boven, L., & Loewenstein, G. (2003). Social projection of transient drive states. *Personality and Social Psychology Bulletin, 29*(9), 1159–1168.
Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes, 115*(2), 157–168.
Zhong, C. B. (2011). The ethical dangers of deliberative decision making. *Administrative Science Quarterly, 56*(1), 1–25.