

Precommitment allows leaders to maintain trust when de-escalating commitment

Ariella Kristal^{1*}, Charles Dorison², and Francesca Gino¹

1. Harvard Business School, Harvard University, Boston, MA 02163
2. Kellogg School of Management, Northwestern University, Evanston, IL 60208

*Corresponding Author: akristal@hbs.edu

ABSTRACT

Even in cases where future prospects look dire, leaders routinely escalate commitment to failing courses of action, costing businesses and taxpayers billions of dollars. Prior research argues that reputational considerations, such as self-presentation motives, are a key driver of such escalation; leaders may escalate to avoid reputational penalties. How can such reputational costs be alleviated, paving the way to value-maximizing de-escalation? In six pre-registered experiments (N = 4,635), we identify an effective strategy to de-escalate while maintaining trust: pre-specifying how a decision will be made before the time of choice (i.e., precommitment). In the presence (vs. absence) of precommitment, third-party observers perceived leaders who de-escalated as having more integrity and were more likely to trust them in a behavioral economic game (Studies 1, 2a, 2b, 3, and 4). This effect held across types of prior investment (time vs. money), sample populations (convenience sample vs. managers vs. nationally representative), contexts (consumer product vs. infrastructure projects), and measurements of trust (self-report vs. financially incentivized behavior). Decision-makers themselves recognized how precommitment would influence their reputation: in the presence (vs. absence) of this precommitment, managers were not only more likely to de-escalate, but also anticipated that they would be perceived as more trustworthy when doing so (Study 5). Whereas previous work on precommitment demonstrated its intrapersonal benefits, we argue and show that, by harnessing precommitment, leaders can maintain trust when deciding to de-escalate.

1. Introduction

Following through on one's commitments is generally viewed as a good decision across contexts, from delivering a service to a client by the promised deadline to completing an assignment for a coworker. Leaders, in particular, are expected to stick to their public commitments. We expect elected officials to deliver on campaign promises, executives to follow through with obligations to employees, and university administrators to serve students.

However, blind adherence to prior commitments is not always a force for good (Salancik 1977). There are cases when de-escalation to a commitment is prudent, as when new information arises or a past commitment appears to be a failing endeavor. For example, the French and British governments continued to invest millions of taxpayers' dollars on a supersonic plane (the Concorde) even after it became abundantly clear that it would never make a profit (Arkes and Ayton 1999, Dawkins and Carlisle 1976). Or consider the costly and high-profile managerial failures of the Shoreham Nuclear Power Plant in New York state, documented by Ross and Staw (1993); construction costs swelled from initial estimates of \$75 million to over \$4 billion before completion. Even after the plant had been built, it took another six years and \$1.5 billion to prepare the plant to become operational before the decision was made not to open it, based on safety concerns. Leaders of the Long Island Lighting Company, the company managing the project, had countless opportunities during the 23-year project to de-escalate commitment to their increasingly failing course of action, but did not.

Additional examples abound. Indeed, decades of research in judgment and decision-making have made clear that, even in cases where future prospects look dire, decision-makers routinely escalate commitment to failing courses of action (Staw and Fox 1977; Brockner et al. 1981; Staw 1981; Bazerman et al. 1984; Arkes and Blumer 1985; Staw and Ross 1987; Brockner

1992; Staw 1997; Arkes and Ayton 1999; Moon 2001; Simonson and Nye 1992; Sleesman et al. 2012, 2018). When made by high-powered leaders, such escalation choices can carry devastating consequences, costing substantial amounts of money, wasting time, and even putting lives at risk. Unfortunately, expertise is no cure for escalation of commitment; in fact, it can exacerbate the problem when there is not sufficiently clear negative feedback (Garland et al. 1990, Staats et al. 2018).

Why does costly escalation persist? Prior research posits a central role for reputational considerations as a key reason why leaders and decision-makers continue to escalate costly commitments in the face of previous investment. Specifically, scholars have long theorized that decision makers often fail to de-escalate commitment because they worry that doing so would engender social penalties from observers (e.g., Fox and Staw 1979, Arkes and Blumer 1985, Kanodia et al. 1989, Bornstein and Chapman 1995, Kertzer and Brutger 2016). Experimental evidence suggests this worry is well-founded. Dorison and colleagues recently found, for instance, that decision makers who de-escalated commitment in a project-management scenario were trusted less by third-party observers than those who remained committed to the project (Dorison et al. 2021). More generally, de-escalating commitment might make a decision-maker look hypocritical (Jordan et al. 2017, Kreps et al. 2017); this concern, in turn, might lead decision makers to escalate commitment to “save face” (Brockner et al. 1981). Returning to the Shoreham Nuclear Power Plant example, Ross and Staw (1993) suggest that reputational concerns regarding shareholders and the public utilities commission might have contributed to increasing escalation. Of note, the Long Island Lighting Company chairman offered vocal support for the project, appearing in advertisements and serving as its public face. It is therefore

unsurprising that the eventual de-escalation and closure of the plant took place only after the chairman was replaced.

2. Aims and theoretical background

Despite the significant material costs that failure to de-escalate commitment can create, reputational concerns may blunt key decision makers from doing so, with lasting damage to their organizations. This suggests that if leaders could maintain their reputations while de-escalating commitment, they might be more likely to de-escalate to failing courses of action.

How might they do so? In this paper, we hypothesize that precommitment (i.e., identifying *how* a decision will be made *before* the time of the choice, Strotz 1956, Elster 1979, Schelling 1984, Bryan et al. 2010) reduces such reputational penalties, thus removing a key barrier to de-escalation. We examine this hypothesis primarily for observers, but also for decision-makers. First, on the observer side, we theorize that specifying conditions for de-escalation in the future makes decision-makers who choose to de-escalate commitment appear more consistent and less hypocritical. In turn, we hypothesize that precommitment increases trust for leaders who de-escalate commitment.

Second, on the decision-maker side, we predict that leaders will be more likely to choose to de-escalate in the presence of precommitment. We are not the first to make this prediction (Boulding et al., 1997, Keil and Robey 1999, Simonson and Staw 1992). However, while there is prior evidence suggesting that there are *intrapersonal* mechanisms through which precommitment may encourage de-escalation (Boulding et al. 1997, Keil and Robey 1999 Simonson and Staw 1992), we extend the prior literature by looking at a novel *interpersonal* mechanism: reputational concerns. We theorize that decision-makers will anticipate being seen as more trustworthy when choosing to de-escalate commitment when they previously outlined

criteria for de-escalating (vs. when they had not provided such criteria). We test these hypotheses in six pre-registered experiments with externally valid samples (e.g., business managers with multiple subordinates and nationally representative samples).

Our work makes three main contributions. First, we contribute to the literature on escalation of commitment by arguing and showing that the interpersonal costs of de-escalation can be reduced. For decades, prior research has focused primarily on examining when and why escalation of commitment is most likely to occur and suggested that group- and organizational-level factors play a significant role (e.g., Sleesman et al. 2012, 2018). We replicate work demonstrating that self-presentation motives influence the decision to escalate commitment, thus also contributing to the dearth of empirical work examining the social determinants of escalation of commitment, as noted by Sleesman and colleagues (2012; 2018). We extend this work by identifying precommitment as a mechanism that allows leaders to de-escalate commitment while mitigating reputational costs.

Second, we contribute to existing psychological research on the benefits of precommitment. Prior work has focused on the intrapersonal benefits of precommitment, such as an increase in an individual's own future wellbeing (Ashraf et al. 2003, 2006; Giné et al. 2010; Milkman et al. 2014; Schwartz et al. 2014; Beshears et al. 2020). We argue that precommitment also has interpersonal benefits. Specifically, it attenuates social penalties faced by decision makers who decide to de-escalate commitment to failing courses of action. Further, we extend prior work by showing that the effects of precommitment extend beyond the self. While previous work shows how precommitment can help individuals make better decisions long-term for themselves (Milkman et al. 2008, Bryan et al. 2010, Rogers et al. 2014), we argue

and find that precommitment helps individuals make better long-term decisions for organizations.

Finally, we contribute to the broader research literature on behavior change and de-biasing (Milkman et al. 2009). A nascent literature reveals the social benefits accrued to leaders who act counter to prescriptive decision theory (Jordan et al. 2016, Everett et al. 2018, Tenney et al. 2019, Dorison et al. 2021, Dorison and Heller 2022). Traditionally, prior work on de-biasing has focused on the cognition of the individual decision maker (e.g., Morewedge et al. 2015, Sellier et al. 2019). While these approaches have found success, our work suggests that a complementary approach focused on reputational incentives might prove fruitful. We discuss these three contributions further in the General Discussion.

2.1. A central role for reputation

People are attuned to reputational incentives (see Schlenker and Weigold 1992, Baumeister and Leary 1995, Lerner and Tetlock 1999, Tetlock 2000, Tetlock 2002). Psychological research on self-presentation suggests that people are interested in controlling the impression they make on others (Schlenker 1980, Tetlock 2002, Sleesman et al. 2012). Self-presentation, as conceptualized in this well-established body of work, is a goal-directed strategy people engage in to create specific images of themselves so as to influence how others perceive them and treat them. People attempt to control their self-presentation not only to be liked (Buss 1983, Hill 1987, Baumeister and Leary 1995), but also to establish a particular reputation and be trusted (Baumeister 1982, Godfrey et al. 1986, Wayne and Liden 1995). In fact, moral character in general, and integrity-based trust in particular, plays a key role in impression formation (McFall 1987, Cohen et al. 2014, Goodwin et al. 2014, Uhlmann et al. 2015).

A large interdisciplinary body of research points to the central role of trust in social relationships and organizations (e.g., Arrow 1974, Mayer et al. 1995, Kramer 1999, Dirks and Ferrin 2002). This prior work makes clear that trust is essential for effective relationships and collaborations, both inside and outside the workplace (Zaheer et al. 1998, Wieselquist et al. 1999, Cook et al. 2005, Schweitzer et al. 2006, Levine and Schweitzer, 2015, Zlatev 2019). Indeed, classic work on trust has even gone as far as calling it “the lubricant of the social system” (Arrow 1974). The desire to gain a reputation for being trustworthy is so strong that people often act in seemingly counterintuitive ways to signal trustworthiness. These include lying to appear honest (Choshen-Hillel et al. 2020), acting in a biased manner against friends so as to appear unbiased (Shaw et al. 2018), changing cooperation strategies when a decision is public rather than private (Jordan et al. 2016), and forgoing economic benefits (Dana et al. 2006).

As described in greater detail below, we take a bidirectional view of the role of reputation, generally, and trust, specifically. To do so, we draw on the canonical Ability-Benevolence-Integrity (ABI) framework of trust in organizations (Mayer et al. 1995). We begin with predictions regarding the reputational consequences of decisions to de-escalate in the presence (vs. absence) of precommitment. This theorizing is most relevant to social judgments made by observers. Then, we turn to predictions regarding the reputational causes of decisions to de-escalate in the presence (vs. absence) of precommitment. Here, this theorizing is most relevant to decisions made by leaders themselves. Across both observers and decision makers, we test the overarching hypothesis that precommitment allows leaders to maintain trust while de-escalating commitment.

2.2. Observers’ perceptions of decision-makers in the presence of precommitment

Prior research commonly defines precommitment as a (somewhat) binding decision to follow a particular course of action in order to increase the likelihood of following through at a future time (Strotz 1956, Elster 1979, Schelling 1984, Bryan et al. 2010). Whereas research on precommitment tends to examine intrapersonal effects, here we examine the interpersonal effects, namely on trust.

To explain why precommitment increases trust in a leader in the context of de-escalation, we draw on the theory of “behavioral integrity,” defined as employee perceptions of managers’ pattern of word-deed alignment. Perceptions of behavioral integrity have been shown to affect employee trust, performance, intent to stay with an organization, and willingness to promote and implement espoused changes (Simons 2002, Simons et al. 2015). Consistency, one element of behavioral integrity, is an essential leadership trait. Acting in a consistent manner is seen as a valuable signal (Falk and Zimmerman 2017, Efron et al. 2018); people who act inconsistently are considered hypocrites, and hypocrites are strongly disliked (Barden et al. 2005, Alicke et al. 2013, Jordan et al. 2017), viewed as unreliable, and less likely to be trusted by others (McAllister 1995). Furthermore, evidence suggests that blaming external circumstances for a change in one’s position does not reduce perceptions of hypocrisy and still leads to negative interpersonal perceptions (Kreps et al. 2017). Therefore, when a decision-maker precommits to criteria or circumstances under which they would de-escalate commitment, following through with that decision may be seen as consistent, not hypocritical, and therefore may be more likely to maintain the trust of observers. This consistency mechanism is most relevant to perceptions of integrity-based trust. In their seminal work on trust, Mayer et al. (1995) write:

“The relationship between integrity and trust involves the trustor’s perception that the trustee adheres to a set of principles that the trustor finds acceptable . . . Such issues as

the consistency of the party's past actions, credible communications about the trustee from other parties, belief that the trustee has a strong sense of justice, and the extent to which the party's actions are congruent with his or her words all affect the degree to which the party is judged to have integrity.”

However, we theorized that these trust benefits may also extend to the two other sub-facets of trust identified by Mayer and colleagues (1995): ability-based trust and benevolence-based trust.

Perhaps complicating matters, prior research has shown that observers are sometimes sympathetic to *any* justification for an action, even a poor one (Langer et al. 1978, Dolinsky and Nawrat 1998). To provide a conservative test of this alternative hypothesis, we contrast precommitment with a different, strong justification for de-escalation: teaching people about the sunk-cost fallacy, the most common economic rationale for ignoring past unrecoverable investments (Thaler, 1980, Arkes and Blumer, 1985). We theorize that leveraging precommitment as a means of increasing de-escalation will be more effective than this economic justification, for two key reasons. First, evidence suggests that teaching people about certain biases in general, and de-escalation of commitment in the face of prior investments in particular, can be ineffective at changing their biased perceptions (Arkes 1991). Second, even if teaching individuals about the sunk-cost fallacy made them recognize why they should not escalate commitment in certain situations, if they were to explain the economic rationale behind their decision to stakeholders, it is unlikely to convince others of their consistency; therefore, we theorize it will be less likely to attenuate the reputational costs to the same extent as appealing to precommitment.

2.3. Decision-making in the presence of precommitment

In addition to examining how observers judge the actions of leaders who de-escalate commitment, we also explore how leaders themselves expect their actions to be perceived by others. The act of precommitment has been shown to help individuals and leaders make better far-sighted decisions for themselves (Milkman et al. 2008, Rogers et al. 2014). For example, through the “Save More Tomorrow” program, employers can enable employees to precommit to increasing their retirement savings rates in advance of salary increases. This workplace precommitment program has been found to dramatically increase employee retirement savings rates – from 3.5 percent to 13.6 percent over 40 months (Thaler and Benartzi, 2004). The effectiveness of precommitment is not limited to savings or a North American context: farmers in Kenya made better use of planting technology (i.e., use of fertilizer early in the planting season) when they committed to purchasing schedules in advance (Duflo et al. 2011).

Prior research also provides tentative evidence that precommitment can serve as an effective tool to encourage de-escalation. For example, Boulding et al. (1997) found that precommitment to a predetermined rule led senior-level managers to higher levels of de-escalation than a no-information control; unfortunately, in this condition, the precommitment was optional. Similarly, a study of 787 bank loans across five United States branches found that an organizational precommitment to increased monitoring when a borrower’s risk level passed a certain threshold led to a significant increase in de-escalation of commitment to risky borrowers (McNamara et al. 2002). Most relevant for the present research, Simonson and Staw (1992) found that having MBA student decision makers specify the conditions under which they would change their decisions made them more likely to de-escalate as compared to a control condition. Simonson and Staw (1992) speculated on two cognitive mechanisms: (1) reducing the likelihood of managers interpreting negative evidence as ambiguous or supportive of continuing the project

and (2) making the possibility of failure more salient, thus prompting the creation of back-up plans or other alternatives. They also consider the possibility that “including in the original plan the possibility of not achieving target performance and having to change the course of action might increase the legitimacy of withdrawal” (Simonson and Staw 1992, p. 425).

Taken together, though prior research offers initial evidence consistent with our theory that leaders are more likely to de-escalate commitment in the presence versus absence of precommitment, no prior work has thoroughly examined why such effects occur, especially in terms of the reputational drivers of such effects.

3. Research Overview

3.1. Overview of Studies

We conducted six pre-registered experiments (collective N = 4,635) to test the overarching hypothesis that precommitment allows leaders to maintain trust when de-escalating commitment. All experiments used hypothetical scenarios in which a project had been mostly completed at the time when the decision to (de-)escalate was to be made. We chose a high level of completion, in line with the previous finding that decision makers are more likely to escalate commitment when projects are close to completion (Conlon and Garland 1993).

Our experiments use a diverse array of scenarios (infrastructure vs. consumer product), samples (convenience vs. managers vs. nationally representative), and measures (self-report vs. financially-incentivized behavior). They unfold as follows. First, we test whether third-party observers make more positive evaluations of leaders who de-escalate in the presence (vs. absence) of precommitment (Studies 1, 2a, 2b, 3, and 4). Second, we test whether managers themselves recognize how precommitment changes their reputational incentives, and whether they are more likely to de-escalate commitment when it is present (Study 5). Taken together, we

demonstrate an effective way of leveraging precommitment to help leaders make better decisions while maintaining trust.

3.2. Open Science Statement

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in all experiments (Simmons et al. 2011). All experiments were pre-registered on aspredicted.org, and all pre-registrations, materials, data, and code are available on researchbox.org.¹

4. Study 1

Study 1 provides a proof-of-concept test that precommitment can increase perceived trust for a leader who de-escalates commitment from a prior unrecoverable investment. For robustness, we also tested whether the effect of precommitment would hold regardless of whether the prior investment was time or money (c.f., Soman 2001). While our primary outcome variable was perceived integrity-based trust, we also assess two other sub-facets of trust: ability-based trust and benevolence-based trust (Mayer et al. 1995).

4.1. Method

4.1.1. Participants. We recruited 601 participants (36.1% male; $M_{\text{age}} = 39.4$, $SD_{\text{age}} = 12.3$, 16.5% with a master's degree, professional degree or doctorate) from Cloud Research to participate in a study in exchange for payment. Based on pilot data, we predetermined our sample size to ensure 80% power to detect a Cohen's $d = 0.23$.

4.1.2. Procedure. All participants read a short vignette about Jess, the manager of a fast-food restaurant that had been trying to come up with a new recipe for a bun on which the chain can sell burgers. Participants were randomly assigned to one of four between-subjects

¹ For all materials, pre-registrations, and code: https://researchbox.org/257&PEER_REVIEW_passcode=KBUEVZ

experimental conditions in a 2 (prior investment: time vs. money) x 2 (precommitment: present vs. absent) design. Depending on prior investment condition, participants were randomly assigned to read about the manager either investing time (“Jess has spent 8 months working on developing this new bun”) or money (“the cost of developing the new recipe has been \$800,000”). In the precommitment present condition, participants read “Jess precommits to pulling the new bun if customers don’t seem to like it.” In the precommitment absent condition, participants read no such text. Finally, all participants learned that the manager de-escalated commitment from selling the bun. Specifically, they read that “When the bun is introduced in stores, it does not sell well. After 3 months, Jess decides to stop selling it.”

4.1.3. Dependent variables. After reading the vignette, participants in all four conditions made social judgments regarding the perceived trustworthiness of the manager (Jess). These judgments served as the primary dependent variables, with integrity-based trust serving as the central construct of interest. Participants answered three sets of Likert items rating the manager on ability-, benevolence-, and integrity-based trust (Mayer et al. 1995, Levine and Schweitzer 2014, Zlatev 2019, Dorison et al. 2021). Participants indicated how much they agreed vs. disagreed with statements on 7-point Likert scales ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Ability-based trust was measured using three items: “Jess is intelligent,” “Jess is competent,” and “Jess is capable” (alpha = 0.95). Benevolence-based trust was measured using three items: “Jess is kind,” “Jess is nice,” and “Jess is selfish” (reverse-scored), (alpha = 0.78). Finally, integrity-based trust was also measured using three items: “Jess has a great deal of integrity,” “I can trust Jess’s word,” and “Jess cares about honesty and truth,” (alpha = 0.88). After collection of these main dependent variables, participants were asked “Would you have made the same decision if you were in Jess’s position?”

4.1.4. Demographics. At the end of the study, participants answered demographic questions, including questions about their age, gender, and education.

4.1.5. Previous knowledge. Finally, we asked participants if they had encountered a scenario like this before and whether they had been taught the economic principle of sunk costs prior to completing this study.

4.2. Results

For both temporal and monetary costs, precommitment was effective at increasing not only integrity-based trust, but also ability- and benevolence-based trust. Results are depicted in Figure 1.

4.2.1. Integrity-based trust. As hypothesized, participants rated the manager as having greater integrity in the precommitment condition ($M = 5.52$, $SD = 1.00$) compared with the control ($M = 4.80$, $SD = 1.05$; $B = 0.72$, $SE = 0.09$, $t(599) = 8.19$, $p < 0.0001$, $d = 0.67$, $95\%CI[0.51,0.84]$). This effect held, and was similar in size, regardless of whether the prior investment was time ($M_{control} = 4.95$, $SD_{control} = 1.09$, $M_{precommitment} = 5.70$, $SD_{precommitment} = 1.04$; $B = 0.75$, $SE = 0.12$, $t(297) = 6.08$, $p < 0.0001$, $d = 0.70$, $95\%CI[0.49,0.94]$) or money ($M_{control} = 4.65$, $SD_{control} = 0.98$, $M_{precommitment} = 5.33$, $SD_{precommitment} = 1.13$ $B = 0.68$, $SE = 0.12$, $t(300) = 5.61$, $p < 0.0001$, $d = 0.70$, $95\%CI[0.47,0.96]$).

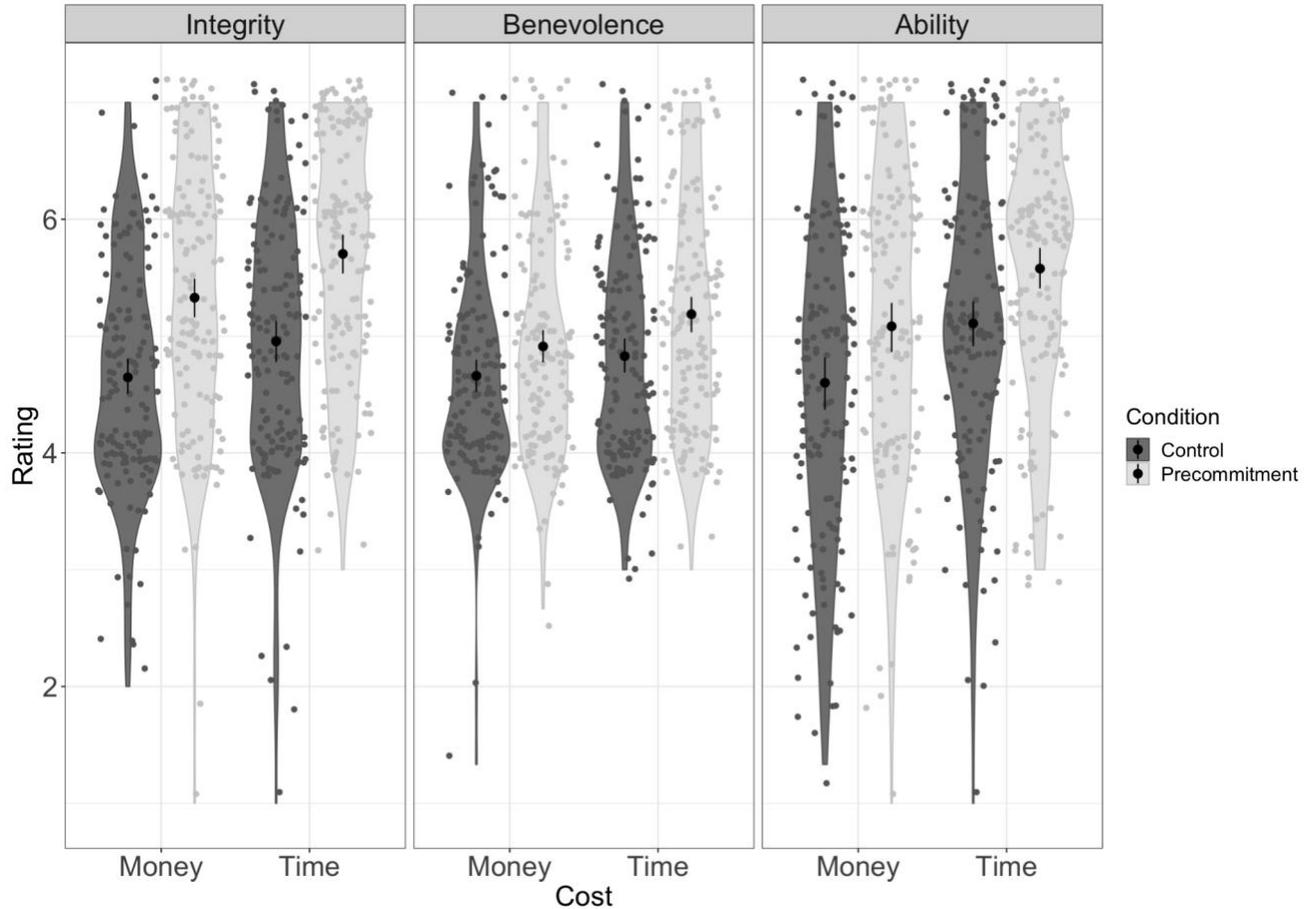
4.2.2. Ability-based and benevolence-based trust. Participants also rated Jess higher on measures of ability-based trust in the precommitment condition ($M = 5.33$, $SE = 1.21$) compared with the control ($M = 4.85$, $SD = 1.21$; $B = 0.48$, $SE = 0.10$, $t(599) = 4.66$, $p < 0.0001$, $d = 0.38$, $95\%CI[0.22,0.54]$). This effect held regardless of whether the prior investment was time ($M_{control} = 5.11$, $SD_{control} = 1.21$, $M_{precommitment} = 5.58$, $SD_{precommitment} = 1.08$; $B = 0.47$, $SE = 0.13$, $t(297) = 3.55$, $p = 0.0004$, $d = 0.41$, $95\%CI[0.18,0.64]$) or money ($M_{control} = 4.60$, $SD_{control} = 1.36$,

$M_{\text{precommitment}} = 5.08$, $SD_{\text{precommitment}} = 1.29$, $B = 0.48$, $SE = 0.15$, $t(300) = 3.17$, $p = 0.002$, $d = 0.37$, 95% CI[0.13,0.61]).

Finally, the pattern of results for benevolence was consistent with those above.

Participants again rated Jess higher on measures of benevolence-based trust in the precommitment condition ($M = 5.05$, $SE = 0.97$) compared with the control ($M = 4.75$, $SD = 0.90$; $B = 0.31$, $SE = 0.08$, $t(599) = 3.99$, $p < 0.0001$, $d = 0.33$, 95% CI[0.17,0.49]). This effect again held regardless of whether the prior investment was time ($M_{\text{control}} = 4.83$, $SD_{\text{control}} = 0.91$, $M_{\text{precommitment}} = 5.19$, $SD_{\text{precommitment}} = 1.00$; $B = 0.36$, $SE = 0.11$, $t(297) = 3.25$, $p = 0.001$, $d = 0.38$, 95% CI[0.15,0.61]) or money ($M_{\text{control}} = 4.66$, $SD_{\text{control}} = 0.90$, $M_{\text{precommitment}} = 4.91$, $SD_{\text{precommitment}} = 0.93$, $B = 0.25$, $SE = 0.11$, $t(300) = 2.39$, $p = 0.018$, $d = 0.28$, 95% CI[0.05,0.50]).

Figure 1. Integrity, Ability, and Benevolence-Based Trust by Condition and Cost



Notes. When a decision-maker de-escalates commitment to a failing course of action, third-party observers rate decision-makers higher on integrity-, ability-, and benevolence-based trust in the presence of precommitment when the previous investment was either money or time. Error bars represent 1 SE and dots represent raw data (Study 1).

4.3. Discussion

Study 1 provided a proof-of-concept test that precommitment can increase perceived trustworthiness for a manager who de-escalates commitment after a prior investment. Of note, the effects held not only across different prior investments (time and money), but also across three sub-facets of trust (integrity, ability, and benevolence). In Studies 2-4, we continue to assess the robustness of the effect of precommitment on observers' trust judgments and behaviors.

5. Study 2

Study 2 has two goals. First, it conceptually replicates Study 1 by examining whether precommitment allows leaders to maintain trust when de-escalating commitment. Second, and more critically, it seeks to rule out an alternative explanation: that any justification (even a bad one) could lead to these interpersonal benefits (Langer et al. 1978). As a strict test of this alternative hypothesis, we included a condition in which the leader gave a strong justification: an economic justification based on principles of sunk costs (described below and in the method section).

As in Study 1, participants played the role of observer and made social judgments of a hypothetical leader who de-escalated commitment to a failing course of action. This leader was either a governor (Study 2a) or CEO (Study 2b). As in Study 1, we predicted that observers would perceive the leader who gave a precommitment justification more positively than a leader who gave no explanation. Building on Study 1, we also included a third condition: the economic justification condition. We predicted that the precommitment condition would outperform not only the control condition, but also this economic justification condition. Finally, Study 2b has a third goal: to replicate this effect among managers themselves, who are responsible for making these types of decisions and therefore may be less prone to penalizing others who de-escalate commitment (c.f., Larrick et al. 1990).

5.1. Study 2a

5.1.1. Method

5.1.1.1. Participants. We recruited one-thousand two hundred and one participants (51.2% male; $M_{\text{age}} = 40.5$, $SD_{\text{age}} = 12.9$, 15.6% with a master's degree, professional degree, or doctorate) from Amazon Mechanical Turk to participate in a study in exchange for payment. Based on pilot data, we predetermined our sample size to ensure 80% power to detect a Cohen's

$d = 0.25$, with p -values adjusted for multiple comparisons (as prescribed by Brooks and Johanson 2011).

5.1.1.2. Procedure. All participants read the following vignette, which we adapted from the widely-used “radar blank plane” scenario developed by Arkes and Blumer (1985; see also Feldman and Wong 2018, Olivola 2018, Dorison et al. 2021).

“As the Governor of a large U.S. state, Governor Smith decided to allocate \$40 billion of the state’s budget to a major transportation infrastructure project. When the project is 90% complete (\$36 billion spent), a new study reveals that innovations in other transportation technologies have reduced demand for the project. The new technologies move people much faster and are far more economical than the project the state is building.”

Participants were then randomly assigned to one of three between-subjects experimental conditions. The conditions varied according to the explanation the governor gave for his decision. In the control condition, the governor did not give an explanation for the choice to de-escalate. Our main treatment condition was the *precommitment condition*. In this condition, the governor explained that he had chosen to de-escalate because of a precommitment to discontinuing the project under a certain set of circumstances (which came to pass in this scenario). Specifically, participants read the following justification from the governor: “At the time we allocated the original money, we committed to review the project and stop spending if a faster and more economical option became available. Because this predetermined benchmark was met, we made the choice to stick with what we previously decided and discontinue the project.” Unlike in Study 1, this text did not explicitly include the word “precommit.”

Finally, in this study we included a third condition: the *economic explanation condition*, in which the governor gave an explanation based on economic principles. Specifically, participants read: “A sunk cost is a cost that has already been committed and cannot be recovered. Because nothing can be done about sunk costs, we should ignore them when making decisions about various aspects of life, including policy decisions. Therefore, we made the choice to discontinue the project.” The explanation was copied from a widely used textbook on principles of microeconomics (Mankiw 2020, p. 261). This is traditionally taught as the main reason why individuals should not escalate commitment to a failing course of action across undergraduate economics and psychology courses, as well as master’s level courses in public policy and business administration. We theorized that while this explanation might be successful, it would be less so than the precommitment explanation.

5.1.1.3. Dependent variables. After reading the vignette, participants in all three conditions made social judgments regarding the perceived trustworthiness of the governor and whether they would vote for the governor. These judgments served as the primary dependent variables. Participants answered two sets of Likert items rating the governor on benevolence- and integrity-based trust (Mayer et al. 1995, Levine and Schweitzer 2014, Zlatev 2019). As in Study 1, participants indicated how much they agreed vs. disagreed with statements on 7-point Likert scales ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Benevolence-based trust was measured using three items: “The governor is kind,” “The governor is nice,” and “The governor is selfish” (reverse-scored), ($\alpha = 0.86$). Integrity-based trust was also measured using three items: “The governor has a great deal of integrity,” “I can trust the governor’s word,” and “The governor cares about honesty and truth,” ($\alpha = 0.94$). In this study, we replaced perceptions of ability-based trust with a question in which participants could indicate if they were more or less

likely to re-elect the governor, if given the chance (on a binary response scale of more likely vs. less likely).

5.1.1.4. Demographics. At the end of the study, managers answered demographic questions, including questions about their age, gender, and education.

5.1.1.5. Previous knowledge. Finally, we asked managers if they had encountered a scenario like this before and whether they had been taught the economic principle of sunk costs prior to taking this study.

5.1.2. Results

Precommitment outperformed the economic explanation and the control condition. Precommitment was effective at increasing perceptions of integrity-based trust and had no impact on benevolence-based trust, whereas an economic rationale had no effect on integrity-based trust and a significantly negative effect on benevolence-based trust. For both of these dependent variables, we conducted three pairwise contrasts, in which the dependent variable was a sub-facet of trust and the independent variable was condition. We used a Tukey Honest Significance Difference test to control for family-wise error and report adjusted p-values for pairwise contrasts below. Results are depicted in Figure 2.

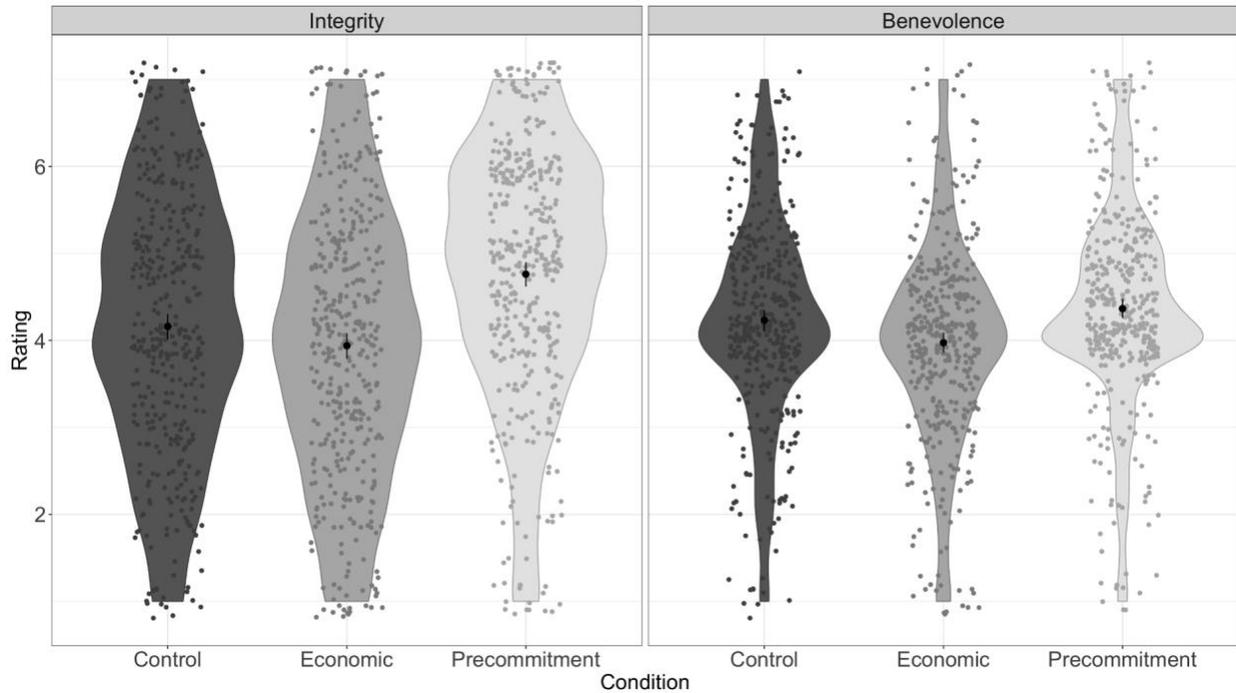
5.1.2.1. Integrity-based trust. As predicted, we found evidence that a precommitment justification significantly increased perceptions of integrity-based trust ($M = 4.76$, $SD = 1.47$) as compared to the no-explanation condition ($M = 4.16$, $SD = 1.51$; $\text{diff} = 0.60$, adjusted $p < 0.0001$, $d = 0.40$, 95% CI[0.26,0.55]) and the economic justification condition ($M = 3.94$, $SD = 1.55$; $\text{diff} = 0.82$, adjusted $p < 0.0001$, $d = 0.55$, 95% CI[0.41,0.69]). We failed to detect an effect of the economic justification on perceptions of integrity-based trust as compared to the no-

explanation control condition (diff = -0.22, $p = 0.10$, $d = 0.15$, 95% CI[-0.29,-0.01]), although the economic justification condition was directionally inferior to the control condition.

5.1.2.2. Benevolence-based trust. We failed to detect an effect of a precommitment justification on perceptions of benevolence-based trust ($M = 4.37$, $SD = 1.13$) as compared with the no-explanation condition ($M = 4.23$, $SD = 1.16$; diff = 0.13, adjusted $p = 0.227$, $d = 0.12$ 95% CI[-0.02,0.26]); however, we found a significant effect compared with the economic justification condition ($M = 3.97$, $SD = 1.15$; diff = 0.39, adjusted $p < 0.0001$, $d = 0.34$, 95% CI[0.20,0.49]). This occurred because the economic justification undermined perceived benevolence-based trust as compared to the no-explanation control condition (diff = -0.26, adjusted $p = 0.004$, $d = -0.22$, 95% CI[-0.36,-0.08]).

5.1.2.3. Likelihood to re-elect the governor. In the control condition, fewer than half of the participants reported being more likely to re-elect the governor (41.2%) based on their choice (significantly lower than 50%, $t(402) = -3.59$, $p = 0.0004$). More centrally, precommitment again outperformed the economic justification. While participants in the precommitment justification condition were significantly *more* likely to re-elect the governor compared with the control condition (56.6%; diff = 0.15, adjusted $p < 0.0001$, $d = 0.31$, 95% CI[0.18,0.45]), participants in the economic justification were significantly *less* likely to re-elect the governor compared with the control condition (32.6%, diff = -0.09, adjusted $p = 0.033$, $d = -0.18$, 95% CI[-0.34,-0.03]). The precommitment justification condition also significantly outperformed the economic justification condition (diff = 0.24 $p < 0.0001$, $d = 0.50$, 95% CI[-0.35,0.64]) when compared directly.

Figure 2. Integrity and Benevolence-Based Trust by Condition



Notes. When an elected official de-escalates commitment to a failing course of action, third-party observers rate the leader higher on integrity and benevolence-based trust in the presence of precommitment. Error bars represent 1 SE and dots represent raw data (Study 2a).

5.2. Study 2b

Study 2a provided initial evidence that a precommitment justification outperforms an economic justification (and no justification). In Study 2b, we sought to replicate Study 2a in a new domain and with a sample with more decision-making experience in professional settings. Specifically, Study 2b included evaluations of a private-sector CEO (instead of a governor) and included employees with management experience (instead of a lay sample). We theorized that, even with the new domain and with the more experienced sample, precommitment would again outperform an economic justification to improve perceived trust of a leader who de-escalates commitment.

5.2.1. Method

5.2.1.1. Participants. One-thousand two hundred and thirty-one full-time employees with management experience were recruited from Prolific Academic to participate in a study in exchange for payment. Based on pilot data, we predetermined our sample size to ensure 80% power to detect a Cohen's $d = 0.25$, with p -values adjusted for multiple comparisons (as prescribed by Brooks and Johanson 2011). After excluding data from managers who did not pass an attention check and two comprehension checks, we were left with a sample of 1,038 managers (61.8% male; $M_{\text{age}} = 40.0$, $SD_{\text{age}} = 9.8$, 32.4% with a master's degree, professional degree, or doctorate).² Of note, almost a third of the participants in the sample held a master's, professional, or doctorate degree.

5.2.1.2. Procedure. All managers read a vignette about a CEO who de-escalated commitment to a failing course of action. We again adapted the widely-used "radar-blank plane" scenario from Arkes and Blumer (1985) but changed the context to test robustness across specific wordings/products. Managers read:

"The CEO of Acme Communication Company decided to allocate \$40 billion of the company's research budget to a project. The purpose was to build a novel communication technology. When the project is 90% complete (\$36 billion already spent), another firm begins marketing a similar communication technology that is much faster and more economical than the technology Acme is building. At this point, the CEO decides to discontinue the project and not invest the last 10% of the research funds."

As in Study 2a, managers were then randomly assigned to the same one of three between-subjects experimental conditions (*control, precommitment, and economic explanation*).

² The first 143 failed the attention check before random assignment. The following 50 failed comprehension checks, but the pattern of results hold with and without including them in the sample.

5.2.1.3. Dependent variables. After reading the vignette, managers in all three conditions made social judgments regarding the perceived trustworthiness of the CEO. These judgments served as the primary dependent variables. Managers answered three sets of Likert items rating the CEO on ability-, benevolence-, and integrity-based trust (Mayer et al. 1995, Levine and Schweitzer 2014, Zlatev 2019). Managers indicated how much they agreed vs. disagreed with statements on 7-point Likert scales ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Ability-based trust was measured using three items: “The CEO is intelligent,” “The CEO is competent,” and “The CEO is capable,” (alpha = 0.94). Benevolence-based trust was also measured using three items: “The CEO is kind,” “The CEO is nice,” and “The CEO is selfish” (reverse-scored), (alpha = 0.71). Integrity-based trust was also measured using three items: “The CEO has a great deal of integrity,” “I can trust the CEO’s word,” and “The CEO cares about honesty and truth,” (alpha =0.91).

5.2.1.4. Demographics. At the end of the study, managers answered demographic questions, including questions about their age, gender, and education.

5.2.1.5. Previous knowledge. Finally, we asked managers if they had encountered a scenario like this before and whether they had been taught the economic principle of sunk costs prior to taking this study.

5.2.2. Results

We found that precommitment increased perceptions of not only integrity-based trust, but also ability-based and benevolence-based trust. In contrast, an economic rationale had a small beneficial effect on integrity-based trust, but no effect on ability-based or benevolence-based trust. For all three dependent variables, we conducted three pairwise contrasts, in which the dependent variable was a sub-facet of trust and the independent variable was condition. We used

a Tukey Honest Significance Difference test to control for family-wise error and report adjusted p -values for pairwise contrasts below. We depict the main results in Figure 3.

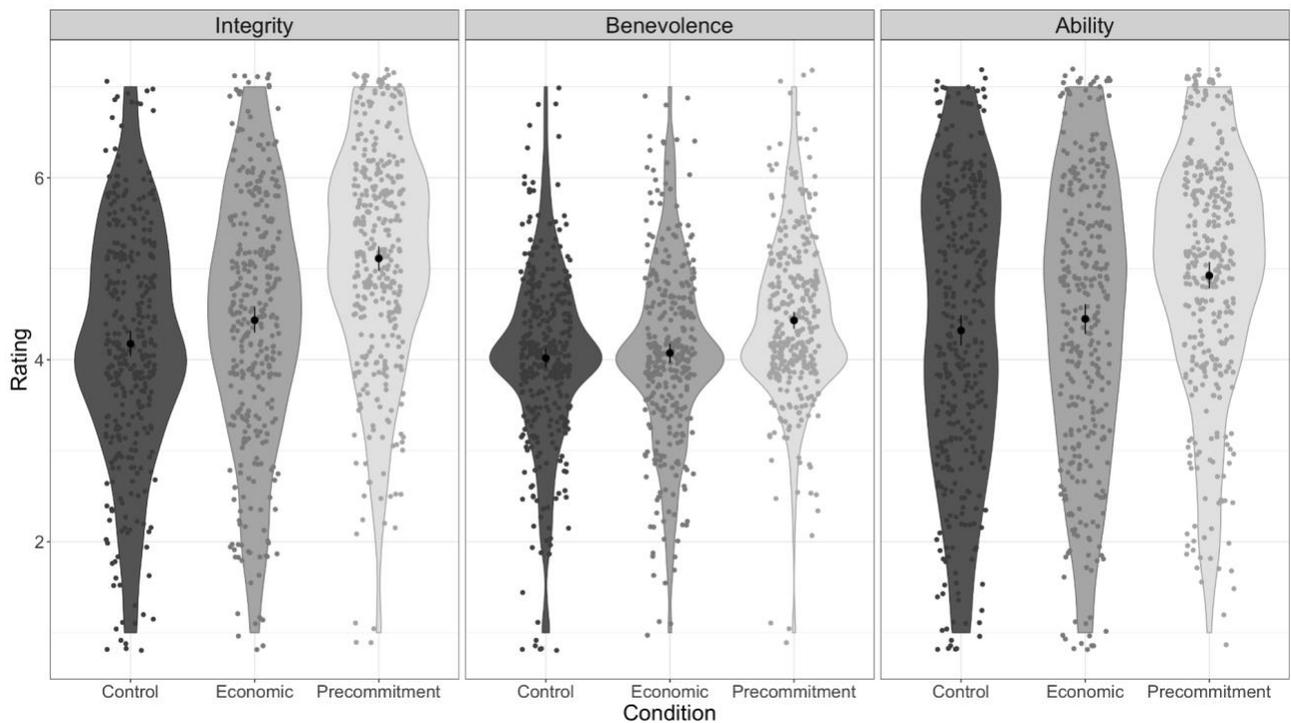
5.2.2.1. Integrity-based trust. The precommitment justification significantly increased perceptions of integrity-based trust ($M = 5.11$, $SD = 1.25$) as compared to the no-explanation condition ($M = 4.18$, $SD = 1.29$; $\text{diff} = 0.94$, adjusted $p < 0.0001$, $d = 0.74$, 95%CI[0.58,0.89]) and the economic justification condition ($M = 4.43$, $SD = 1.39$; $\text{diff} = 0.68$, $p < 0.0001$, $d = 0.52$, 95%CI[0.36,0.67]). In contrast to Study 2a, the economic justification slightly increased perceptions of integrity-based trust as compared to the no-explanation control condition ($\text{diff} = 0.26$, $p = 0.03$, $d = 0.19$, 95%CI[0.04,0.34]).

5.2.2.2. Ability-based and benevolence-based trust. The precommitment justification also significantly increased perceptions of ability-based trust ($M = 4.93$, $SD = 1.35$) as compared to the no-explanation condition ($M = 4.32$, $SD = 1.58$; $\text{diff} = 0.60$, adjusted $p < 0.0001$, $d = 0.41$, 95%CI[0.26,0.56]) and the economic justification condition ($M = 4.45$, $SD = 1.56$; $\text{diff} = 0.48$, $p < 0.0001$, $d = 0.33$, 95%CI[0.17,0.48]). We failed to detect an effect of economic justification on ability-based trust as compared to the no-explanation control condition ($\text{diff} = 0.13$, $p = 0.51$, $d = 0.08$, 95%CI[-0.07,0.23]).

Finally, mirroring the results above, the precommitment justification significantly increased perceptions of benevolence-based trust ($M = 4.43$, $SD = 0.89$) as compared to the no-explanation condition ($M = 4.02$, $SD = 0.98$; $\text{diff} = 0.41$, adjusted $p < 0.0001$, $d = 0.44$, 95%CI[0.29,0.59]) and the economic justification condition ($M = 4.08$, $SD = 1.01$; $\text{diff} = 0.36$, $p < 0.0001$, $d = 0.38$, 95%CI[0.22,0.53]). We failed to detect an effect of economic justification on benevolence-based trust as compared to the no-explanation control condition ($\text{diff} = 0.06$, $p = 0.72$, $d = 0.06$, 95%CI[-0.09,0.21]).

5.2.2.3. Comparison across sub-facets of trust. We conducted exploratory analyses to compare effects of precommitment across the sub-facets of trust. For simplicity, we dropped the economic justification condition for this analysis to isolate the effect of precommitment vs. control. A series of pairwise 2 x 2 mixed ANOVAs revealed that the effects were largest for integrity. Specifically, the effect of precommitment (vs. control) on perceived integrity was significantly larger than the effects on ability and benevolence (both pairwise contrasts p 's < 0.001). We did not observe a significant difference between ability and benevolence.

Figure 3. Perceptions of Integrity-, Benevolence-, and Ability-Based Trust in De-Escalators by Justification



Notes. When a leader de-escalates commitment to a failing course of action, third-party observers rate the leader higher on integrity-, ability-, and benevolence-based trust in the presence of precommitment. Error bars represent 1 SE and dots represent raw data (Study 2b).

5.3. Discussion

Study 2 provided evidence that different justifications for de-escalating commitment can lead to different levels of trust from observers. Consistent with our predictions, when de-

escalating commitment to a failing course of action, a leader who gave a justification based on precommitment was perceived as more trustworthy than a leader who gave an economic justification (or did not give an explanation). Furthermore, we ruled out the argument that any justification is better than no justification (Langer et al. 1978) by comparing precommitment to an economic justification (that is, in fact, taught to most students of economics and most management students) and we found support for our hypothesis that precommitment leads to higher levels of trust.

Of note, the benefits of precommitment were consistent in both a lay sample (Study 2a) and a sample of employees with management experience (Study 2b). However, one intriguing inconsistency merits note. Specifically, when it comes to perceived integrity-based trust, while the economic justification performed worse than the no-explanation control for the lay sample, it directionally outperformed the control condition for the sample of employees with management experience. We discuss this intriguing finding further in the General Discussion.

6. Study 3

Study 3 continues to examine the reputational consequences of precommitment in the context of de-escalation of commitment. It builds on the previous studies in four critical ways. The first two changes focus on increasing external validity. First, while prior studies used hypothetical vignettes adapted from prior literature, Study 3 uses stimuli based on a real infrastructure project scenario unfolding in California at the time of the study. Second, while prior studies used convenience samples (of either lay people or managers), Study 3 uses a nationally representative sample of adults in the United States. The third and fourth changes focus on boundary conditions. Third, this study not only explores the reputational consequences of precommitment at the time of de-escalation, but also explores how precommitment affects

reputation at the inception of a project. It could be the case that precommitment yields reputational costs for the leader at the time of initial investment, possibly because it makes the leader look uncommitted or unsure. Finally, whereas prior studies laid out clear contingencies for exactly how precommitment would lead to de-escalation, such roadmaps may be less clear in many cases in organizations. Thus, leaders may be able to precommit to simply re-evaluate at a later time. We added a third supplemental condition to our two primary conditions to test whether this weaker form of precommitment would still yield reputational benefits.

To measure reputational consequences, we again use the ability-benevolence-integrity framework of trust (Mayer et al. 1995). In addition, we added a binary measure assessing whether participants would want to reappoint the decision-maker to their current leadership role.

6.1. Method

6.1.1. Participants. We recruited a nationally representative sample of seven-hundred fifty-three adults in the U.S. from Prolific Academic to participate in a study in exchange for payment (48.2% male; $M_{\text{age}} = 45.1$, $SD_{\text{age}} = 16.3$, 21.0% with a master's degree, professional degree, or doctorate). The sample was recruited to be representative of age, sex, and ethnic group proportions based on the US Census Bureau population group estimates from 2015 (Prolific.co 2021). Based on pilot data, we predetermined our sample size to ensure 80% power to detect a Cohen's $d = 0.20$.

6.1.2. Procedure and Measures. All participants read about the California High Speed Rail and learned about the Secretary of Transportation in charge of the project. Details were pulled from popular news sources such as the *New York Times* (Nagourney 2018), the *Associated Press* (Ronayne 2021), and *Capital Public Radio* (a *National Public Radio* affiliate, Adler and Miller 2019). Participants learned that this project was approved in 2008 with a budget of \$40

billion with the goal of connecting Los Angeles to Sacramento in under three hours. After learning about the project, participants were randomly assigned to one of three between-subjects experimental conditions: two primary conditions (control, precommitment to de-escalate) and one supplemental condition (precommitment to re-evaluate). In the control condition, participants did not read anything else after learning about the details of the project. In the precommitment to de-escalate condition, participants read:

“At the time of the approval, the Secretary of Transportation **precommitted to stopping the project or scaling it back** if time and budget overruns became too great.”

In the precommitment to re-evaluate condition, participants read about a weaker version of precommitment:

“At the time of the approval, the Secretary of Transportation **precommitted to re-evaluating the project** if time and budget overruns became too great.”

At this point in the survey (before learning any information about the success or failure of the project), participants were asked to rate the Secretary of Transportation on the three ability-, benevolence-, and integrity-based trust scales. We refer to these ratings as “Time 1” ratings.

On the following page, after completing these three measures, participants learned that the project was not going well and that the Secretary of Transportation had chosen to scale back the project. Specifically, participants read:

“Now, in 2021, \$5.9 billion has already been spent, the project construction is desperately lagging. The project is estimated to cost over \$100 billion in total and will not be completed until 2023. The Secretary of Transportation decides to dramatically scale back the project and focus on only connecting two minor cities in the middle of California

(Bakersfield and Merced), thus abandoning the hopes of connecting it to Sacramento or Los Angeles.”

Participants were also then shown a map that delineated both the original plans and the scaled-back plans. Once again, at this second point in time, participants were asked to rate the Secretary of Transportation on the three ability-, benevolence-, and integrity-based trust scales. Furthermore, participants were then asked, “Should the Secretary of Transportation be reappointed for another term?” We refer to these ratings as “Time 2” ratings.

Finally, participants were asked what they would do in the Secretary of Transportation’s position and why, followed by questions about demographics, whether they had encountered a scenario like this before, and whether they had been taught the economic principle of sunk costs prior to taking this study.

6.2. Primary Results

Three key results emerged. First, we did not detect any reputational costs of precommitment to de-escalate at the time of initial investment (if anything, it yielded slight reputational benefits). Second, we replicated the robust reputational benefits of precommitment at the time of eventual de-escalation. Finally, we found a similar pattern of results for our supplemental, weaker precommitment condition as we did with our primary, stronger precommitment condition.

Our analysis proceeds in lockstep with the three results outlined above. We first report the outcome measures for our primary two conditions (control vs. precommitment) at the time of initial investment (T1) and the time of eventual de-escalation (T2). We then report secondary results for the supplemental re-evaluate condition. Because we planned to compare these three

conditions, in both sections we used a Tukey Honest Significance Difference test to control for family-wise error and report adjusted p-values for pairwise contrasts below.

6.2.1. Time 1. We first assessed whether precommitment might yield reputational costs at the time of initial investment. In contrast to this idea, precommitment led to significantly increased perceptions of integrity-based trust ($M = 4.23$, $SD = 1.35$) at Time 1 as compared to the control condition ($M = 3.97$, $SD = 1.11$; $\text{diff} = 0.28$, adjusted $p = 0.023$, $d = 0.23$, 95% CI[0.05,0.40]). We also failed to detect reputational costs on other measures of trust. Precommitment significantly increased perceptions of ability-based trust ($M = 4.72$, $SD = 1.38$) as compared to the no-explanation condition ($M = 4.50$, $SD = 1.17$; $\text{diff} = 0.25$, adjusted $p = 0.066$, $d = 0.19$, 95% CI[0.01,0.37]). We failed to detect an effect of precommitment on perceptions of benevolence-based trust ($M = 4.04$, $SD = 1.05$) compared to the control condition ($M = 4.19$, $SD = 0.84$; $\text{diff} = -0.12$, adjusted $p = 0.316$, $d = -0.13$, 95% CI[-0.31,0.05]).

6.2.2. Time 2. We next assessed whether we replicated our prior results at the time of eventual de-escalation. We did; precommitment significantly increased perceptions of integrity-based trust ($M = 4.17$, $SD = 1.56$) at Time 2 as compared to the control condition ($M = 2.81$, $SD = 1.25$; $\text{diff} = 1.37$, adjusted $p < 0.0001$, $d = 0.97$, 95% CI[0.78, 1.18]). Precommitment also yielded reputational benefits for both ability-based trust ($M = 4.17$, $SD = 1.63$) at Time 2 as compared to the control condition ($M = 2.97$, $SD = 1.27$; $\text{diff} = 1.21$, adjusted $p < 0.001$, $d = 0.82$, 95% CI[0.65,1.02]), and for benevolence-based trust at Time 2 ($M = 3.91$, $SD = 1.12$) as compared to the control condition ($M = 3.66$, $SD = 0.89$; $\text{diff} = 0.25$, adjusted $p = 0.014$, $d = 0.25$, 95% CI[0.07,0.42]). Participants in the precommitment condition were also the most likely to reappoint the Secretary of Transportation (44%) compared to the control condition (where 16% indicated they would; $\text{diff} = 0.28$, adjusted $p < 0.0001$, $d = 0.63$, 95% CI[0.44, 0.83]).

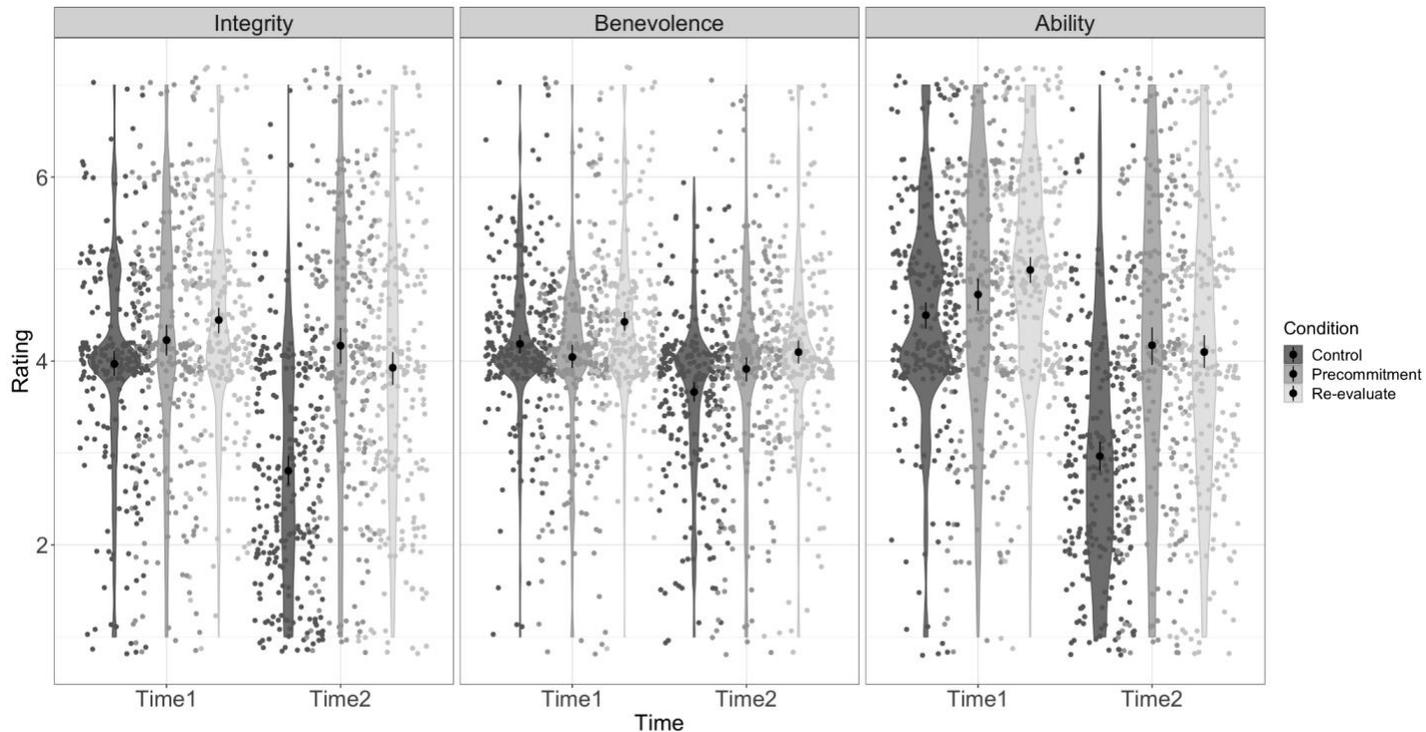
6.2.3. Differences at T1 vs. T2. Finally, we directly compared the effects of precommitment across time. To do so, we conducted a series of pairwise 2 (between subjects: control vs. precommitment) x 2 (within subjects: time 1 vs. time 2) mixed ANOVAs. Results revealed that the effect of precommitment on perceived trust were significantly greater at the time of eventual de-escalation; this was true not only for integrity-based trust, but also for ability-based and benevolence-based trust (all three pairwise contrasts p 's < 0.001; full results available online).

6.3. Secondary Results

Finally, in a set of secondary analyses, we examined whether a weaker form of precommitment could yield the same reputational benefits. While we briefly summarize the results here, they are visualized in Figure 4 and full details are available in our online materials.

Generally, the results with this weaker form of precommitment (i.e., simply to re-evaluate at a later point) are extremely similar to the results with the stronger form of precommitment (i.e., to de-escalate at a later point). More specifically, the re-evaluate condition significantly outperformed the control condition and sometimes even the precommitment condition at Time 1. At Time 2 the re-evaluate condition also outperformed the control condition on all three measures, but we did not detect significant differences between the re-evaluation and precommitment conditions.

Figure 4. Perceptions of Integrity-, Benevolence-, and Ability-Based Trust in De-Escalators by Justification and Time Period



Notes. When a public official de-escalates commitment to a failing course of action, third-party observers rate the leader higher on integrity-, ability-, and benevolence-based trust in the presence of precommitment at Time 2 (after de-escalation) without incurring penalties at Time 1 (compared with a control). Precommitment to re-evaluate also outperforms the control group at both times. Error bars represent 1 SE and dots represent raw data (Study 3).

6.4. Discussion

Drawing on an externally-valid scenario and with a nationally-representative participant sample, Study 3 replicated the finding that precommitment increases perceived trust for leaders who de-escalate commitment to a failing course of action. Further, Study 3 provided evidence that such benefits (1) do not come with concomitant reputational costs at the time of initial investment and (2) persist with a weaker form of precommitment (i.e., simply a precommitment to re-evaluate, rather than de-escalate).

7. Study 4

Study 4 had three primary goals. First, Studies 1-3 provided consistent evidence that precommitment increases perceived trust of a leader who de-escalates commitment. In Study 4,

we move from self-reported measures of trust to a financially-incentivized, behavioral measure (Berg et al. 1995). Second, while Studies 1-3 demonstrated that precommitment increases perceptions of trust, it remained unclear to what extent these benefits offset the reputational costs of de-escalating itself (vs. escalating; as documented by Dorison et al. 2021). To make this assessment, Study 4 includes evaluations of both leaders who de-escalate commitment and leaders who escalate commitment. Finally, the mechanism through which precommitment increases perceived trust remained untested in Studies 1-3. Here, we assessed whether perceptions of consistency and hypocrisy underpin such changes.

In Study 4, we manipulated the presence or absence of precommitment before an initial investment choice. We hypothesized that while in the absence of escalation (i.e., the control condition), participants would show a preference for the decision maker who escalated, the presence of precommitment (i.e., the treatment condition) would reduce or even eliminate this preference. As in Study 2b, we again recruited a sample of employees with management experience to provide a stricter test of our theorizing.

7.1. Method

7.1.1. Participants. We aimed to recruit 440 managers through the online platform Prolific Academic in order to be powered to detect an effect of at least Cohen's $d = 0.30$ with 80% power. While 466 managers took our survey in exchange for monetary payment, our analysis is limited to the 418 managers (73.2% male, $M_{\text{age}} = 33.8$, $SD_{\text{age}} = 8.8$, $M_{\text{management experience}} = 6.0$ years, $SD_{\text{management experience}} = 4.3$, 27.0% with a master's degree, professional degree or doctorate) who passed two comprehension checks before random assignment. Of note, as in Study 2b, almost a third of our participants held a master's, professional, or doctorate degree.

7.1.2. Procedure. All managers read instructions about a trust game and answered two comprehension-check questions regarding procedure (Berg et al. 1995, Jordan et al. 2016, Dorison et al. 2021). Managers were told they were going to be “Senders” and that they had the option to send \$1.50 to one of two “Receivers,” who would then have the option of keeping the \$1.50 as bonus money, or splitting it evenly and returning \$0.75 to the “Sender,” in which case the managers would receive a \$0.75 bonus. The managers were told the “Receivers” were participants in a previous study who read a similar vignette to that presented in Study 2b. Managers were then presented with one “Receiver” who chose to escalate and another who chose to de-escalate. In this study, managers read the original scenario developed by Arkes and Blumer (1985):

“As the president of an airline company, the Receiver decided to allocate \$10 million of the company’s research budget to a project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed (\$9 million already spent), another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane the Receiver's company is building.”

Managers were then randomly assigned to one of two between-subjects experimental conditions, with the explanation the CEO gave for this choice varied between conditions. In the control condition, all the managers read, “The question the Receiver now faces is: should the company invest the last 10% of the research funds to finish the radar-blank plane? Receiver 1 decided **TO INVEST** the remaining funds. Receiver 2 decided **NOT TO INVEST** the remaining funds.” Receiver 1’s and Receiver 2’s decisions to invest (vs. not) were counterbalanced. Managers in the *precommitment condition* read the following,

“Of note, at the time the Receiver’s company allocated the original money, they committed to review the project and stop investing if another firm began marketing a radar-blank plane that was faster and more economical than theirs.”

Managers in the precommitment condition then read the same last lines as the managers in the control condition.

7.1.3. Dependent variables. At the end of the scenario descriptions, managers in both conditions made the following incentive-compatible choice: “Based on their decision, which receiver would you prefer to pass money to?” Our main dependent variable was the binary choice of whether the manager passed the money to the participant who escalated or de-escalated.

7.1.4. Mediating variables. Managers then answered two 7-item questions asking which Receiver they thought was more hypocritical (-3 = Receiver 1 is much more hypocritical, 0 = no difference, 3=Receiver 2 is much more hypocritical) and which receiver they thought was more consistent (-3 = Receiver 1 is much more consistent, 0 = no difference, 3 = Receiver 2 is much more consistent). Hypocrisy was then reverse coded ($r = 0.74$). We averaged these items to create an indexed score, with positive values referring to the de-escalator and negative values referring to the escalator.

7.1.5. Demographics. At the end of the study, managers completed short demographic questions, including questions about their age, gender, and education.

7.1.6. Previous knowledge. Finally, we asked managers if they had encountered a scenario like this before and whether they had been taught the economic principle of sunk costs prior to taking this study.

7.2. Results

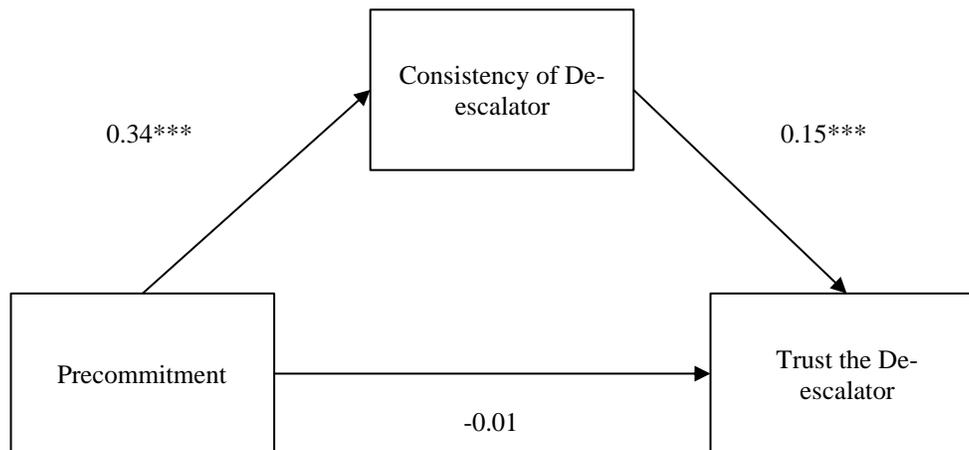
7.2.1. Trust. We first assessed whether we replicated the findings of Dorison and colleagues (Dorison et al. 2021), in which decision makers who de-escalate commitment (vs. escalate) are trusted less by third-party observers. We found this is the case: in the absence of precommitment (i.e., the control condition), only 28.0% of managers chose to trust the receiver who de-escalated commitment (i.e., did not invest the remainder of the funds).

More centrally, our key theoretical question was whether this preference would be attenuated by precommitment. It was; in the presence of precommitment, 40.8% of managers chose to pass money to the receiver who de-escalated commitment (logistic regression results in $B = 0.57$, $SE = 0.21$, $z(409) = 2.73$, $p = 0.006$). Of note, this represented a 46% relative increase in trusting the decision maker who de-escalated commitment compared to the control condition.

7.2.2. Perceptions of consistency. As described above, our measure of consistency was an indexed score, with positive values referring to the de-escalator and negative values referring to the escalator. In the absence of commitment, the average consistency rating favored the escalator ($B = -0.90$, $SE = 0.11$, $t(408) = -8.64$, $p < 0.001$). In line with our hypothesis, precommitment significantly increased perceptions of consistency of the de-escalator ($B = 0.98$, $SE = 0.16$, $t(406) = 6.32$, $p < 0.001$).

7.2.3. Mediation analysis. Finally, we examined whether the changes in perceived consistency underpinned changes in trust behavior. To test this hypothesis, we fit a mediation model using the Lavaan package in R (Rosseel 2012). The independent variable was condition (0 = control, 1 = precommitment), the mediator was perceived consistency, and the dependent variable was trust (1 = trust de-escalator, 0 = trust escalator). In line with our hypothesis, perceived consistency mediated the effect of condition on trust (indirect effect = 0.051, $p < 0.001$). The results are depicted in Figure 5.

Figure 5. Mediation Model



Note. Perceptions of consistency mediate the impact of precommitment on trusting the receiver who de-escalates commitment (Study 4).

7.3. Discussion

Study 4 provided financially-incentivized behavioral evidence that observers are more likely to trust leaders who de-escalate in the face of precommitment as compared with no justification for de-escalation. We provide further evidence that precommitment increases observers' perceptions of consistency when a leader de-escalates, and that this increased perception of consistency mediates the increased trust of de-escalators.

8. Study 5

Study 5 turns the attention from observers to decision makers. In doing so, we examine whether managers are more likely to de-escalate commitment in the presence (vs. absence) of precommitment. Further, we examine whether such effects are mediated by how managers anticipate they will be perceived by others—specifically, in the form of anticipated perceptions of trust.

8.1. Method

8.1.1. Participants. We aimed to recruit six hundred full-time employed managers who had at least two subordinates currently reporting to them. We did so through the online platform Prolific Academic (Carton and Lucas 2018, Palan and Schitter 2018). Based on pilot data, we predetermined our sample size to ensure 80% power to detect a Cohen's $d = 0.25$. While 609 managers took our survey in exchange for monetary payment, our analysis is limited to the 602 managers (59.5% male, $M_{\text{age}} = 33.7$, $SD_{\text{age}} = 9.9$, 33.8% with a master's degree, professional degree, or doctorate) who passed one attention check. As in Studies 2b and 4, approximately one third of the sample held a master's, professional, or doctorate degree.

8.1.2. Procedure. All managers read a vignette about a CEO who faced a choice of whether to escalate or de-escalate commitment to a failing course of action. We again adapted the widely used "radar-blank plane" scenario designed by Arkes and Blumer (1985). Managers read:

"As the president of an airline company, you have decided to allocate \$10 million of the company's research budget to a project. The purpose was to build a plane that would not be detected by conventional radar, in other words, a radar-blank plane. When the project is 90% completed (\$9 million already spent), another firm begins marketing a plane that cannot be detected by radar. Also, it is apparent that their plane is much faster and far more economical than the plane your company is building."

Managers were then randomly assigned to one of two between-subjects experimental conditions (control, precommitment). In the control condition, managers read no further information. In the precommitment condition, managers read the following text at the end of the scenario:

“Of note, at the time your company allocated the original money, you committed to review the project and stop investing if another firm began marketing a radar-blank plane that was faster and more economical than yours.”

8.1.3. De-escalation behavior. After reading the scenario, managers read: “The question you now face is: should you invest the last 10% of the research funds to finish the radar-blank plane? What would you do?” They could choose to either escalate commitment (i.e., invest the remaining funds) or de-escalate commitment (i.e., not invest the remaining funds). This binary choice served as our primary dependent variable. We predicted that managers would be more likely to de-escalate commitment in the precommitment condition as compared to the control condition.

8.1.4. Anticipated trust. We theorized that one reason precommitment could prompt de-escalation is because it changes managers’ anticipated reputational incentives, particularly in the form of trust in the eyes of others. To assess anticipated trust, we again drew on Mayer and colleagues’ canonical three-part model of trust: the ability-benevolence-integrity framework. (Mayer et al. 1995, Levine and Schweitzer 2014, Zlatev 2019). For all items, managers indicated how they anticipated others would perceive them on 5-point Likert scales ranging from 1 (Much more _____ if I decided NOT TO INVEST) to 5 (Much more _____ if I decided TO INVEST). Ability-based trust was measured using three items (intelligent, competent, and capable; $\alpha = 0.84$). Benevolence-based trust was also measured using three items (nice, kind, selfish (reverse-coded); $\alpha = 0.67$). Integrity-based trust was also measured using three items (likely to keep my word, integrity, likely to care about truth and honesty; $\alpha = 0.85$). Within each index, questions were answered in a randomized order. Across the indices, the ability-, benevolence-,

and integrity-based trust items were answered on separate screens and in a randomized order. We averaged each set of items to create three separate indexed scales.

8.1.5. Preference for consistency. As an exploratory individual difference variable, managers completed a four-item “preference for consistency” scale (adapted from Cialdini et al. 1995). Managers rated how much they agreed vs. disagreed with statements on 7-point Likert scales from 1 (Strongly Disagree) to 7 (Strongly Agree). The four items were: “The appearance of consistency is an important part of the image I present to the world,” “It is important to me that others view me as a stable person,” “I make an effort to appear consistent to others,” and “It doesn’t bother me much if my actions are inconsistent” (reverse-coded; $\alpha = 0.75$). We theorized that the effect of precommitment on de-escalation may be stronger for decision makers higher in preference for consistency.

8.1.6. Demographics. After completing all dependent variables, managers completed short demographic items, including questions about their age, gender, and education.

8.1.7. Previous knowledge. Finally, at the end of the experiment, we asked managers if they had encountered a scenario like this before and whether they had been taught the economic principle of sunk cost prior to taking this experiment.

8.2. Results

8.2.1. Effect of precommitment on de-escalation behavior. Our first hypothesis was that managers would be more likely to de-escalate commitment in the presence of precommitment. This hypothesis was supported. While only 18.3% of managers in the control condition chose to de-escalate commitment, 39.7% of managers in the precommitment condition chose to de-escalate commitment, more than doubling the percentage of managers who de-escalated

commitment (a 117% increase). A logistic regression indicated that this represented a statistically significant increase ($B = 1.08$, $SE = 0.10$, $z(601) = -5.67$, $p < 0.0001$).

8.2.2. Effect of precommitment on anticipated trust. Next, we examined whether the presence of precommitment influenced decision makers' anticipated reputational consequences of de-escalation. We found a significant impact of precommitment on anticipated ability-based ($B = -0.32$, $SE = 0.08$, $t(600) = -4.18$, $p < 0.0001$) and integrity-based trust ($B = -0.75$, $SE = 0.09$, $t(600) = -8.60$, $p < 0.0001$), but no significant impact on anticipated benevolence-based trust ($B = -0.06$, $SE = 0.03$, $t(600) = 1.66$, $p = 0.10$). Of note, a 2 (between: precommitment vs. control) x 2 (within: integrity vs. ability) mixed ANOVA revealed that the effect on integrity was significantly larger than the effect on ability ($F(1,600) = 54.93$, $p < 0.0001$). These findings are in line with prior research highlighting the importance of integrity in perceptions of leaders (Simons 2002, Simons et al. 2015).

8.2.3. Mediation analysis. A key remaining question is to what extent, if at all, the changes in anticipated trust underpin changes in de-escalation behavior. We pre-registered that integrity- and ability- based trust, but not benevolence-based trust would mediate the effect of precommitment. To test this hypothesis, we fit a parallel mediation model using the Lavaan package in R (Rosseel 2012). The independent variable was condition (0 = control, 1 = precommitment). The three parallel mediators were ability-, benevolence-, and integrity-based trust. The dependent variable was de-escalation behavior (0 = escalate, 1 = de-escalate). Consistent with predictions, anticipated integrity and ability mediated the effect of condition on de-escalation (indirect effects = 0.15 and 0.03, respectively, both p 's < 0.001); however, anticipated benevolence did not (indirect effect = -0.002, $p = 0.429$). Of note, while the indirect effects through integrity and ability both reached traditional levels of statistical significance, a

pairwise contrast testing the two indirect effects against each other revealed that the effect through integrity was approximately *five times* the size of the indirect effect through ability ($b = 0.15$ vs. 0.03), and this difference was significantly different from zero (contrast between the two paths: $b = 0.12$, $z = 5.13$, $p < 0.001$).

8.2.4. Exploratory moderation analysis. Finally, we tested the exploratory hypothesis that the effect of precommitment on de-escalation would be moderated by preference for consistency. Using a logistic regression, we found preliminary evidence to support this hypothesis ($B = 0.41$, $SE = 0.20$, $z(601) = 2.09$, $p = 0.04$). For those low in need for consistency, there was still a treatment effect of precommitment on likelihood of de-escalation ($B = 0.16$, $SE = 0.05$, $t(360) = 3.30$, $p = 0.001$); however, the treatment effect was approximately double in size for those high in need for consistency ($B = 0.30$, $SE = 0.06$, $t(238) = 5.43$, $p < 0.001$).

8.3. Discussion

Study 5 provided evidence that precommitment increases the tendency for decision makers to de-escalate commitment. It also provided evidence that one key reason such effects occur is because decision makers believe that precommitment changes their perceived reputational incentives, especially in the form of anticipated integrity-based trust. We also found evidence that leaders who have a higher need for consistency are even more sensitive to precommitment than those lower in need for consistency.

9. General Discussion

Escalation of commitment is costly (Desai and Chulkov 2009). Previous research has set forth four main determinants of escalation: project attributes, psychological factors, social pressures, and organizational variables (Staw and Ross 1989); however, to date, very little empirical research has explored how social pressures drive escalation of commitment (Sleesman

et al. 2012, Sleesman et al. 2018). Whereas the reputational costs of de-escalation have been widely theorized in the management and social science literatures (Arkes and Blumer 1985, Kanodia et al. 1989, Fearon 1997, Busemeyer and Pleskac 2009, Mahlendorf and Wallenburg 2013, Kertzer and Brutger 2016), there is a dearth of evidence regarding the role of reputation in escalation of commitment and how reputational concerns can be overcome. Our research extends this prior work by documenting a novel strategy for leaders to optimize both material outcomes and reputational rewards simultaneously.

In this work, we replicate the general tendency for observers to judge those who de-escalate commitment to a failing course of action more harshly than those who (often wrongfully) escalate commitment (as shown in Study 4, as well as in Dorison et al. 2021). Furthermore, we extend these findings by introducing a way for leaders to de-escalate commitment to a failing course of action while maintaining trust: by leveraging the power of precommitment. Precommitment to withdrawing from a failing course of action under certain conditions makes the eventual withdrawal more palatable in the eyes of observers. Across six pre-registered experiments, we demonstrated that leaders correctly anticipate that de-escalation will make them appear less trustworthy because they are acting inconsistent with previously-stated goals, but that a precommitment to conditions for de-escalation attenuates the negative reputational consequences of de-escalation.

Specifically, we first demonstrated that leaders' reputational concerns are valid. In five studies, we found that observers are in fact likely to penalize leaders who de-escalate without a justification and are more likely to trust leaders who de-escalate in the face of precommitment as compared with leaders who provide no justification for de-escalation or who provide an

economic justification (Studies 1-4). Furthermore, as leaders anticipated, increased perceptions of consistency mediated the increased trust in de-escalators who precommitted.

Next, we found support for our hypothesis that leaders are more likely to de-escalate commitment to a failing course of action in the presence (vs. absence) of precommitment (Study 5). Precommitment changed reputational incentives by making a de-escalator appear more consistent, which led to increased perceptions of trust. We also found that need for consistency moderated the impact of precommitment, such that leaders who have a higher need for consistency were even more likely to de-escalate commitment in the face of precommitment than leaders with a lower need for consistency (Study 5).

Our work identifies how central integrity is in influencing the perceptions of decision-makers who decide to escalate or de-escalate commitment. When it comes to benevolence-based trust, the presence or absence of precommitment had a minimal impact on leaders' anticipated reputational consequences of de-escalation and the actual consequences of de-escalation. Given the demonstrated importance of integrity on perceptions of moral character (McFall 1987, Uhlmann et al. 2015, Zlatev 2019) and perceptions of leaders (Simons 2002, Simons et al. 2015), this finding is perhaps unsurprising, though we did not explicitly theorize it in advance. Though we do not find it here, there are circumstances where benevolence- and integrity-based trust can even go in the opposite direction (Levine and Schweitzer 2015).

The small, and at times not significant, impact of precommitment on benevolence is also indicative that precommitment is not merely improving observers' overall perceptions of the decision maker but rather is operating specifically on integrity (and, to a lesser extent, ability). Therefore, we can partially rule out the possibility that precommitment leads to a general "halo effect" on perceptions of decision-makers.

In our study, we varied our sample to include lay observers (such as those who might vote for a public official in an election) as well as managers (who make promotion decisions and evaluate subordinates' decision-making and work-related outcomes). While we have demonstrated that precommitment is an effective strategy at mitigating the negative reputational effects of de-escalation among both types of observers, we do note an interesting difference with respect to an alternative justification we tested in Study 2. Specifically, we find that a general lay audience penalizes politicians who give economic justifications for de-escalation compared with a control no-explanation condition, whereas the managerial sample either does not differentiate, or sometimes even rewards, the economic justification compared with the control. Nevertheless, precommitment outperforms both justifications among both samples.

9.1. Theoretical Contributions

Our work makes at least three primary theoretical contributions. First, we extend prior work on escalation of commitment by contributing evidence to the underexplored idea that self-presentation motives influence the decision to escalate commitment. Though a substantial amount of theory expounds on the role of social or reputational factors that lead to escalation (Arkes and Blumer 1985, Kanodia et al. 1989, Staw and Ross 1989, Fearon 1997, Bussemeyer and Pleskac 2009, Mahlendorf and Wallenburg 2013, Kertzer and Brutger 2016), empirical work lags behind (as suggested by Sleesman et al. 2012, Sleesman et al. 2018) and so does the theorizing for these relationships. Here, we turned our attention to observers' perceptions, and argued that perceptions of consistency explain why precommitment changes how much trust we place in those who de-escalate commitment.

Second, we highlight situations in which decision makers who make choices that are costly from the perspective of material outcomes (i.e., escalate commitment to a failing course of

action) are trusted more by third-party observers. Consequently, these findings contribute to the nascent literature recognizing the social benefits accrued to leaders who act counter to prescriptive decision theory (Jordan et al. 2016, Everett et al. 2018, Tenney et al. 2019, Dorison et al. 2021, Dorison and Heller 2022). Of note, such findings have clear implications for behavior change: the present studies provide evidence that interventions aimed at reducing excessive escalation should focus on the social and organizational context in addition to the cognition of the individual decision maker (see also Dorison et al. 2021).

Finally, we demonstrate that precommitment not only has intrapersonal benefits (as demonstrated in the psychology and economics literatures: Strotz 1956, Elster 1979, Schelling 1984, Milkman et al. 2008, Bryan et al. 2010, Rogers et al. 2014), but also interpersonal benefits for leaders who use precommitment strategies in their decision-making. Specifically, we show that precommitment attenuates social penalties faced by decision makers who decide to de-escalate commitment to failing courses of action.

9.2. Practical Implications

From the Vietnam War to the Shoreham Nuclear Power Plant, to the Concorde, to the recent U.S. Air Force's F1 fighter jet and the ongoing California High Speed Rail project, history is littered with expensive, time-consuming, and even deadly examples of escalation of commitment. While teaching people about the economic principle of escalating commitment to a failing course of action due to sunk costs may lead decision makers to know what they *should* do (Larrick et al. 1990), it does not eliminate the reputational consequences of de-escalation, and using it as a justification has little impact on observers' perceptions. By contrast, precommitment can be a valuable way to encourage de-escalation; however, it is underused.

To determine the appeal of precommitment as a de-escalation tool, we conducted a survey with 156 managers active on LinkedIn from around the world (54% male, 28% in the United States, with an average of 10.2 years of management experience, $SD = 7.9$ years).

Specifically, we told managers:

“Often, it makes sense to follow projects through to completion, such as when they’re going well and meeting key deadlines and objectives. Other times, however, it makes sense to discontinue them, such as when new information arises or when they no longer appear to be on track.

We’d like you to think about a project you will be responsible for at work over the coming weeks or months. In planning this project, which of the following options would you prefer?”

We then provided the option to choose “At the beginning of the project, precommit to criteria under which you will discontinue it” or “Monitor the project as it goes along and make a decision to discontinue at a later date if you deem it necessary.” Only 21% said that they would precommit. We also collected open-ended responses asking managers why they selected their chosen option. One of the most-cited reasons that managers gave for being hesitant to precommit is wanting the flexibility to be able to change their mind as the project progressed. However, our Study 3 indicates that precommitment to re-evaluation when specific conditions come to pass, even without the commitment to de-escalation, may provide some trust-benefits without locking managers into specific actions.

We are aware of one high-profile example of precommitment in government spending. In 2010, the U.K. Government created the Behavioural Insights Team (BIT) to incorporate behavioral science into public policy making. Recognizing the role of inertia in government

institutions, BIT insisted on incorporating a two-year sunset clause that would require the government to make an active choice to keep BIT going if the team did not produce a tenfold return on the costs of investments; otherwise, it would shut down ([BIT 2012](#)). Unfortunately, examples like this are more the exception than the norm.

To the extent that leaders and organizational decision makers can incorporate precommitment to stopping criteria when planning projects (particularly large-scale, costly projects), they will be better able to de-escalate commitment to failing projects while also maintaining trust.

9.3. Limitations and Future Directions

Because of our interest in the reputational consequences of de-escalation, across these six experiments, we have demonstrated the effect of precommitment from a previous time period on the perceptions of a subsequent decision. There are several future avenues that would be fruitful for further research. Additional research could also shed light on whether an initial precommitment needs to be made public at the time or only when a decision-maker is justifying de-escalation. It is also worth investigating the impact of precommitment on different types of third-party observers. In our studies, most of our observers were “disinterested” third-parties; however, future research should investigate the impact of precommitment on specific stakeholders, such as employees who invested time and energy in the project, or investors or taxpayers who invested funds.

We also examined one form of precommitment in our studies, but it can take many forms. Leaders can precommit to criteria under which they will completely discontinue a project, or they may take a more cautious approach in a complex environment and instead precommit to criteria under which a third party reviews the project (since they are less likely to escalate

commitment to a failing course of action, see Staw 1976; Bazerman et al. 1982, 1984). Further research is needed to determine the optimal type of precommitment that enables leaders to navigate complex, unpredictable environments while also enabling them to maintain integrity in the eyes of others if and when they decide to de-escalate commitment to a failing course of action. In Study 3, we began to explore how a precommitment to re-evaluate the project would be perceived (if certain outcomes came to pass) and it seemed promising, but further research needs to explore how specific that precommitment to re-evaluate must be. For example, a general commitment to “review the project periodically” may be insufficient to generate the reputational gains discussed in this paper, because the commitment is not to conditions under which the project would be reconsidered, but rather to vague monitoring.

One of the main limitations of the online studies conducted in this paper is that they use stylized examples when it is made clear that there is a failing course of action. Consequently, future research needs to explore the role of precommitment in naturalistic field settings, over a longer time horizon. The intertemporal nature of the problem of escalation of commitment and the solution of precommitment makes it a challenging but worthwhile topic of study.

10. Conclusion

Though failure to de-escalate commitment can be costly, leaders often choose it because they worry their reputation will suffer if they change course. In this paper, we argued and found that precommitment allows leaders to maintain trust when de-escalating commitment. Leaders anticipate greater trustworthiness from their stakeholders when they do vs. do not precommit by showing greater consistency, and this makes them more willing to de-escalate. Observers, in a similar way, are more trusting of leaders who precommit, viewing them as more consistent than

those who do not. By using precommitment, leaders can signal consistency and maintain trust when (rightly) deciding to de-escalate.

References

- Adler B, Miller N (2019) Governor Newsom Proposes Scaling-Back of California Bullet Train Project in First State of the State Address. Retrieved December 1, 2021, <https://www.capradio.org/articles/2019/02/12/governor-newsom-proposes-scaling-back-of-california-bullet-train-project-in-first-state-of-the-state-address/>
- Alicke M, Gordon E, Rose D (2013) Hypocrisy: what counts?. *Philos. Psych.* 26(5), 673-701.
- Arkes HR, Blumer C (1985) The psychology of sunk cost. *Organ. Behav. and Human Decision Processes.* 35(1),124-140.
- Arkes HR, Ayton P (1999) The sunk cost and Concorde effects: Are humans less rational than lower animals?. *Psych. Bull.* 125(5), 591.
- Arkes HR (1991) Costs and benefits of judgment errors: Implications for debiasing. *Psych. Bull.* 110(3), 486–498.
- Arrow KJ (1974) *The Limits of Organization* (W. W. Norton & Company, New York)
- Ashraf N, Gons N, Karlan DS, Yin W (2003) A review of commitment savings products in development countries. *Asian Development Bank* (July), <http://hdl.handle.net/11540/1969>.
- Ashraf N, Karlan D, Yin W (2006) Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quart. J. of Econ.* 121(2), 635-672.
- Barden J, Rucker DD, Petty RE (2005) “Saying one thing and doing another”: Examining the impact of event order on hypocrisy judgments of others. *Personality and Soc. Psych. Bull.* 31(11), 1463-1474.
- Baumeister RF, Leary MR (1995) The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psych. Bull.* 117(3), 497.
- Baumeister RF (1982) A self-presentational view of social phenomena. *Psych. Bull.* 91(1)3.
- Bazerman MH, Beekun RI, Schoorman FD (1982) Performance evaluation in a dynamic context: A laboratory study of the impact of a prior commitment to the ratee. *J. of Appl. Psych.* 67(6), 873.
- Bazerman MH, Giuliano T, Appelman A (1984) Escalation of commitment in individual and group decision making. *Organ. Behav. and Human Performance.* 33(2), 141-152.
- Behavioural Insights Team (2012). Welcome to the BIT Blog!. Accessed January 7, 2022, <https://www.bi.team/blogs/welcome-to-the-bit-blog/>
- Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games and Econ. Behav.* 10(1), 122-142.
- Beshears J, Choi JJ, Harris C, Laibson D, Madrian BC, Sakong J (2020) Which early withdrawal penalty attracts the most deposits to a commitment savings account?. *J. of Public Econ.* 183, 104144.
- Bornstein BH, Chapman GB (1995) Learning lessons from sunk costs. *J. of Experiment. Psych.: Appl.* 1(4), 251.

- Boulding W, Morgan R, Staelin R (1997) Pulling the plug to stop the new product drain. *J. of Marketing Res.* 34(1), 164-176.
- Brockner J, Rubin JZ, Lang E (1981) Face-saving and entrapment. *J. of Experiment. Soc. Psych.* 17(1), 68-79.
- Brockner J (1992) The escalation of commitment to a failing course of action: Toward theoretical progress. *Acad. of Management Rev.* 17(1), 39-61.
- Brooks GP, Johanson GA (2011) Sample size considerations for multiple comparison procedures in ANOVA. *J. of Modern Appl. Statist. Methods.* 10(1), 10.
- Bryan G, Karlan D, Nelson S (2010) Commitment devices. *Annual Rev. of Econ.* 2(1), 671-698.
- Busemeyer JR, Pleskac TJ (2009) Theoretical tools for understanding and aiding dynamic decision making. *J. of Math. Psych.* 53(3), 126-138.
- Buss LW (1983) Evolution, development, and the units of selection. *Proc. of the National Acad. of Sci.* 80(5), 1387-1391.
- Carton AM, Lucas BJ (2018) "How can leaders overcome the blurry vision bias? Identifying an antidote to the paradox of vision communication." *Acad. of Management J.* 61(6), 2106-2129.
- Choshen-Hillel S, Shaw A, Caruso EM (2020) Lying to appear honest. *J. of Experiment. Psych.: General.*
- Cialdini RB, Trost MR, Newsom JT (1995) Preference for consistency: The development of a valid measure and the discovery of surprising behavioral implications. *J. of Personality and Soc. Psych.* 69, 318-328.
- Cohen TR, Panter AT, Turan N, Morse L, Kim Y (2014) Moral character in the workplace. *J. of Personality and Soc. Psych.* 107(5), 943.
- Conlon D. E., Garland, H. (1993) The role of project completion information in resource allocation decisions. *Acad. of Management J.* 36(2), 402-413.
- Cook KS, Hardin R, Levi M (2005) *Cooperation without trust?* (Russell Sage Foundation)
- Dana J, Cain DM, Dawes RM (2006) What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Org. Behav. and Human Decision Processes.* 100(2), 193-201.
- Dawkins R, Carlisle TR (1976) Parental investment, mate desertion and a fallacy. *Nature.* 62(5564), 131-133.
- Desai MS, Chulkov DV (2009) Escalation of commitment in MIS projects: A meta-analysis. *Internat. J. of Management & Inform. Systems.* 13(2), 29-38.
- Dirks K, Ferrin D (2002) Trust in Leadership: Meta-Analytic Findings and Implications for Research and Practice. *The J. of App. Psych.* 87. 611-28. 10.1037//0021-9010.87.4.611.
- Dorison CA, Umphres C, Lerner JS (2021) Escalation of commitment to a failing course of action signals trustworthiness. *Harvard University Working Paper.*
- Dorison CA, Heller BH (2022) Observers penalize decision-makers whose risk preferences are unaffected by loss-gain framing. *J. of Experiment. Psych.: General. Forthcoming.*
- Duflo E, Kremer M, Robinson J (2011) Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *Amer. Econ. Rev.* 101(6), 2350-90.
- Elster J (1979) *Ulysses and the Sirens: Studies in Rationality and Irrationality*, (Cambridge University Press, Cambridge, UK).
- Effron D, O'Connor K, Leroy H, Lucas B (2018) From inconsistency to hypocrisy: When does "saying one thing but doing another" invite condemnation?. *Res. in Organ. Behav.* 38.

- Everett JA, Faber NS, Savulescu J, Crockett MJ (2018) The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *J. of Experiment. Soc. Psych.* 79, 200-216.
- Falk A, Zimmermann F (2017) "Consistency as a signal of skills." *Management Sci.* 63(7): 2197-2210.
- Fearon JD (1997) Signaling foreign policy interests: Tying hands versus sinking costs. *J. of Conflict Resolution.* 41(1), 68-90.
- Feldman G, Wong KFE (2018) When action-inaction framing leads to higher escalation of commitment: A new inaction-effect perspective on the sunk-cost fallacy. *Psych. Sci.*, 29(4), 537-548.
- Fox FV, Staw BM (1979) The trapped administrator: Effects of job insecurity and policy resistance upon commitment to a course of action. *Admin. Sci. Quart.* 449-471.
- Garland H, Sandefur CA, Rogers AC (1990) De-escalation of commitment in oil exploration: When sunk costs and negative feedback coincide. *J. of Appl. Psych.* 75(6), 721.
- Giné X, Karlan D, Zinman J (2010) Put your money where your butt is: a commitment contract for smoking cessation. *Amer. Econ. J.: Appl. Econ.* 2(4), 213-35.
- Godfrey DK, Jones EE, Lord CG (1986) Self-promotion is not ingratiating. *J. of Personality and Soc. Psych.* 50(1), 106.
- Goodwin GP, Piazza J, Rozin P (2014) Moral character predominates in person perception and evaluation. *J. of Personality and Soc. Psych.* 106(1), 148.
- Hill CA (1987) "Affiliation motivation: people who need people... but in different ways." *J. of Personality and Soc. Psych.* 52(5), 1008.
- Jordan JJ, Hoffman M, Nowak MA, Rand DG (2016) Uncalculating cooperation is used to signal trustworthiness. *Proc. of the National Acad. of Sci.* 113(31), 8658-8663.
- Jordan JJ, Sommers R, Bloom P, Rand DG (2017) Why do we hate hypocrites? Evidence for a theory of false signaling. *Psych. Sci.* 28(3), 356-368.
- Kanodia C, Bushman R, Dickhaut J (1989) Escalation errors and the sunk cost effect: An explanation based on reputation and information asymmetries. *J. of Accounting Res.* 27(1), 59-77.
- Keil M, Robey D (1999) Turning around troubled software projects: An exploratory study of the deescalation of commitment to failing courses of action. *J. of Management Inform. Systems.* 15(4), 63-87.
- Kertzer JD, Brutger R (2016) Decomposing audience costs: Bringing the audience back into audience cost theory. *Amer. J. of Political Sci.* 60(1), 234-249.
- Kramer RM (1999) Trust and distrust in organizations: Emerging perspectives, enduring questions. *Ann. Rev. of Psych.* 50, 569-598.
- Kreps TA, Laurin K, Merritt AC (2017) Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *J. of Personality and Soc. Psych.* 113(5), 730.
- Langer EJ, Blank A, Chanowitz B (1978) The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *J. of Personality and Soc. Psych.* 36(6), 635.
- Larrick RP, Morgan JN, Nisbett RE (1990) Teaching the use of cost-benefit reasoning in everyday life. *Psych. Sci.* 1(6), 362-370.
- Lerner JS, Tetlock PE (1999) Accounting for the effects of accountability. *Psych. Bull.* 125(2), 255.

- Levine EE, Schweitzer ME (2014) Are liars ethical? On the tension between benevolence and honesty. *J. of Experiment. Soc. Psych.* 53, 107-117.
- Levine EE, Schweitzer ME (2015) Prosocial Lies: When Deception Breeds Trust. *Organ. Behav. and Human Decision Processes.* 126 88-106.
- Mahlendorf MD, Wallenburg CM (2013) Public justification and investment in failing projects: the moderating effect of optimistic outcome expectations. *J. of Appl. Soc. Psych.* 43(11), 2271-2286.
- Mankiw NG (2020) *Principles of Econ.* Cengage Learning.
- Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. *Acad. of Management Rev.* 20(3), 709-734.
- McAllister DJ (1995) Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Acad. of Management J.* 38(1), 24-59.
- McFall L (1987) Integrity. *Ethics.* 98(1), 5-20.
- McNamara G, Moon H, Bromiley P (2002) Banking on commitment: Intended and unintended consequences of an organization's attempt to attenuate escalation of commitment. *Acad. of Management J.* 45(2), 443-452.
- Milkman KL, Chugh D, Bazerman MH (2009) How Can Decision Making Be Improved?. *Perspect. on Psych. Sci.* 4 (4), 379-383
- Milkman KL, Minson JA, Volpp KG (2014) Holding the hunger games hostage at the gym: An evaluation of temptation bundling. *Management Sci.* 60(2), 283-299.
- Milkman KL, Rogers T, Bazerman MH (2008) Harnessing our inner angels and demons: What we have learned about want/should conflicts and how that knowledge can help us reduce short-sighted decision making. *Perspect. on Psych. Sci.* 3(4), 324-338.
- Moon H (2001) The two faces of conscientiousness: duty and achievement striving in escalation of commitment dilemmas. *J. of Appl. Psych.*,86(3), 533.
- Morewedge C, Yoon H, Scopelliti I, Symborski C, Korris J, Kassam K (2015) Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behav. and Brain Sci.* 2. 129-140.
- Nagourney A (2018) A \$100 Billion Train: The Future of California or a Boondoggle?. *New York Times* (July 30), <https://www.nytimes.com/2018/07/30/us/california-high-speed-rail.html>
- Olivola CY (2018) The interpersonal sunk-cost effect. *Psych. Sci.* 29(7), 1072-1083.
- Palan S, Schitter C (2018) Prolific. ac—A subject pool for online experiments. *J. of Behav. and Experiment. Finance.* 17, 22-27.
- Prolific.co (2021) Representative Sample FAQ. Retrieved December 17, 2021, <https://researcher-help.prolific.co/hc/en-gb/articles/360019238413-Representative-samples-FAQ>
- Rogers T, Milkman KL, Volpp KG (2014) Commitment devices: using initiatives to change behavior. *J. of the Amer. Medical Assoc.* 311(20), 2065-2066.
- Ronayne K (2021) California bullet train's latest woe: Will it be high speed? *Associated Press* (October 13), <https://apnews.com/article/business-gavin-newsom-san-francisco-california-high-speed-rail-authority-state-budgets-d6667f05a1ae22989bb63ceb8aaf4f7c>
- Ross J, Staw BM (1993) Organizational escalation and exit: Lessons from the Shoreham nuclear power plant. *Acad. of Management J.* 36(4), 701-732.
- Rosseel Y (2012) lavaan: An R Package for Structural Equation Modeling. *J. of Statist. Software.* 48, 1-36.

- Salancik GR (1977) Commitment is too easy!. *Organ. Dynam.* 6(1), 62-80.
- Schelling TC (1984) Self-command in practice, in policy, and in a theory of rational choice. *Amer. Econ. Rev.* 74(2), 1-11.
- Schlenker BR, Weigold MF (1992) Interpersonal processes involving impression regulation and management. *Annual Rev. of Psych.* 43(1), 133-168.
- Schlenker BR (1980) *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations* (Brooks/Cole, Monterey, CA).
- Schwartz J, Mochon D, Wyper L, Maroba J, Patel D, Ariely D (2014) Healthier by precommitment. *Psych. Sci.* 25(2), 538-546.
- Schweitzer ME, Hershey JC, Bradlow ET (2006) Promises and lies: Restoring violated trust. *Organ. Behav. and Human Decision Processes.* 101(1), 1-19
- Sellier AL, Scopelliti I, Morewedge CK (2019) Debiasing Training Improves Decision Making in the Field. *Psych. Sci.* 30(9), 1371-1379.
- Shaw A, Choshen-Hillel S, Caruso EM (2018) Being biased against friends to appear unbiased. *J. of Experiment. Soc. Psych.* 78, 104-115.
- Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psych. Sci.* 22(11), 1359-1366.
- Simons T, Leroy H, Collewaert V, Masschelein S (2015) How leader alignment of words and deeds affects followers: A meta-analysis of behavioral integrity research. *J. of Business Ethics.* 132(4), 831-844.
- Simons T (2002) Behavioral integrity: The perceived alignment between managers' words and deeds as a research focus. *Organization Sci.* 13(1), 18-35.
- Simonson I, Nye P (1992) The effect of accountability on susceptibility to decision errors. *Organ. Behav. and Human Decision Processes.* 51(3), 416-446.
- Simonson I, Staw BM (1992) Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *J. of Appl. Psych.* 77(4), 419.
- Sleesman DJ, Conlon DE, McNamara G, Miles JE (2012) Cleaning up the big muddy: A meta-analytic review of the determinants of escalation of commitment. *Acad. of Management J.* 55(3), 541-562.
- Sleesman DJ, Lennard AC, McNamara G, Conlon DE (2018) Putting escalation of commitment in context: A multilevel review and analysis. *Acad. of Management Ann.* 12(1), 178-207.
- Soman D. (2001) The mental accounting of sunk time costs: Why time is not like money. *J. of Behav. Decision Making.* 14(3), 169-185.
- Staats B, Diwas SKC, Gino F (2018) Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Sci.* 64(2): 804-824.
- Staw BM, Fox FV (1977) Escalation: The determinants of commitment to a chosen course of action. *Human Relations,* 30(5), 431-450.
- Staw BM, Ross J (1987) Behavior in escalation situations: Antecedents, prototypes, and solutions. *Research in Organ. Behav.*
- Staw BM, Ross J (1989) Understanding Behavior in Escalation Situations. *Sci.* 246(4927), 216-220.
- Staw BM (1976) Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organ. Behav. and Human Performance.* 16(1), 27-44.
- Staw BM (1981) The escalation of commitment to a course of action. *Acad. of Management Rev.* 6(4), 577-587.

- Staw BM (1997) The escalation of commitment: An update and appraisal. Shapira Z, ed. *Organizational decision making* (Cambridge University Press, Cambridge, UK), 191–215.
- Strotz RH (1956) Myopia and inconsistency in dynamic utility maximization. *The Rev. of Econ. Studies.* 23(3), 165-180.
- Tenney ER, Meikle NL, Hunsaker D, Moore DA, Anderson C (2019) Is overconfidence a social liability? The effect of verbal versus nonverbal expressions of confidence. *J. of Personality and Soc. Psych.* 116(3), 396.
- Tetlock PE (2000) Cognitive biases and organizational correctives: Do both disease and cure depend on the politics of the beholder?. *Administrative Sci. Quart.* 45(2), 293-326.
- Tetlock PE (2002) Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psych. Rev.* 109(3), 451.
- Thaler RH, Benartzi S (2004) Save more tomorrow™: Using behavioral economics to increase employee saving. *J. of Political Econ.* 112(S1), S164-S187.
- Thaler RH (1980) Toward a positive theory of consumer choice. *J. of Econ. Behav. and Organ.* 1, 39-60.
- Uhlmann EL, Pizarro DA, Diermeier D (2015) A person-centered approach to moral judgment. *Perspect. on Psych. Sci.* 10(1), 72-81.
- Wayne SJ, Liden RC (1995) Effects of impression management on performance ratings: A longitudinal study. *Acad. of Management J.* 38(1), 232-260.
- Wieselquist J, Rusbult C, Foster C, Agnew C (1999) Commitment, Pro-Relationship Behavior, and Trust in Close Relationships. *J. of Personality and Soc. Psych.* 77. 942-66.
- Zaheer A, McEvily B, Perrone V (1998) Does Trust Matter? Exploring the Effects of Inter-Organizational and Inter-Personal Trust on Performance. *Organ. Sci.* 9. 141-159. 10.1287/orsc.9.2.141.
- Zlatev JJ (2019) I may not agree with you, but I trust you: Caring about social issues signals integrity. *Psych. Sci.* 30(6), 880-89