HARVARD | BUSINESS | SCHOOL

Going beyond the 'Self' in Self-Control: Interpersonal Consequences of Commitment Strategy Use

Ariella Kristal
Julian Zlatev

Working Paper 22-033

Going beyond the ‘Self’ in Self-Control: Interpersonal Consequences of Commitment Strategy Use

Ariella Kristal
Harvard Business School

Julian Zlatev
Harvard Business School

Working Paper 22-033

Funding for this research was provided in part by Harvard Business School.

Going beyond the 'self' in self-control:

Interpersonal consequences of commitment strategy use

Ariella S. Kristal[1] Julian J. Zlatev

Harvard University

[1] Corresponding Author: Ariella S. Kristal (akristal@g.harvard.edu)

Abstract

Commitment strategies are effective mechanisms individuals can use to overcome self-control problems. Across seven studies (and three supplemental studies), we explore the negative interpersonal consequences of commitment strategy use. In Study 1, using an incentivized trust game, we demonstrate that individuals trust people who use a commitment strategy less than those who use willpower to achieve their goals. In Study 2, across four domains we show that people judge commitment strategy users less favorably than willpower users in pursuit of the same goals, particularly when it comes to integrity-based trust. In Studies 3a, 3b, 4a and 4b, we explore potential mechanisms underlying this effect. Finally, in Study 5, we demonstrate that people's anticipation of these negative consequences of commitment strategy use contributes to their reticence to adopt such strategies. Thus, we establish the role of willpower as a positive signal in impression formation, as well as the negative interpersonal consequences of relying on external aides when faced with temptation.

Going beyond the 'self' in self-control:

Interpersonal consequences of commitment strategy use

"It isn't what you do, but how you do it." - John Wooden

When working on this very project, one of the authors used internal willpower to design studies, analyze data, and write up results without succumbing to the temptation of watching funny cat videos or memes on the internet. The other author used a combination of a web-blocker to limit access to distracting websites, public commitments to accomplishing certain tasks by specified deadlines, and a timer set for twenty-five minute intervals of intense focus. While we both ended up completing our tasks, this paper examines whether the different methods we used are judged differently by outside observers. In other words, consistent with revered college basketball coach John Wooden's saying quoted above, we explore whether people's goal achievement behavior is judged not just on *what* they achieved, but *how* they went about achieving it.

Self-control, the ability to prioritize beneficial long-term goals over immediate desires, is valued in a functioning society. Demonstrating the capacity for self-control at a young age has been shown to predict future educational competence (Mischel, Shoda, & Peake, 1988), employment outcomes (Daly et al., 2015), health outcomes (Moffitt et al., 2011), and even credit ratings (Moffitt, et al., 2011). Among adults, stronger self-control is associated with better health outcomes such as lower rates of obesity (VanEpps, Downs & Loewenstein, 2016); better financial outcomes such as lower levels of debt (Gathergood, 2012); as well as interpersonal benefits such as stronger relationships (Finkel & Campbell, 2001, Tangney et al, 2004), greater

honesty (Gino et al, 2011) and increased trustworthiness in the eyes of others (Righetti & Finkenauer, 2011).

However, there are many ways in which people can achieve the outcome of being self-controlled (Inzlicht et al., 2021). Of particular importance, some forms of self-control involve using internal resolve to resist temptation in the moment, often referred to as using willpower (Duckworth, Milkman, & Laibson, 2019). Another form of self-control involves using external aides to avoid the temptation altogether. While it is possible to demonstrate self-control by using brute-force internal willpower, recent research has shown that this is less effective than taking steps to avoid being placed in a tempting situation to begin with (Adriaanse et al., 2014; de Ridder et al., 2012; Galla & Duckworth, 2015; Hofmann et al., 2012). Often referred to as "commitment strategies," these external approaches are an arrangement one enters into to constrain future behaviors either by changing one's physical environment or the decision environment (i.e. changing payoffs associated with different behaviors). The use of commitment strategies has been shown to increase goal achievement across a number of domains including smoking reduction, weight loss, academic achievement, and financial savings (for a review see Bryan, Karlan & Nelson, 2010).

While previous work has identified the myriad individual-level benefits of using commitment strategies compared to internal willpower, little work has examined the potential interpersonal implications of choosing to use one of these forms of self-control over another. In particular, we examine how an individual's choice of self-control strategy influences the inferences people draw about that individual. We find that, despite the usefulness of commitment strategies, observers consistently form more negative impressions of targets who choose to use commitment strategies than targets who choose to use internal willpower. We additionally find

evidence that this occurs because using willpower is typically viewed as requiring more effort than using commitment strategies. Because the exertion of effort is itself moralized, and morality is an important component of integrity-based trust in particular, this leads commitment strategy users to be seen as less trustworthy than willpower users.

Willpower and Commitment Strategies

Though research in the behavioral sciences sometimes treats willpower and self-control interchangeably, willpower is in fact only one of many self-control strategies one could use while trying to avoid temptation in order to pursue long-term goals (Fujita, 2011; Gillebaart & de Ridder, 2015; Hofmann & Kotabe, 2012; Kristal & Zlatev, 2021). In particular, willpower is conceptualized as an internal "brute force" approach used in the moment to resist temptation directly (Duckworth, Milkman, & Laibson, 2019; Katzir et al., 2021). A classic example of willpower in action is the 'marshmallow test' designed by Mischel and colleagues (Mischel & Ebbesen, 1970; Mischel, Ebbesen, & Zeiss, 1972). In that work, children were seated in a room and asked to avoid eating a small number of marshmallows in order to receive a larger number later. The ability to sit in front of these marshmallows without eating them, knowing a larger reward was coming later, is what we and others refer to as willpower.

The 'in-the-moment' nature of willpower can be contrasted with other self-control approaches that are instead planned in advance to avoid being placed into a tempting situation altogether (for a review of alternatives to willpower, see Duckworth, Milkman, & Laibson, 2019). In order to clarify this distinction between "intrapsychic" approaches to self-control that require willpower and situational strategies that are external to the self, we draw on models of emotion regulation (Gross, 1998) and subsequent extensions into the domain of self-regulation (Duckworth, Gendler and Gross, 2014, 2016). This work documents five points at which

regulation can occur: (a) selection of the situation (e.g. choosing not to keep cookies in the house), (b) modification of the situation (putting the cookies out of reach), (c) deployment of attention (don't look at the cookies), (d) change of cognitions (focus on how many calories are in the cookies), and (e) modulation of responses (resist the cookies). The first two represent what we mean by external situational strategies that occur in advance of the temptation while the final three represent what we mean by internal approaches meant to resist temptation in the moment.

One particularly effective external alternative to "willpower" is the use of commitment strategies (Bryan, Karlan & Nelson, 2010; Rogers, Milkman & Volpp, 2014). Commitment strategies involve taking an action prior to exposure to a temptation in order to increase the cost of, or preclude the ability to act upon, the future temptation. This idea has been inherent in the psychological and economic literature for decades under various names such as "precommitment," (Strotz, 1956), "self-binding" (Elster, 1979), "anticipatory self-command," "strategic self-frustration," (Schelling, 1984), and "commitment devices" (Bryan, Karlan, & Nelson, 2010).

Commitment strategies themselves can vary on several dimensions. They can be hard commitments, making it nearly impossible to engage in an activity (e.g., Homer's Odysseus, knowing he would be tempted to certain death by the song of the sirens, had his sailors bind him to the mast of his ship so he could hear the song without being physically able to move) or they can be a softer form of commitment, making it less desirable to engage in an activity (e.g., taking Antabuse, a pill that makes someone vomit as soon as they consume alcohol or making a public commitment to a deadline). Commitment strategies can rely on enforcement from a wide variety of sources, including technology (e.g., using an app to block distracting websites while working) or close others (e.g., paying a friend $5 if you fail to show up to your gym session with them). A

central component of a commitment strategy is that it is external to the self, something one can use as opposed to relying on an internal force of will (Duckworth, Gendler & Gross, 2016; Kristal & Zlatev, 2021).

In addition to the domains discussed earlier, commitment strategies have proven effective at increasing savings rates (Thaler & Benartzi, 2004), facilitating smoking cessation (Giné, Karlan, & Resnick, 2010), increasing gym attendance (Milkman, Minson, & Volpp, 2014), reducing alcohol consumption (Goldstein, 2001), increasing academic performance (Ariely & Wertenbroch, 2002) and increasing productivity (Marotta & Acquisti, 2017), among others. The fact that commitment strategies are an effective way to deal with self-control problems is well-established. However, most relevant to the present paper, this prior work has looked only at the "first-order" effects of commitment strategies, that is, the immediate effect on the individual engaging in them. We build on this work by focusing instead on the "second order" effects, or the ways in which beliefs about commitment strategies may affect how people judge those who use them. We hypothesize that individuals will perceive targets who use commitment strategies less favorably than targets who use willpower to achieve the same goal. We outline our reasoning behind this hypothesis in the following sections.

Self-Control and Effort

An important distinction between the use of willpower versus commitment strategies in overcoming self-control problems is the point at which the primary decision takes place. While commitment strategies require implementation of a plan prior to facing temptation, willpower requires active rejection of the temptation as it occurs. Thus, when using willpower, the primary decision of how to act takes place in a "hot" state where the benefits of the immediate reward are more psychologically salient than the benefits of the long-term goal. The effortful nature of this

act of resisting temptation in the moment has been shown to lead to fatigue (Inzlicht & Schmeichel, 2012).

There is evidence that people perceive using willpower to be an effortful act.[2] Even linguistically, "willpower" connotes the need for power to control one's will, as does "brute-force," (Duckworth, Milkman & Laibson, 2019) or "effortful inhibition" (Galla & Duckworth, 2015), both common terms used in conjunction with willpower. This suggests that there exists a lay association between willpower use and exerting effort.

Although the literature has not, until recently, made a fine distinction between willpower and other forms of self-control, there is evidence that strategies that do not require willpower may be less effortful (Adriaanse et al., 2014; de Ridder et al., 2012; Galla & Duckworth, 2015; Hofmann et al., 2012). Avoiding or minimizing the temptation via commitment strategies, however, allows the primary decision to be made in a "cold" state, thus creating conditions that eschew the need for effortful engagement or inhibition in a "hot" state (e.g. by selecting or modifying the situation, Duckworth, Gendler & Gross, 2016; Gross, 1998). The fact that the decision point for implementing a commitment strategy occurs well before any temptation can actually occur should lead people to see its use as less effortful than the use of willpower.

Valuing Effort

If using willpower is seen as more effortful than using commitment strategies, the natural next question is why this leads to stronger inferences of trustworthiness and moral character evaluations. To explain this, we draw on a large body of research demonstrating that, for many and in many cultures, the exertion of effort is itself moralized. In Western societies in particular,

[2]In the present work we do not necessarily take a stance regarding how objectively effortful various self-control strategies actually are. The construct of cognitive effort has recently been criticized as imprecise and difficult to define (Thomson & Oppenheimer, 2022). Instead we focus on the degree to which people perceive self-control strategies to be effortful, and the consequences of this for impression formation.

hard work is highly valued (Weber, 1905) and displaying a strong work ethic can signal that one is honest, virtuous, and accountable for their actions (Amos, Zhang & Read, 2019). Related work finds that people prefer donating to fundraising campaigns when the process is thought to be more effortful (e.g., participating in a marathon vs. a walk-a-thon; Olivola & Shafir, 2013). Effort more broadly carries an inherent value for many (Inzlicht, Shenhav & Olivola, 2018). For example, people view goals that feel more effortful as more valuable (Brehm et al., 1983) and consumers rely on effort as a heuristic for quality, preferring products that required more effort to be produced (Kruger et al., 2004). Effort's association with positive outcomes is so strong that people value even clearly unnecessary or irrelevant effort (Kim & Labroo, 2011; Labroo & Kim, 2009; Labroo, Lambotte & Zhang, 2009; Schrift, Netzer & Kivetz, 2011), even going as far as moralizing effort that has no objective purpose (Celniker et al, 2022).

Additional evidence ties the value people place on effort directly to the ability to overcome temptation or delay gratification. For example, delay of gratification has been proposed as a central tenet of the Protestant Work Ethic (Miller et al., 2002), and the effective use of self-control – both in close others and strangers – is associated with increased perceptions of trustworthiness (Righetti & Finkenauer, 2011). In addition, previous work has demonstrated that adults view overcoming inner conflict as more morally laudable than doing the right thing in the absence of this conflict (Berman & Small, 2018; Starmans & Bloom, 2016). Beyond the Western context, in China, children also recognize that delaying gratification can have reputational benefits, and therefore are more likely to do so when they know their behavior will be observed, particularly by an authority figure such as a teacher (Ma et al., 2020).

In sum, people place a high value on effort, both broadly and when specifically linked to self-control behaviors. Additionally, we argue that willpower is perceived as a more effortful

self-control strategy than the use of commitment strategies. It should be noted that psychological phenomena rarely, if ever, have a single explanation (Kirmani, 2015; Pham, 2013). As such, in Studies 3 and 4 we test our proposed effort mechanism along with other plausible mechanisms, which we discuss in greater detail further in the paper.

In the next section we outline how the premises we have put forth so far relate to our outcome measure of interest: trustworthiness.

Self-Control and Trustworthiness

We focus our examination of how people react to others' use of self-control techniques on evaluations of trustworthiness. Trust is a critical component of sustained social interactions. Relationships characterized by high levels of trust are more persistent, productive, and amicable, whereas distrust can erode or end once-fruitful collaborations (Cook, Hardin & Levi, 2005; Schweitzer, Hershey, & Bradlow, 2006; Wieselquist et al., 1999; Zaheer, McEvily, & Perrone, 1998) While past research has identified many facets of trust (Mayer, Davis, & Schoorman, 1995; Rempel, Holmes, & Zanna, 1985), we focus on two components that are particularly relevant to impression formation (Dorison, Umphres, & Lerner, 2021; Levine & Schweitzer, 2015; Zlatev, 2019): integrity-based trust and benevolence-based trust.

Integrity is commonly defined as adherence to a consistent set of principles (Mayer, Davis, & Schoorman, 1995). In particular, McFall (1987) argues that one can be said to have integrity if her beliefs meet two criteria. First, one's beliefs must be internally coherent (both with respect to one's other beliefs as well as one's actions and motivations). Second, one's beliefs must relate to socially valued topics. In other words, they must involve values that people would generally agree are important. It is this second criteria that is particularly relevant to the present work. To the extent that using willpower is seen as more effortful than using

commitment strategies, and that expending effort is a socially valued principle, it stands to reason that the choice to use willpower will be seen by others as a higher-integrity decision than the choice to use a commitment strategy.

Benevolence–the other fact of trustworthiness we explore here–is typically defined as the degree to which a trustee has good intentions toward a trustor (Mayer, Davis, & Schoorman, 1995). The proximal link between effort and benevolence—and hence the more distal link between self-control strategy choice and benevolence—is less clear. While people may have other reasons for believing a willpower user is benevolent, we do not have strong a priori beliefs about what they might be. Thus, we expect the impact on benevolence perceptions to be less pronounced than on integrity perceptions.

Contributions

While prior work has proposed that self-control more broadly is moralized (Mooijman et al., 2017), and that those exhibiting self-control in general are seen as more trustworthy (Righetti & Finkenauer, 2011), this work has not distinguished between different self-control strategies available to an individual. This is important because our findings demonstrate that people don't view all self-control strategies equivalently, underscoring the importance of examining them separately. Moreover, prior work has largely focused on the outcome of engaging in self-control (i.e., whether someone's attempt to behave in a self-controlled way succeeds or fails). In contrast, the present work focuses on the process by which self-control is used.

Our paper also contributes to the nascent literature on the importance of distinguishing between different self-control strategies (Duckworth, Gendler, & Gross, 2016; Fujita, 2011; Katzir et al., 2021) and highlights that not all strategies are perceived to be equally laudable. While this work begins to explore the interpersonal consequences of using different types of self-

control strategies, it also lays the groundwork for a new lens of exploring why people may fail to use certain strategies at their disposal. Despite the effectiveness of commitment strategies for achieving self-control, there is emerging evidence that people often fail to use them (Gine, Karlan & Zinman, 2010; Marotta & Acquisti, 2017; Moser, Schoenebeck & Resnick, 2019; Royer, Stehr, & Sydnor, 2015). The interpersonal consequences may be one piece of this puzzle, a point we return to in our final study and the General Discussion.

Research Overview

Across seven studies (and three supplemental studies), we explore whether people who choose to use commitment strategies are judged more negatively than people who choose to use internal willpower to overcome self-control problems. Specifically, we hypothesize that people who choose commitment strategies over willpower are trusted less and viewed as lower in integrity. In Study 1, we demonstrate this effect in an incentive-compatible trust game. In Study 2, we examine the generalizability of this phenomenon by demonstrating its persistence across a variety of self-control situations. In Studies 3a and 3b, using a causal chain design, we investigate two mechanisms: perceptions of effort and perceptions of impulsivity. Studies 4a and 4b employs a manipulate-the-mediator approach to further explore our proposed effort mechanism. Finally, in Study 5, we examine an implication of this work by exploring whether the negative interpersonal consequences of commitment strategy use we find in our previous studies contributes to people's documented reluctance to use commitment devices.

All studies received relevant IRB approval and were pre-registered, and all materials, data, and code are available at https://researchbox.org/470&PEER_REVIEW_passcode=NDECVI. We sought to maximize

power by collecting at least 100 participants per condition (as per the recommendation provided in Simmons, Nelson & Simonsohn, 2018).

Study 1

Study 1 seeks to test whether people are more likely to trust those who use willpower as opposed to commitment strategies to achieve their goals. The pre-registration for Study 1 is available at https://aspredicted.org/VYY_Q3G.

Study 1: Method

Participants. We recruited 400 participants and pre-registered that we would exclude participants who failed the comprehension checks. Our final sample consists of 383 participants (48.8% female, 34.2% non-white; M_{age}= 35.1, SD_{age}= 12.7) recruited from Prolific Academic who took part in an online experiment in exchange for monetary payment.

Procedure. Participants first read instructions for the trust game (Berg, Dickhaut, & McCabe, 1995), which involves two players, referred to here as the blue player and red player. Participants took on the role of the blue player, who received a $0.50 endowment and had to pass it to the red player.[3] Once the participant passes the money to the red player, the value of that money tripled to $1.50 and the red player has the choice to keep the entire $3 or split it with the participant (thereby each receiving a bonus of $0.75). After answering two comprehension questions about the trust game, participants read about two other online workers who could serve as the red player in their game. Specifically, participants read:

> *We are recruiting other Mechanical Turk workers to serve as the RED player.*
>
> *Some of those Mechanical Turk workers choose to use **willpower to resist** tempting websites (like Facebook or YouTube) when working online.*

[3] We recruited a separate group of participants to make decisions as the red player in order to match participants for bonus payments to avoid deception.

*Some of those Mechanical Turk workers choose to **download an app that blocks** tempting websites (like Facebook or YouTube) when working online.*

**

You have the option to pass the $0.75 to a RED player who uses willpower or to pass the $0.75 to a RED player who uses the app.

Remember - you want to send the money to the person who you trust is more likely to split it with you.

Which option do you choose?

After participants decided who they wanted to pass the money to, they provided their demographic information.

Study 1: Results

Effect of willpower on trust. Our primary dependent measure was which of the two red players participants preferred to pass the money to: the red player who used willpower or the red player who used the commitment strategy.

Of the 383 participants in our final sample, 263 (69%) chose to pass the money to the red player who used willpower, and 120 (31%) chose to pass the money to the red player who used the commitment strategy. A chi-squared test indicated that participants preferred to pass the money to the player who used willpower by a significant margin ($\chi^2 = 52.65$, $p < 0.0001$). In other words, people were significantly more likely to trust those who use willpower compared with those who use an external commitment strategy.

Study 1: Discussion

We demonstrated in an incentive-compatible study that people are more likely to trust those who use willpower as opposed to commitment strategies to achieve their goals (see Study

S1 in the SOM for a replication). However, there are a few limitations of this study, which we address in Study 2. First, Study 1 only looked at one type of commitment strategy within a single context. Second, Study 1 examined trust more broadly, but did not look at the two components of trust, integrity and benevolence, discussed above. We next seek to explore if our predictions hold across various scenarios and when looking specifically at integrity-based trust.

Study 2

Study 2 adds to the findings in Study 1 in three important ways. First, it moves beyond a single domain to gauge how individuals perceive a target who chooses to pursue their goal using willpower or a commitment strategy across a variety of common goal-setting contexts. Second, as opposed to looking at trust more generally, it examines multiple components of trust—benevolence-based and integrity-based trust—as well as impressions of the target's morality. Finally, it looks at whether people hold different beliefs about the efficacy of various goal achievement methods. The pre-registration for Study 2 is available at https://aspredicted.org/blind.php?x=p5ej6f.

Study 2: Method

Participants. Six-hundred two participants (49.5% female; 25% non-white; M_{age} 40.5, SD_{age} 12.8) from Amazon Mechanical Turk took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to read one of four vignettes regarding two individuals facing self-control problems. In each vignette, one individual tried solving the problem by relying on willpower, while the other did so by using a commitment strategy (namely, paying a friend $5 upon failing to exercise self-control). The four scenarios included

avoiding drinking alcohol, avoiding junk food, increasing gym attendance, and obtaining a flu shot (see Table 1for the full text of each scenario).

After reading the scenario, participants rated each individual in the vignette on a 1-7 scale regarding how much *integrity* they had ("has a great deal of integrity," "can trust their word," and "cares about honesty and truth," alpha = 0.93), how *benevolent* they were ("is kind," "is nice," and "is selfish" (reverse-coded), alpha = 0.75), and how *moral* they were ("is moral" and "is ethical," alpha = 0.97; items are from Levine & Schweitzer, 2015). Participants then indicated on a bipolar 7-point scale who would be more likely to achieve their goal (1 = Definitely the person using willpower, 4 = both equally likely, and 7 = Definitely the person using the commitment strategy).

Table 1. Scenarios used in Study 2.

Alcohol Avoidance	Junk Food Avoidance	Flu Shot	Gym Attendance
Max and Tom both want to drink less alcohol. At a party on New Years Eve, they can each decide whether they want to use pure willpower to avoid drinking alcohol or to set up a system where they pay a friend $5 if they drink alcohol. *Max chooses to use pure willpower to avoid drinking alcohol at the party.*	*Christine and Jane both want to stop eating as much unhealthy food. They can each decide whether they want to use pure willpower to avoid eating junk food in their house or to set up a system where they pay a friend $5 if they eat junk food.* *Christine chooses to use pure willpower to avoid eating junk food at home.* *Jane chooses to pay a friend $5 if she eats junk food at home.*	*Angela and Katy both want to get a flu shot this year. They can each decide whether they want to use pure willpower to go to the clinic and get the shot or to set up a system where they pay a friend $5 if they fail to get the shot.* *Angela chooses to use pure willpower to get the flu shot.* *Katy chooses to pay a friend $5 if she*	*Andrew and Brett both want to exercise more, specifically by going to the gym three times a week. They can each decide whether they want to use pure willpower to go to the gym or to set up a system where they pay a friend $5 each time they plan to go to the gym but do not make it.* *Andrew chooses to use pure willpower to go to the gym three times a week.*

Tom chooses to pay a friend $5 if he drinks alcohol at the party.		*doesn't get the flu shot.*	*Brett chooses to pay a friend $5 each week he does not make it to the gym three times.*

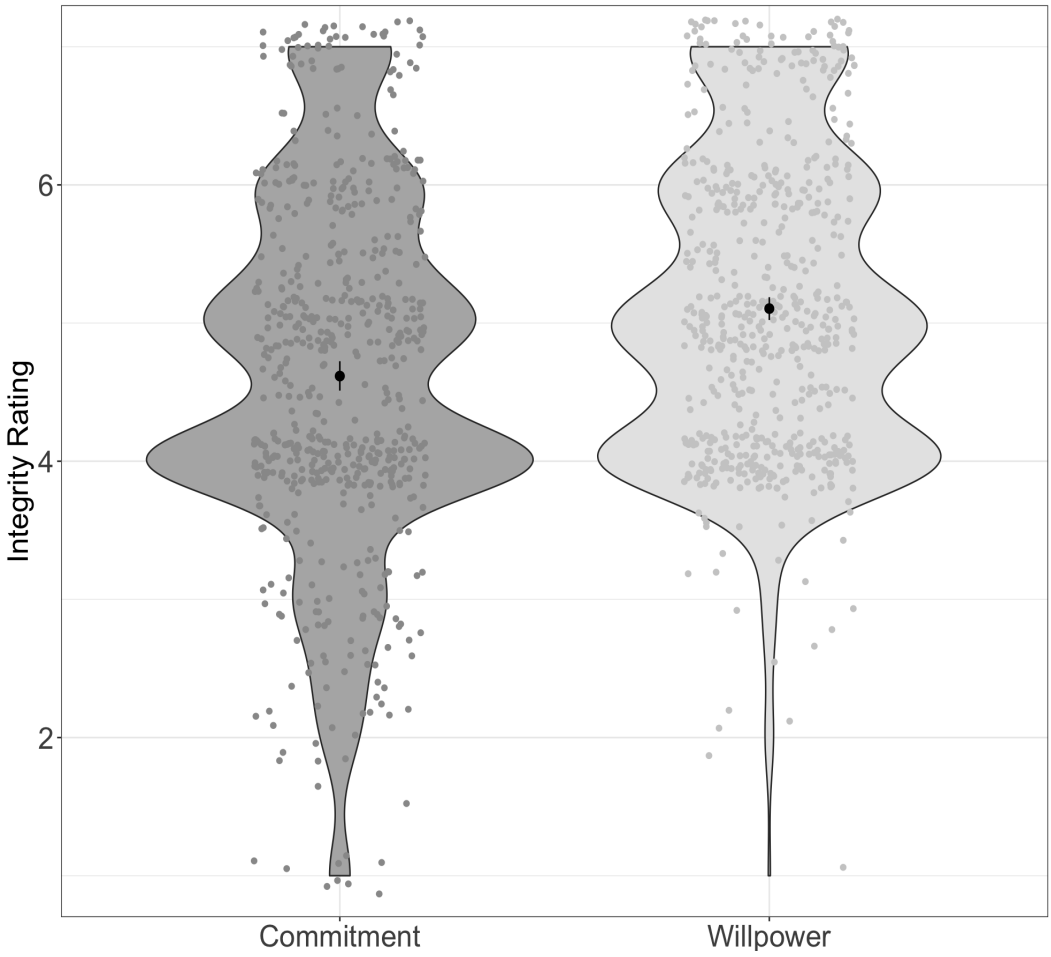
Study 2: Results

To test our predictions, we employed a series of mixed-effects linear regressions with random intercepts for participant and scenario fixed effects.

Effect of willpower on integrity-based trust. As predicted, we found that participants rated targets who used a commitment strategy to achieve their goal as having significantly lower integrity-based trust than targets who used willpower ($\beta = 0.40$, $SE = 0.05$, $t(601) = 8.04$, $p < 0.0001$; see Figure 1). This difference was significant within each individual scenario: flu shot scenario ($\beta = 0.53$, $SE = 0.11$, $t(300) = 4.76$, $p < 0.0001$), gym attendance scenario ($\beta = 0.39$, $SE = 0.11$, $t(300) = 3.48$, $p < 0.001$) alcohol avoidance scenario ($\beta = 0.39$, $SE = 0.11$, $t(304) = 3.51$, $p = 0.001$), and junk food avoidance scenario ($\beta = 0.28$, $SE = 0.12$, $t(292) = 2.40$, $p = 0.017$).

Effect of willpower on benevolence-based trust. We additionally found that targets who used a commitment strategy were rated significantly lower on benevolence-based trust than targets who used willpower ($\beta = 0.17$, $SE = 0.04$, $t(601) = 4.28$, $p < 0.0001$). However, this difference was only significant in the flu shot scenario ($\beta = 0.29$, $SE = 0.11$, $t(300) = 2.58$, $p = 0.010$) and was marginally significant in the alcohol avoidance scenario ($\beta = 0.22$, $SE = 0.11$, $t(304) = 1.961$, $p = 0.051$). We failed to detect a difference in benevolence-based trust by condition in either the gym attendance scenario ($\beta = 0.08$, $SE = 0.12$, $t(300) = 0.75$, $p = 0.448$) or the junk food avoidance scenario ($\beta = 0.08$, $SE = 0.12$, $t(292) = 0.68$, $p = 0.498$).

Figure 1. Overall impact of target's strategy use on ratings of integrity in Study 2.



Study 2: Discussion

Study 2 provided a conceptual replication of Study 1, using different measures of trust and a wider variety of scenarios that differed from Study 1 in terms of both the type of commitment strategy as well as the domain. Furthermore, we examined different components of trust, finding that the effect is more pronounced on integrity-based trust than benevolence-based trust. We also found a robust effect of self-control type on ratings of morality (see SOM), which is not surprising given how closely morality and integrity are linked (Uhlmann, Pizarro, & Diermeier, 2015). In all, Study 2 demonstrates that those who use commitment strategies are

viewed by others as lower on integrity-based trust compared with those who use willpower to achieve their goals.[4]

Studies 3a-3b

In Studies 3a and 3b we seek to explore potential mechanisms underlying this phenomenon. To recap, previous work has demonstrated that people place a high value on the exertion of effort (Celniker et al, 2022; Inzlicht, Shenhav & Olivola, 2018 Kim & Labroo, 2011; Labroo & Kim, 2009; Labroo, Lambotte & Zhang, 2009; Schrift, Netzer & Kivetz, 2011). Additionally, we suggest that internal willpower is viewed as more effortful than commitment strategy use. As a result, we propose that perceived effort plays a role in people's evaluations of trustworthiness as a function of the type of self-control used.

However, as mentioned earlier, most psychological phenomena are multiply-determined (Kirmani, 2015; Pham, 2013). As such, differential perceptions of effort is not the only possible reason why people trust willpower users over commitment strategy users. In particular, two other plausible mechanisms for this finding exist. First, people may view someone who uses willpower to achieve a goal as more committed to that goal than someone who uses a commitment strategy. Second, people may believe that someone who requires the use of a commitment strategy might be more impulsive than someone who is able to "brute force" a goal using willpower.

Differences in beliefs about goal commitment and impulsiveness both represent plausible alternative explanations for the link between self-control use and trustworthiness. It should be noted, however, that none of these explanations are mutually exclusive. As a result, in Studies 4a

[4] To rule out an alternative explanation that people have a specific positive association with the term "willpower," we ran a supplemental study (Study S2 in the SOM) asking participants to evaluate someone who chooses to use a commitment strategy vs. someone who "chooses not to use a commitment strategy" (i.e., without any mention of the use of willpower). We find similar results to Study 2, suggesting that participants are not simply reacting to the potentially positively-charged concept of willpower, but distrust commitment strategy users even in the absence of any mention of willpower.

and 4b we test all three of these potential mechanisms using a design that allows us to examine which combination (if any) may be contributing to our findings. To do so, we employ a causal chain design (e.g., Jung, Gonzalez, & Critcher, 2020) which avoids some of the pitfalls of the measured-mediation approach (Pirlott & MacKinnon, 2016; Smith, 2012; Spencer, Zanna, & Fong, 2005). Specifically, Study 3a tests the causal effect of strategy type (willpower vs. commitment) on our three potential mediators by experimentally manipulating which strategy a target uses and then measuring perceptions of the target's goal commitment, impulsiveness, and effort. Next, Study 3b examines the second part of the causal chain by manipulating the potential mechanisms experimentally and measuring their effect on integrity- and benevolence-based trust. It should be noted that both of these steps are required in order to satisfy the conditions for a successful causal chain design (Spencer, Zanna, & Fong, 2005). As a result, we only include in Study 3b the potential mechanisms that were significantly affected by strategy type in Study 3a. The pre-registrations for these studies are available at https://aspredicted.org/BXC_J3T (Study 3a) and https://aspredicted.org/2J4_S4V (Study 3b).

Study 3a: Method

Participants. Three-hundred ninety-nine participants (49.4% female; 30.6% non-white; M_{age} 35.3, SD_{age} 12.8) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. Participants all read the following scenario.[5]

"Rory and Sam are both online workers. They both spend 6 hours a day completing tasks. Rory uses internal willpower to stay focused and complete MTurk tasks. Sam uses

[5] Due to an oversight, participants read about two Mechanical Turk workers in this study although the study was run on Prolific Academic. However, there is no reason to believe this affected the results in any way.

an app that blocks distracting websites (like Facebook, Instagram, etc.) across multiple devices to stay focused and complete MTurk tasks."

After reading the scenario, participants were randomly assigned to either rate Rory or Sam on the three potential mechanisms discussed earlier on a scale from 1 (Strongly disagree) to 7 (Strongly agree).

To assess impulsivity we adapted the Attention subscale of the Barratt Impulsiveness Scale (BIS-11; Patton, Stanford, & Barratt, 1995), which consists of the following items: "X has racing thoughts," "X concentrates easily," "X has outside thoughts when thinking," and "X is restless at lectures or talks," (alpha = 0.89).

To assess effort, we adapted the Work Effort Scale from Brown and Leigh (1996), which consists of the following items: "When there's a job to be done, they will devote all their energy to getting it done," "When they work, they will do so with intensity," "They will work at their full capacity in all of their job duties," "They will strive as hard as they can to be successful in their work," and "When they work, they will exert themselves to the fullest." (alpha = 0.96).

To assess goal commitment, we adapted the four highest-loading items from the Goal Commitment Scale used in Hollenbeck and Klein (1987), which consists of the following items (the first three are reverse coded): "Quite frankly, X doesn't care if they complete MTurk tasks," "It wouldn't take much to make X abandon their goal of completing MTurk tasks," "It is unrealistic to expect X to reach their goal of completing MTurk tasks," and "X thinks completing MTurk tasks is a good goal to shoot for," (alpha =0.77). Presentation of dependent variable and mechanism scales were counterbalanced between subjects.

Study 3a: Results

Effect of willpower on perceptions of effort. Participants indicated they believed the target who used a commitment strategy exerted significantly less effort than the target who used willpower ($\beta = -0.21$, SE = 0.10, $t(397) = -2.10$, $p = 0.036$).

Effect of willpower on perceptions of impulsivity. Participants additionally indicated they believed the target who used a commitment strategy was significantly more impulsive than the target who used willpower ($\beta = 1.28$, SE = 0.08, $t(397) = 16.67$, $p < 0.0001$).

Effect of willpower on perceptions of goal commitment. There was no significant difference between ratings of goal commitment as a function of self-control strategy used ($\beta = 0.08$, SE = 0.10, $t(397) = 0.82$, $p = 0.41$).

Study 3a: Discussion

In the first part of a causal chain design, we tested three different potential mechanisms for our proposed effect: perceptions of effort, impulsivity, and goal commitment. We find that, compared to willpower users, people who use commitment strategies are seen as both exerting less effort towards reaching their goal and having stronger negative impulses that could impede goal achievement. However, we did not find evidence that they were viewed any differently on the degree to which they were committed to their goal.

In Study 3b we test the second part of the causal chain by manipulating the potential mechanisms experimentally and measuring the effect on our outcome measures of interest. As mentioned earlier, we only include potential mechanisms in the second part of the causal chain that were significantly affected by the commitment strategy manipulation in the first part (Study 3a). As a result, in Study 3b, we manipulate effort and impulsivity (but not goal commitment) independently, and measure their impact on perceptions of integrity- and benevolence-based trust.

Study 3b: Method

Participants. Eight-hundred participants (49.6% female; 25.3% non-white; M_{age} = 37.2, SD_{age} 12.7) from Prolific Academic took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to read a scenario that highlighted one of two mechanisms. Participants assigned to the effort condition read the following scenario:

> *"Rory and Sam are both online workers. They both spend 6 hours a day completing tasks. Rory is the type of person who [DOES/DOES NOT] work with intensity and WILL/WILL NOT] strive as hard as they can to be successful in their work. Sam is the type of person who [DOES NOT/DOES] devote all their energy to getting their work done and [WILL NOT/WILL] strive as hard as they can to be successful in their work."*

Participants assigned to the impulsivity condition read the following scenario:

> *"Rory and Sam are both online workers. They both spend 6 hours a day completing tasks. Rory is the type of person who [DOES/DOES NOT] have racing thoughts and [CANNOT/CAN] concentrate easily. Sam is the type of person who[DOES NOT/DOES] have racing thoughts and [CAN/CANNOT} concentrate easily."*

After reading the scenario, participants were randomly assigned to either rate Rory or Sam on the same *integrity* and *benevolence* scales described in Study 2 (as the dependent variables).

Study 3b: Results

To test our predictions, we employed a series of linear regressions to analyze the two different mechanism conditions separately.

Effect of effort on perceptions of integrity. We found a significant effect of effort such that targets who exerted more effort were rated significantly higher on perceptions of integrity than targets who exerted less effort ($\beta = 1.44$, $SE = 0.07$, $t(399) = 20.86$, $p < 0.001$).

Effect of effort on perceptions of benevolence. We additionally found that targets who exerted more of effort were rated significantly higher on perceptions of benevolence than targets who exerted less effort ($\beta = 1.01$, $SE = 0.08$, $t(399) = 11.70$, $p < 0.001$).

Effect of impulsivity on perceptions of integrity. Turning to the other potential mechanism, we also found a significant effect of impulsivity such that less impulsive targets were rated significantly higher on perceptions of integrity than more impulsive targets ($\beta = -0.41$, $SE = 0.10$, $t(397) = -4.19$, $p < 0.001$).

Effect of impulsivity on perceptions of benevolence. We did not detect a difference in benevolence-based trust as a function of impulsivity level ($\beta = -0.08$, $SE = 0.10$, $t(397) = -0.76$, $p = 0.449$).

Study 3: Discussion

Across Studies 3a and 3b, we employed a causal chain design (Spencer, Zanna, & Fong, 2005) to test three different potential mechanisms for the relationship between self-control strategy use and trustworthiness: perceptions of effort, impulsivity, and goal commitment. In line with our hypotheses, we found evidence that when people use commitment strategies they are thought to be exerting less effort toward reaching their goal. We additionally found evidence for a second mechanism: commitment strategy users are also seen as having stronger negative impulses that could impede goal achievement. We did not find any evidence that the third potential mediator—differences in goal commitment—was underlying this phenomenon.

In sum, these results suggest that the effect of self-control strategy use on perceptions of trustworthiness is multiply-determined, a point we return to in the General Discussion. In order to see how robust these results are, in Studies 4a and 4b we employ an alternative method of experimental mediation to examine potential mechanisms.

Studies 4a-4b

Study 4 further investigates the potential mechanisms underlying the negative impressions people form about individuals who use commitment strategies instead of willpower to achieve their goals While Studies 3a-3b used a causal chain design, in Studies 4a-4b we employ an alternative approach also recommended by Spencer, Zanna, and Fong 2005): a moderation-of-process design. We use both of these methods in tandem to demonstrate the robustness of the evidence for these mechanisms. In a moderation-of-process design, the independent measure and mechanism are manipulated orthogonally using a factorial design. Because we were examining two potential mechanisms, we manipulate both independently in the same study and evaluate them separately.

The pre-registration for Study 4a is available at https://aspredicted.org/blind.php?x=n5qd5y and the pre-registration for Study 4b is available at https://aspredicted.org/X2M_GFD.

Study 4a: Method

Participants. Five hundred eighty-four participants (57.5% female; 22.4% non-white; M_{age} 41.1, SD_{age} 12.6) from Amazon Mechanical Turk took part in an online experiment in exchange for monetary payment.

Procedure. Participants read one of the four vignettes used in Study 2 about two individuals facing self-control problems and wanting to eat less junk food. In the vignette, one

individual tried solving the problem relying on willpower while the other individual used a commitment strategy (namely, paying a friend $5 upon failing to exercise self-control).

Participants were randomly assigned to one of two between-subjects effort conditions. In the willpower effort condition, participants read that the person using willpower exerted more effort to achieve the goal, while in the commitment strategy effort condition, participants read that the person using the commitment strategy exerted more effort to achieve the goal. ("X puts in **a lot of effort** to reach the goal of avoiding eating junk food at home" and "Y puts in **very little effort** to reach the goal of avoiding eating junk food at home.")

After reading the scenario, participants rated each individual in the vignette on a 1-7 scale regarding how much *integrity* they had ("has a great deal of integrity," "can trust their word," and "cares about honesty and truth," alpha = 0.95), how *benevolent* they were ("is kind," "is nice," and "is selfish" (reverse-coded), alpha = 0.74), and how *moral* they thought they were ("is moral" and "is ethical," alpha = 0.95; items are from Levine & Schweitzer, 2015). Participants then indicated on a bipolar 7-point scale which of the two individuals would be more likely to achieve their goal.

A combination of two results would be consistent with effort as the driver of the effect of strategy choice on integrity impressions. First, giving people explicit information about the amount of effort each target exerted should nullify the effect of strategy choice on ratings of integrity. In other words, when participants are randomly assigned to learn that either the willpower or commitment strategy user exerted greater effort, we expect a null result of strategy choice on impressions of integrity. Second, the amount of effort exerted should now play a role in ratings of integrity. In other words, regardless of strategy used, we expect people will rate the high effort target as having greater integrity than the low effort target.

Study 4a: Results

To test our predictions, we employed a series of mixed-effects linear regressions with random intercepts for participants.

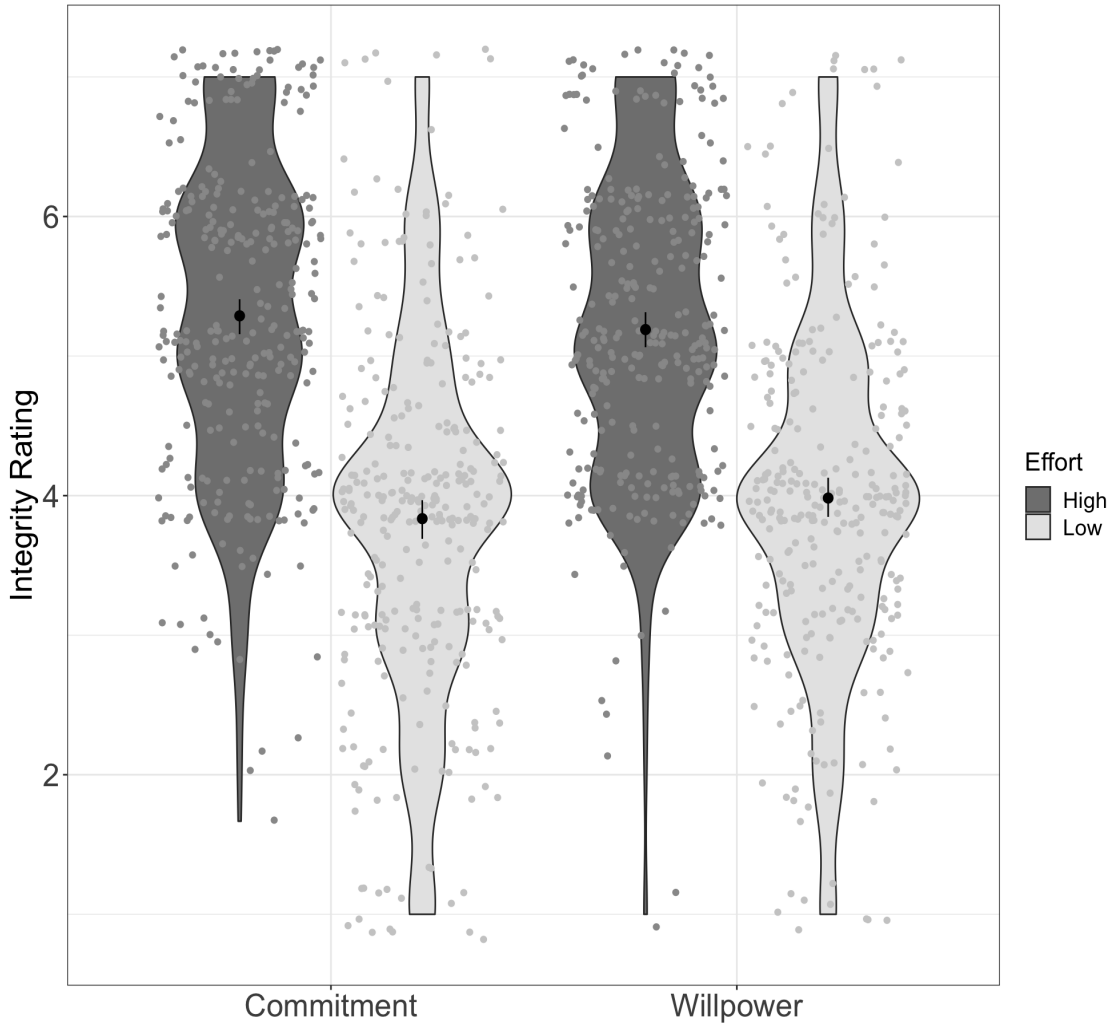
Effect of willpower on integrity. Consistent with our expectations, we find no effect of strategy choice on ratings of integrity ($\beta = 0.03$, SE = 0.06, $t(583) = 0.479$, $p = 0.632$). In other words, we failed to detect a significant main effect of willpower on integrity when collapsing across effort conditions.

Effect of effort on integrity. Consistent with the second pattern of results described earlier, we find a significant effect of perceptions of effort on integrity, such that, high-effort targets were rated higher on integrity than low-effort targets ($\beta = 0.98$, SE = 0.05, $t(1166) = 19.09$, $p < 0.0001$). There was a marginally significant interaction between willpower and effort on integrity ($\beta = 0.18$, SE = 0.10, $t(1164) = 1.77$, $p = 0.08$).

Effect of willpower on benevolence. Similarly, once we account for effort exerted, we also no longer find a significant impact of willpower on benevolence ($\beta = 0.00$, SE = 0.03, $t(583) = 0.043$, $p = 0.97$).

Effect of effort on benevolence. Hitherto, the effects of willpower on benevolence were modest at best, and here, we find that perceptions of high effort do not seem to impact perceptions of benevolence. Using a hierarchical linear model with scenario fixed effects and participant random slopes, we fail to detect a significant effect ($\beta = -0.04$, SE = 0.03, $t(584) = -1.12$, $p = 0.263$). There was, however, a significant interaction between willpower and effort on benevolence ($\beta = 0.41$, SE = 0.15, $t(581) = 2.70$, $p = 0.007$).

Figure 2. Overall integrity ratings for targets deploying high vs. low effort to achieve their goal in Study 4a.



Study 4b: Method

Participants. Six hundred and five participants (57.9% female; 26.0% non-white; M_{age} 40.1, SD_{age} 12.5) from Amazon Mechanical Turk took part in an online experiment in exchange for monetary payment.

Procedure. Participants read the same vignette as in Study 4a. In the vignette, one individual tried solving the problem relying on willpower while the other individual used a commitment strategy (namely, paying a friend $5 upon failing to exercise self-control).

Participants were randomly assigned to one of two effort conditions. In the willpower impulsive condition, participants read that the person using willpower had to overcome stronger impulses to achieve the goal, while in the commitment strategy effort condition, participants read that the person using the commitment strategy had stronger negative impulses. ("X does NOT have to overcome strong impulses to reach the goal of avoiding eating junk food at home," and "Y does have to overcome strong impulses to reach the goal of avoiding eating junk food at home.")

After reading the scenario, participants rated each individual in the vignette on a 1-7 scale regarding how much *integrity* they had ("has a great deal of integrity," "can trust their word," and "cares about honesty and truth," alpha = 0.92) and how *benevolent* they were ("is kind," "is nice," and "is selfish" (reverse-coded), alpha = 0.71); items are from Levine & Schweitzer, 2015). Participants then indicated on a bipolar 7-point scale which of the two individuals would be more likely to achieve their goal.

Study 4b: Results

To test our predictions, we employed a series of mixed-effects linear regressions with random intercepts for participants.

Effect of willpower on integrity. Unlike in Study 4a, here we still find an effect of strategy choice on ratings of integrity, such that, willpower using targets were rated higher on integrity than commitment strategy-using targets ($\beta = 0.156$, SE = 0.05, $t(603) = 3.08$, $p = 0.002$). In other words, we still detect a significant main effect of willpower on integrity when collapsing across impulsivity conditions.

Effect of impulsivity on integrity. Unlike in Study 3, here we fail to detect a significant effect of perceptions of impulsivity on integrity, ($\beta = -0.03$, SE = 0.05, $t(603) = -0.66$, $p = 0.51$).

There was no significant interaction between willpower and impulsivity on integrity (β = 0.13, SE = 0.13, t(602) = 1.04, p = 0.30).

Effect of willpower on benevolence. Similarly, once we account for impulsivity, we also no longer find a significant impact of willpower on benevolence (β = -0.03, SE = 0.04, t(603) = -0.68, p = 0.50).

Effect of impulsivity on benevolence. Hitherto, the effects of willpower on benevolence were modest at best, and here, we find that perceptions of high impulsivity do not seem to impact perceptions of benevolence. Using a hierarchical linear model with scenario fixed effects and participant random slopes, we fail to detect a significant effect (β = -0.05, SE = 0.04, t(603) = -1.43, p = 0.15). We did not detect a significant interaction between willpower and impulsivity on benevolence (β =0.13, SE = 0.15, t(602) = 0.90, p = 0.37).

Study 4: Discussion

Studies 4a and 4b provide evidence that effort plays a key role in the different impressions people form of willpower and commitment strategy users; on the other hand, the role of impulsivity—while evident in Studies 3a and 3b—was less clearcut. Providing explicit information on how much effort each person actually exerted in achieving their goal eliminated the effect of strategy choice on integrity-based trust. Additionally, effort played a role in integrity impressions independently of strategy chosen. Neither of these patterns were true for impulsivity.

One unexpected finding revolved around the interaction between effort and strategy choice in Study 4a, which was marginal for integrity-based trust and significant for benevolence-based trust. We hesitate to draw any strong conclusions from these specific results, as they were not predicted a priori and the specific patterns found are not easily explained by existing theories. Nevertheless, it may be fruitful for future research to explore this further.

Study 5

Studies 1 through 4 examined this phenomenon from a third-party perspective; that is, these studies looked at others' impressions of an individual's strategy choice. Study 5 shifts to the decision maker's perspective, and seeks to investigate whether individuals anticipate these negative interpersonal consequences of commitment strategy use and change their behavior as a result. Specifically, we look at whether people are more likely to employ a commitment strategy when the decision will be kept private compared with when the decision will be made public. In doing so we examine a potential implication of the findings in this paper, namely they may contribute to reluctance to take up opportunities to use commitment strategies, particularly when doing so is made public to others. The pre-registration for Study 5 is available at https://aspredicted.org/blind.php?x=SZS_DPQ.

Study 5: Method

Participants. Three-hundred one participants (45.2% female; 22.6% non-white; M_{age} 44.5, SD_{age} 12.8) from Amazon Mechanical Turk took part in an online experiment in exchange for monetary payment.

Procedure. Participants were randomly assigned to read about one of three commitment strategies: a commitment contract website, a web blocker application, and a lock box. In the commitment contract scenario, participants read:

> "*There is a website that allows you to make a "commitment contract" to stick to your goals. You can create a goal and then give the website your credit card details. If you fail to reach your goal then the website automatically donates money to either to a charity you support or a charity you do not support. You can select the amount of money to put*

on the line and whether you want your money to go to a charity you support or do not support if you fail to reach your goal."

In the web blocker scenario, participants read:

"*There is an app you can download that can block tempting websites (like Facebook or YouTube) across all of your devices during working hours that you set.*"

In the lock box scenario, participants read the following and then were shown a photo of the lock box:

"*Below is a photo of a product that helps people with self-control. For example, you can put cookies into the lock box, and then set a timer, so that you can only access the items in the box for a certain amount of time each day.*"

Participants then answered two questions (the order was counterbalanced across participants): "How likely are you to use this _____ **if you knew that <u>other people would find out</u> that you used it**?" and "How likely are you to use this _____ **if you knew that <u>nobody would find out</u> that you used it**?" These were answered on a five-point Likert scale ranging from "Very Likely" (which we coded as 2) to "Very Unlikely" (which we coded as -2). Then participants were faced with a final, forced choice question: "Assuming you had to use_____, under which condition would you prefer to" and could select between when the decision would be kept private or made public. The blank space included one of the following, depending on the assigned scenario: "this website" (commitment contract), "this app" (web blocker), or "this lock box".

Study 5: Results

Effect of privacy on decision to use a commitment strategy. When the decision would be kept private, participants were fairly unlikely to use the commitment strategy (M = -0.37, SD = 1.37). Broken down by strategy, we see the same pattern for the commitment contract scenario

(M = -0.67, SD = 1.26), the web blocker scenario (M = -0.16, SD = 1.32), and the lock box (M = -0.27, SD = 1.38). When the decision would be made public, participants were significantly less likely to use the commitment strategy (M = -0.63, SD = 1.27, t(598) = -2.47, p = 0.013). Broken down by strategy, we see the same pattern for the commitment contract (M = -0.76, SD = 1.20), for the web blocker (M = -0.46, SD = 1.28ß), and for the lock box (M = -0.67, SD =1.31). Using a hierarchical linear model controlling for commitment strategy type and including participant random intercepts, we see similar results; namely, participants were significantly less likely to indicate that they would use a commitment strategy when the decision would be public compared to when it would be private (β = -0.20, SE = 0.04, t(300) = -4.75, p < 0.0001). Table 2 summarizes these results and examines whether each rating differed significantly from the midpoint of the scale.

Table 2. Effect of privacy decision to use a commitment strategy, by scenario, and difference from midpoint in Study 5.

Scenario	**Private** Mean (SD)	**Difference from Midpoint**	**Public** Mean (SD)	**Difference from Midpoint**
All	-0.37 (1.37)	$t_{(300)}$ = -4.74, p <0.0001	-0.63 (1.27)	$t_{(300)}$ = -8.59, p <0.0001
Commitment contract	-0.67 (1.26)	$t_{(99)}$ = -5.30, p <0.0001	-0.76 (1.20)	$t_{(99)}$ = -6.34, p <0.0001
Web blocker	-0.16 (1.38)	$t_{(100)}$ = -1.21, p =0.230	-0.46 (1.31)	$t_{(100)}$ = -3.56, p <0.001
Lock box	-0.27 (1.32)	$t_{(99)}$ = -1.95, p =0.054	-0.67 (1.28)	$t_{(99)}$ = -5.11, p <0.0001

Condition under which participants would use a commitment strategy (forced choice). Overall, 78% of participants preferred to use the commitment strategy when the decision would be kept private compared to made public, the majority by a significant margin (χ^2 = 94.89, p < 0.0001). In the commitment contract scenario, 71% of participants indicated they would prefer to

use it if their choice were kept private ($\chi^2 = 17.64$, $p < 0.0001$), while in the lock box and web blocker app scenarios, the percent who would use it if their choices were kept private were 82% ($\chi^2 = 40.96$, $p < 0.0001$) and 81% respectively ($\chi^2 = 39.30$, $p < 0.0001$). In all three scenarios, this was the majority by a significant margin.

Study 5: Discussion

Although there were subtle differences between different types of commitment strategies, across the three scenarios examined in this study, overall, participants were hesitant to use commitment strategies regardless of whether their decision would be public or private. However, consistent with the prediction that people anticipate the negative interpersonal consequences of commitment strategy use, participants expressed stronger hesitancy if the decision would be made public compared to if it were kept private. This study provides initial evidence that people are at least somewhat aware of the negative interpersonal consequences of commitment device use; however, the low likelihood of using a commitment strategy even when its use would be kept private indicates that there are additional concerns fueling the underuse of effective commitments strategies.

General Discussion

The understanding that people can use commitment strategies to overcome self-control problems has existed for thousands of years. Since at least the time of Homer and *Odysseus*, the focus of this domain has been mainly on the efficacy of these strategies for the person choosing to engage in them. This prior work has demonstrated, for example, that Odysseus made the right decision by tying himself to the mast rather than attempting to use willpower to resist the sirens in the moment (e.g., Ashraf, Karlan, & Yin, 2006).

However, what has until now gone unexamined is whether there are interpersonal consequences to these otherwise helpful strategies. What did Odysseus's crew think about his decision to eschew willpower in favor of precommitment? Our findings suggest that individuals who use commitment strategies are often seen as less trustworthy (Study 1), particularly with regard to integrity-based trust. These negative perceptions of commitment strategy users vis-a-vis willpower users persist even in cases where individuals recognize the efficacy of the commitment strategies. We further find that this effect is at least partially driven by the fact that people perceive those who use commitment strategies as employing less effort toward achieving their goal than those who use willpower. Accounting for this differential perception in effort exerted mitigates the negative effects of commitment strategy use.

Theoretical Implications

Our work contributes to three main literatures. First, we add to the self-control literature. This has been an active field of research for over half a century, with some psychologists, such as Walter Mischel, recognizing the effectiveness of different types of self-control strategies (Mischel & Mischel, 1987). However, much of the ensuing work in this domain has considered "self-control" as synonymous with "willpower" (Duckworth & Kern, 2011; Inzlicht, Schmeichel, & Macrae, 2014; Lian, Yam, Ferris, & Brown, 2017; Metcalfe & Mischel, 1999). Only in the past few years has there been an attempt to distinguish between internal willpower and other external self-control strategies (Ainslie, 2021; Duckworth, Gendler & Gross, 2016; Duckworth, Milkman, & Laibson, 2019; Inzlicht et al., 2021, Kristal & Zlatev, 2021). This paper demonstrates that people evaluate these external self-control strategies differently (i.e., less positively) than willpower, providing corroborating evidence for their distinctiveness.

Second, we contribute to the literature on effort and, more specifically, lay beliefs about effort. While effort has previously been shown to have intrinsic value (Inzlicht, Shenhav, & Olivola, 2018), even when it is not productive (Celinker et al., working paper; Inzlicht, Shenhav & Olivola, 2018), we introduce these ideas in a new domain, namely, that of self-control. In Studies 3a and 3b, we employ a causal chain design to demonstrate that people judge those who use willpower as exerting more effort towards reaching their goal compared with those who use a commitment strategy, and that exerting effort leads to higher perceptions of integrity. Further, in Study 4a, we demonstrate that when we explicitly stipulated which target exerted more effort, the impact of willpower on integrity disappeared. This implies that, in the absence of additional information, people infer that using willpower exerts more effort than using a commitment strategy. We also found suggestive evidence that inferences about the degree to which people have negative impulses may also underlie the relationship between strategy use and integrity (Studies 3a and 3b). However, we did not find further evidence for this mechanism in a follow-up study using a moderation-of-process design (Study 4b).

Furthermore, we contribute to a growing literature examining what signals people use to evaluate others' trustworthiness (e.g., Dorison, Umphres, & Lerner, 2021; Zlatev, 2019). Though previous work has demonstrated that increased self-control can lead to greater perceptions of trustworthiness (Righetti & Finkenauer, 2011), this work did not distinguish between different types of self-control strategies, namely the use of internal approaches (e.g., willpower) compared to external approaches (e.g., commitment strategies). We demonstrate that, beyond simply using self-control or not, the way in which one demonstrates self-control impacts how trustworthy they appear to others, primarily driven by implicit perceptions of effort.

Finally, as we see in Study 5, people are hesitant to use commitment strategies particularly when their use will be made public. Although we demonstrate that there is an element of anticipated interpersonal consequences driving this effect, we still find self-reported resistance to using commitment strategies even when their use will be kept private. Therefore, we identify only one of many potential barriers preventing people from using commitment strategies effectively.

Practical Implications

A conclusion that one might draw from the voluminous literature demonstrating the advantages of commitment strategies is that these tactics are unequivocally positive, and thus always preferable to the alternative. We show that, in fact, there are real consequences to being seen as the type of person who chooses to use a commitment strategy to overcome temptation. Being seen as untrustworthy can harm one's social relationships (Rempel, Holmes, & Zanna, 1995) as well as economic outcomes (Balliet & Van Lange, 2013). As a result, while using commitment strategies may have first-order benefits, it is important to keep in mind that they may simultaneously have second-order drawbacks. The current work, however, does suggest one way to lessen the negative impact of addressing self-control problems with commitment strategies. In particular, commitment strategies that appear effortful may be valued more by others. This presents a promising path toward identifying win-win self-control tactics that are both effective and widely accepted.

This work also suggests that judgments about a target's morality and trustworthiness are based less on that target's competence and more on her internal preferences surrounding how to deal with internal conflict. Evidence for this comes from a supplemental study (Study S3 in the SOM) demonstrating a target's choice of strategy played a role in trustworthiness evaluations

even when people were given explicit information about whether or not the target succeeded in overcoming a self-control problem. This result bolsters the claim that people use others' autonomous decisions as indicative of their internal traits. This is in line with previous work finding that people judge others' morality less on the outcomes of their decisions and more on what led to those outcomes and what the decisions say about the type of person they are (Critcher, Helzer, & Tannenbaum, 2020; Tannenbaum, Uhlmann, & Diermeier, 2011; Uhlmann, Pizarro, & Diermeier, 2015).

This has important implications for the structure of programs and initiatives whose goal is to increase the uptake of precommitment strategies. Especially because, as we have demonstrated in Study 5, people are sensitive to whether their decision to use a commitment strategy would be made public. For example, people may avoid signing up for a program such as Stickk, where people set up commitment contracts to help reach their goals, not because they don't think it would be helpful, but rather because they are afraid of what others might think of their choice to do so. Designing these programs to minimize the interpersonal consequences of engaging in them would help increase their use.

Limitations and future directions

Across seven studies, we demonstrate that the type of self-control strategy one uses influences interpersonal perceptions of trust. Though we obtain these results in both self-reported and behavioral measures, a key limitation of this research is that it is primarily scenario-based. Future research should investigate the interpersonal consequences of commitment strategy use in more behavioral settings.

We also consciously decided to limit our research to an American sample, because we predict that lay beliefs regarding the link between effort and self-control would be particularly

strong in countries with high levels of Protestant Work Ethic and Judeo-Christian values. However, recent work has found that delay of gratification is valued as a form of reputation management among children in China (Ma et al., 2020), suggesting that the desire to be seen as self-controlled by others spans these cultural differences. As a result, future research should examine if this effect holds, or how it may differ, in other cultures that may have different norms around when and how effort is valued.

We also limited our research to goal achievement situations where individuals face want/should conflicts related to self-control failures (Bazerman et al., 1998; Milkman, Rogers & Bazerman, 2008). As such, we did not explore the effect of strategy use in dealing with potential moral failures. However, it seems likely that the interpersonal costs of using a commitment strategy would be even stronger in the moral domain (e.g., someone using a commitment strategy to prevent themselves from cheating on their spouse or abusing animals). It is additionally possible that the impulsivity mechanism would play a larger role in the moral domain than the effort mechanism.

Moreover, though we test a series of different types of strategies, and do observe some variance between them (e.g. in Study 5 where we can explicitly compare commitment contracts to lock boxes and web blockers), we do not systematically evaluate different types of commitment strategies. Additional research is needed to understand how various features of commitment strategies are perceived by potential users (e.g., how unusual a strategy is or whether it has a social component), as well as third-party observers, to determine what elements of a commitment strategy are seen as more or less favorable.

Finally, despite our demonstration that people are more likely to use a commitment strategy when its use is kept private, Study 5 indicates that people are still fairly unlikely to

express interest in commitment strategies altogether. Future research needs to explore the other, non-interpersonal factors influencing commitment strategy use.

Conclusion

Self-control is an important feature of everyday decision-making but, until recently, many researchers and individuals conflated willpower (a specific strategy) with self-control (a desired outcome). We demonstrate that not only is willpower but one means of self-control, but also those who use it are rated more favorably than those who use alternative commitment strategies. This benefit seems to be derived from the assumption of how much effort a person is putting in to achieving their self-control goal. By examining the role of interpersonal judgments in self-control strategy use, we can begin to understand why people may not use these beneficial strategies and how to promote effective strategy use.

References

Adriaanse, Marieke A., Floor M. Kroese, Marleen Gillebaart, and Denise TD De Ridder. "Effortless inhibition: Habit mediates the relation between self-control and unhealthy snack consumption." *Frontiers in psychology*, 5, 444.

Amos, C., Zhang, L., & Read, D. (2019). Hardworking as a heuristic for moral character: Why we attribute moral values to those who work hard and its implications. *Journal of Business Ethics*, 1-16.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological science*, 13(3), 219-224.

Ashraf, N., Karlan, D., & Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, 121(2), 635-672.

Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of management review,* 23(2), 225-241.

Bryan, G., Karlan, D., & Nelson, S. (2010). Commitment devices. *Annu. Rev. Econ.,* 2(1), 671-698.

Celniker, J., Gregory, A., Koo, H., Piff, P. K., Ditto, P. H., & Shariff, A. (2022). The Moralization of Unproductive Effort. *Journal of Experimental Psychology: General*

Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Moral character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology*, *87*, 103906.

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the

workplace. *Journal of Personality and Social Psychology,* 107(5), 943.

Cook, K. S., Hardin, R., & Levi, M. (2005). *Cooperation without trust?.* Russell Sage Foundation.

Daly, M., Delaney, L., Egan, M., & Baumeister, R. F. (2015). Childhood self-control and unemployment throughout the life span: Evidence from two British cohort studies. *Psychological science,* 26(6), 709-723.

De Ridder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review,* 16(1), 76-99.

Dorison, C. A., Umphres, C. K., & Lerner, J. S. (2021). Staying the course: Decision makers who escalate commitment are trusted and trustworthy. *Journal of Experimental Psychology: General*.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2014). Self-control in school-age children. *Educational Psychologist,* 49(3), 199-217.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science,* 11(1), 35-55.

Duckworth, A. L., Milkman, K. L., & Laibson, D. (2019). Beyond willpower: Strategies for reducing failures of self-control. *Psychological Science in the Public Interest,* 19(3), 102-129.

Elster, J. (1979). Ulysses and the sirens: Studies in rationality and irrationality.

Finkel, E. J., & Campbell, W. K. (2001). Self-control and accommodation in close relationships: an interdependence analysis. *Journal of personality and social psychology*, 81(2), 263.

Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of

impulses. *Personality and social psychology review*, *15*(4), 352-366.

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of personality and social psychology*, 109(3), 508.

Gathergood, J. (2012). Self-control, financial literacy and consumer over-indebtedness. *Journal of economic psychology*, 33(3), 590-602.

Gillebaart, M., & de Ridder, D. T. (2015). Effortless self-control: A novel perspective on response conflict strategies in trait self-control. *Social and Personality Psychology Compass*, *9*(2), 88-99.

Giné, X., Karlan, D., & Zinman, J. (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4), 213-35.

Gino, F., Schweitzer, M. E., Mead, N. L., & Ariely, D. (2011). Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational behavior and human decision processes*, 115(2), 191-203.

Goldstein, A. (2001). *Addiction: From biology to drug policy.* Oxford University Press.

Gross, J. J., & Duckworth, A. L. (2021). Beyond willpower. Behavioral and Brain Sciences, 44.

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: an experience sampling study of desire, conflict, and self-control. *Journal of personality and social psychology*, 102(6), 1318.

Hofmann, W., & Kotabe, H. (2012). A general model of preventive and interventive self-control. *Social and Personality Psychology Compass*, *6*(10), 707-722.

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-

regulation. *Annual Review of Psychology*, *72*, 319-345.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in cognitive sciences*, 22(4), 337-349.

Jung, M. H., Gonzalez, F. J., & Critcher, C. R. (2020). The vicarious construal effect: Seeing and experiencing the world through different eyes. *Journal of Personality and Social Psychology*, *118*(4), 617.

Katzir, M., Baldwin, M., Werner, K. M., & Hofmann, W. (2021). Moving beyond inhibition: Capturing a broader scope of the self-control construct with the Self-Control Strategy Scale (SCSS). *Journal of Personality Assessment*, 1-44.

Kim, S., & Labroo, A. A. (2011). From inherent value to incentive value: When and why pointless effort enhances consumer preference. *Journal of Consumer Research,* 38(4), 712-742.

Kirmani, A. (2015). Neatly tied with a bow. *Journal of Consumer Psychology*, 2(25), 185-186.

Kristal, A.S, & Zlatev, J. J. (2021). Willpower is a form of, but not synonymous with, self-control. Commentary on Ainslie, George. "Willpower With and Without Effort." *Behavioral and Brain Sciences:* 1-81.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). *The effort heuristic. Journal of Experimental Social Psychology, 40(1), 91-98.*

Labroo, A. A., & Kim, S. (2009). The “instrumentality” heuristic: Why metacognitive difficulty is desirable during goal pursuit. *Psychological Science,* 20(1), 127-134.

Labroo, A. A., Lambotte, S., & Zhang, Y. (2009). The “name-ease” effect and its dual impact on importance judgments. *Psychological Science,* 20(12), 1516-1522.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust.

Organizational Behavior and Human Decision Processes, 126, 88-106.

Ma, F., Zeng, D., Xu, F., Compton, B. J., & Heyman, G. D. (2020). Delay of Gratification as Reputation Management. *Psychological Science*, 0956797620939940.

Marotta, V., & Acquisti, A. (2017). Online distractions, website blockers, and economic productivity: A randomized field experiment. *Preliminary Draft*.

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

McFall, L. (1987). Integrity. *Ethics,* 98(1), 5-20.

Milkman, K. L., Minson, J. A., & Volpp, K. G. (2014). Holding the Hunger Games hostage at the gym: An evaluation of temptation bundling. *Management science*, 60(2), 283-299.

Milkman, K. L., Rogers, T., & Bazerman, M. H. (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. *Perspectives on Psychological Science,* 3(4), 324-338.

Miller, M. J., Woehr, D. J., & Hudspeth, N. (2002). The meaning and measurement of work ethic: Construction and initial validation of a multidimensional inventory. *Journal of Vocational Behavior*, 60(3), 451–489

Mischel, H. N., & Mischel, W. (1987). The development of children's knowledge of self-control strategies. In *Motivation, intention, and volition* (pp. 321-336). Springer, Berlin, Heidelberg.

Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of personality and social psychology*, 54(4), 687.

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... & Sears, M. R. (2011). A gradient of childhood self-control predicts health, wealth, and Public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693-2698.

Mooijman, M., Meindl, P., Oyserman, D., Monterosso, J., Dehghani, M., Doris, J. M., & Graham, J. (2018). Resisting temptation for the good of the group: Binding moral values and the moralization of self-control. *Journal of Personality and Social Psychology*, 115(3), 585.

Olivola, C.Y. (2011). When noble means hinder noble ends: the benefits and costs of a preference for martyrdom in altruism. D.M. Oppenheimer, C.Y. Olivola (Eds.), *The Science of Giving: Experimental Approaches to the Study of Charity*, 49-62.

Olivola, C. Y., & Shafir, E. (2013). The martyrdom effect: When pain and effort increase prosocial contributions. *Journal of behavioral decision making*, 26(1), 91-105.

Pham, M. T. (2013). The seven sins of consumer psychology. *Journal of consumer psychology*, 23(4), 411-423.

Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of experimental social psychology*, 66, 29-38.

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of personality and social psychology*, *49*(1), 95.

Righetti, F., & Finkenauer, C. (2011). If you are able to control yourself, I will trust you: the role of perceived self-control in interpersonal trust. *Journal of Personality and Social Psychology*, 100(5), 874.

Schelling, T. C. (1984). Self-command in practice, in policy, and in a theory of rational choice. *The American Economic Review*, 74(2), 1-11.

Schrift, R. Y., Netzer, O., & Kivetz, R. (2011). Complicating choice. *Journal of Marketing Research,* 48(2), 308-326.

Schweitzer, M. E., Hershey, J. C., & Bradlow, E. T. (2006). Promises and lies: Restoring violated trust. *Organizational behavior and human decision processes,* 101(1), 1-19.

Smith, E. R. (2012). Editorial. *Journal of Personality and Social Psychology, 102*(1), 1–3

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of personality and social psychology*, *89*(6), 845.

Starmans, C., & Bloom, P. (2016). When the spirit is willing, but the flesh is weak: Developmental differences in judgments about inner moral conflict. *Psychological science*, 27(11), 1498-1506.

Strotz, R.H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23, 165–180.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality*, 72(2), 271-324.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, 112(S1), S164-S187.

Thomson, K., & Oppenheimer, D. (2023). The (effort) elephant in the room: What is effort, anyway?. *Perspectives on Psychological Science.*

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science,* 10(1), 72-81.

VanEpps, E. M., Downs, J. S., & Loewenstein, G. (2016). Advance ordering for healthier

eating? Field experiments on the relationship between the meal order–consumption time delay and meal content. *Journal of Marketing Research*, 53(3), 369-380.

Walker, L. J., & Hennig, K. H. (2004). Differing conceptions of moral exemplarity: just, brave, and caring. *Journal of personality and social psychology*, *86*(4), 629.

Watson, S. J., & Milfont, T. L. (2017). A short-term longitudinal examination of the associations between self-control, delay of gratification and temporal considerations. *Personality and Individual Differences,* 106, 57-60.

Wieselquist, J., Rusbult, C. E., Foster, C. A., & Agnew, C. R. (1999). Commitment, pro-relationship behavior, and trust in close relationships. *Journal of personality and social psychology,* 77(5), 942.

Zaheer, A., McEvily, B., & Perrone, V. (1998). Does trust matter? Exploring the effects of interorganizational and interpersonal trust on performance. *Organization science,* 9(2), 141-159.

Zlatev, J. J. (2019). I may not agree with you, but I trust you: Caring about social issues signals integrity. *Psychological science*, *30*(6), 880-892.