UNIVERSITY OF CALIFORNIA SAN DIEGO

Harm, Tradeoffs, and Behavioral Consequences in Social Choices

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Management

by

Kathryn Hillegass

Committee in charge:

Professor On Amir, Chair
Professor Ayelet Gneezy
Professor Wendy Liu
Professor Piotr Winkielman

2024

The Dissertation of Kathryn Hillegass is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

I dedicate this dissertation to James.

> *So, I love you because the entire universe conspired to help me find you.*
>
> ~ Paulo Coelho

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Words cannot express the depth of my gratitude to all the people and institutions that have supported me over the past several years.

First, I would like to thank my advisor, On Amir. From day one, On extended more trust and autonomy than I deserved. On, you provided the resources and wisdom to allow me to run with my ideas. You were my Occam's razor, finding simplicity when my tendency was to overcomplicate. Above all, you were always present and available for others. If I take anything from this program, I hope I can emulate your steadfast example of just being there for the people who need you. Thank you.

To my committee, thank you for teaching, guiding, and supporting me on this endeavor. I was lucky to take a class from each of you during my first year, which helped shape my research identity. Wendy, you taught me how to really *read* a paper, and how to develop new research ideas. Ayelet, you inspired me early on to pursue questions that have real impact on people's well-being. Piotr, you modeled a beautiful example of a person who engages in intellectual debate with the most serious of philosophers, without taking yourself too seriously. Thank you.

To the rest of the faculty and staff at the Rady School of Management and UCSD Psychology Department, thank you for entertaining my ten thousand questions. Thanks to Uma Karmarkar for giving me my first introduction to the lab and setting such a high bar for what it means to be a careful researcher. Thanks to Ania Jaroszewicz for your thoughtfulness and intentionality as a co-author and a human. Yuval Rottenstreich, thank you for the foundational grounding in Judgement and Decision Making concepts. Craig McKenzie, thank you for providing us the context and insight to understand the evolution of the field and the ongoing debates about how we form our preferences and judgments (although I still think dogs know

calculus). Thanks to the rest of the quantitative marketing, management and behavioral economics faculty for your investment in me as a researcher. Piotr, Tim Brady and Mike McCullough, thanks for improving my literacy in cognitive and social sciences and being so thoughtful in your curriculum development. Thanks to John Wixted for pushing us to go beyond using the right statistical test, but to understand statistics *in my bones.* Thanks to Lauren Eckhardt who has made invaluable improvements to the UCSD Behavioral Lab which has facilitated so much research and exploration.

To the many faculty outside of UCSD who helped shaped this journey, thank you. Rachel Gerson and Jackie Silverman, I am so grateful for the time and effort you have invested in me over the past three years. I am lucky to call you both co-authors and friends. To Leif Nelson, Don Moore, and Joe Simmons, thank you for hosting one of the greatest classes I have ever attended, for holding our field accountable and for teaching me to hold my own research to a higher standard.

To the *one sec* team, David Grüning, Frederik Riedel, Donatus Wolf, thank you for bringing such a cool product to market that has helped so many people curb their social media addiction. I am in awe of the scale of work you do with such a small team. Thank you for your commitment to science and humanity and for trusting us to run an experiment on your app.

To all the amazing PhD students at Rady, past and present, I am humbled to be your colleague, friend, and someday co-author. Everyone talks about how lonely the PhD journey can be. I never felt that in your company. Thanks for the many memories in the office and in my home. A special thanks to Giulia Maimone and Ariel Fridman for being my first introduction to Rady. You two have set a beautiful example of what it means to be a great researcher and an even better colleague and mentor. To my cohort mates, Brianna Chew and Carolina Raffaelli, we

hit the jackpot. I could not have picked better people to distract all day long across the cubicle. Michelle Kim, there is not a person in the world I would rather bounce ideas off than you.

Pursuing a PhD as a new mother presents its own set of challenges, and I could not be more grateful for the unconditional support of so many people who helped make this journey feasible. To the faculty, particularly my advisor, thank you for accepting me as an aspiring researcher when I showed up with a one year old and another on the way. To my classmates, particularly Michelle, Bri, Carolina, and Seung Kim, thank you for loving Frankie and Patrick and being our extended family in California. To my in-laws and parents, thank you for swooping in to graciously rescue us from every childcare crisis. Finally, to Marinella and Agata, the two women who deserve so much credit for my ability to thrive at UCSD, thank you for your unending support, love and patience as you helped us raise good humans. I love you all.

To the love of my life, James, we survived. Thank you for the continued joy you bring to my life and your eternal optimism about our future. Thank you for testing every survey and reading every draft. Thanks for inspiring me with new ideas and being my sounding board. Thank you for every dish you cleaned, every shirt you folded, and every Saturday morning you took the kids out of the house. All your continuous acts of service have not gone unnoticed, and I know I could not have completed this journey without your unwavering support. I love you.

Finally, I would like to thank the United States Army for affording me the time and financial support to pursue this degree. I am immensely grateful for this rare opportunity and know that I will cherish this time at UCSD as one of the unique highlights of my career. I hope to give back by using my research skills to help America's youth make informed judgements and decisions as they contemplate the call to serve.

Chapter 1, in full, is currently being prepared for submission for publication of the material with On Amir. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material with On Amir. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material with Silverman, Jackie; Gershon, Rachel; Grüning, David; Riedel, Frederik; Wolf, Donatus. The dissertation author was the primary researcher and author of this material.

VITA

2007 Bachelor of Science in Arabic and French, United States Military Academy

2016 Master of Arts in Security Studies, Georgetown University

2024 Doctor of Philosophy in Management, University of California San Diego

2007 – Present United States Army Officer

ABSTRACT OF THE DISSERTATION

Harm, Tradeoffs, and Behavioral Consequences in Social Choices

by

Kathryn Hillegass

Doctor of Philosophy in Management

University of California San Diego, 2024

Professor On Amir, Chair

This dissertation comprises three papers exploring the tradeoffs individuals navigate in social contexts, particularly when balancing personal benefits against moral, social, and self-regulatory concerns. Each chapter examines distinct contexts where individuals' decisions are

affected by social pressures, shedding light on how resources, risk perceptions, and behavioral feedback shape these tradeoffs.

Chapter 1 investigates why people do not advocate for their time, finding that individuals are less likely to complain when their time is violated, even if the harm is equivalent to monetary losses. Across multiple experiments, this chapter reveals that time, unlike money, has no defined rules, lacking an automatic moral trigger, making complaints about time seem petty and less justified. This reluctance underscores the need for employers and platforms to respect workers' time, as individuals may not advocate for it as strongly as they would for financial resources.

Chapter 2 examines motivated unrealistic optimism in risk-taking, showing that individuals perceive risks they impose on others as less severe than those they would accept for themselves. This phenomenon leads people to prioritize personal benefits over potential harm to others, driven by biased, motivated assessments of risk. These findings help explain selfish behavior in moral decision-making, with implications for policies and everyday decisions that require weighing private cost and benefits against the costs and benefits to others.

Chapter 3 assesses the influence of streak-based feedback in digital self-regulation, using a field experiment with an app designed to disrupt habitual social media use. Results show that while highlighting streaks motivates users to build positive patterns and break negative ones, it also increases social media engagement overall. This contributes to emerging work on maladaptive gamification, revealing how consumer preference for gamified feedback can lead to goal-inconsistent behavior.

Together, these chapters contribute to our understanding of decision-making tradeoffs in social contexts, illustrating how norms, perception, and feedback mechanisms influence behavior in ways that may harm our individual and collective well-being.

Chapter 1

STOLEN TIME: HOW RULES LEAD TO DIFFERENTIAL RESPONSES TO VIOLATIONS OF TIME VERSUS MONEY

Kathryn Hillegass & On Amir

Rady School of Management, University of California, San Diego, La Jolla, CA 92093, USA

ABSTRACT

Despite time being our most precious resource, it is psychologically puzzling that people do not seem to advocate for their time. Across seven experiments (N= 3699, Cloud Research and university lab), we find that people complain less when time is violated compared to money, even when we hold constant the financial harm, because while stealing money violates a rule, stealing time does not. Consequently, when people's money is stolen, they are relatively insensitive to the level of harm because breaking a rule is binary. However, when people's time is stolen, no rule is broken, so people are more likely to consider various factors such as the level of harm which predicts how petty it feels to complain. Given the widely accepted rules about stealing money and the absence of rules about stealing time, this research highlights the responsibility of employers and platforms to respect workers' time, given their likely reluctance to advocate for it themselves.

There is one kind of robber whom the law does not strike at, and who steals what is most precious to men: time.

~ Napoleon

Introduction

Modern parlance is so filled with adages about the preciousness of time, we might expect humans to guard this resource with their life. Instead, we often let others dictate how we spend our time, and we fail to speak up against its misuse. To illustrate this in concrete terms, consider Tina, who completes online surveys to supplement her income. Today, Tina sets a goal of making $20 in 2 hours. Tina picks up a gig advertised as $1 for 5-minutes, equivalent to $12/hour. However, the survey takes a full 10 minutes, effectively halving her hourly rate to $6/hour and derailing her goal. Would Tina complain?

We refer colloquially to Tina's experience as *stolen time*. Time theft - forcing others to lose their time without appropriate compensation - may seem inconsequential in isolation but has a disproportionate impact on people of lower socioeconomic status (Pierce et al. 2024). Given the decades of research focusing on how people (mis)value their time, there is a surprising gap in our understanding about why people are so resistant to advocating when their time preferences are violated. We explore how people react to theft framed in terms of time versus money in labor contexts where time's value is explicit. In Tina's example, if the terms were reversed – a 10-minute gig offered for $2 but which only paid $1– she'd be more inclined to complain, despite the net result being equivalent.

Although we frequently exchange our time for wages, we are not adept at accounting for time. The inability to replace or accumulate time means that people often don't track it as meticulously as they do with money, (Soman, 2001; Rajagopal & Rha, 2009), and are prone to duration neglect, inaccurately attending to its passing (Fredrickson and Kahneman, 1993). Moreover, as time does not provide monotonic utility and its value may be ambiguous, it is both difficult to process the valence of gains or losses (Loewenstein and Elster, 1992; Norton et al 2012) and easy to rationalize its loss because the opportunity costs are less straightforward (Okada & Hoch 2004). Time is ambiguous and mutable, and its mutability implies that it is relatively easy to come up with counterfactuals of how time could have unfolded differently (Kahneman and Miller, 1986), so we might set weaker expectations about time.

Time and money are also construed differently: time triggers feelings while money triggers thinking; time activates emotional concepts, while money activates economic utility goals (Liu and Aaker 2008; Vohs et al 2006); time makes people more susceptible to heuristics, whereas money makes people more analytical (Saini & Monga 2008). Consequently, people's evaluation of how much they deserved to be paid for time spent on a given task varies significantly with the elicitation method (Morvinski et al. 2023), and workers are willing to accept significantly different wage rates when thinking about fixed-time versus fixed-money gigs, reflecting how people treat time differently when they construe it as a resource unit compared to an affective unit (Smitizsky et al, 2021). However, little is known about how people react to violations of agreed-upon exchange rates, specifically, when framed in terms of time or money.

We focus on complaints as our key behavioral measure. Complaints play a crucial role in advocacy, information dissemination, and improving market quality, (Richins & Verhage, 1985; Kowalski, 1996), and yet frequently, people do not complain when they have a justification for doing so. Complaints serve as a mechanism for addressing and rectifying harms. To complain, first, the situation must fall short of a person's expectations (Kowalski, 1996) and second, the benefits of complaining (the likelihood the complaint achieves a desired outcome) must outweigh the costs (Singh & Wilkes, 1996; Kim et al, 2003).

We propose that one reason we do not complain about time is the lack of rules prohibiting its misuse. On the other hand, money has deeply engrained rules which lead to its automatic protection. Stealing money is a criminal offense, punishable by fines and jail time. Stealing is also a moral sin. The world's major religions command people not to steal, recommending a range of punishments from restitution to chopping off the hands of a thief. Children learn from an early age that stealing is bad, even if it is just a dollar from their parent's wallet. We all understand that stealing is wrong.

Rules are automatic, binary dividers of actions into right and wrong. If one stops to process the merits of the rule, it is not a rule after all (Raz 1975; Amir & Ariely 2007). Rules tend to be overgeneralized (Sunstein 2004), meaning that they are applied to situations that may be inappropriate (e.g. Jean Valjean in *Les Miserables*). Financial markets and monetary transactions are a domain dominated by rules because they elicit market pricing norms rather than social norms (Fiske 1992; Heyman & Ariely 2004; McGraw and Tetlock 2005; Amir & Ariely, 2007). Money triggers rules.

Stealing money violates a cardinal rule. Norm theory posits that when something clearly violates expectations, it is abnormal, and therefore rapidly and automatically elicits stronger affective responses. (Kahneman and Miller, 1986). When theft is presented in monetary terms, it activates the rule that stealing is wrong and therefore leads to an automatic response to rectify the violation. Whereas other mediums, like pens at the office or tokens in a game, are perceived as more flexible (Mazar et al., 2008). We posit that time violations trigger reflexive, social considerations because of the lack of rules about time.

What makes time unique is that, unlike pens, humans routinely exchange their time for money in the marketplace. In 2022, over 78 million workers, 55.6% of the U.S. workforce, were paid hourly (Bureau of Labor Statistics, 2022). While these workers are paid for their labor contributions, hourly workers' contractual pay is predominantly based on time and not output. If people are paid for their time in the marketplace, we might expect that people advocate for their time like they do their money because this should elicit market-pricing norms rather than social norms (Fiske 1992). However, we suspect that even in these market contexts, people will be unwilling to advocate for their time because it does not trigger the same rule-based response. If true, this is particularly alarming as society continues to shift toward a gig economy where workers have less protections and are twice as likely (26%) as W-2 hourly wage earners (11%) to make less than $10.00 per hour, while 14% of gig workers do not even make minimum wage (Economic Policy Institute, 2022).

We thus hypothesize that when shorting someone money (relative to forcing someone to work more time), people are more dissatisfied because it violates a clear existing rule and leads to more automatic response. When money is violated, people are relatively scope insensitive (Kahneman & Knetsch, 1992) and do not consider the severity of the harm or the costs of complaining. A rule is broken, so they complain. However, when time is stolen and no rule has been violated, people are reflexive, so they start to take other factors into consideration to decide if the benefits of complaining outweigh the costs.

One such factor that seems to come into play when considering complaining about time is the social costs of seeming negative (Kowalski, 1996). For example, time is often perceived as more personally controllable and discretionary than money (Aaker et al, 2011; Zauberman & Lynch 2005; Donnelly et al 2021), so complaining about time theft could signal a lack of control or competence (Whillans et al 2021). Additionally, complaining about time may also seem petty, violating a social norm that we should not overly attend to trivial details, like minutes and cents (Kim et al, 2019). While past works finds that both time and money can trigger pettiness concerns, Kim and colleagues do not directly compare the two resources. However, given that money primes reduce sensitivity to social rejection (Zhou et al, 2009), we hypothesize that complaining about time invokes more pettiness concerns than money.

In this paper, we explore how these rules-based difference between time and money lead to the exploitation of people's time (relative to money) even when the financial harm is held constant. In other words, when the theft is framed in terms of time, no rule has been broken, so people must elaborate on whether the benefits of complaining outweigh the costs. For example, when time is stolen, people consider morality and harm (Schein & Gray, 2018), which subsequently informs how likely they are to seek redress and how petty it feels to complain. Consistent with our rules-based theory, slowing down a person's automatic response to a monetary violation and forcing more reflection on the circumstances may make people less likely to complain about stolen money.

While we posit that the lack of rules about time is a driving reason why people do not advocate for their time, it is clearly not the sole factor. Across our experiments, we will attempt to control for other aforementioned differences between time and money (see page 4) to show that while those factors may contribute to differences in willingness to stand up for time, they alone do not explain the full effect. Across seven pre-registered experiments, we find that people complain less when time is violated compared to money. In experiments 1 and 2, we show that this is driven in part by differing sensitivity of noticing the violation and in part by different expectations between time and money. However, neither of these accounts explains the preponderance of difference in complaining between time and money.

In the remaining experiments, we focus our investigation on our rules-based theory. In experiment 3, we provide evidence that norms about money are more fixed than norms about time. In experiment 4, we show how violations of time trigger scope-sensitive evaluations of harm, while violations of money seem scope insensitive. In experiment 5, we induce a reflection on the relative harm of stealing time and find that it makes people reality more likely to complain about time and makes complaining about time seem less petty. In experiment 6, we show that the differences in complaining about violations of time are money are more driven by properties unique to money, rather than time. Finally, in experiment 7, we interrupt the rule-based automatic processing of monetary violations and find that this makes people more likely to consider other factors, thereby reducing the likelihood of complaining about money. Taken together, we provide evidence that people's relative unwillingness to advocate for their time has less to do with inherent perceptual properties of time, and more to do with the lack of monetary-like rules associated with its misuse.

Experiment 1:

We use a lab-in-the-field, incentive compatible design to test whether people complain more when working extra time without pay or when paid less than agreed. The field setting we use is Cloud Research, an interface which streamlines the recruitment and vetting of Amazon Mechanical Turk gig workers (MTurkers) who routinely complete surveys for compensation. We chose this venue because it is a natural place where people accept gigs based on a fixed (list price) and variable (bonuses) compensation with a specified time estimate attached to each gig.

When a survey runs longer than listed, this directly impacts the Mturker's ability to complete additionally surveys, lowering their hourly wage, thus leading to direct financial harm.

Method.

Participants. We recruited 201 participants through Cloud Research (M_{age} = 39.87, 50.75% females). We offered all participants the opportunity to participate in a follow-on survey for bonus money. 184 participants agreed to the follow-on survey, and those who agreed were then randomized into one of two violated resource conditions: *time* or *money*. This experiment was pre-registered on aspredicted.org (#148487) and received an IRB exemption (#808158). All pre-registrations, materials, data, and analysis scripts are available on our Research Box: https://researchbox.org/3014&PEER_REVIEW_passcode=IWCJBM.

Procedure. After completing a one-minute filler study for compensation, participants were given the opportunity to proceed to the end of the survey or to complete a follow-on survey for up to 10 minutes with a guaranteed bonus between \$1 and \$2 with a guaranteed rate of \$12/hour. Only those who agreed to participate in the follow-on survey were randomly assigned to treatment, to avoid any selection effects by condition. All participants were told they were chosen to take a survey where they would watch 5 advertisements and respond to a series of timed questions about the ads. In the *time* condition, participants were told the survey would take 5 minutes and they would receive a bonus of \$1 upon completion. In the *money* condition, participants were told the survey would take 10 minutes and they would receive a bonus of \$2 upon completion. Then all participants completed a timed filler survey on brand attitudes designed to take exactly ten minutes to mask the true purpose of the study. Upon completion, all participants were asked to indicate how long the survey took to complete on a sliding scale between 0 and 20 minutes. Then, all participants were told they would be bonused $1, which was in fact bonused to them within one hour. Twenty-four hours after completion of the survey, all participants were bonused an additional $1 to compensate them for the time or money stolen.

Measures. After answering how long the survey took and being told they would be bonused $1, we measured participants' satisfaction with the survey using a 5-point scale with faces ranging from a red angry face (1) to a green happy face (5). Then, we added a free-text comment bar to the end of the survey.

Our primary dependent variable of interest was a binary coding of whether they complained about the violation of their time or money in the comment box. Per our pre-registration, we coded any complaint about the survey being longer than promised or being paid less than promised as a complaint (1) and any other comment (e.g. a positive comment about the filler survey) or a blank response as a non-complaint (0). As a confirmation, we had Chat-GPT4 code each response. [1] Where our classification mismatched with Chat-GPT4, we had a research assistant blind to our hypotheses adjudicate.

Results.

As predicted, using a logistic regression model (GLM) covariate free analysis, we found that people complained less in the *time* condition (M=0.045, SD=0.21) compared to *money* (M=0.25, SD=0.43), B = -1.95, SE = .56, z(182) = -3.45, $p<0.001$, 95% CI [.-3.21, -0.94], d

[1] We modified the chat-GPT4 prompt from our pre-registration to give it more contextual clues to distinguish between complaints about time being stolen (e.g. taking longer than promised) and complaints about the pace of the survey (e.g. wishing they had more time to be thoughtful). The prompt was as follows: "*Please review all of the comments in column B. In column C, please classify the comments as 1 if the comment is complaining about not getting paid enough or the survey taking too long (e.g. I was supposed to be paid $2 instead of $1 or this was only supposed to be 5 minutes but was way longer or It said $2. or $1, really? or This survey was too long). Please classify as a 0 if it is blank, the comment is positive (e.g. thanks great survey) or if the comment is about wishing they had more time to complete it (e.g. I wish I had more time to answer or the survey was very fast).*" See Research Box for full data and methods.

=0.59. Unsurprisingly, we also found an inverse relationship of satisfaction on complaints, $B = -1.10$, $SE=0.21$, $z(182) = -5.20$, $p<0.001$, 95% CI [-1.55, -0.71], meaning that the more dissatisfied someone was, the more they were likely to complain. However, using an ordinary least squares (OLS) regression, we found no significant direct effect of condition on people's levels of satisfaction, $B=0.06$, $SE=.16$, $t(182) = 0.41$, $p=0.69$, 95% CI: [-0.25, 0.38]. Likewise, when we added satisfaction as a covariate to our main model, we find that people are still much less likely to complain when in the *time* condition, $B=-2.64$, $SE=0.72$, $z(181) = -3.69$, $p<.001$, 95% CI [-4.23, -3.69]. Age and gender had no statistically significant effects on either measure.

We also find evidence that people are less likely to notice working longer (compared to being paid less) and more likely to accept what we told them about how long the survey would take, regardless of their actual experience. Although the actual minutes spent on the survey did not statistically vary between the money ($M=11.43$, $SD=5.35$) and time ($M=11.13$, $SD= 2.36$) conditions, $t(133.05)=0.49$, $p=0.63$, participants in the stolen time condition ($M=6.35$, $SD=2.16$) reported spending more than three minutes less on the survey than participants in the stolen money condition ($M=9.52$, $SD=1.30$), $t(150.25)=11.95$, $p<0.001$. Using an OLS regression, participants in the *time* condition reported that the total time they spent on the survey was three minutes less than participants in the *money* condition, $B=-3.17$, $SE=0.26$, $t(182) = -12.20$, $p <0.001$, 95% CI [-3.68, -2.66], $d= 1.80$. When we control for reported time spent on the survey in an OLS regression predicting complaining, the direct effect of condition on complaints is still significant, $B=-1.36$, $SE =0.66$, $z(181) = -2.073$, $p=0.038$, 95% CI [-2.80, -0.18], suggesting that noticing the violation is not the sole reason why people do not complain about time.

Discussion.

In a real-world setting where people routinely conduct timed surveys for money, less than 5% of our participants complained when the survey went twice as long and their effective wage rate was cut in half. These workers, like Tina, complain less when the survey goes twice as long, than when paid half as much. We also find evidence that this unwillingness to complain about time is partially driven by people not attending to how much time has passed, but that perceived time is not the sole factor.

Across a series of studies that vary in magnitude, context, and controls (see web appendix), we consistently replicate a large effect size that people are less likely to complain about working longer without compensation compared to being paid less than agreed, using a simple weighted average based on sample size, D=.93 95% CI: [.74, 1.13]. See Figure 1.1.

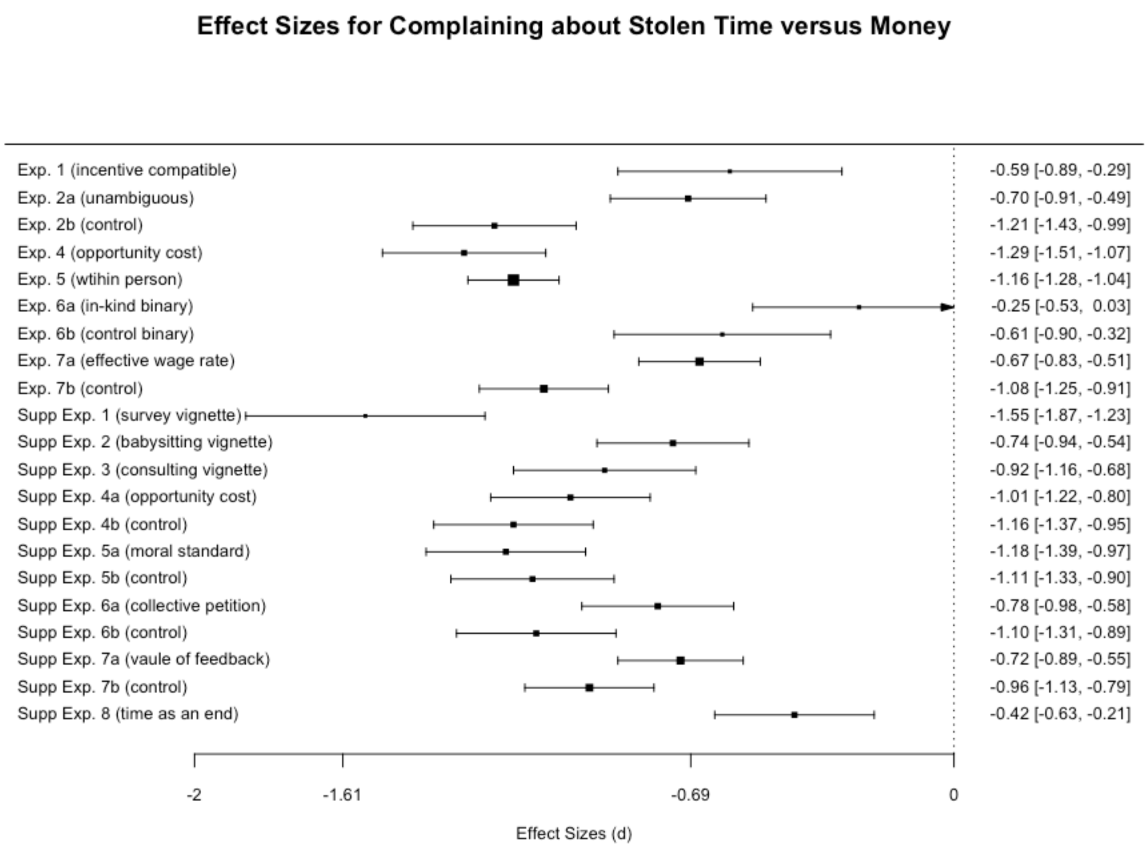

Figure 1.1. Effect sizes with 95% confidence intervals for complaining about time versus money.

Experiment 2:

In this experiment, we conceptually replicate the results of our incentive compatible study using a hypothetical vignette, and we attempt to rule out the fuzziness or mutability of time as the sole explanation for why people respond differently to contract violations of time and money. One might suspect that the reason people do not advocate for their time is that when they enter a contract, money feels like a promise within the control of the employer, while time seems like a more ambiguous estimate outside the control of the employer. Therefore, people set wider expectations about how long something will take, and so going over feels like less of a contract violation. In this experiment, we manipulate both the ambiguity of time and whether people set their expectations about time and money ex ante.

Method.

Participants. We recruited 800 participants through Cloud Research (M_{age} = 42.57, 48.5% females) for a 2 resource violated (time versus money) x 2 ambiguity of time description (ambiguous versus unambiguous) x 2 expectations (explicitly set vs control) between subject design. We pre-registered that we would collapse the expectation setting condition in our main analysis if explicit expectation setting or its interaction with resource stolen did not significantly predict people's willingness to complain. Fifteen participants (1.88%) failed the attention check. This experiment was pre-registered on aspredicted.org (#169498)

Procedure. In the *time* condition, participants imagined they agreed to participate in a survey where they would earn $1 for five minutes of their time. In the *money* condition, participants imagined they agreed to participate in a survey where they would earn $2 for ten minutes of their time. Both wage rates are equivalent of $12/hour.

In the ambiguous time description, we gave no more amplifying information about what participants were expected to do in the study. Whereas in the unambiguous time description condition, participants were told that they would watch five ads on an auto-advance time so they would not be able to click through at their own pace. In the explicit expectation setting condition, participants rated how long they expected the task to take and how much they expected to be paid (point estimate, minimum and maximum).

Then, all participants were told they had completed the survey exactly as instructed, the survey took a full ten minutes to complete, and they were paid $1. Therefore, all participants were effectively paid an equivalent rate of $6/hour for the survey, half the rate they agreed to work for.

Measures. Participants responded to a binary measure of whether they would complain., 1 = "*I would complain to the surveyor*" or 2 = "*I would NOT complain to the surveyor*." All participants reported their age and gender.

Results.

First, we replicated our primary hypothesis that participants are much less likely to complain when forced to work longer than agreed to without pay than when paid less than agreed. Participants were 6.18 times more likely, in terms of odds, to complain about in the *money* condition (M=0.68, SD=0.47) than *time* (M=0.26, SD=0.44) using a binary GLM estimating the main effect of resource stolen, B = 1.82, SE =0.16, 95% CI [1.51, 2.13], z = 11.48, $p < 0.001$.

To better understand how people's expectations may differ between time and money, we subset our data to include only participants in the expectation setting condition. Remember, in the *time* condition (n=194), participants read they agreed to a five minute survey for \$1, while in the *money* condition (n=194), participants agreed to a ten minute survey for \$2, so collapsing across resource conditions, we compare the mean expectation for time to 7.5 minutes and the mean expectation for pay to \$1.5. Using a one-sample t-test, participants expectations about how much they will be paid (M=\$1.53, SD=.77) was not statistically different than \$1.5, $t(387)=0.85$, $p=0.40$. However, participants expectations about how long the survey will take (M=8.31, SD=4.44) was greater than 7.5 minutes, $t(387)=3.58$, $p<0.001$. These inflated expectations about how long the survey would take were relatively higher for participants who expected to take a 5 minute survey (6.44 minutes which is 28.86% higher than the advertised time, $t(193)=5.40$, $p<.001$) compared to those who expected to take a 10 minute survey (10.17 minutes which is not statistically different than the advertised time , $t(193)=0.55$, $p=0.58$).

Given these different expectations between time and money, we find differences in whether the violation fell outside of the expectations based on resource condition. See Figure 1.2 In the *money* condition, the actual pay was lower than the reported expectations for 88% of participants (M=0.88, SD=0.33). Meanwhile, in the *time* condition, the actual time spent on the survey was higher than the reported expectations for only 67% of participants (M=0.67, SD=0.47).

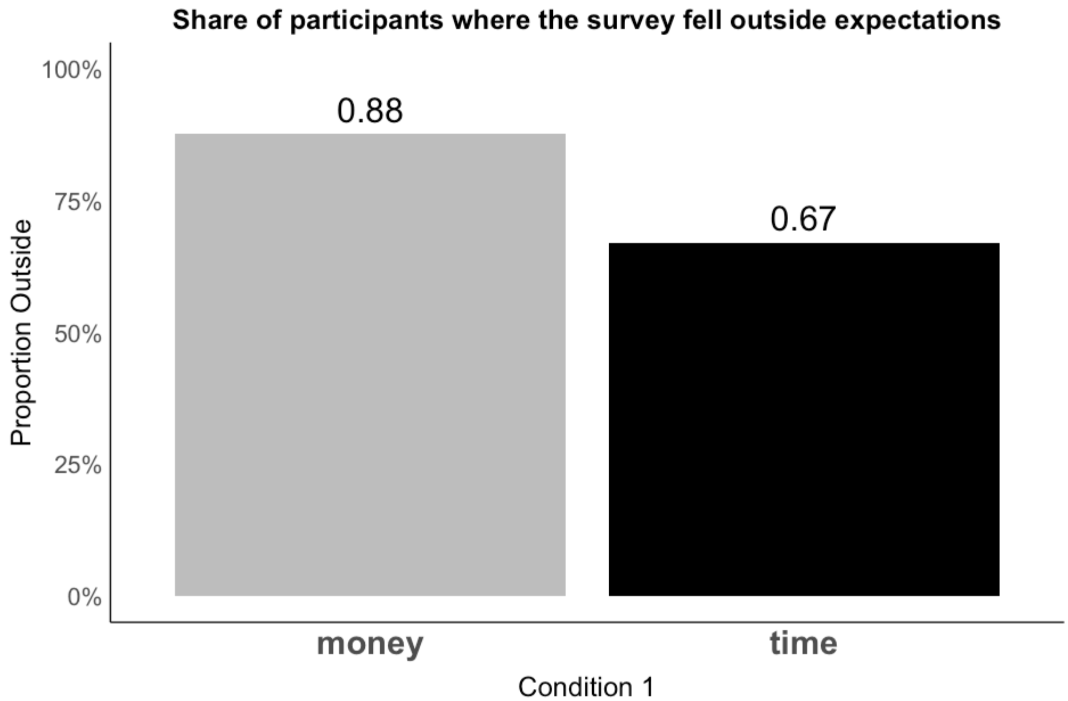

Figure 1.2. Share of participants where the survey fell outside their apriori expectations.

Adding a binary variable for whether the survey fell outside of expectations, we include this as a covariate to our base model of complaining to see if falling outside of expectations mediates the effect of resource on complaining. A mediation analysis reveals the average causal mediation effect (ACME) was $B=-0.061$, 95% CI [-0.10, -0.03], $p < .001$. The average direct effect (ADE) was $B=-0.3743$, 95% CI [-0.47, -0.29], $p < .001$. The total effect of resource condition on complaining was $B=-0.44$, 95% CI [-0.53, -0.35], $p < .001$. Therefore, approximately 14% of the total effect of resource on complaining was mediated by the survey falling outside of expectations, suggesting that even when controlling for expectations being violated, most of the direct effect remains. Additionally, when we added the expectation setting condition as a predictor to our model of complaining, there was no direct effect of expectation setting ($z= -1.45$, $p = .15$), nor an interaction with resource stolen ($z= .53$, $p = .60$). Therefore, consistent with our prediction and pre-registration, we collapsed across expectation setting and control conditions for all further analysis.

Making the time description unambiguous partially moderates the differences in complaining between *time* and *money*. Including the ambiguity of time condition as a predictor to our base model, we find that reducing the ambiguity of time has a direct effect which makes participants complain less, B = -.46, SE = .22, z= -2.07, p = .039, 95% CI [-.89, -.03]. However, interacting the description with the *time* condition, participants complain more when the time description is unambiguous, B = .91, SE=.32, z= 2.85, p=.004, 95% CI [.29, 1.54]. Nevertheless, the difference in complaining between the two resources was still large (B = -2.29, SE =0.24, z = -9.70, p <.001, 95% CI [-2.77, -1.84]), suggesting that the ambiguity of time moderates only 13.63% of the effect and does not fully explain why people do not stand up for their time. See Figure 1.3.

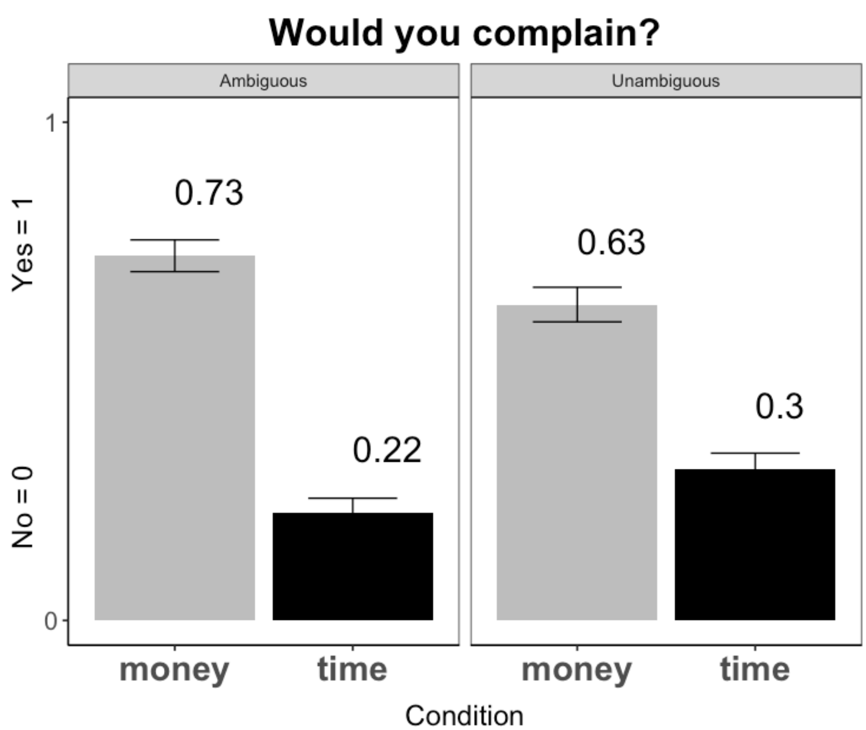

Figure 1.3. Proportion of participants who complained based on resource and ambiguity of time.

Discussion.

In experiment 1, we find that the unwillingness to complain about time is partially driven by people not attending to time. In experiment 2, a hypothetical study in which we control for differences in attention to time, we find that unwillingness to complain about time is partially explained by the ambiguity of time. For one, people set wider expectations about time than money, so violations of time are less likely to fall outside a person's expectations. Additionally, when we eliminate the ambiguity of time by attributing the reason the survey takes longer directly to survey design, people are slightly more likely to complain. While people's unwillingness to advocate for their time as they would their money is clearly a multiply determined phenomena, neither lack of attention to time, differing expectations about time and money, nor the ambiguity of time fully explains this effect. In the studies that follow, we seek to move beyond differences in the perception of time, and instead focus on our alternative explanation of money triggering rules, while time does not.

Experiment 3.

In this experiment, we sought to better understand norms about time and money, specifically to test whether descriptive (what people do) and/or injunctive norms (what people ought to do) about stealing or about complaining were stronger for time than money. Given the difference in ambiguity between time and money observed in experiment 2, we tested norms about stealing money in a context that was also more ambiguous: bonus pay rather than base pay. We thought adding ambiguity about money and increasing the plausibility of it being stolen on a platform like Mturk would be a more conservative test of the differences between time and money.

Method.

Participants. We recruited 103 participants (M_{age} = 37.89, 48.54% female) for a 2 (resource: time vs. money) within person design and counterbalanced the order in which participants responded to questions about each resource. Three participants failed the attention check. This experiment was pre-registered on aspredicted.org (#188,143).

Procedure. All participants first responded to a block of questions about norms related to either time or money. Each time or money block began with two injunctive normative statements about the acceptability of stealing time or money and then the acceptability of complaining about time or money theft. After the injunctive norm questions, participants read about a scenario, and then answered a descriptive normative question about how many people would steal time or money followed by a descriptive normative question about how many people would complain about the stolen time. After the first block about one resource, participants would respond to the second block about the alternate resource.

Measures. For the block of questions about stealing time, participants rated their agreement with the following statements on a scale of 1 to 7: "Most surveyors think it is acceptable to design their surveys to take longer than advertised" (injunctive norm about stealing time) and "Most survey participants think it's acceptable to complain about surveys taking longer than advertised" (injunctive norm about complaining about stolen time).

Then participants read, " Imagine that you saw a survey posted for $1 for 5 minutes with no advertised bonus." Participants responded on a scale of 0 to 100, "Out of every 100 surveys advertised like this, how many do you think are designed to actually take ten minutes or more?" (descriptive norm about stealing time). Then participants read, "Now imagine that you accepted this gig and the survey took five minutes longer than advertised and there was no bonus paid." Participants responded, "Out of every 100 survey participants, how many would complain about the survey taking longer than advertised?" (descriptive norm about responding to stolen time).

For the block of questions about stealing money, participants rated their agreement with the following statements on a scale of 1 to 7: "Most surveyors think it's acceptable to not pay the bonuses they advertise" (injunctive norm about stealing money) and "Most survey participants think it's acceptable to complain about not receiving an advertised bonus" (injunctive norm about complaining about stolen money).

Then participants read, "Imagine that you saw a survey posted for $1 for 10 minutes with a guaranteed bonus of $1 upon completion." Participants responded on a scale of 0 to 100, "Out of every 100 surveys advertised like this, how many do you think are designed to not pay the $1 bonus? (descriptive norm about stealing money). Then participants read, "Now imagine that you accepted this gig for $1 for 10 minutes with a guaranteed bonus of $1 upon completion and the bonus was never paid." Participants responded, "Out of every 100 survey participants, how many would complain about not receiving a bonus?" (descriptive norm about responding to stolen money).

Results.

In terms of injunctive norms, participants thought it was equally acceptable to complain about violations of time ($M=4.95$, $SD=1.55$) and money ($M=4.96$, $SD=1.92$), $t(198)=0.04$, p=0.97. However, they thought it was much less acceptable to steal money (M=2.85, SD=1.65) than time (M= 4.01, SD=1.65), t(198)=4.97, p< 0.001. See Figure 1.4.

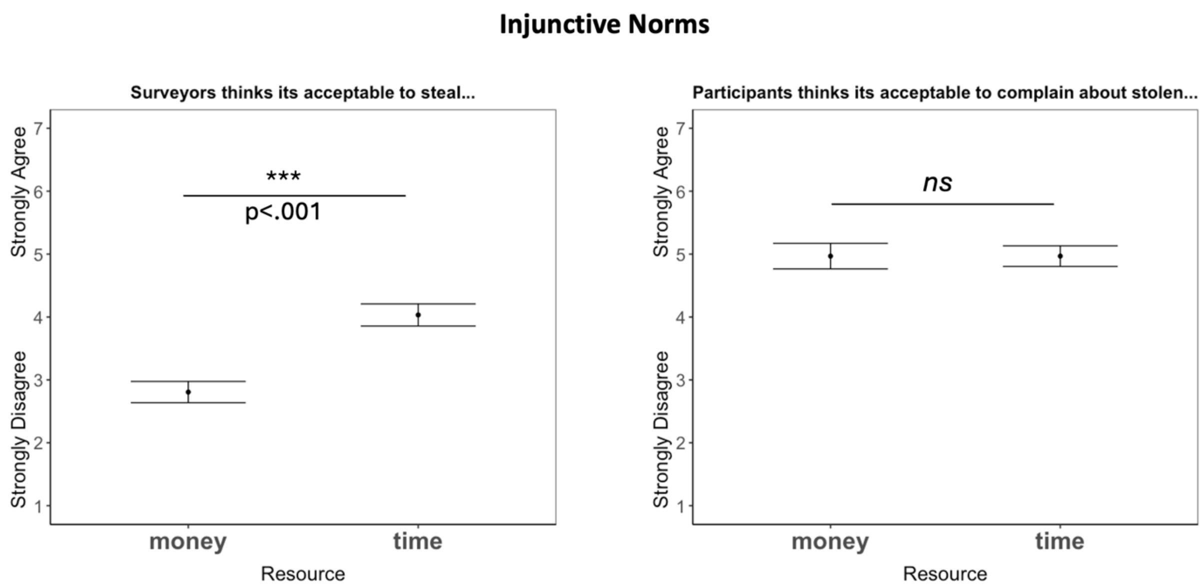

Figure 1.4. Injunctive norms about the acceptability of stealing and complaining about stolen time versus money.

In terms of descriptive norms, participants estimated that more people would intentionally design a study to steal time (M=32.6, M=24.1) versus money (M=25.0, SD=28.0), t(198)=2.06, p=0.040. However, participants estimated that more people would complain about money (M=59.9, SD=29.8) than time (M=50.8, SD=25.7), t(198)=2.32, p=0.021. See Figure 1.5.

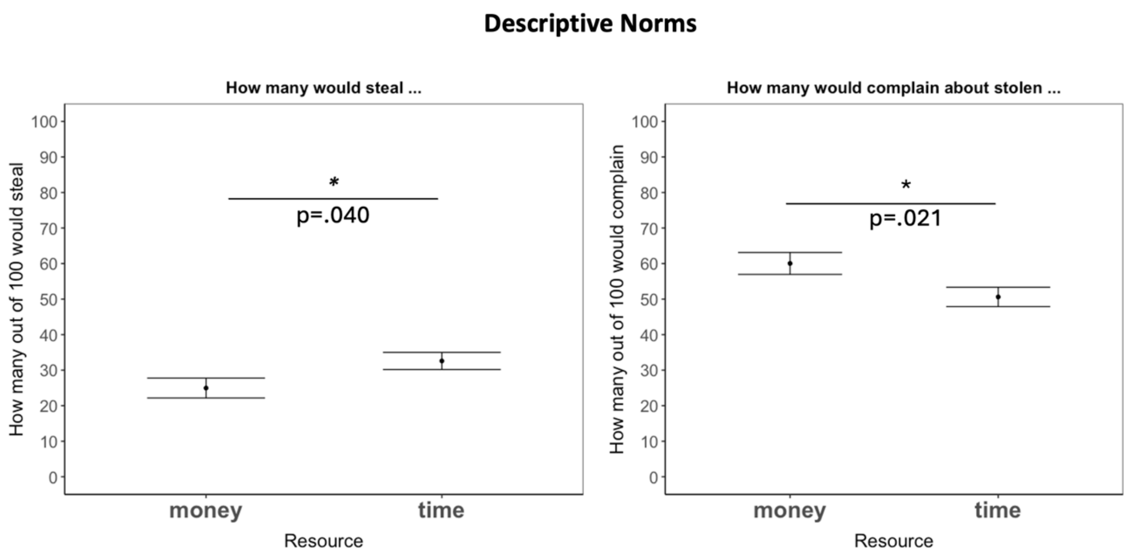

Figure 1.5. Descriptive norms about how many people would steal or complain about stolen time versus money

To understand whether norms are more fixed for time or money, we analyze the relationship between norms about stealing and norms about complaining. Using a Pearson correlation, we find no significant correlation between the acceptability of stealing money and the acceptability of complaining about money, r(99)=-.054, p=0.59. Meanwhile, we find that a significant positive correlation between the acceptability of stealing time and the acceptability of complaining about stolen time, r(99)=.21, p=0.032. Using a Fisher's z-test to compare the two correlation coefficients, we find that the two resource coefficients are marginally different from one another, z=1.91, p=0.06. See Figure 1.6.

In terms of descriptive norms, once again using a Pearson correlation, we find no significant correlation between the percentage of people who would steal money and the percentage who would complain about it, r(98)=0.11, p=0.29. Meanwhile, we find that a significant positive correlation between the percentage of people who would steal time and the percentage who would complain about stolen time, r(98)=.31, p=0.0016. Using a Fisher's z-test to compare the two correlation coefficients, we find that the two resource coefficients are not statistically different from one another, z=1.52, p=0.13.

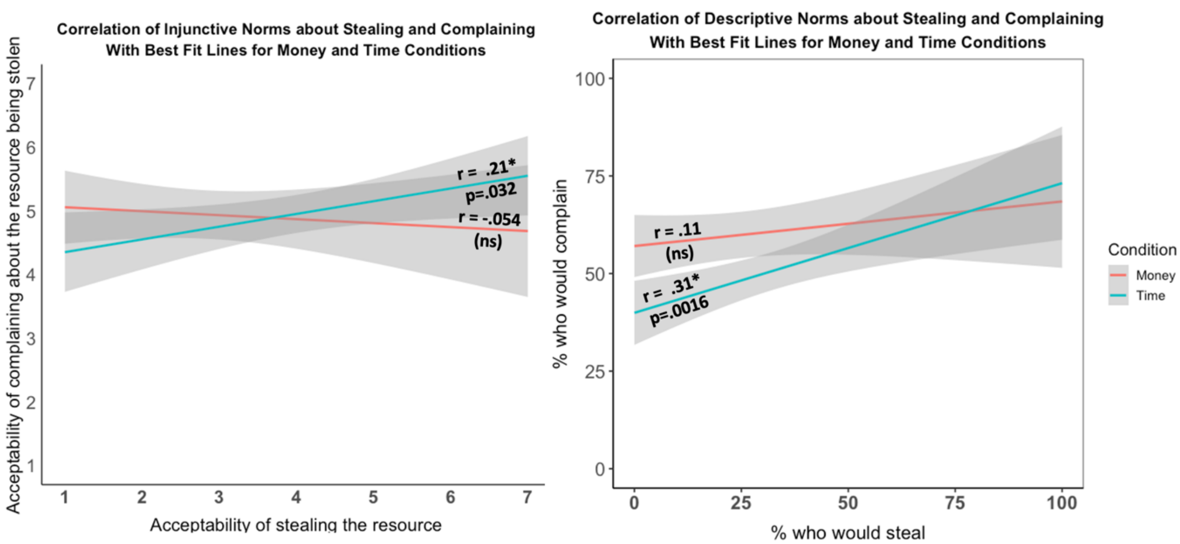

Figure 1.6. Correlation between norms about stealing and complaining by resource.

Discussion.

People seem to have a moral ambivalence toward stealing time. Participants report that it is clearly unacceptable to steal money but that it is neither unacceptable nor acceptable to steal money. However, they report that it is equally acceptable to complain about time or money. Meanwhile, people report that is more common to steal time, and less common to complain about it.

We interpret these results to support our rules-based account for money, and lack of rules for time. Both injunctive and descriptive norms about complaining about money are independent of the acceptability and frequency of people stealing. People are not considering how acceptable or commonplace it is to steal before the decide to complain. Instead, in the wake of stolen money, people automatically complain. This suggests a rules-based approach to handling monetary theft. On the other hand, people's inferences about the acceptability of stealing time and the frequency of complaining about it are positively correlated with how often people steal time and how acceptable it is. This suggests that people are considering these factors and weighing evidence when deciding if they should complain about time. This indicates a non-rules-based, reflexive approach to handing infractions of time.

Experiment 4:

Experiment 4 builds on the theory that money triggers rules while time does not. We explore whether people's willingness to complain about money which is driven by a binary right and wrong, whereas people's willingness to complain about time is driven more by other considerations. We suspected that the perception of harm and the social costs of seeming petty would impact people's willingness to complain about time. Additionally, given past work which shows that the opportunity costs of time are less straightforward (Okada & Hoch 2004), we attempt to control for that in this experiment by making salient the opportunity costs of stealing time by highlighting a follow-on paid task foregone.

Method.

Participants. We recruited 400 participants through Cloud Research (M_{age} = 40.73, 52.25% females) for a 2 resource violated (*time* versus *money*) between subject design. Two participants (0.5%) failed the attention check. This experiment was pre-registered on aspredicted.org (#174832).

Procedure. All participants read that they had ten minutes before their next engagement and were planning to use that time to earn some money on an online survey platform. In the *time* condition, they found two five minute surveys that each pay $1, while in the *money* condition, they found one ten minute survey paying $2 (both equivalent to $12/hour).

Then, all participants were told they had completed the (first) survey exactly as instructed, the survey took a full ten minutes to complete, and they were paid $1. In the *time* condition, participants were reminded that because the first survey ran long, they were unable to do the second survey. Once again, all participants were effectively paid an equivalent rate of $6/hour for the survey, half the rate they agreed to work for.

Measures. Participants responded to a binary measure of whether they would complain., 1 = "*I would complain to the surveyor*" or 2 = "*I would NOT complain to the surveyor*." Participants also reported on a seven point scale to what extent they thought they were harmed and to what extent it would be petty to complain. We randomized whether these two questions were asked before or after our primary measure of complaining and planned to collapse across order if there was no main effect of order. All participants reported their age and gender.

Results.

First, we replicated our primary hypothesis that participants are much less likely to complain when forced to work longer than when paid less than agreed, even when we make explicit the monetary opportunity cost of working longer. Participants complained more in the *money* condition (M=.77, SD=.42) than *time* (M=.23, SD=.42), B = -2.43, SE =0.24, z(396) =-10.19, p <.001, 95% CI [-2.91, -1.97].

When we include pettiness and harm in our model, these predictors partially mediate the effect of resource violated on complaining, B = -1.90, SE =0.27, z(394) = -7.10, p <.001, 95% CI [-2.44, -1.38]. People are less likely to complain the pettier it is (B = -.35, SE =0.07, z(394)= -4.62, p <.001, 95% CI [-.50, -.20]) and more likely to complain the more harmful they think it is (B = .52, SE =0.08, z(394)= 6.75, p <.001, 95% CI [.37, .67]). The distribution of responses to how harmful it is to pay someone less or make someone work longer and how petty it would be to complain about these respective infractions suggest that while it is overall perceived as relatively more harmful to pay people less (M = 4.42, SD = 1.85) compared to making them work longer (M= 3.05, SD = 1.84), there is a stronger norm that it is not considered petty at all to complain about monetary violations (M= 2.38, SD = 1.79) compared to time (3.77, SD = 1.96). See Figure 1.7.

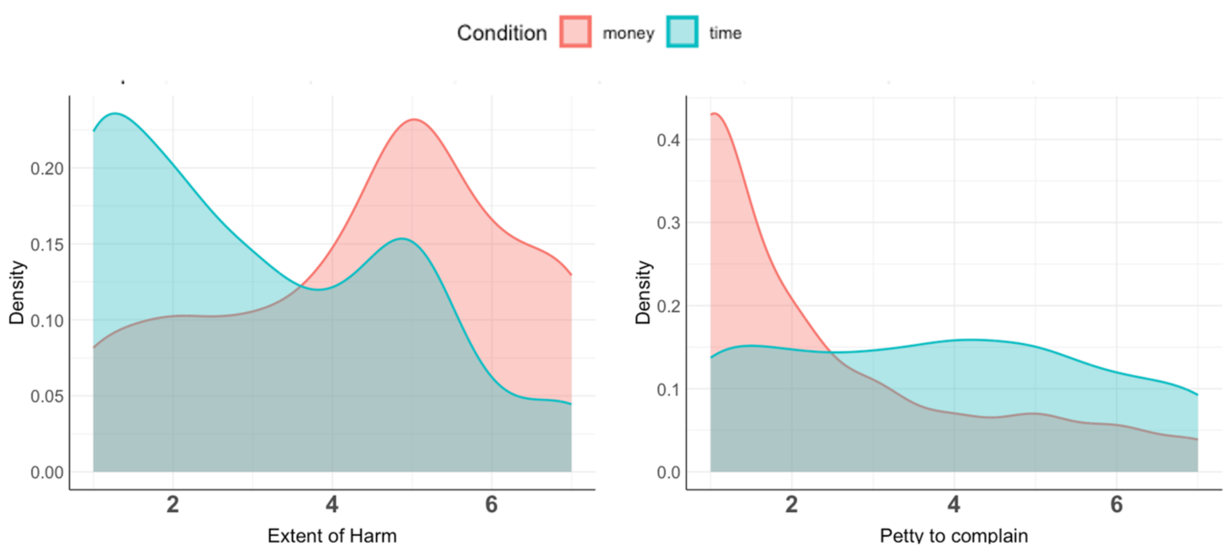

Figure 1.7. Subjective ratings of harm and pettiness in Experiment 3.

Exploring the relationship between the perceived harm of stealing and the perceived pettiness of complaining in a LM regressing pettiness on harm and its interaction with condition, we find that when facing stolen money, perceived harm does not predict how petty it seems to complain, B=-0.01, SE=0.07, 95% CI [-0.15, 0.13], t(394)=-0.14, p=0.88. However, when facing stolen time, perceived harm marginally predicts the perceived pettiness of complaining, B=-0.18, SE=0.10, 95% CI [-0.38, 0.02], t(394)=-1.79, p=0.074. See Figure 1.8.

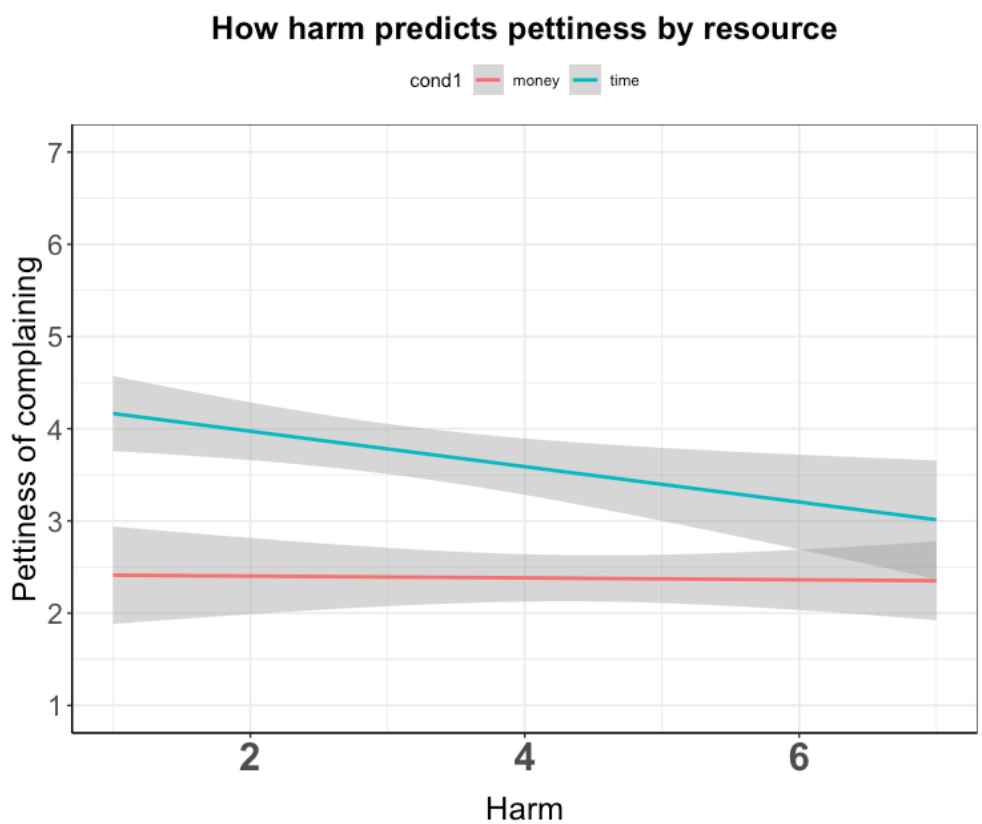

Figure 1.8. Harm predicts pettiness of complaining between time and money.

Discussion.

Given that pettiness is defined as attention to trivial details (Kim et al, 2019), we would expect the less harmful something is, the pettier it would seem to complain. However, in the *money* condition, regardless of how harmful it seemed, it was not petty to complain. This indicates a binary perception of right and wrong, with no regard for the perceived social cost of complaining. However, in the *time* condition, the level of harm predicts perceived pettiness, suggesting a non-binary, reflective consideration of how harmful the stealing was and the subsequent social costs of complaining.

Experiment 5.

In this experiment, we build on the different relationship between harm and pettiness across resources revealed in Experiment 4. If objective harm is held constant between our time and money conditions, then arguably whether it is money or time stolen is just a matter of framing. Framing effect arises from these immediate, automatic normative perceptions (Tversky & Kahneman, 1981; Kahneman and Miller, 1986). Framing effects are known to diminish when individuals are more cognitively engaged with a problem (Stanovich & West, 2008). Similarly, bounded ethicality suggests that people often overlook unethical behaviors that they would condemn upon deeper reflection (Chugh et al., 2005; Kern and Chugh, 2009). Consequently, encouraging people to compare the repercussions of time and monetary theft might heighten the perceived harm of *stealing time*, thereby reducing pettiness perceptions and leading people to be more likely to stand up for their time. We hypothesize that having participants reflect on the similar impacts of making people work longer without pay compared to paying less would make the time violation seem more harmful and therefore less petty. Here, we test whether a deliberate reflection increases the perception of the relative harm of violating people's time, and therefore reduces the perceived pettiness of complaining about time.

Method.

Participants. We recruited 599 participants through Cloud Research (M_{age} = 40.30, 52.25% female) and used all responses in our analysis. Participants were randomly assigned to one of two between subject conditions (reflection of harm vs control). This experiment was pre-registered on aspredicted.org (#123185).

Procedure. This study employed a mixed factorial design with two between-subject conditions (reflection of harm vs control) and two within-subject (*time* vs *money*). Instead of using a first-person vignette, we asked participants to imagine two people who routinely do surveys through an online platform as a supplemental income source. Participants read about Taylor who agreed to a five minute survey for $1, but which took ten minutes to complete by design and was still paid just $1. Then participants rated how likely Taylor would be to complain on a 7-point scale from not likely at all to extremely likely. Next, participants read about Riley who agreed to a ten minute survey for $2, which took ten minutes to complete by design, but was paid just $1. Then participants rated how likely Riley would be to complain. We counterbalanced the order of the vignettes, so half the participants saw the *time* vignette first, and half saw *money* first. After reading both vignettes and indicating the likelihood of complaining for both workers, participants in the reflection of harm condition were asked to describe in an open text response in what way these two scenarios were similar in terms of exploitation of workers.

Measures. Participants rated who was harmed more in the scenarios on a bipolar unnumbered scale recoded as 1 (Taylor was harmed more) to 7 (Riley was harmed more) with a specified midpoint at 4 (They were equally harmed). Next, participants rated who should be more likely to complain on a similar bipolar unnumbered scale from 1 (Taylor) to 7 (Riley) with a midpoint 4 (They should be equally likely to complain). Then, participants rated how petty they would feel complaining if they were Taylor (i.e., about *time*), and then how petty they would feel complaining if they were Riley (i.e., about *money*) on 7-point scales from 1 (not petty at all) to 7 (extremely petty). Once again, we counterbalanced the order of these questions. Finally, participants reported their age and gender.

Results.

First, we confirm that the relative unwillingness to complain about time replicates in a within-subject design where participants see both scenarios sequentially, regardless of order. Using a linear model to compare likelihood of complaining based on the resource violated and the order in which participants read the two vignettes, we find that people indicate that the worker who worked longer is much less likely to complain, $B = -2.02$, $SE = .10$, $t(1195) = -20.04$, $p < .001$, 95% CI [-2.22, -1.81], and there were no significant order effects, $B = .06$, $SE = 10$, $t(1195) = .57$, $p = .57$, 95% CI [-.14, .26].

Then we test whether deliberately reflecting on how these two scenarios are similar affects participants' perceptions of the relative harm of each violation. Using a covariate-free linear regression to predict the relative harm of the two vignettes based on the reflection condition, we find that the deliberate reflection does make people think that making someone work more without pay is relatively more harmful compared to participants in the control, $B = -.40$, $SE = .11$, $t(597) = -3.79$, $p < .001$, 95% CI [-.61, -.19]. See Figure 1.9.

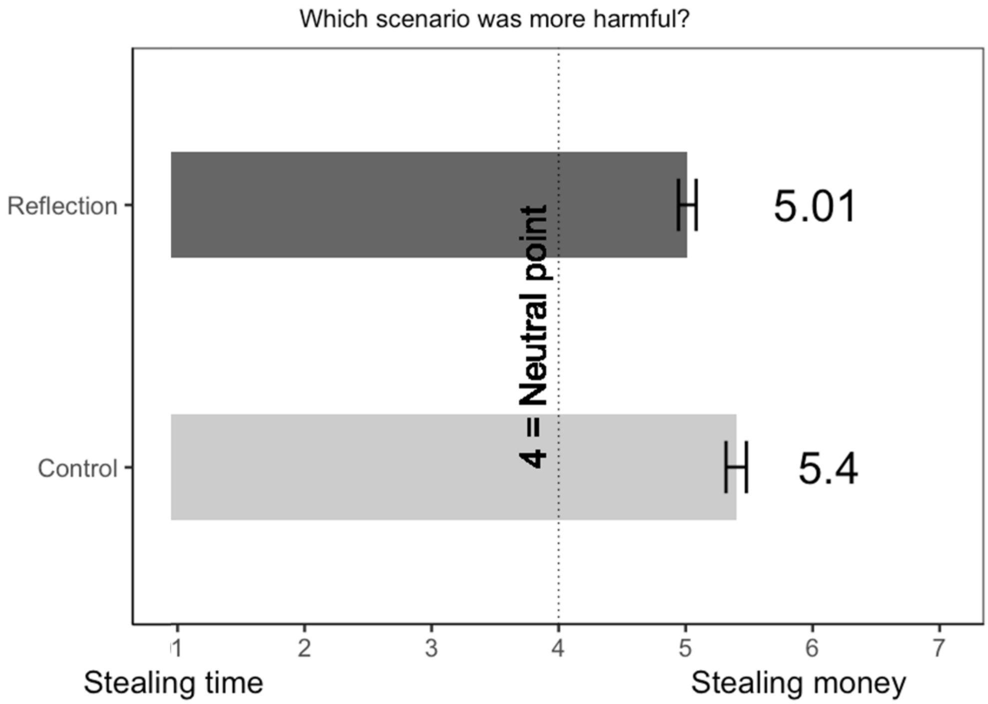

Figure 1.9. Relative harm based on the reflection manipulation.

Using a GLM regression to predict which worker should be more likely to complain based on the reflection condition and the relative harm, we find that both the reflection manipulation itself, B = .16, SE = .08, t(596)=2.02, p =0.04, 95% CI [.0044, .31], and the relative harm of the two violations, B = .76, SE= .03, t(596) = 25.59, p<.001, 95% CI [.70, .82], make participants more likely to think that the worker whose time was stolen should be relatively more likely to complain.

We find once again that relative harm is not predictive of pettiness when considering complaining about stolen money, r(600)=-0.0013, p=0.97, but highly predictive of pettiness when considering complaining about stolen time, r(600)=0.34, p<.001. See Figure 1.10. We also find marginally significant evidence in a linear model that the harm reflection itself has a direct effect on making complaining about time feel less petty, B = -.30, SE = .16, 95% CI [-.61,

.00040], t(597) = -1.96, p=.0508, while it has no significant effect on how petty it feels to complain about money, B=0.09, SE=0.17, 95% CI [-0.24, 0.42], t(600) = 0.52, p=0.60.

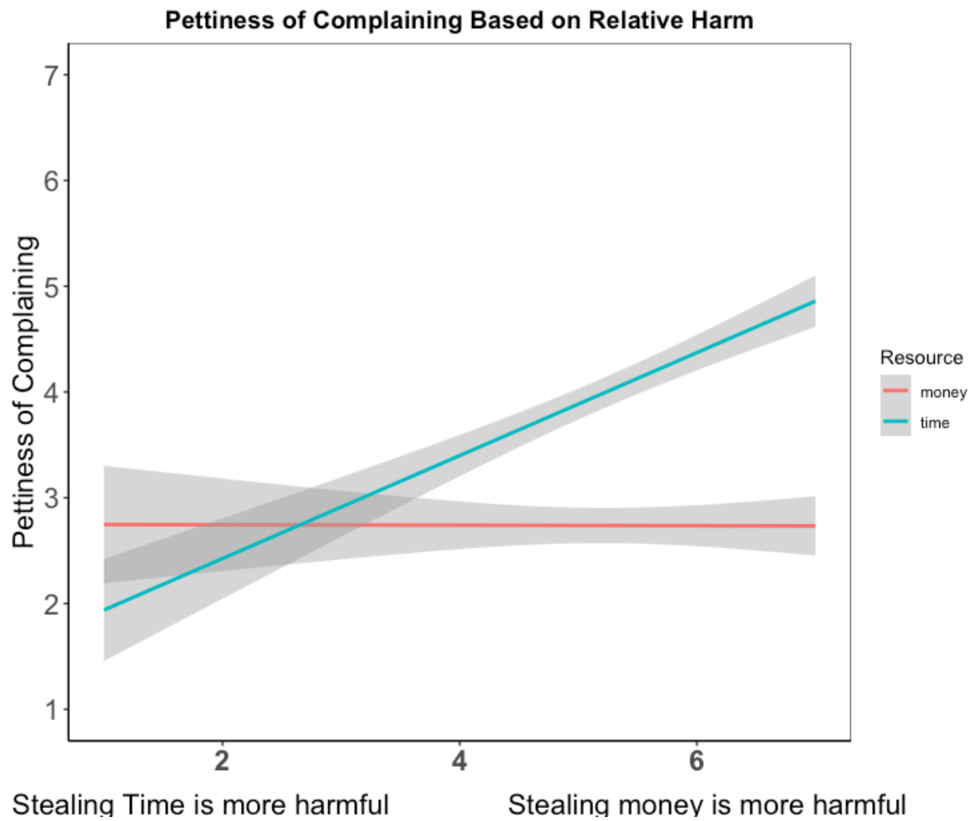

Figure 1.10 Pettiness of complaining by resource based on relative harm.

Discussion.

In experiment 4, we found a correlation relationship between harm and pettiness when processing a violation of time, but no relationship when processing a violation of money. We find evidence of that same pattern in this last experiment. Additionally, in experiment 5, we build on that by experimentally manipulating harm by prompting participants to deliberately reflect on how these two scenarios are similar. This manipulation makes people more likely to perceive the harm in violating someone's time, and therefore makes it less petty to complain about time, while it has no impact on the perceived pettiness of complaining about money.

Experiment 6.

We further test whether the different propensity to complain is driven primarily by the rules associated with money or the lack of rules associated with time. Is it uniquely petty to complain about time or is it just not petty to complain about money because there is a commonly accepted rule that stealing money is harmful? We test this by introducing another monetarily equivalent resource– being paid through merchandise, in kind – that should not evoke a universal rule.

Method.

Participants. We recruited 396 unique participants from a southwestern university (M_{age} = 20.8, 57.1% female) and used all responses in our analysis. Participants were randomly assigned to one of four conditions in a 2 (resource: *time* vs. *money*) x 2 (payment: cash vs in-kind) between-subject design. This experiment was pre-registered on aspredicted.org (#160697).

Procedure. Participants were told to imagine that they had started their own consulting business with small local businesses. We used wage rates that would be appropriate for a small, local business given the experience of our undergraduate population taking classes in the business school. In the *money* condition, participants agreed to provide a 6-hour introductory package for $300 (3-hour on site visit, 1-hour review of the company's public facing social media, 2-hour preparing the assessment). In the *time* condition, participants agreed to provide a 5-hour introductory package for $250 (the only difference was a 2-hour planned site visit). Both wage rates are equivalent of $50/hour. Then, all participants were told they had completed the introductory package exactly according to the agreement, but participants in the *time* condition were told that the client took you on a 3-hour tour which was longer than you agreed to. All participants were then explicitly told they were paid $250 for working six hours ($41.67/ hour). In the in-kind payment condition, participants agreed to work for and were subsequently paid in an equivalent amount of store merchandise.

Measures. Participants provided their intent to complain by answering which of the following describes their most likely response on an ordered three factor scale: 1 = *I would tell the client I am unsatisfied with my compensation*; 0 = *I would not say anything about my compensation to the client*; -1 = *I would tell the client I am satisfied with my compensation*. They then rated their level of satisfaction on a 6-point scale with no midpoint, randomizing the order of eliciting these two measures, which we confirmed was not significant. Next participants rated how morally acceptable the client's actions were and how petty it would be to say something about it on a 7-point scale. Finally, participants answered a manipulation check about how the outcome was different from their initial expectation: 1 = *I worked longer than I agreed to*; 2 = *I was paid less than I agreed to*; 3 = *It did not violate my expectations*.

Results.

Using our pre-registered ordered probit model, we find that once again participants complain more about *money* than *time*, B=-.40, SE=.12, t(393) = -3.25, p = .001, 95% CI [-.64, -.16]. When we include the payment method and its interaction with resource, we find that payment method moderates the difference between *time* and *money*, B = .55, SE =.26, t(391) =

2.14, p =.033, 95% CI [.05, 1.05]. Furthermore, the pairwise unadjusted marginal means comparison between time and money when paid in kind is no longer significant, B= .22, SE =.16, z =1.32 , p = 0.19, 95% CI [-.11, .54]. In terms of participant satisfaction with the compensation, using a linear model, we find the interaction of payment method on resource is significant, B = -.46, SE = .20, t(392) = -2.29, p = .023, 95% CI [-.85, -.06], and once again, the pairwise marginal means comparison between time and money when paid in kind is no longer significant, B = -.16, SE = .14, t(392) =-1.17 , p = 0.24, 95% CI [-.44, .11]. See Figure 1.11.

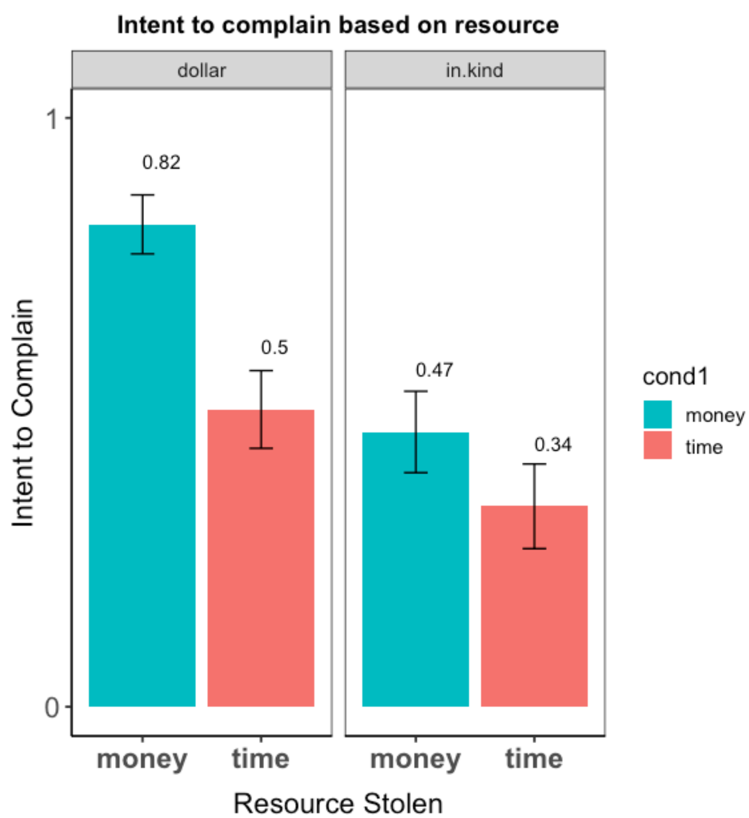

Figure 1.11. Intent to complain based on resource and payment method.

Finally, we explore the results of our manipulation check to see how participants encoded these infractions. Using a dummy-coded variable for those who said they were paid less than agreed to, we use the probit model to predict complaining based on the encoding dummy variable for those in the *time* conditions. We find that people who encoded the time violation as being paid less than agreed to rather than just working longer than agreed to were much more likely to complain, B= .40, SE = .18, 95% CI [0.05, 0.74], t(194)= 2.27, p=.024. However, because the manipulation check was asked after our DV, we cannot rule out whether the decision to complain made people encode the infraction as being paid less, or whether the encoding as being paid less made people more likely to complain.

Discussion.

The difference in complaining is diminished when people are not paid in cash. This supports the rules-based account that people immediately recognize being underpaid in cash as wrong (Amir & Ariely, 2007), but when they are paid in kind or shorted in time, there is no obvious rule violation. Without the activation of this rule, people are more likely to consider the costs and benefits of complaining. Therefore, people's tendency to not complain when they are overworked is driven less by the inherent properties of time itself, but rather the absence of the monetary-like rules.

Experiment 7.

The exploratory evidence in experiment 6 reveals that people who see their stolen time as stolen money are more likely to complain. Similarly, past literature suggests that when people deliberately convert their time to money, they are more likely to treat their time the way they treat their money (Pfeffer and DeVoe, 2012). Therefore, in this experiment we had people explicitly calculate their promised wage rate after reading about the gig they had accepted (which was \$12/hour for all participants). Then, once they were told about how much they were actually paid, they once again had to calculate their wage rate (which was \$6/hour for all participants).

If unwillingness to advocate for time is driven purely by people not accounting for their time like money, we would expect this explicit wage rate condition to increase complaints about stolen time. However, if the difference in complaining between resources is driven more by the automatic activation of a rule that stealing money is bad, this same reflection would make people in the money condition less likely to automatically act on the rule, and more likely to elaborate on the broader circumstances of the infraction, thereby lowering their likelihood of complaining.

Method.

Participants. We recruited 1203 participants (M_{age} = 41.94, 44.97% female) on Mturk for a 2 (resource: time vs. money) x 2 (explicit wage rate calculation vs control) design. This experiment was pre-registered on aspredicted.org (#110954).

Procedure. We use the same survey scenario as in our previous hypothetical experiments. In the explicit wage rate intervention, after reading about the gig they agreed to do, they were asked the following: "Based on the advertised \$/min rate, if you continued doing surveys for this same pay rate, what would be your effective hourly wage? (Effective hourly wage = how much money you would earn in an hour)." Then upon learning of how much they were actually paid and how longer the survey took, participants once again calculated their wage rate: "Based on how long the survey actually took (10 minutes) and what you were actually paid (\$1), if you continued doing surveys for this same pay rate, what would be your effective hourly wage? (Effective hourly wage = how much money you would earn in an hour)."

Measures. Participants answered the binary DV of whether they would complain, and then we asked how likely they would be to complete another survey from this surveyor (1 = extremely unlikely to 7 extremely likely). Then, participants answered questions about perceived pettiness of complaining, morality of the surveyor, impact of complaining, and their history of complaining. Finally, they rated their agreement with the following statement on a 7-point scale: *Time is money when working on Mturk.*

Results.

In a covariate free binary GLM regression, we find that participants are much less likely to complain when their time is stolen, $z(1199) = -11.15$, $p < .001$, but the explicit wage rate intervention made participants even *less* likely to complain, $z(1199) = -3.50$, $p = .00046$, and there was a significant interaction, $z(1199) = -.73$ $p = 0.0043$. However, when comparing pairwise means between time and money across experimental conditions, we find that there was no significant difference in complaints about time from the control group ($M = .23$, $SD = .42$) to the explicit wage rate condition ($M = .26$, $SD = .43$), $t(597.49) = 0.69$, $p = 0.49$. However, when participants money was stolen, they were significantly less likely to complain when in the explicit wage rate condition ($M = .57$, $SD = .49$) than the control group ($M = .71$, $SD = .46$), $t(597.08) = 3.55$, $p < 0.001$. See Figure 1.12.

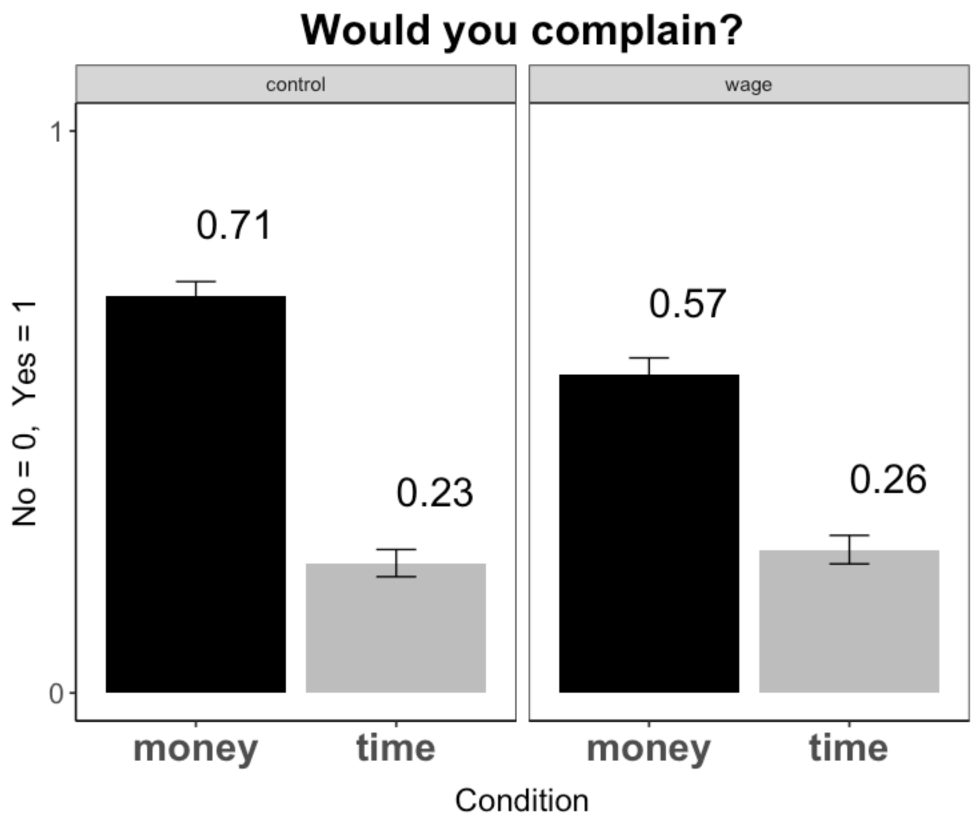

Figure 1.12. Explicit Wage rate manipulation lowers likelihood of complaining about stolen money.

To explore why the explicit wage rate condition made people in the money condition less likely to complain, we explored the intervention's effect on our mediators of interest. We find that the main effect of the wage rate manipulation on perceived morality is positive and significant, $t(1199) = 2.26$, $p=0.02$, while the interaction with time is negative, $t(1199) =-2.07$, $p=0.039$.

Finally, we found that despite there still being a persistent direct effect from resource stolen to complaining, all participants seem to recognize that time is money on these platforms ($M=6.37$, $SD=0.98$), with the median response of all participants regardless of condition being a 7, ”completely agree”. See Figure 1.13.

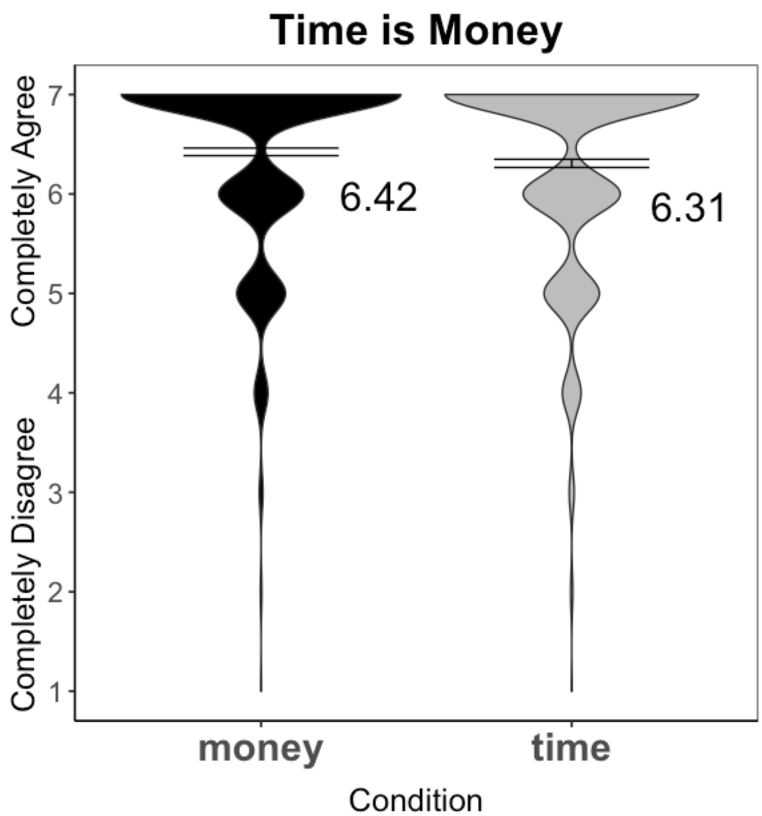

Figure 1.13. Participants agreement with “Time is money when working on Mturk.”

Discussion.

This experiment shows that gig workers on these platforms know time is money, and yet they still will not advocate for their time, even when they are forced to explicitly calculate their wage rate. Instead, participants whose time was stolen was less likely to complain when they had to explicitly calculate their wage rate. In the absence of the rules-based account, this result may seem puzzling. However, our theory predicts that when a rule is violated (i.e. stealing money is wrong), people will have an automatic response to the injustice. When we interrupt that automatic response by forcing participants to calculate the impact on effective wage rate, they were no longer simply following a rule. Instead, they seemed to conclude that only paying $6/hour wasn't *that bad* (likely compared to their average pay on Mturk), so stealing the $1 seemed relatively less amoral.

General Discussion.

This research highlights a fundamental asymmetry in how people advocate for their time versus their money. Across seven experiments, we show that people are significantly less likely to complain when their time is stolen compared to their money, even when the magnitude of harm is held constant. This discrepancy stems from the lack of clear norms surrounding time violations, which contrasts sharply with the automatic rule-based responses evoked by monetary theft.

Our findings build on theories of norms, rules and time money-inconsistencies. Monetary violations trigger deeply ingrained societal rules, leading to automatic responses that bypass extensive deliberation. In contrast, time violations lack such established norms and are processed reflectively, with individuals weighing factors like harm severity and social costs. This reflexive processing can make complaints about time theft seem petty, even when the harm is objectively comparable to monetary theft. For example, in Experiment 4, we found that the perceived harm of time theft predicts how petty it feels to complain, a relationship absent in monetary contexts.

This reluctance has profound implications for hourly and gig workers, who already face systemic disadvantages. The absence of time-related norms may also explain why time theft often manifests subtly, as in requests for "just a minute" or the expectation of unpaid overtime. It underscores the need for employers, platforms, and policymakers to proactively respect workers' time, given their diminished likelihood to advocate for themselves.

By situating this work in labor contexts—where time's value is explicitly tied to wages—we present a conservative test of these effects. If individuals are unwilling to advocate for their time in gig work, where compensation is explicitly tied to time, they may be even less likely to stand up for their time in social or less structured contexts where the value of time is more ambiguous. Beyond the workplace, these findings illuminate interpersonal dynamics where undervaluing others' time—whether by bosses, colleagues, or family members—can go unchecked.

Future research should explore interventions that render time theft as tangible and actionable as monetary theft by introducing rules about time. It should also investigate the boundary conditions of these effects, particularly when the magnitude of time theft increases. Future research should also test whether the lack of norms about time makes people more likely to take another's time in the first place. Future research should explore the variability in protections of workers' time across different cultures and across different freelance platforms such as Mturk and Prolific to see how this impacts time theft and complaining. Understanding how norms evolve around time in the gig economy will be crucial for addressing systemic inequalities and fostering equitable exchanges.

In sum, while time is often deemed our most precious resource, it lacks the normative and societal scaffolding that protects money. Addressing this asymmetry could empower individuals to advocate for their time, fostering a culture where time theft is recognized and rectified. Remember, the next time someone monopolizes your time, take a moment to reflect on the harm they are causing, and perhaps you will not feel as petty telling them you have other things to do.

ACKNOWLEDGEMENTS

Chapter 1, in full, is currently being prepared for submission for publication of the material with On Amir. The dissertation author was the primary investigator and author of this paper.

Supplemental Experiment 1

In Supplemental Experiment 1, we establish the effect in a survey vignette with low stakes (5 minutes or $1).

Method.

Participants. We recruited 201 participants through Cloud Research (M_{age} = 39.87, 50.75% female) and used all responses in our analysis. Participants were assigned to one of two conditions: time or money. This experiment was pre-registered on aspredicted.org (#103827).

Procedure. In the *time* condition, participants imagined that they agreed to participate in a survey where they would earn $1 for five minutes of their time. In the *money* conditions, participants imagined they agreed to participate in a survey where they would earn $2 for ten minutes of their time. Both wage rates are equivalent of $12/hour. Then, all participants were told they had completed the survey exactly as instructed, but the survey took a full ten minutes to complete, and they were paid $1. Therefore, all participants were effectively paid an equivalent rate of $6/hour for the survey, half the rate they agreed to work for. The "exactly as instructed" detail was meant to highlight that the survey took ten minutes because it was designed to take ten minutes, not because the participants were just slow to finish.

Measures. First, participants indicated on a binary scale how they would respond with 1 being "I would contact them about how much I was paid" and 0 being "I would not contact them about how much I was paid". Next participants rated how upset they were on a Likert scale from 1 to 5, with 1 being "I am not upset at all" to 5 being "I am extremely upset." We also conducted a manipulation check where we asked participants how much they should have been paid for a 10-minute survey based on the agreed upon wage. We pre-registered that we would exclude any participants that do not say $2, but the results do not change if we included everyone in the sample. We also counterbalanced whether the manipulation check happened before or after the main dependent variables (DV) to see if making people explicitly calculate the time/money conversion before the main dependent variables influences their responses.

Results.

Participants complained more in the *money* condition (M=.80, SD=.40) than *time* (M=.19, SD=.39) using a binary GLM estimating the main effect of resource, B = -2.85, SE =0.36, z(199) = -7.98, p <.001, 95% CI [-3.57, -2.17], with an observed Cohen's *d* = 1.54, 95% CI [1.22, 1.85]. Using an ordinary least squares (OLS) regression, we also found that participants were significantly less upset in the time condition, B = -.91, SE = .14, t(199) = -6.26 p<.001, 95% CI [-1.19, -.62]. The timing of the manipulation check had no significant main effect or interaction with either dependent variable (ps>.62). Age and gender had no significant effect with either measure (ps>.25) (all data and materials are available in Research Box).

Supplemental Experiment 2

In Supplemental Experiment 2, we replicate the effect in a babysitting vignette with medium stakes.

Method.

Participants. We recruited 398 unique participants through the university lab (M_{age} = 21.0, 51.7% female) and used all responses in our analysis. Participants were randomly assigned to one of two resource conditions (time vs. money) and two employer conditions (mom vs dad).

Procedure. Participants were given a babysitting vignette. In the *money* condition, participants read that they agreed to babysit for a family for two hours for $25/hour. In the *time* condition, participants read that they agreed to babysit for a family for two hours for $20/hour, but the parents came home 30 minutes late. In both conditions they are paid \$40, which is \$10 less than their agreed-on wage.

Measures. Participants answered which of the following describes their most likely response: 1 = *I will immediately tell the parents they still owe me money* or 0 = *I will not mention this to the parents unless it happens again*. Next participants rated on a 7-point Likert scale of agreement: *I will babysit for this family again*.

Results.

Participants complained more in the *money* condition (M=.80, SD=.40) than *time* (M=.46, SD=.50) using a binary GLM estimating the main effect of resource stolen, B = -1.52, SE =0.23, z(394) = -6.69, p <.001, 95% CI [-1.98, -1.08], with an observed Cohen's *d* = .74, 95% CI [.53 .94]. Parent had no significant main effect (p=.35) nor interaction with resource (p=.51) on responding.

Supplemental Experiment 3

In Supplemental Experiment 3, we replicate the effect in a consulting vignette with higher stakes (one hour or \$50).

Method.

Participants. We recruited 301 unique participants through Cloud Research (M_{age} = 38.0, 52% female) and used all responses in our analysis. Participants were randomly assigned to one of two resource conditions (time vs. money).

Procedure. Participants were told to imagine that they had started their own consulting business with small local businesses. In the *money* condition, participants agreed to provide a 6-hour introductory package for \$300 (3-hour on site visit, 1-hour review of the company's public facing social media, 2-hour preparing the assessment). In the *time* condition, participants agreed to provide a 5-hour introductory package for $250 (the only difference was a 2-hour planed site visit). Both wage rates are equivalent of $50/hour. Then, all participants were told they had completed the introductory package exactly according to the agreement, but participants in the time condition were told that the client took you on a 3-hour tour which was longer than you agreed to. All participants were then explicitly told they were paid $250, which based on working six hours, was $41.67/ hour.

Measures. Participants answered which of the following describes their most likely response: 1 = *I would tell the client they owe me more money* or 0 = *I would not tell the client they owe me more money*. Next participants rated how petty it would be to say something about it on a 7-point Likert scale.

Results.

Participants complained more in the *money* condition (M=.86, SD=.35) than *time* (M=.46, SD=.50) using a binary GLM estimating the main effect of resource stolen, B = -1.94, SE =0.28, z(299) = -6.85, p <.001, 95% CI [-2.52, -1.40], with an observed Cohen's *d* = .92, 95% CI [.68, 1.16]. When we include pettiness and its interaction with resource in the model, it fully mediates the effect of resource on complaining, B = .54, SE =0.74, z(297) = .736, p =.46, 95% CI [-.88, 2.04]. People are less likely to complain the pettier it is (B = -54, SE =0.13, z(297)= -4.16, p <.001, 95% CI [-.80, -.29]) and even more so in the time condition (B = -.58, SE =0.21, z (297)= -2.76, p =.006, 95% CI [-1.00, -.18]).

Supplemental Experiment 4

In this experiment, we attempt to make salient the opportunity costs, since we know that people are more likely to consider the opportunity costs of money rather than time (Okada & Hoch 2004).

Method.

Participants. We recruited 806 participants (M_{age} = 41.51, 52.73% female) for a 2 (resource: time vs. money) x 2 (opportunity cost intervention vs control) design. This experiment was pre-registered on aspredicted.org (#107109).

Procedure. We use the same survey scenario as Experiment 4A. In the opportunity cost intervention, after the vignette description prior to asking if they would complain, participants read: "You are tight on time today as you have many errands to run and household chores to finish before someone comes over later. As soon as you finish the survey, you plan to try and finish everything you had planned to do before your friend arrives." The purpose of this intervention was to remind participants that they had something planned for both their time and money. We predicted that reminding participants of the opportunity costs of their time would make them more likely to complain about their time.

Measures. In addition to the main binary measure of complaining, we measured several exploratory variables to help explain why participants tend not to complain about time. First, we measured how upset people were on a scale of 1 to 5. Then, participants rated on a 7 point Likert scale how much they agreed with the following statements in a randomized order: *it was an honest mistake* (intentionality), *it would be petty to say something* (pettiness), *I generally expect tasks to take longer than people say they will* (expectation), *things like this happen to me all the time* (commonplace), *paying people less than agreed to is unacceptable* (morality of stealing money)*, keeping people longer than agreed to is unacceptable* (morality of stealing time).

Results.

Per our pre-registered covariate free analysis, we find that participants are much less likely to complain when their time is stolen, $z(802) = -9.64$, $p < .001$, but neither the opportunity cost main effect, $z(802) = -.97$, $p = 0.33$, nor its interaction with resource stolen, $z(802) = 0.83$, $p = 0.41$, was significant. While we were not able to moderate the unwillingness to complain about stolen time, we did learn more about the mechanism through our exploratory variables. When we included all measured covariates in our model, we found that the only significant predictors were upsetness, $z(791) = 6.13$, $p < .001$, pettiness, $z(791) = -6.54$, $p < .001$, expectation, $z(791) = -2.78$, $p <= 0.0054$, and intentionality, $z(791) = 2.16$, $p = .03$.

Supplemental Experiment 5

We suspected that stealing money automatically triggers a moral evaluation, while stealing time does not trigger this same evaluation. However, by forcing people to evaluate the morality of the act, people may be more likely to perceive the harm, which would make them more likely to complain.

Method.

Participants. We recruited 801 participants ($M_{age} = 39.96$, 55.18% female) for a 2 (resource: time vs. money) x 2 (moral standard intervention vs control) design. This experiment was pre-registered on aspredicted.org (#108593).

Procedure. We use the same survey scenario as Experiment 4A. In the moral standard intervention, after the vignette prior to asking if they would complain, participants answered the following question: “How morally acceptable was the behavior of the person who designed and advertised the survey?” (1 = completely unacceptable to 7 = completely acceptable).

Measures. In addition to the main binary DV of complaining, we measured how satisfied people were with the gig on a scale of 1 to 7. Then, participants answered the same questions about pettiness and commonplace as in experiment 5.

Results.

In our pre-registered covariate free binary GLM regression, we find that participants are much less likely to complain when their time is stolen, $z(797) = -9.10$, $p < .001$, but neither the introduction of the moral standard, $z(797) = -.04$, $p = 0.97$, nor its interaction with resource stolen, $z(797) = -.33$ $p = 0.74$, were significant. Once again, we were not able to moderate the difference in complaining about stolen time compared to money. When we included all covariates in our model, we did confirm that perceived pettiness was once again the most robust predictor of complaining, $z(795) = -8.93$, $p < .001$.

Given that asking people to evaluate the morality of the surveyor seemed to have no effect, we were curious if people just do not find it wrong to steal time. Using the responses from participants in the moral standard condition who rated the acceptability of the surveyor's behavior, we find that participants rate the acceptability of stealing money ($M_{money} = 1.73$, $SD = 1.22$) as much lower than the acceptability of stealing time ($M_{time} = 3.04$, $SD = 1.53$), $t(412) = 9.63$, $p < .001$, although both are negatively valanced below the scale midpoint.

As an exploratory analysis we find evidence that morality and perceived pettiness partially mediate the relationship between resource stolen and complaining using a serial mediation analysis with 10,000 bootstrapped samples (Hayes 2022, model 6). The results revealed a significant indirect effect of resource stolen on complaining through morality and pettiness (serial paths $a_1 \times d_{21} \times b_2 = -0.44$, $SE = 0.094$, 95% CI: [-.66, -.29]; see Fig. 2)). Furthermore, the direct effect of resource stolen on complaining in the presence of the mediators was also significant (b = -1.79. p<.001). Hence, there is a partial serial mediation of morality and pettiness on the relationship of stealing time versus money and willingness to complain.

Supplemental Experiment 6

In this experiment, we test whether introducing a descriptive norm that others are also complaining may make complaining about time feel more acceptable.

Method.

Participants. We recruited 801 participants (M_{age} = 40.77, 57.43% female) for a 2 (resource: time vs. money) x 2 (normative intervention vs control) design. This experiment was pre-registered on aspredicted.org (#109217).

Procedure. We use the same survey scenario as Experiment 4A. In the normative intervention, after the vignette, participants read: "You learn that there is a collective petition from other people who completed this gig to file a complaint to the surveyor."

Measures. After responding to the binary DV of complaining, we asked how likely the would be to complete another survey from this surveyor (1 = extremely unlikely to 7 extremely likely). Then, participants answered the same pettiness question as experiment 5 plus two other exploratory variables: *how likely is your complain to result in action by the surveyor?* (impact) and *how morally acceptable is the behavior of the surveyor?* (moral). Finally, we asked participants if they have ever filed a complaint in the past, since past complaining behavior is a likely predictor of future complaints (Yes = 1, No = 0, Don't remember = 5)

Results.

Per our pre-registered binary GLM regression controlling for individual difference in pettiness, gender, age, and history of complaining, we find once again that participants are much less likely to complain when their time is stolen, $z(791) = -6.55$, $p < .001$, that the descriptive norm makes participants more likely to complain, $z(791) = .77$, $p=0.011$, but the interaction with resource stolen was insignificant, $z(791) = .67$, $p=0.099$. In terms of the covariates, pettiness was the largest predictor, $z(791) = -11.35$, $p<.001$, followed by history of complaining, $z(791) = 7.49$, $p<.001$, and age had a relatively small significant effect on complaints, $z(791) = -0.02$, $p=.016$.

Furthermore, the normative intervention made participants think that their complaint would have more impact, $t(799) = 4.16$, $p<.001$, but the intervention had no direct effect on morality or pettiness.

Supplemental Experiment 7

Given pettiness concerns seem to be depressing complaints, we borrow from Kim et al (2019) and provide participants a reason attending to the minutes and cents is not trivial. We test whether emphasizing the value of feedback moderates pettiness, thereby increasing willingness to complain about time.

Method.

Participants. We recruited 1202 participants through Cloud Research ($M_{age} = 39.98$, 54.91% female) and used all responses in our analysis. Participants were randomly assigned to one of four conditions in a 2 (stolen resource: time vs. money) x 2 (feedback: solicited vs control) between-subject design. This experiment was pre-registered on aspredicted.org (#113283).

Procedure. We use the same base vignette as experiment 4A and vary once again whether we steal 5 minutes (*time)* or $1 (*money)*, keeping fixed the advertised effective wage rate ($12/hour) and the post-stealing wage rate of $6/hour.

In the feedback solicited condition, we told participants that the surveyor added a message to the survey saying, “Please consider providing feedback on this survey. 99% of researchers say they highly value both positive and negative feedback on their surveys. Feedback helps surveyors improve their research design and also improves the worker experience.” Participants in the feedback control did not see this prompt.

Measures. Participants were asked how likely they were to provide positive feedback to the surveyor or negative feedback to the surveyor on a scale of 1 (extremely unlikely) to 100 (extremely likely), counterbalancing the order of the two questions. Next participants rated how likely they would be to complete another survey form this surveyor on a 7-point Likert scale. Participants completed a manipulation check by filling in the following blank: “Based on the advertised $/min rate, how much money should the researcher have paid you for this ten minute survey?”. Additionally, participants rated whether it would be petty to say something about this *(petty)* and how likely it is the complaint would result in action by the surveyor *(impact)* on 7-point Likert scales. All participants reported their age, gender, and whether they had filed a complaint in the past (*history*).

Results.

People are more likely to provide negative feedback to the surveyor when their money is stolen (M=70.95, SD = 31.27) compared to their time(M=44.78, SD = 31.05), t(1200) = -14.56, p <.001. Using our pre-registered analysis including all covariates in a linear regression, we confirm the direct effect that time versus money makes people less likely to provide negative feedback, B = -17.86, SE=2.12, t(1192)= -8.44, p<.001, 95% CI [-22.01, -13.71], Cohen’s *d*=.84. The only significant covariates in this model are pettiness, B = -8.93, SE = .39, t(1192) = -22.77, p<.001, 95% CI [-9.70, -8.16] and history of complaining, B = 7.55, SE=1.61, t(1192) = 4.70,

$p<.001$, 95% CI [4.40, 10.70]. Perceived pettiness continues to drive down the likelihood that people will provide negative feedback, while complaining in the past makes you more likely to complain in the future. We find a non-significant direct effect of the feedback manipulation, $B = -2.62$, $SE = 2.07$, $t(1192) = -1.26$, $p = .21$, 95% CI [-6.68, 1.45], but a significant interaction with time, $B = 6.07$, $SE = 2.90$, $t(1192) = 2.09$, $p = .037$, 95% CI [.37, 11.76].

Finally, using a linear regression to predict people's willingness to do another survey from this researcher, participants whose time was stolen were much more likely to agree to do another survey from the same researcher, $B = .89$, $SE = .13$, $t(1192) = 6.72$, $p < .001$, 95% CI [.63, 1.15], suggesting that there is less backlash from violating time than money. The feedback manipulation did have a positive direct effect on willingness to do another survey, $B = .29$, $SE = .13$, $t(1192) = 2.19$, $p = .028$, 95% CI [.03, .54], but no significant interaction with the resource stolen, $B = 0.07$, $SE = .18$, $t(1192) = 0.37$, $p = .71$, 95% CI [-.29, .42].

Supplemental Experiment 8

Supplemental Experiment 8 examined how people's willingness to complain differs when time is the end and money is the means.

Method

Participants. We recruited 354 participants in the university lab over the course of one week (age: $M = 20.75$, $SD = 1.86$; gender: 54.52% female, 44.63% male, 0.85% nonbinary) for a 2 resource violated (*time* versus *money*) between subject design. This experiment was pre-registered on aspredicted.org (#175684).

Procedure. All participants were told that they had rewarded themselves with a 75-minute massage for $150. However, when they went to pay the masseuse, in the *money* condition, they charge $180 for the 75-minute massage, instead of $150 as listed. Whereas in the time condition, they charge $150, but the massage was only 60-minutes long, instead of 75-minutes as listed.

Measures. Participants responded to a binary measure of whether they would complain., 1 = "*I would complain*" or 2 = "*I would NOT complain*." Participants also reported on a seven point Likert scale to what extent they thought they were harmed and to what extent it would be petty to complain in a randomized order. All participants reported their age and gender.

Results

Participants complained more in the *money* condition (M=.85, SD=.36) than *time* (M=.68, SD=.47) using a binary GLM estimating the main effect of resource stolen, B = -1.01, SE =0.27, z(352) = -3.81, p <.001, 95% CI [-1.55, -.50], with an observed Cohen's *d* = -.42. 95% CI [-.74, -.30]. When we include pettiness and harm in our model, these predictors partially mediate the effect of resource violated on complaining, B = -.81, SE =0.28, z (350) = -2.91, p <.001, 95% CI [-1.37, -.27]. People are less likely to complain the pettier it is (B = -.33, SE =0.08, z(350)= -4.20, p <.001, 95% CI [-.49, -.18]) and more likely to complain the more harmful they think it is (B = .18, SE =0.09, z (350)= 2.04, p =.041, 95% CI [.01, .35]).

REFERENCES

Aaker, J. L., Rudd, M., & Mogilner, C. (2011). If money does not make you happy, consider time. Journal of consumer psychology, 21(2), 126-130.

Bureau of Labor Statistics. (2022). Characteristics of minimum wage workers, 2022. U.S. Department of Labor. https://www.bls.gov/opub/reports/minimum-wage/2022/#:~:text=In%202022%2C%2078.7%20million%20workers,wage%20of%20%247.25%20per%20hour.

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy, 17(1), 74-95.

Donnelly, G. E., Wilson, A. V., Whillans, A. V., & Norton, M. I. (2021). Communicating resource scarcity and interpersonal connection. Journal of Consumer Psychology, 31(4), 726-745.

Economic Policy Institute. (2022). Surveying gig work: Pay, benefits, and workforce data insights. Economic Policy Institute. https://www.epi.org/publication/gig-worker-survey/.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. Journal of personality and social psychology, 65(1), 45.

Hayes, A. F. (2018). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach (2nd ed.). The Guilford Press.

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management, 22*(1), 57–70. https://doi.org/10.1016/0095-0696(92)90019-S

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. Psychological review, 93(2), 136.

Kern, M. C., & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. Psychological Science, 20(3), 378-384.

Kim, C., Kim, S., Im, S., & Shin, C. (2003). The effect of attitude and perception on consumer complaint intentions. Journal of consumer marketing, 20(4), 352-371.

Kim, T., Zhang, T., & Norton, M. I. (2019). Pettiness in social exchange. Journal of Experimental Psychology: General, 148(2), 361.

Kowalski, R. M. (1996). Complaints and complaining: Functions, antecedents, and consequences. Psychological bulletin, 119(2), 179.

Liu, W., & Aaker, J. (2008). The happiness of giving: The time-ask effect. Journal of consumer research, 35(3), 543-557.

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

Morvinski, C., Saccardo, S., & Amir, O. (2023). Mis-nudging morality. Management Science, 69(1), 464-474.

Loewenstein, G., & Elster, J. (1992). Utility from memory and anticipation. Choice over time, 213-234.

Mogilner, C., & Norton, M. I. (2016). Time, money, and happiness. Current Opinion in Psychology, 10, 12-16.

Mogilner, C., Whillans, A., & Norton, M. I. (2018). Time, money, and subjective well-being. Handbook of well-being.

Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. Journal of consumer psychology, 22(3), 453-460.

Okada, E. M., & Hoch, S. J. (2004). Spending time versus spending money. Journal of consumer research, 31(2), 313-323.

Pierce, J. R., Giurge, L. M., & Aeon, B. (2024). Time theft: exposing a subtle yet serious driver of socioeconomic inequality. Business & Society, 00076503231224710.

Pfeffer, J., & DeVoe, S. E. (2012). The economic evaluation of time: Organizational causes and individual consequences. Research in Organizational Behavior, 32, 47-62.

Rajagopal, P., & Rha, J. Y. (2009). The mental accounting of time. Journal of Economic Psychology, 30(5), 772-781.

Richins, M. L., & Verhage, B. J. (1985). Seeking redress for consumer dissatisfaction: The role of attitudes and situational factors. Journal of Consumer Policy, 8(1), 29-44.

Saini, R., & Monga, A. (2008). How I decide depends on what I spend: Use of heuristics is greater for time than for money. Journal of Consumer Research, 34(6), 914-922.

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by redefining harm. Personality and Social Psychology Review, 22(1), 32-70.

Singh, J., & Wilkes, R. E. (1996). When consumers complain: A path analysis of the key antecedents of consumer complaint response estimates. Journal of the Academy of Marketing science, 24, 350-365.

Smitizsky, G., Liu, W., & Gneezy, U. (2021). On the value (s) of time: Workers' value of their time depends on mode of valuation. Proceedings of the National Academy of Sciences, 118(34), e2105710118.

Soman, D. (2001). The mental accounting of sunk time costs: Why time is not like money. Journal of behavioral decision making, 14(3), 169-185.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. Journal of personality and social psychology, 94(4), 672.

Trupia, M. G., Engeler, I. (2022). When the Unexpected Happens: how People Respond to Unbudgeted Time Savings. Association for Consumer Research Conference October 2022.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. science, 211(4481), 453-458.

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. science, 314(5802), 1154-1156.

Whillans, A. V., Yoon, J., Turek, A., & Donnelly, G. E. (2021). Extension request avoidance predicts greater time stress among women. Proceedings of the National Academy of Sciences, 118(45), e2105622118..

Zauberman, G., & Lynch Jr, J. G. (2005). Resource slack and propensity to discount delayed investments of time versus money. Journal of Experimental Psychology: General, 134(1), 23.

Zhou, X., Vohs, K. D., & Baumeister, R. F. (2009). The symbolic power of money: Reminders of money alter social distress and physical pain. Psychological Science, 20(6), 700-706.

Chapter 2

SELFISH RISK PERCEPTIONS:
THE PSYCHOLOGY OF IMPOSING HARM ON OTHERS.

Kathryn Hillegass & On Amir

Rady School of Management, University of California, San Diego, La Jolla, CA 92093, USA

ABSTRACT

This paper explores how unrealistic optimism shapes the risks individuals are willing to impose on others. While existing research on decision-making often focuses on self-risk, many real-life decisions, like sending a symptomatic child to daycare, impose risks on others. We hypothesize that people display motivated unrealistic optimism, perceiving the risks they impose on others as less severe than the risks they would assume for others. Across four experiments involving both hypothetical and incentive-compatible scenarios, we find consistent evidence that individuals perceive risks differently depending on who bears the consequence. Participants rated imposed risks as less severe than assumed risks, resulting in a higher willingness to impose risk for personal benefit. Our findings extend traditional theories of unrealistic optimism by linking it to behavioral outcomes. We also find evidence of selfish preferences driving this motivated, unrealistic optimism rather than a non-motivated factors The results underscore how biased risk assessments can lead to behavior that prioritizes personal benefit at the expense of others. These findings have broad implications, suggesting that motivated unrealistic optimism may contribute to both everyday and policy-level decisions that undervalue risks imposed on others.

Introduction

Humans face near constant risk and uncertainty in their day to day lives, to include decisions about personal finance, transportation, preventative health measures, and social situations. Most of the work in judgement and decision making focuses on single agent risk – thinking about *how much risk am I willing to accept for my own benefit.* However, many decisions people make in life not only direct impact themselves, but oftentimes impose risk on other people. For example, during cold and flu season, wearing a mask does not just prevent oneself from getting sick, but the decision to wear a mask prevents others from getting sick as well. While there is a long history of work on preventative health measures and risk assessment particularly in the wake of the COVID-19 pandemic (Scheier & Carver, 1985; Hoorens & Buunk 1993; Klein & Kunda 1993; De Bruin & Bennett 2020; Savadori & Lauriola 2021; Vieites et al 2021), there is less work on how we consider the risks we impose on others in the face of uncertainty.

This paper investigates decisions that directly benefit us while putting another's interest at risk (e.g. *sending a symptomatic child to daycare, drinking or texting while driving, copying someone's homework*). It contrasts these decisions with the decisions others make that put our interests at risk for their sole benefit. When faced with these scenarios, people must assess how much risk they are willing to assume for someone else's benefit versus how much risk they are willing to impose on someone else for their own benefit. We refer to this type of interpersonal risk as *assuming risk for others*. This paper explores how we subjectively assess, decide, and hedge against these types of interpersonal risks, and documents a pattern of inflating the subjective risk we assume for others while minimizing the subjective risks we impose on others.

We extend the well-established phenomenon of unrealistic optimism (Weinstein 1980) whereby people tend to think bad things are less likely to happen to them. We posit that when faced with a decision to impose risk on others, people tend to exhibit motivated unrealistic optimism. The bulk of previous work on unrealistic optimism focuses strictly on perceptions of single-agent risk with implied behavioral outcomes. By measuring both risk perceptions and people's willingness to impose this risk on others, we find evidence of motivated unrealistic optimism driven by people's selfish preferences. We investigate this in two hypothetical scenarios of uncertainty and two incentive compatible scenarios of fixed risk. Across all studies, we find that people tend to assume that the risks they impose on others are less severe than the risks they assume for another's benefit. This differential perception of risk systematically drives people to impose more risk on others than they are willing to accept themselves.

Risk perception

Researchers have long investigated how people systematically deviate from the standard, probabilistic based, expected utility theory when facing risky prospects (Kahneman & Tversky 1972; 1979). Prospect Theory establishes that people tend to systematically underweight high probabilities relative to certain outcomes (Kahneman & Tversky 1979; Tversky & Fox, 1995), which explains why people are more likely to opt out of a warranty to protect against screen cracks despite knowing that dropping one's phone is very common. Conversely, people tend to overweight low probabilities, which explains why people buy lottery tickets despite very low odds of winning. Therefore, when soliciting subjective risk of an activity, we would expect participants to inflate low probabilities and underestimate high probabilities.

Beyond these systematic computational biases, another large stream of research explores how negative feelings of fear and dread influence risk perceptions. Risk-as-feelings details how these immediate, visceral anticipatory emotions override cognitive assessments of risk (Loewenstein et al 2001). For example, when people read about the death of a single person, the negative affective response leads to a global increase in the perceived risk of both related and unrelated hazards, (Johnson & Tversky 1983), regardless of (miss)attribution (Winkielman et al 1997). Visceral factors – such as fear and pain - tend to narrow attention toward the self relative to others, leading to distorted perceptions of risk (Loewenstein 1996). Likewise, dread is a driving determinant of risk perception (Fischhoff et al 1978) for low-probability, catastrophic events. For example, it feels very risky for a child to cross the street, but objectively, there are very few accidents (Howarth, 1988). Similarly, in the three months following the 9/11 attack, Americans avoided dreaded air travel and choose to drive more instead, which resulted in more fatal traffic accidents than usual due to the higher probabilistic, but more mundane risk of a driving (Gigerenzer 2004). Therefore, we expect that in interpersonal risk contexts, the more a decision invokes immediate emotions, the more they are likely to use the emotional assessment in lieu of a computational assessment.

Beyond our immediate affective responses, anticipated emotions also influence risk-related decisions. Anticipated regret theory predicts that people will choose suboptimal outcomes to avoid "wrong decision" that could cause regret, particularly when people expect to get feedback on their choice and the foregone alternatives (Loomes & Sugden 1982; Bell 1983; Zeelenberg 1999). Similarly, decision affect theory predicts that in the face of risky prospects, people compare counterfactuals and tend to maximize subjective expected pleasure, choosing options that will make them feel best and avoiding options that will make them feel worse

(Mellers et al 1997; Mellers et al 1999). Together, these findings are consistent with the affect heuristic (Slovic & Peters, 2006; Slovic et al 2007) whereby people use their rapid and automatic characterization of a stimulus as good or bad as a highly evaluable criterion (Hsee 1996) to inform subsequent judgements. Their anticipated emotions (predicted emotions) are factored into their expected utility. In decisions involving *assuming risk for others*, the affect heuristic predicts that the anticipated emotions associated with a poor outcome will loom larger when assuming risk for others than when imposing it on someone else.

While the affect heuristic proposes that these anticipated emotions are an input to a broader cognitive assessment about expected utility, the risk-as-feelings account proposes that when affective assessments diverge from cognitive assessment, anticipatory emotional responses (emotions felt during the decision process) will ultimately guide behavior (Loewenstein et al 2001). For example, fear and heart rate reactivity increase in the immediate lead up to a risky outcome being resolved (e.g. impending electric shock) (Monat 1976), likely for adaptive reasons so that the organism focuses on the most imminent threat first (Loewenstein et al 2001). The risk-as-feelings theory proposes a competing explanation to Kahneman and Tversky (1979) for why people tend to buy lottery tickets, suggesting that instead of inflating probabilities, people tend to disregard them completely, paying instead for the feeling of hope that their money problems disappear, regardless of the odds. While we do not disentangle the role of anticipated and anticipatory emotions in this paper, we assume that both accounts will lead people to impose more risk on others than they are willing to impose themselves.

Another factor that influences risk perception is the size of the reward. Conventional wisdom posits that big risks lead to big rewards, an expression that reflects a positive correlation between risk and reward. However, in people’s minds, this correlation is less evident. In fact, researchers find that while there is no systematic relationship between perceived existing risk and benefits across dozens of activities such as riding a motorcycle or hunting, people's willingness to accept those risk depends on the perceived benefit they expect to get from undertaking the hazardous activity (Fischhoff et al 1978). In fact, researchers demonstrated an inverse relationship between perceived risk and benefits, whereby the more negatively a person feels about an activity such as using pesticides, the more risky and less beneficial the activity seems (Alhakami & Slovic, 1994; Finucane et al 2000). Therefore, people are using negative (positive) affective cues to both over-inflate (under-inflate) the risks and under-inflate (over-inflate) the benefits, demonstrating the lack of independence between these two variables. This aligns with work demonstrating that the assumption of probability-outcome independence may hold for monetary values, but not affective values (Rottenstreich & Hsee 2001). They find that affect-rich outcomes yield more pronounced overweighting of small probabilities and underweighting of large probabilities. For example, people would pay a lot less money to avoid a 1% chance of being electrocuted but would pay a lot more to avoid a 99% chance of being electrocuted, relative to what they would pay to avoid financial loss. Given that people value their own benefits more than the benefits of others and are likely to experience more positive affect when facing a positive personal outcome, we would expect that people will be willing to downplay the risk when they are imposing it on someone else for their own benefit compared to when they are accepting it for another's benefit.

Unrealistic Optimism

Unrealistic optimism is the belief that good things are more likely to happen to them than to others, while negative events are less likely. In the seminal work on the subject, college students rate their own chances of experiencing positive and negative events, and consistently demonstrate unrealistic optimism, especially for events with greater perceived controllability (Weinstein 1980). In fact, this ability to distort incoming information in a manner than enhances self-esteem, personal efficacy, and optimism may be an adaptive mechanism that improves people's well-being (Taylor & Brown 1988) mental and physical health outcomes (Sharot 2012), and work-related behaviors (Puri & Robinson 2007). Unrealistic optimism is consistent with the better-than-average effect (Alicke & Klotz, 1995) whereby people evaluate themselves more favorably on a diverse range of measures than the average person. Subsequent researchers use a randomly generated benchmark to find that the unrealistic optimism is not just a matter of self-presentation, but a genuine belief that their performance is superior to others (Williams & Gilovich, 2000).

Unrealistic optimism can manifest in an absolute sense when comparing people's singular judgements about their own risk to an external representation of their actual risk, or in comparative optimism when people subjectively judge their own risk directly against the risk of someone else (Shepperd et al 2013). While unrealistic optimism is a robust finding across a broad variety of contexts, there are multiple competing accounts, both motivated and non-motivated for why this phenomenon emerges. Motivated accounts of unrealistic optimism tend to be related to desirability bias, defined as a causal influence of preferences on expectations about the future (see review Krizan and Windschitl 2007). Desirability bias is considered a subset of motivated reasoning, whereby people bias their evidence accumulation in favor of a desired conclusion (Kunda 1990). For example, regardless of being told accurate or inflated averages of other people's unhealthy behaviors (i.e. drinking alcohol, eating sugar), people report they partake in these behaviors less frequently than the average person (Klein and Kunda, 1995).

Motivated reasoning may lead to differential scrutiny, whereby people impose different confidence thresholds when thinking about what is considered risky for themselves or others (Gilovich et al., 1993; Trope & Liberman, 1996; Krizan & Windschitl, 2007). However, Krizan and Windschitl find limited evidence of a causal relationship of preferences on expectations, citing more evidence in studies that examined outcome predictions, and less evidence in studies eliciting likelihood judgements (2007). In studies where participants have an explicit (e.g. accuracy bonus) or implicit (e.g. fear of failing an attention check) motives to be precise about likelihood judgments, accuracy motives may suppress desirability biases (Tetlock, 1983; Bar-Hillel & Budescu, 1995; Windschitl & Wells, 1996; Krizan & Windschitl 2007).

Sense of control, defined as the belief that one has options to influence the aversiveness of an event (Thompson 1981), is another motivational account that supports unrealistic optimism (Chambers et al., 2003; Weinstein, 1980; Zakay 1984). Unrealistic optimism was stronger in people with more internal rather than external locus of control (Hoorens & Buunk 1993). In terms of control, even just the illusion of control can increase unrealistic optimism. For example, people think they are more likely to win a coin toss when they toss the coin (rather than the experimenter) (Langer & Roth, 1975). Previous work has attempted to disentangle unrealistic optimism from the illusion of control and finds that while people think they are less likely to get into an accident as a driver compared to other people, they think they are just as likely to get into an accident as a passenger (McKenna, 1993). This is driven by perceptions of control, because when ability to control the outcome is eliminated, people see these negative outcomes as more equally likely. However, given the small sample in McKenna’s work, we suspect that this effect is multiply determined by both unrealistic optimism and the illusion of control and he was underpowered to detect the effect of unrealistic optimism when manipulating the level of control.

If we find unrealistic optimism in a fixed risk scenario where the risk manipulation is truly random, that would provide evidence for unrealistic optimism regardless of control. Additionally, while McKenna's work supports our generally theory about risk, he only measures perceived likelihood of a negative outcome in uncertain situations and does not extend this work to any behavioral outcomes.

On the other hand, theories on non-motivated sources of bias such as focalism and egocentrism cast doubt on the motivated accounts of unrealistic optimism. The focalism account suggests that when comparing relatively likelihood of person A being guilty or innocent compared to person B, people tend to focus on the facts that support the focal hypothesis (person A is *guilty* rather than person A is *innocent*) akin to a positive test strategy (Klayman & Ha, 1987; Moore & Kim, 2003). Focalism can explain belief reversals about who is more likely to be guilty when assessing comparative guilt or innocence compared to when people separately assess the subjective likelihood that person A or person B is guilty or innocent (Fox and Levav, 2000). While focalism is generally considered a non-motivated account, motivated reasoning also shapes which hypotheses people form in the first place (Kunda 1990). Alternatively, egocentrism drives people to consider more thoroughly their own risk-mitigating actions while neglecting other's risk mitigation efforts (Kruger & Burrus, 2004), an effect which is strengthened when the activity being assessed is something that people have frequent experience with (Weinstein 1980; Chambers et al, 2003). However, frequency can have an opposite effect, reducing unrealistic optimism, when elicited through absolute rather than comparative judgment (Price et al., 2002).

Another difference that is less evident between these motivated and non-motivated accounts of unrealistic optimism is whether people think they are less likely to have a bad thing happen to them because of lowered unconditional probability or lowered conditional probability.

For example, a person may think they are less likely to get cancer in general (unconditional), or less likely to get cancer given exposure to carcinogen (conditional). Previous research finds that both lay people and experts show similar biases in terms of overestimating the likelihood of unconditional probability (e.g. dying from a given condition) and condition probability (e.g. dying from a condition given one has the condition) (Wright et al 2002). While most comparative optimism studies do not often distinguish between unconditional and conditional probability, the latter is more directly tied to perceptions of vulnerability to threat. People tend to underestimate their own vulnerability to negative life events (e.g. divorce, disease, addiction) compared to an average, unidentified person (Perloff & Fetzer 1986). This effect attenuates when compared to a close friend or relative but holds when a comparing a specific other person to an average group member (Klar et al 1996). Taken together, these accounts suggest that when people think of a specific person, they can imagine the specific risk-mitigation measures that person may take to avoid victimization but fail to consider those measures when the target is non-specific.

While previous theories attempt to explain why this unrealistic optimism may be adaptive, Taylor & Brown (1988) note that while this optimism be adaptive for the individual, in the long run, this positive illusion "may lead a person to trample on the rights and values of others" (p. 204). For instance, people high in trait optimism are less bothered by illness symptoms (Scheier & Carver 1985), which may have positive implications for people's ability to cope. We suspect that such optimism may lead people to be less likely to take preventative measures such as wearing a mask when sick, which imposes harm on others. However, to our knowledge, this supposition that unrealistic optimism may lead to selfish risk taking has never been experimentally tested.

Present Research

We investigate people's willingness to impose risk on others and their beliefs about the subjective probability of a negative outcome. We test willingness to impose risk on others in games with fixed risk and scenarios of uncertainty, measuring both perceptions of risk and behavioral outcomes. Risky scenarios refer to situations where the likelihood of something happening can be expressed as a probability, whereas uncertain scenarios are those in which there is vague or ambiguous information that makes assigning a probability impossible. In general, risk perceptions and risk aversion are relatively tractable to estimate, but it is harder to precisely measure uncertainty aversion, and risk aversion and uncertainty aversion do not necessarily go hand in hand (Epstein 2004). Previous work on unrealistic optimism measures this in comparative judgements of uncertainty. Additionally, much of this previous work measures self-assessment, with less focused on behavioral outcomes, except for some notable examples (i.e. betting on self-assessments in Williams & Gilovich, 2000). Self-reports tend to be more sensitive to changes in risk than behavioral measures such as meliorative actions or risk mitigation actions (Loewenstein & Mather 1990). By testing this in both risky and uncertain contexts with self-reports of risk and behavioral measures, we highlight the robustness of this comparative optimism when imposing risks on others.

How might measuring interpersonal risk taking be different than other measures of unrealistic optimism? For one, humans do not like harming others and often act prosocially at their own expense (Bénabou & Tirole 2006). According to Moral Foundation Theory, harm aversion and concern for fairness seem to be universal human values, transcending political ideology (Graham et al., 2009). People are willing to spend more to reduce electric shocks to a stranger than to reduce their own shocks (Crockett et al., 2010). Similarly, it may feel more costly to impose risk on others than it is to benefit at someone else's expense. Given that harm caused by action is generally considered worse than harm caused by inaction (Cushman et al 2006), we expect this harm aversion to be particularly relevant to scenarios where we ask people to actively decide how much risk they are imposing on someone else (experiments 3 and 4). If harm aversion is a stronger determinant of behavior than unrealistic optimism, we would expect people to be willing to assume more risk for others than they would impose on others for their personal benefit. We might also expect people to rate risk they impose on others as greater because they find those outcomes as more aversive. Similarly, if unrealistic optimism is driven by an increased perception of invulnerability or immunity to threat (Perloff & Fetzer 1986), we would expect people to rate the hazards another person imposes on them to be comparatively less risky than the hazards they impose on others.

If unrealistic optimism is driven exclusively by the illusion of control, then we would expect people to impose more risk on others than they would assume themselves only in situations which they feel like they can control. Similarly, if unrealistic optimism is driven by non-motivated factors such as focalism or egocentrism, we also would not expect to find it in situations where the risk is fixed, and outcomes are randomly determined. Both accounts would predict finding more unrealistic optimism in situations of uncertainty where they can justify that they take better precautions against a given hazard (experiment 1 and 2), but not situations of fixed risk with outcomes determined by a random device (experiments 3 and 4). However, motivated unrealistic optimism driven by desirability predicts that people will systematically impose more risk on others than they would assume themselves because they bias their evidence accumulation to make the risks imposed on others seem less severe. If desirability bias is driving this effect, we should expect to see it in both cases of risk and uncertainty.

Experiment 1: Daycare Vignette

Method.

Participants. We recruited 401 participants through Cloud Research (M_{age} = 41.52, 42% female, 57% male). We report analyses of the 400 participants who passed the attention check.

Procedure and design. Participants were randomly assigned to a beneficiary condition in which either their child was sick (self) or someone else's child at their child's daycare was sick (other). They were also randomly assigned to a high risk condition in which participants were reminded of seasonal RSV symptoms or a low risk condition in which they were reminded of seasonal common cold symptoms. This experiment was preregistered on aspredicted.org (#161595).

Participants assigned to the condition in which their child was sick read, " Now, imagine that you were already at work, but your partner who also works, called to say that your child woke up with a cough. They ask for your advice on whether to bring the child to daycare because they have an important meeting at work today that they cannot miss. While you would love to help, it would be impossibly hard to miss work today too." Meanwhile, participants assigned to the condition in which another child was sick read, "Now, imagine that you were already at work, but another parent from your daycare called to say that their child woke up with a cough.

They ask for your advice on whether to bring the child to daycare because they have an important meeting at work today that they cannot miss. While you would love to help, it would be impossibly hard to miss work today too."

Participants then answered, "Would you recommend bringing the child to daycare?" on a seven point scale ranging from *Definitely not* to *Yes, definitely.* Then they answered, "Would you recommend bringing the child to the doctor and/or Urgent Care?" on the same seven point scale. Then they rated "How much risk does this child pose to the other children at daycare?" on a seven-point scale from *Not Risky at All* to *Extremely Risky*. Finally, participants answered demographic questions about their age and gender and then rated how relatable this scenario was to their day-to-day life on a seven point scale.

Results.

Participants were less likely to recommend sending another person's child to daycare (M=2.90, SD = 1.72) than their own child (M=3.47, SD=1.89). In a linear model controlling for beneficiary, risk and their interaction, participants were more likely to recommend bringing their own child to daycare than the other child, B= 0.71, SE =.26, 95% CI [0.19, 1.23], t(396)=2.70, p=.007. Participants were also more likely to recommend sending the child daycare when reminded of the low risk symptoms, B= 0.50, SE =.25, 95% CI [0.01, 1.02], t(396)=1.99, p=.047. There was no significant interaction between risk and beneficiary, B= -0.26, SE =.36, 95% CI [-.97, .45], t(396)=-.71, p=.48.

Participants rated it riskier to send another child to daycare (M=5.24, SD=1.31) than their own child (4.96, SD=1.32). In a linear model controlling for beneficiary and risk, the riskiness of sending the child to daycare as less risky when it was their own child, B= -0.27, SE =.13, 95%

CI [-0.53, -0.02], t(397)=-2.10, p=.036, and less risky in the low risk condition, B= -0.32, SE =.13, 95% CI [-.58, -0.07], t(397) = -2.42, p=.014. See Figure 2.1.

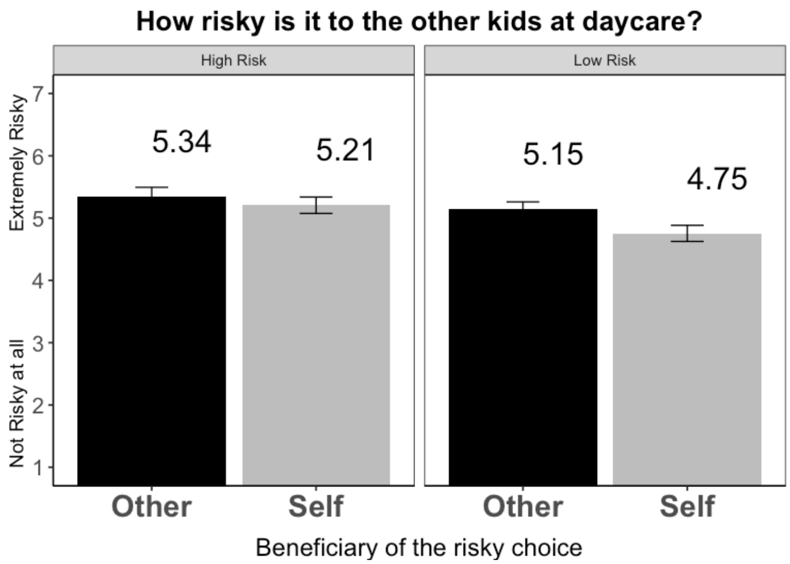

Figure 2.1. Subjective risk of sending child to daycare based on beneficiary and risk level in Experiment 1.

Participants rated the child as requiring less urgent care when the risk was low, B= -0.78, SE =.18, 95% CI [-1.14, -0.42], t(399)=-4.30, p<.001, but no statistically significant difference in urgency between self and other, B= -0.30, SE =.18, 95% CI [-.67, 0.06], t(397) = -1.64, p=.10. Finally, we pre-registered exploratory analysis controlling for gender and relatability. We found that males in general were more likely to recommend sending the child regardless of beneficiary condition, B= 0.43, SE =.18, 95% CI [0.07, 0.79], t(398)=2.33, p=.02, and assessed the risk as lower regardless of beneficiary condition, B= -0.27, SE =.13, 95% CI [-0.53, -0.00], t(398)=2.00, p=.046. We also found that in general, the more relatable this was, the more likely participants were to send their kid to daycare, B=.22, SE=.04, 95% CI [0.03, 0.20], t(399)=2.71, p =.0069.

We also found that the more relatable it was, it seemed marginally less risky, B= -0.05, SE =.03, 95% CI [-0.11, 0.00], t(399)=-1.76 p=.079.

Discussion.

Experiment 1 supports the two primary hypotheses of this paper. First, people are more willing to impose risk on others than they are to assume risk themselves for another's benefit. Second, people assume the risks they impose on others are less severe than the risks others impose on them. These findings support unrealistic optimism as the mechanism for why we impose more risk on others than we are willing to assume ourselves. However, in this experiment, participants always report subjective risk after they report whether they would send the child to daycare, so the elevated subjective risk to others might just be a post-hoc justification. In Experiment 2, we test the robustness of these findings in a different context and vary the order of soliciting willingness to accept the risk and the subjective risk level.

Experiment 2: Scooter Vignette

Method.

Participants. We recruited for two weeks at the University Lab, recruiting 619 participants (M_{age} = 20.64, 45% female, 54% male). We exclude the 7.27% who failed the attention check (n = 574).

Procedure and design. Participants were randomly assigned to one of two between subject beneficiary condition (self or other), and two between subject order conditions (soliciting subjective risk before or after the primary dependent variable). Then participants read and responded to three vignettes in a fixed order: selling a scooter on an online marketplace, texting while driving, cheating on homework. In addition to the within subject vignettes, we also varied the risk level of the scenario (low vs high) within subject. This experiment was preregistered on aspredcited.org (#172607) Per our pre-registration, we only report the results from the first vignette, but all exploratory results are available in the appendix.

In the first scenario, participants read that they were either the buyer or the seller of an electric scooter on an online platform like Facebook Marketplace. When participants were the buyer, the beneficiary of *not* disclosing the risk was the other person. Conversely, when participants were the seller, they themselves were the beneficiary of *not* disclosing the risk. In the low risk condition, participants read that the scooter had a faulty headlight that the buyer may notice on first inspection. In the high risk condition, the defect was a loose wheel. Participants answered, "How risky is it to not disclose the defect?" on a seven-point scale from *Not Risky at All* to *Extremely Risky*, and "If the buyer does not ask directly, must the buyer disclose the defect?" on a six-point scale ranging from *Definitely not* to *Yes, definitely.* The order of these questions depended on the order condition participants were assigned to.

Results.

We conducted a manipulation check to confirm that participants in fact rated the high risk condition (a loose wheel) as riskier than the low risk condition (a faulty headlight), B=0.39, SE=0.12, 95% CI [0.17, 0.62], $t(572) = 3.42$, $p < .001$. First, participants thought it was less necessary to disclose information about the defective scooter when they were the seller (M=4.65, SD=1.38) rather than the buyer (M=4.92, SD =1.31). In a linear model controlling for beneficiary and risk condition, participants reported that the seller was less obligated to disclose the risk when they were the seller compared to the buyer, B=-0.28, SE = .11, 95% CI [-0.50, -0.06], t(571)=-2.47, p=0.014.

Second, conceptually replicating the results of experiment 1 when we asked participants about subjective risk after we asked whether the seller must disclose. In this case, participants thought selling the defective scooter was less risky when they were the seller (M=4.72, SD=1.58) versus the buyer (M=5.18, SD=1.36), t(287)=2.65, p=.0084. Importantly, when we interact the beneficiary condition with order, we find a significant interaction of beneficiary and order, B=0.52, SE=.23, 95% CI [0.06, 0.97], t(570)=2.24, p=.025. Therefore, when participants evaluate the risks before they respond to whether the seller must disclose, this difference in perceived risk between being the buyer (M=5.04, SD=1.22) or seller (M=5.11, SD=1.35) disappears, t(272)=-.40, p=.69. See Figure 2.2.

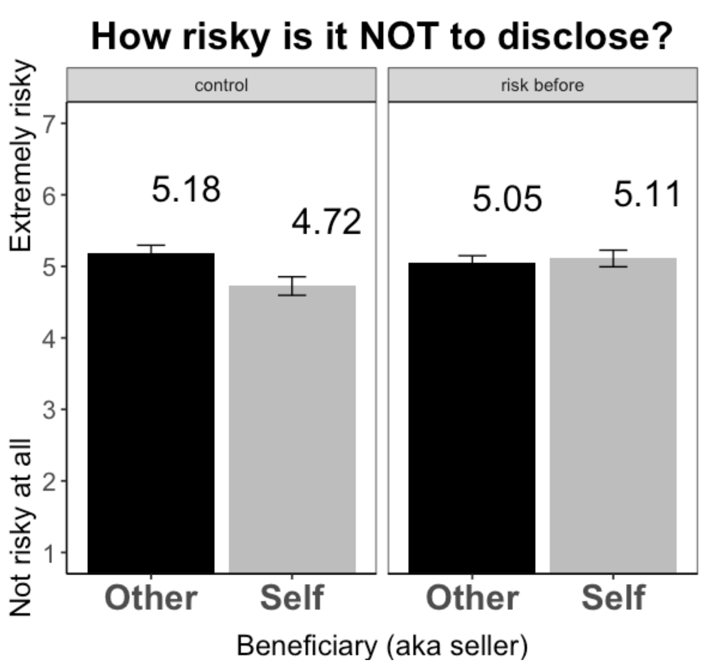

Figure 2.2. Subjective risk of not disclosing the defect based on seller and order in Experiment 2.

Discussion.

Experiment 2 supports the hypothesis that people are more willing to impose risk on others than they are to assume risk themselves for another's benefit. We also find that people tend to downgrade the risk they impose on others when asked after the fact. However, when they evaluate the risks prior to their decision, we do not find evidence of differential risk between self and other. This provides evidence of a motivated determinant of risk. Once people make a choice, they have an outcome they want to support, which may bias which evidence factors into their subjective risk calculation. When participants have not made a decision yet and not endorsed one particular outcome, their subjective risk calculation is less biased.

In these first two experiments, participants are evaluating the subjective risk of inherently uncertain, ambiguous situations. While we can speak to comparative optimism at the group level for self-versus-other, we cannot speak to absolute optimism since we do not know the true risk level for these negative outcomes. There may in fact be information asymmetries or individual differences in susceptibility to risk that make the downgrading of subjective risk accurate for certain individuals. Therefore, in the next experiment we develop an incentive compatible game with fixed risk to test whether people are more willing to impose risk on others and whether they still downgrade the subjective risks they are imposing on others.

Experiment 3: Incentive Compatible Spin the Wheel Game

Method.

Participants. We recruited 805 participants through Connect (M_{age} = 39.31, 47.2% female, 51.2% male). All participants passed the attention check.

Procedure and design. Participants were randomly assigned to one of two between subject beneficiary condition (self or other), one of two risk conditions (high or low) and two between subject order conditions (soliciting subjective risk before or after the wager). This experiment was preregistered on aspredicted.org (#186199).

To start, we endow all players and their randomly assigned partner $10 each to play a single spin the wheel game with letters A to Z. We instruct participants that one randomly assigned pair of participants will be paid based on the results of their game. The rules of the game are as follows: one player must decide how much their team is willing to wager ($0 to $10) on the wheel spin, but when they decide unilaterally, they commit both players to the same wager. Only one of the two players, the designated beneficiary, will automatically earn three times the wager, regardless of the wheel spin. The critical manipulation is the beneficiary condition. When participants are the beneficiary, they get to decide how much their team will gamble ($0 to $10). When the other person is the beneficiary, the participant gets to set a limit on how much their randomly assigned partner can gamble ($0 to $10). Then, the participant spins the wheel.

If the wheel lands on a winning letter, both players get to keep their wager, and the designated beneficiary earns the additional automatic three times bonus. If the wheel lands on a losing letter, both players lose their wager, but the designated beneficiary still keeps the automatic three times bonus. In the low risk condition, spinning any consonant wins, so the probability of losing is 5/26 ($P_{losing} = .19$). In the high risk condition, a vowel wins, so the probability of losing is 21/26 ($P_{losing} = .81$). We specified in the game that "y" was not a vowel.

With this setup, we ensure that one player faces a guaranteed benefit ($(3 - P_{losing})*W)$) while another player faces only a negative expectancy gamble ($0- P_{losing}*W$) compared to their initial endowment. We hold the guaranteed beneficiary multiplier constant and relatively high compared to the negative expectancy gamble between conditions so changes in wager are driven primarily by sensitivity to the risk imposed on the non-beneficiary. We include one comprehension check with a sample wager and spin and have participants answer how much each player would take home. Participants had to pass this comprehension check before they could proceed to play.

Finally, to test whether the distortion of subjective risk varies between self and other, we ask participants what they think the chances of losing are on a slider scale from 0 to 100. Similar to Experiment 2, we vary whether we ask about the subjective odds before or after the wager.

Results.

First, participants wagered more when they were the beneficiary ($M=\$6.46$, $SD=3.23$) then when they set the limit as the non-beneficiary ($M=\$5.08$, $SD=3.42$), $t(803)=5.87$, $p<.001$. There was also a main effect of risk condition, whereby people wagered more when the risk was low ($M=\$6.21$, $SD=3.23$) versus high ($M=\$5.33$, $SD=3.41$), $t(803) = 3.70$, $p<.001$. There was no significant interaction between risk condition and beneficiary, $p=.37$. See Figure 2.3.

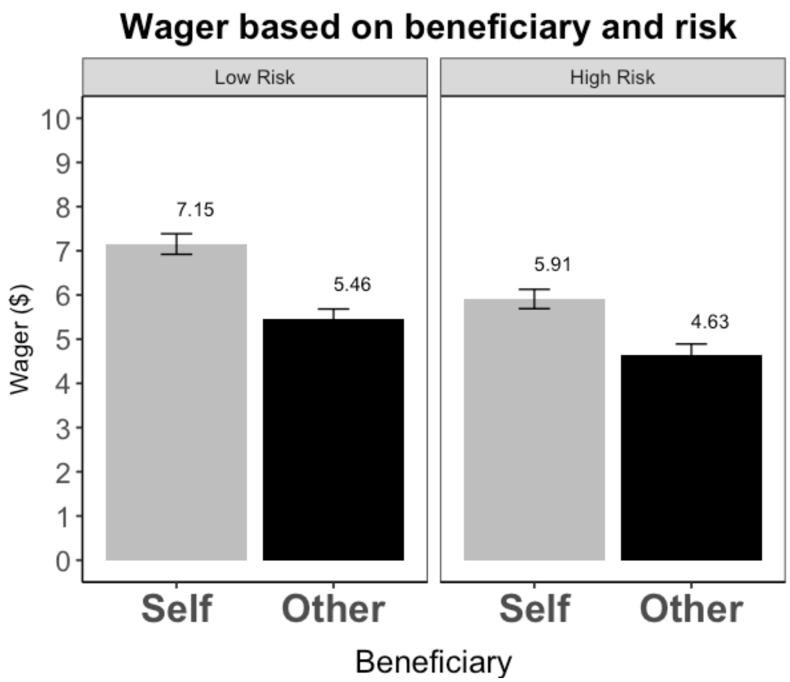

Figure 2.3. Wager based on beneficiary and risk in Experiment 3.

In terms of the subjective chances of losing the spin, we find evidence of unrealistic comparative optimism between self and other. Based on a linear regression controlling for risk condition, we find that participants rate the chances of losing as approximately 6% higher when the other person is the beneficiary and they are being asked to assume risk for others, B=6.05, SD = 1.39, 95% CI [3.33, 8.78], t(802) = 4.36, p <.001. See Figure 2.4.

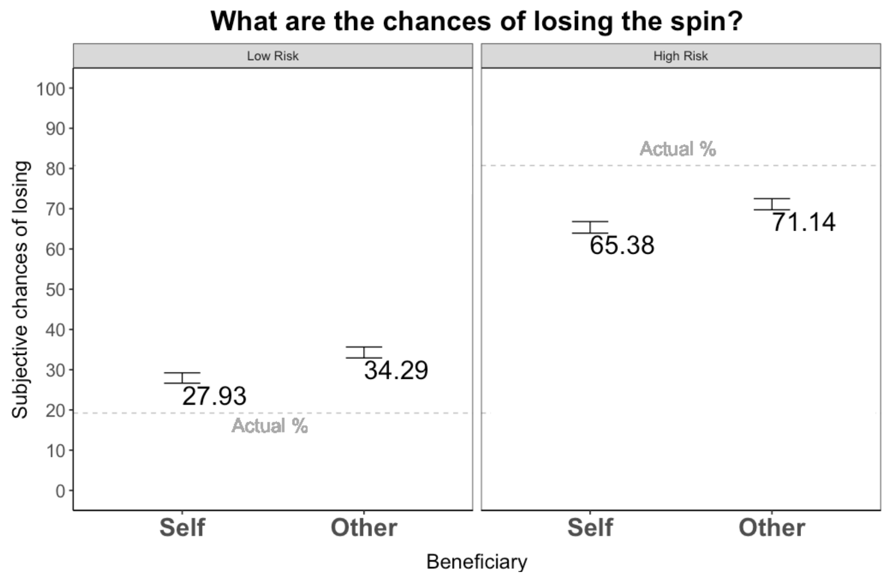

Figure 2.4. Subjective odds of losing the spin based on beneficiary and risk in Experiment 3.

Discussion.

The results about how participants distort their subjective chances of losing the spin suggest that unrealistic optimism is not just an artifact of uncertainty where a person may have privileged information about their individual risk reducing factors. We find a similar pattern of results in a context when risk is fixed and calculable, which we argue supports a more motivated, rather than non-motivated, explanation of unrealistic optimism.

To understand what is driving these motivated biases, we conduct exploratory analysis on the bias. We observed in Figure 2.5 that the bias between subjective chances of losing and actual chances of losing was lowest in the low risk / self condition and the high risk/other condition. In other words, people were most accurate in these two conditions. In fact, we find a significant interaction of beneficiary and risk on bias (measured as the absolute value of the actual chances of losing minus the reported chances), $B= -9.64$, $SD=2.52$, 95% CI [-14.59, -4.70], $t(801)=-3.83$, $p<.001$.

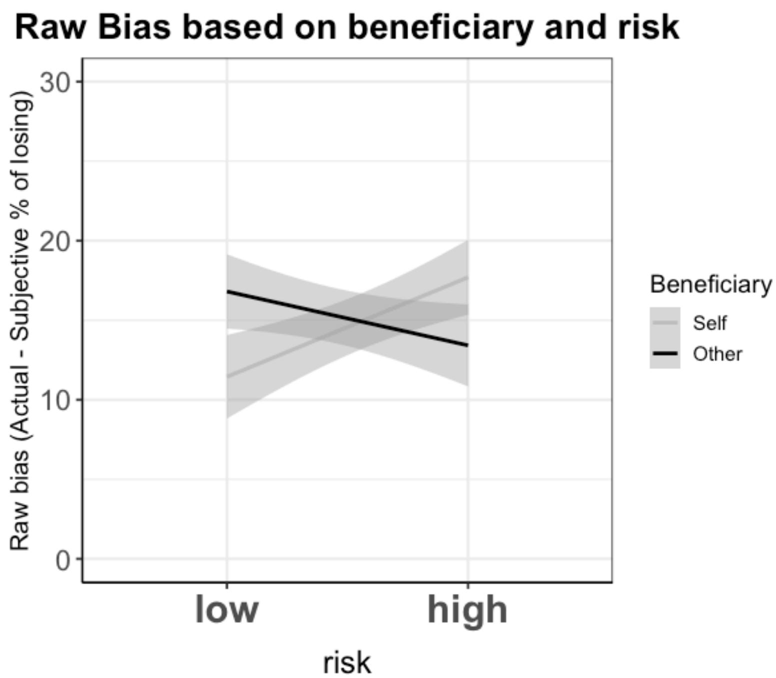

Figure 2.5. Interaction plot of raw bias based on beneficiary and risk in Experiment 3.

Upon further investigation, we find evidence that this debiasing in these two conditions seems to be driven by selfishness. We measure selfishness as how much money a participant risks when they are the beneficiary (their wager), and how much money they refuse to risk ($10-wager) when the other person is the beneficiary. Therefore, someone who is a 10 on this selfish scale risks all their partner's money for their own benefit or risks none of their own money for their partner's benefit. We see in Figure 2.6 that for selfless participants, there is no significant difference between bias for self and other, regardless of risk. However, selfish participants seem to construe the odds in a motivated manner. When assuming risk for others, people tend to amplify their biases to avoid assuming low risks for others, and reduce their biases to justify not assuming high risks for others. In a linear regression of selfishness, beneficiary condition, and risk on bias, we find a significant interaction of selfishness and other on bias, $B=1.47$, $SD=.56$, 95% CI [.38, 2.47], $t(797) = 2.64$, $p=.008$, whereby when assuming risk for others, selfishness makes people overinflate the odds of losing. Additionally, we find a three way interaction of selfishness, other, and high risk on odds, B=-2.40, SD =.77, 95% CI [-3.90, -0.89], t(797) = -3.13, p=.002, suggesting that when the risks are high, this bias reduces. The opposite pattern is true when opposing risk on others. When imposing risk on others, selfishness reduces people's perceptions they will lose the spin. However, when the risk is high and losing is likely, selfish people's biases increase to justify the fact they are still risking someone else's money.

Figure 2.6. Interaction plot of raw bias based on selfishness, beneficiary and risk in Experiment 3.

We cannot disentangle whether selfish preferences are driving this motivated reasoning, or whether the bias interpretation of the risk drives the selfish behavior. When we interact order with this model and find no significant interaction with any of our predictors.

Experiment 4: Incentive Compatible Die Roll Game

Method.

Participants. We recruited 1000 participants through Connect (M_{age} = 42.79, 50% female, 48.6% male). All participants passed the attention check.

Procedure and design. Participants were randomly assigned to one of two between subject beneficiary condition (self or other), one of two risk conditions (high or low). This experiment was preregistered on aspredicted.org (#161641).

To start, we endow all players and their randomly assigned partner $10 each to play a single die roll game. We instruct participants that one randomly assigned pair of participants will be paid based on the results of their game. The rules of the game are as follows: each participant must decide how much their team is willing to wager ($0 to $10) on the die roll, but when they decide unilaterally, they commit both players to the same wager. Only one of the two players, the designated beneficiary, would automatically earn three times the wager, regardless of the wheel spin. Unlike in Experiment 3, each participant decided how much the team would wager, regardless of beneficiary condition, so there was no limit setting manipulation. Then, the participant rolls the die.

If the die rolled a winning number, both players get to keep their wager, and the designated beneficiary earns the additional automatic three times bonus. If the die rolled a losing number, both players lose their wager, but the designated beneficiary still keeps the automatic three times bonus. In the low risk condition, rolling a six loses, so the probability of losing is 1/6 (P_{losing} = .17). In the high risk condition, rolling a six wins, so the probability of losing is 5/6 (P_{losing} = .83).

Similar to the setup in experiment 3, we ensure that one player faces a guaranteed benefit while another player faces only a negative expectancy gamble compared to their initial endowment. We again include one comprehension check with a sample wager and die roll and have participants answer how much each player would take home. Participants had to pass this comprehension check before they could proceed to play.

Instead of directly soliciting the odds of losing which we found subject to motivated reasoning in experiment 3, we asked participants to rate how important certain factors were in their decision making on a seven point scale: *my total payout*, *my partner's total payout*, *doing what is right*, and the *odds of rolling a losing number*.

Results.

First, participants wagered more when they were the beneficiary (M=\$6.66, SD=3.64) than when they were the non-beneficiary (M=\$4.31, SD=3.74), $t(998)=10.03$, $p<.001$. There was also a main effect of risk condition, whereby people wagered more when the risk was low (M=\$6.36, SD=3.63) versus high (M=\$4.58, SD=3.89), $t(998) = 7.52$, $p<.001$. We also find a marginally significant interaction between risk and beneficiary, indicating that people are less sensitive to the high risk when they are the beneficiary, B=.85, SE=.45, 95% CI[-0.03, 1.74], t(996), $p=.059$. See Figure 2.7.

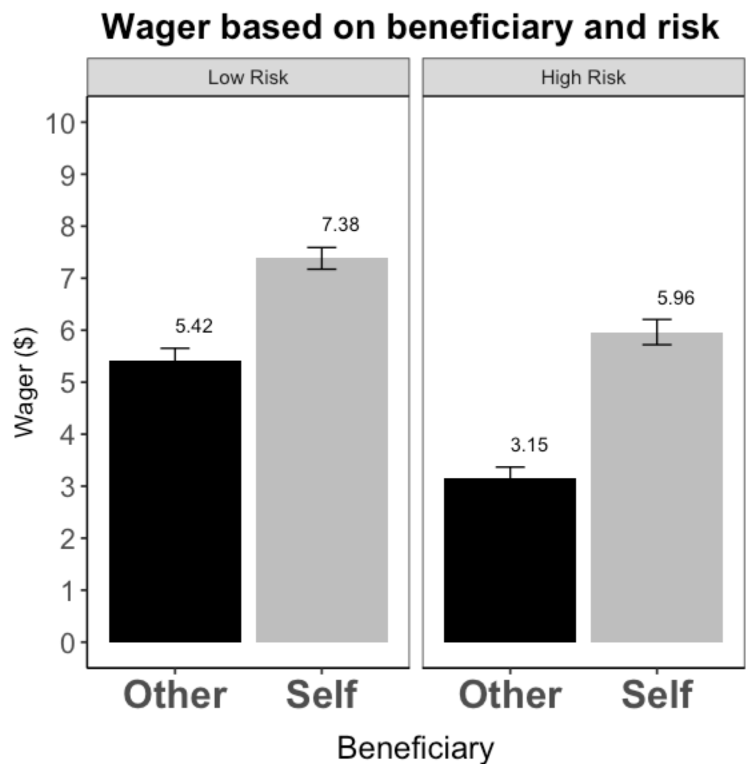

Figure 2.7. Wager based on beneficiary and risk in Experiment 4.

When analyzing participants' self-reports of what factored most into their decision making, the importance of *the odds of losing* seems to best explain why people are less sensitive to high risk when they are the beneficiary. In a linear model predicting the odds of losing based on risk and beneficiary condition, we find a significant interaction showing that participants are least likely to factor the odds of losing when the risks are high and they are the beneficiary, $B=-.79$, $SE=.22$, 95% CI [-1.22, -0.35], $t(996) = -3.54$, $p<.001$. See Figure 2.8.

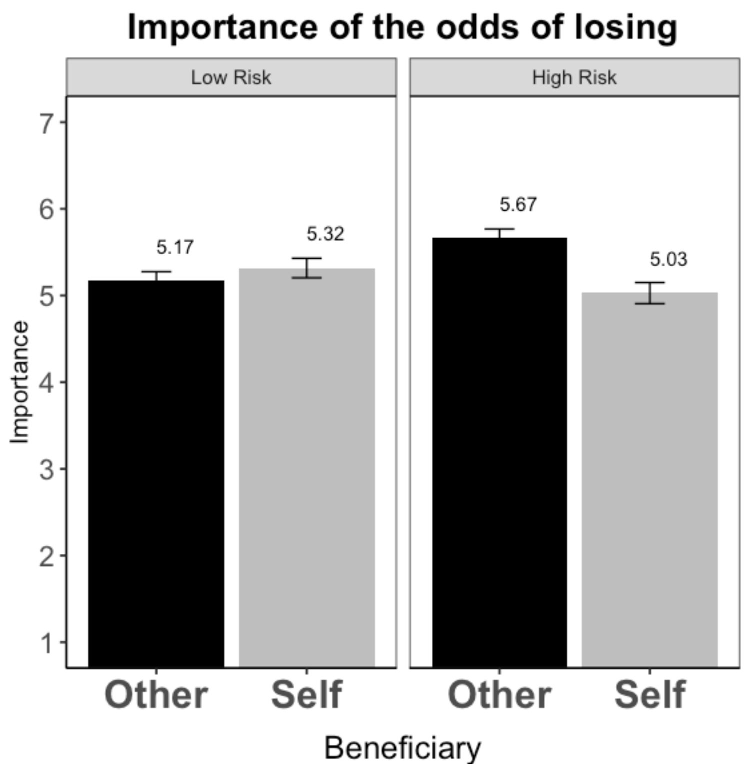

Figure 2.8. Importance of the odds of losing based on beneficiary and risk in Experiment 4.

Discussion.

Once again, we find that people are willing to impose more risk on others than they would accept themselves. We also find limited evidence that this gap is larger as the risks are higher, suggesting that people may be willing to assume small risks for others, but are much more sensitive to rising risks compared to when imposing risk on others. What are the effects of this differential sensitivity to risk? To answer this question, we will examine the expected value to the non-beneficiary based on the median wager by condition. When the risks are low and participants bet \$7.38 of someone else's money for their personal benefit, they are imposing an expected \$1.23 of loss on the other person. Conversely, under the same risk conditions, they only risk \$5.42 of their own money which is an expected \$0.90 they are willing to lose. However, when the risks are high and participants bet \$5.96 of someone else's money for their personal gain, they are imposing an expected \$4.97 of loss on another person. On the other hand, they would risk only $3.15 of their own money for another person's benefit, indicating they are willing to lose only $2.63. See Figure 2.9. The interaction of risk and beneficiary on the expected value of the non-beneficiary is significant, B=-2.02, SE=.28, 95% CI [-2.56, -1.48], t(996), p<.001, reflecting that when the risks are high, people will impose 1.89 times as much cost on someone else than they would be willing to accept themself.

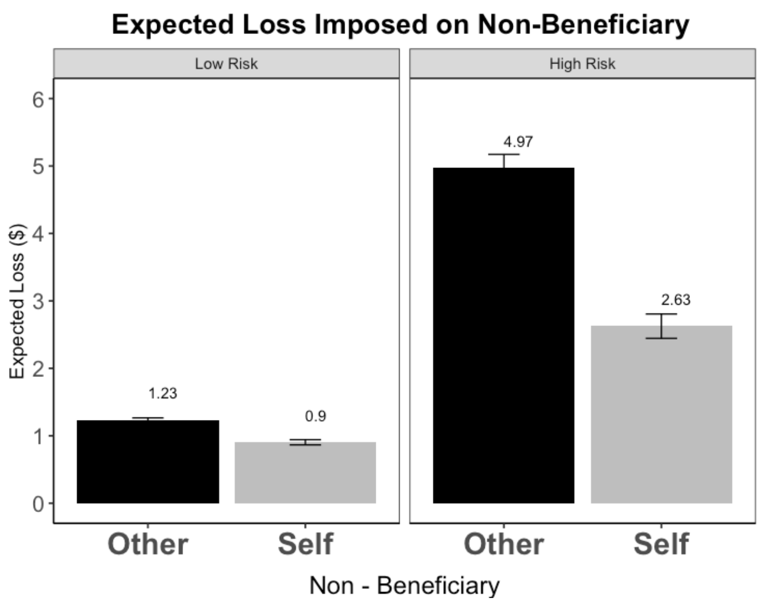

Figure 2.9. Expected loss imposed on the non-beneficiary in Experiment 4.

While experiment 3 provides evidence that people are systematically biasing the odds of losing when imposing high risk on others, experiment 4 provides evidence that people seem to be underweighting the subjective probability when imposing high risks on others. Both explanations speak to an overarching mechanism consistent with a motivated interpretation of risk, biasing the search for evidence to support people's selfish intentions to impose risk on others for their own benefit.

General Discussion.

In our investigation of how people make decisions involving interpersonal risk, we find consistent evidence that people tend to impose more risk on others than they would accept themselves. We find that people are more likely to underweight the high risks they impose on others, in both risky and uncertain situations. This differential perception of risk seems to be exacerbated the more selfishly people behave. Reflecting on the risk level before people deciding may mitigate unrealistic optimism. Finally, we find evidence that people are more sensitive to risk level when they are assuming it for someone else rather than imposing it on others. Taken together, this supports the account of motivated unrealistic optimism, whereby people strategically bias their perceptions of the likelihood of a negative outcome to support their selfish desires.

We contribute to the literature on perceptions of risk by showing direct behavioral consequences of unrealistic optimism. Our work also sheds light on the motivated mechanism behind unrealistic optimism, an account on which previous work has cast doubt. While we do not rule out other drivers of unrealistic optimism, by showing that we find our effect in both situations of uncertainty and fixed risk, we contribute to theory of unrealistic optimism by providing further evidence that it is not exclusively driven by the illusion of control (Langer & Roth, 1975) or non-motivated factors (Chambers et al., 2003; Chambers & Windschitl, 2004).

This evidence has implications for how we make interpersonal judgments. We suspect that we judge the mother who sends their symptomatic child to daycare or the coworker who is constantly sniffling in a meeting while not wearing a mask more harshly than we would judge ourselves. Based on our findings, we suspect that chances are if we were equally symptomatic, we too would be equally likely to not take appropriate precautions because we assume we are less symptomatic. Furthermore, we suspect that awareness of this asymmetric risk perceptions may reshape how quickly we are to judge others, which might prevent interpersonal conflict.

We also suspect that these findings have broader implications for the risks we tend to underweight when our actions or policies impose potential harm on others. At the individual level, we likely think our own actions (e.g. purchasing cheaply made clothes made by low wage workers, using non-sustainable products, texting while driving, asking someone for preferential treatment) are less harmful to others or the planet than when someone else does it. At the policy level, when considering where to build a new power plant, people would tend to minimize those risks as long as it's not in their community. When deciding where and when to employ troops in combat, people would likely minimize those risks if they have no skin in the game. Extrapolating our findings to these broader contexts underscores the dangers of how unrealistic optimism can lead to us to behave in antisocial ways (Taylor & Brown, 1988).

Limitations and Future Directions.

While we present evidence that people assume more risk for others because of unrealistic optimism, we cannot rule out that unwillingness to assume risk for others is driven by strategic pessimism (Krizan & Windschitl, 2007). While unrealistic optimism is a robust, highly replicated phenomenon, people are not always optimistic about the odds of good things happening to them. As previously mentioned, people exhibit less unrealistic comparative optimism when the hazards are uncontrollable and when comparing themselves to people more like themselves (Helweg-Larsen & Shepperd, 2001). Individual differences in anxiety also moderate unrealistic optimism. People high in anxiety tend to exaggerate probability of a negative event and show attentional biases toward threatening stimuli (Mitte 2007; Eysenck et al

2007). Anxiety reduces the function of the top-down, goal-directed attention system, and increased the influence of the bottom-up stimulus driven system (Corbetta & Shulman, 2002; Posner & Petersen, 1990).

Whether people exhibit optimism or pessimism may also be driven by frequency and outcome desirability. When comparing their personal likelihood of experiencing a series of events which ranged in both desirability and rarity compared to the average person, people tend to overestimate their chances of experiencing common events, while underestimate their chances of rare events (Kruger & Burrus, 2004), which is consistent with the general tendency to overweight low probabilities and underweight high probabilities. However, this manifests in unrealistic optimism for common desirable (e.g. graduating with a 3.0 GPA) events and rare undesirable events (e.g. gaining five pounds in the next five days), but in unrealistic pessimism for common undesirable events (e.g. gaining five pounds by age 50) and rare desirable events (e.g. graduating with a 4.0 GPA).

This suggests that whether someone exhibits unrealistic optimism or pessimism might be driven by how easy it is to imagine one outcome versus another (Kahneman & Miller, 1986). We suspect that when imposing risk on others, it is more likely that people focus on the positive outcome for themselves and less likely that they consider the negative outcome. Conversely, when assuming risk for another, the immediate visualization is of the negative outcome for oneself and not on the positive benefit to the other. Focusing on this prefactual may lead to differential scrutiny of unfavorable evidence when considering assuming risk for someone else's benefit (Kruglanski, 1996). Future studies should compare people's perception of risk when considering imposing risk on or assuming risk for another versus a control condition in which they are not personally invested. Depending on which interpersonal risk condition looks more similar to the control would help isolate whether differences in perceived risk are driven by unrealistic optimism as the risk taker or strategic pessimism as the person bearing the risk for another.

Another limitation of our research is we do not directly measure affect, but indirectly manipulate it by varying the benefit to self versus other. We suspect that motivated unrealistic optimism is driven in part by people's desire to feel better about the risks they impose on others. Future research should consider directly measuring or manipulating the affect associated with these decisions involving interpersonal risk. When imposing risk on others, guaranteed benefits likely evoke positive affect while imposing potential harm to someone else evokes conflicting negative affect. Conversely, when assuming risk for others, the negative expectancy gamble likely evokes negative affect because of both the impending loss and the unfairness of the outcomes. However, when the risks are low, we suspect that the positive affect associated with the warm glow of giving (Andreoni, 1990) counteracts the negative affect and leads to more prosocial behavior (Bénabou & Tirole 2006). Warm glow may explain why people in experiment 4 tend to both wager and weight the odds of losing more similarly in the low risk conditions, but this effect wears off as risk level rises. Future research should explore how people affectively respond to these interpersonal gambles.

Conclusion.

Many of the decisions we make in life are not strictly single-agent decisions where I consider my own risk against my own benefit. Our experiments document how unrealistic optimism drives us to systematically impose more risk on others than we are willing to accept ourselves. Having people assess the risks before they act may mitigate this bias.

ACKNOWLEDGEMENTS

Chapter 2, in full, is currently being prepared for submission for publication of the material with On Amir. The dissertation author was the primary investigator and author of this paper.

REFERENCES

Ajzen, I., & Madden, T. J. (1986). Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of experimental social psychology*, *22*(5), 453-474.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of personality and social psychology*, *68*(5), 804.

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal*, *100*(401), 464-477.

Alhakami, A. S., & Slovic, P. (1994). A psychological study of the inverse relationship between perceived risk and perceived benefit. *Risk analysis*, *14*(6), 1085-1096.

Averill, J. R. (1973). Personal control over aversive stimuli and its relationship to stress. *Psychological bulletin*, *80*(4), 286.

Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning*, *1*(1), 71-103.

Bell, D. E. (1983). Risk premiums for decision regret. *Management Science*, *29*(10), 1156-1166.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. American economic review, 96(5), 1652-1678.

De Bruin, W. B., & Bennett, D. (2020). Relationships between initial COVID-19 risk perceptions and protective health behaviors: a national survey. *American Journal of Preventive Medicine*, *59*(2), 157-167.

Chambers, J. R., Windschitl, P. D., & Suls, J. (2003). Egocentrism, event frequency, and comparative optimism: When what happens frequently is “more likely to happen to me”. *Personality and Social Psychology Bulletin*, *29*(11), 1343-1356.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological bulletin*, *130*(5), 813.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, *107*(40), 17433-17438.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological science*, *17*(12), 1082-1089.

Epstein, L. G. (2004). A definition of uncertainty aversion. In *Uncertainty in economic theory* (pp. 171-208). Routledge.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion*, *7*(2), 336.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of behavioral decision making*, *13*(1), 1-17.

Fischhoff, B., Slovic, P., Lichtenstein, S., Read, S., & Combs, B. (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy sciences*, *9*, 127-152.

Folkman, S. (1984). Personal control and stress and coping processes: a theoretical analysis. *Journal of personality and social psychology*, *46*(4), 839.

Fox, C. R., & Levav, J. (2000). Familiarity bias and belief reversal in relative likelihood judgment. *Organizational Behavior and Human Decision Processes*, *82*(2), 268-292.

Gigerenzer, G. (2004). Dread risk, September 11, and fatal traffic accidents. *Psychological science*, *15*(4), 286-287.

Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of personality and social psychology*, *64*(4), 552.

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

Hoorens, V., & Buunk, B. P. (1993). Social comparison of health risks: Locus of control, the person-positivity bias, and unrealistic optimism. *Journal of Applied Social Psychology*, *23*(4), 291-302.

Howarth, C. I. (1988). The relationship between objective risk, subjective risk and behaviour. *Ergonomics*, *31*(4), 527-535.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational behavior and human decision processes*, *67*(3), 247-257.

Johnson, D. D., & Fowler, J.H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317-320.

Johnson, E., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45,* 20-31.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological review*, *93*(2), 136

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive psychology*, *3*(3), 430-454.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 363-391.

Klar, Y., Medding, A., & Sarel, D. (1996). Nonunique invulnerability: Singular versus distributional probabilities and unrealistic optimism in comparative risk judgments. *Organizational Behavior and Human Decision Processes*, *67*(2), 229-245.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological review*, *94*(2), 211.

Klein, W. M., & Kunda, Z. (1993). Maintaining self-serving social comparisons: Biased reconstruction of one's past behaviors. *Personality and Social Psychology Bulletin*, *19*(6), 732-739.

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological bulletin*, *133*(1), 95.

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology*, *40*(3), 332-340.

Kruglanski, A. W. (1996). Motivated social cognition: Principles of the interface. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 493–520). New York: Guilford Press.

Kunda, Z. (1990). The case for motivated reasoning. Psychological bulletin, 108(3), 480.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of personality and social psychology*, *32*(6), 951.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational behavior and human decision processes*, *65*(3), 272-292.

Loewenstein, G., & Mather, J. (1990). Dynamic processes in risk perception. *Journal of Risk and Uncertainty*, *3*, 155-175.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. Psychological bulletin, 127(2), 267.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, *92*(368), 805-824.

McKenna, F. P. (1993). It won't happen to me: Unrealistic optimism or illusion of control?. *British journal of psychology*, *84*(1), 39-50.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychological Science*, *8*(6), 423-429.

Mellers, B., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of experimental psychology: General*, *128*(3), 332.

Mitte, K. (2007). Anxiety and risky decision-making: The role of subjective probability and subjective costs of negative events. *Personality and Individual Differences*, *43*(2), 243-253.

Monat, A. (1976). Temporal uncertainty, anticipation time, and cognitive coping under threat. *Journal of Human Stress*, *2*(2), 32-43.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of personality and social psychology*, *85*(6), 1121.

Perloff, L. S., & Fetzer, B. K. (1986). Self–other judgments and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, *50*(3), 502.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience*, *13*(1), 25-42.

Puri, M., & Robinson, D. T. (2007). Optimism and economic choice. *Journal of financial economics*, *86*(1), 71-99.

Price, P. C., Pentecost, H. C., & Voth, R. D. (2002). Perceived event frequency and the optimistic bias: Evidence for a two-process model of personal risk judgments. *Journal of Experimental Social Psychology*, *38*(3), 242-252.

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological science*, *12*(3), 185-190.

Savadori, L., & Lauriola, M. (2021). Risk perception and protective behaviors during the rise of the COVID-19 outbreak in Italy. *Frontiers in psychology*, *11*, 577331.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: assessment and implications of generalized outcome expectancies. *Health psychology*, *4*(3), 219.

Sharot, T. (2012). *The Optimism Bias: Why we're wired to look on the bright side*. Hachette UK.

Shepperd, J. A., Klein, W. M., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science*, *8*(4), 395-411.

Skinner, E. A. (1996). A guide to constructs of control. *Journal of personality and social psychology*, *71*(3), 549.

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current directions in psychological science*, *15*(6), 322-325.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European journal of operational research*, *177*(3), 1333-1352.

Starr, C. (1969). Social benefit versus technological risk: what is our society willing to pay for safety?. *Science*, *165*(3899), 1232-1238.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychological bulletin*, *103*(2), 193.

Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of personality and social psychology*, *45*(1), 74.

Thompson, S. C. (1981). Will it hurt less if I can control it? A complex answer to a simple question. *Psychological bulletin*, *90*(1), 89.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological review*, *110*(3), 403.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological review*, *102*(2), 269.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, *5*(2), 207-232.

Vieites, Y., Ramos, G. A., Andrade, E. B., Pereira, C., & Medeiros, A. (2021). Can self-protective behaviors increase unrealistic optimism? Evidence from the COVID-19 pandemic. *Journal of Experimental Psychology: Applied*, *27*(4), 621.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of personality and social psychology*, *39*(5), 806.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, *2*(4), 343.

Winkielman, P., & Zajonc & Norbert Schwarz, R. B. (1997). Subliminal affective priming resists attributional interventions. *Cognition & Emotion*, *11*(4), 433-465.

Wright, G., Bolger, F., & Rowe, G. (2002). An empirical test of the relative validity of expert and lay judgments of risk. *Risk Analysis: An International Journal*, *22*(6), 1107-1122.

Zakay, D. (1984). The influence of perceived event's controllability on its subjective occurrence probability. *The Psychological Record*, *34*, 233-240.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of behavioral decision making*, *12*(2), 93-106.

Chapter 3

A FIELD EXPERIMENT ON THE INFLUENCE OF STREAK TRACKING ON SOCIAL MEDIA USAGE

Kathryn Hillegass, University of California San Diego, USA
Jackie Silverman, University of Delaware, USA
Rachel Gershon, University of California Berkeley, USA
David Grüning, Heidelberg University, Germany
Frederik Riedel, one sec app
Donatus Wolf, University of Potsdam, Germany

Rady School of Management, University of California, San Diego, La Jolla, CA 92093, USA

ABSTRACT

This paper investigates the impact of highlighting consumer streaks —sequences of three or more consecutive behaviors— through a field experiment with *one sec,* an app designed to interrupt mindless social media habits. After a consumer tries to open a social media app, *one sec* interrupts the process and forces uses to decide again whether they want to "resist" or "give in" to social media. In a six-week field experiment with 799 users, participants received feedback which highlighted either their streaks of resisting or giving in to social media use, while a control group received no streak-related information. We hypothesized that highlighting positive streaks of resisting social media would encourage users to maintain these streaks, whereas highlighting negative streaks of giving in would motivate users to break them. Results support these hypotheses, showing that users whose streaks were highlighted sought to build positive streaks and break negative streaks. However, highlighting streaks led to more overall usage, which increased the frequency of social media visits, compared to the control. While this behavioral pattern seems at odds with users' initial goal of reducing consumption, consumers reported more favorable attitudes and were less likely to disadopt the technology when their streaks were highlighted. This research contributes to our understanding of how consumers respond to and enjoy gamified feedback even when it leads to counterproductive behaviors. The findings underscore the need for mindful implementation of tracking features in apps aimed at behavior modification.

1. Introduction.

As firms increasingly monitor a diverse range of human behaviors, they face decisions on how to present this data to customers to best meet user demand and support their goals. Firms track both positive, goal-consistent behaviors (e.g. fitness, sleeping, learning, reading, saving) and negative, goal-inconsistent behaviors (e.g. drinking alcohol, smoking, watching TV, consuming sugar, spending). However, the impact of informing consumers about their past negative behaviors remains unclear. For example, if a consumer uses an app to lose weight, but has exceeded their calorie goal for 10 consecutive days, should the app notify the user? This research investigates how providing feedback on streaks of social media use – both positive and negative – affects subsequent social media consumption.

Tracked data offers consumers an objective means to understand their past behavior and provides cues to interrupt bad habits (Verplanken & Wood 2006). The ubiquity of tracking apps reflects both the growth of tracking technologies as a whole and the increasing demand from consumers to quantify their own behavior. Thus, understanding how consumers respond to different types of self-quantification is of utmost interest to firms, consumers, and behavioral scientists.

Despite this explosion in demand for tracking technologies, there are also emerging downsides to tracking. While behaviorally tracked data can improve monitoring and motivate goal-consistent behavior (e.g., pages read, steps walked, vocabulary words learned), it can also feel like work, thereby reducing intrinsic motivation and enjoyment (Kruglanski et al. 1975; Wrzesniewski et al. 2014; Etkin 2016). In fact, adding "gamified" elements to the tracking of positive behaviors can demotivate consumers, such as when their streak is broken (Silverman & Barasch 2023), and can even encourage consumers to spend time on irrelevant or counterproductive activities simply to "do better" within the game statistics, like when language-learners complete lessons with a translator on hand (Mogavi et al. 2022; Tseng et al 2023; Hsee et al. 2003; Kumar et al. 2013). Within this emerging literature, there are still significant gaps in our understanding of how firms with access to consumers' goal-consistent and goal-inconsistent behavior can highlight past behavioral patterns to best support consumer welfare.

One domain where this gap manifests is the growing concern over excessive social media consumption (Haidt 2024). Social media user interfaces (UIs) are designed to be addictive by inducing flow—a state of total immersion resulting in time distortion and endless scrolling (Csikszentmihalyi 2002; Montag et al. 2019). In fact, young adults report spending more than six hours per day on social media (Vannucci et al. 2017) while nearly half of teenager report being on the internet "almost constantly" (Pew 2022). Social media's addictive nature is evident, with 54% of teens reporting difficulty in giving it up (Pew, 2023), 5-25% of adults self-report being addicted to social media (Cheng et al. 2021), and 64% of U.S. adults believing it has a mostly negative impact on the country (Pew, 2020). When it comes to breaking excessive social media habits, the challenge lies in providing users with the tools to break these habits without undermining their motivation.

In this paper, we test if behavioral tracking—specifically, reminding consumers about their past patterns of behavior—can interrupt mindless consumption and reduce problematic social media use while keeping consumers engaged in their behavioral change goal. To test this, we partnered with *one sec*, an app designed to make social media and other distracting apps less appealing through a "friction" intervention that introduces a habit discontinuity (Verpanken et al. 2019). When a *one sec* user attempts to open a social media app, the intervention forces a six-second delay and then prompts users with an active choice to proceed to the app or to cancel the action (See Figure 1). Earlier work found this intervention to be effective at reducing social media usage by 37% (Grüning et al. 2023).

In a six-week field experiment, we test the effectiveness of a behavioral tracking mechanism—highlighting consumers' streaks. We focus on this form of consumer feedback because of its prevalence across existing digital platforms and its promise to motivate and change behavior. Streaks—defined as three or more consecutive behaviors—are a unique pattern that especially attracts attention (Alter & Oppenheimer 2006; Gilovich, Tversky & Vallone 1985; Silverman, Barasch & Small 2023). Prior research finds that maintaining a streak can become a goal in itself when tracking desirable behaviors (Silverman & Barasch 2023). Our study builds on this literature by examining the impact of highlighting positive and negative streaks of past behavior in a setting where people want to reduce an undesirable behavior. Specifically, in the context of *one sec*, we test the effectiveness of highlighting positive streaks of resisting social media and highlighting negative streaks of succumbing to social media among people aiming to reduce their social media usage. We investigate whether highlighting these streaks can foster more mindful social media consumption.

Our research adds to the emerging literature on the role of behavioral tracking in goal pursuit and consumer satisfaction (Etkin 2016; Zimmermann 2021; Silverman & Barasch 2023). In line with other research on how people respond to feedback and non-monetary incentives (Hsee et al. 2003; Fishbach et al 2010; Hsee, Yang, and Ruan 2015; Fong et al 2019), this newer body of research establishes that the feedback people experience due to the monitoring and sharing of their behavior can have notable effects on their attitudes and behavior. Yet while this prior work has largely focused on tracking behaviors which consumers seek to increase (e.g., exercise, learning, reading), we instead study the impact of highlighting patterns of behaviors which consumers view as negative and wish to reduce. In doing so, we build on Silverman and Barasch (2023) by demonstrating a tangible, perverse effect of their theory that streak maintenance is a goal in and of itself. Specifically, we show that the goal to continue positive streaks and break negative streaks may override the original goal of reducing consumption. Importantly, our research also suggests that positive streaks, despite these drawbacks, hold the promise of keeping consumers more engaged over time with technologies designed to curb undesirable behaviors, such as excessive social media consumption.

2. Theory.

2.1 Tracked streaks as incentives.

Incentives play a crucial role in shaping behavior by motivating a particular action, and we posit that streaks are a form of feedback that serve as effective, financially costless incentives. Behavioral tracking feedback introduces goals and incentives by increasing the salience of those measured behaviors and the patterns therein (Etkin 2016; Mehr et al. 2024). Similarly, when consumers are awarded points for ice cream purchases, they deviate from their preferred flavor to maximize otherwise meaningless points (Hsee et al. 2003). In short, points and other tracking metrics have the potential to trump preferences.

If people are willing to change their behavior and put in additional effort to maximize gamified aspects (e.g., gain symbolic points with no monetary value), it follows that they may similarly seek to build positive streaks that are highlighted via tracking technology (Silverman & Barasch 2023). This is unsurprising given that streaks are uniquely powerful. Patterns of three are easy to recognize (Carlson & Shu 2007), and people naturally seek out such repetitive, consistent patterns and ascribe meaning to them, even in settings where these patterns are random (e.g., the “hot hand” in sports; Alter & Oppenheimer 2006; Gilovich, Vallone & Tversky 1985; Ayton & Fisher 2004). Accordingly, streaks of goal-consistent behavior signal higher levels of goal commitment relative to other patterns (Silverman, Barasch & Small 2023), and streaks of identity-relevant behavior are particularly motivating and important to consumers (Weathers and Poehlman 2024). In sum, positive streaks may be interpreted as a form of positive feedback and can affect behavior accordingly

Thus, we propose the following hypotheses regarding positive streaks:

Hypothesis 1A. Consumers will be more likely to continue a positive streak when it is highlighted than when it is not.

While prior work has provided support for this first hypothesis (Silverman & Barasch 2023), note that this has only been established causally within lab experiments. Besides being the first to investigate this prediction experimentally in the field, we propose and test several new additional hypotheses. For one, we predict an important behavioral consequence that prior work was unable to study:

Hypothesis 1B: Consumers will have longer positive streaks when they are highlighted than when not.

Building on prior research which focused on tracking positive behaviors, we posit that tracking negative streaks may also lead to behavioral effects by discouraging the undesirable activity to which the streak is tied. This is because both positive and negative streaks alike provide quantifiable information on goal progress, and therefore influence future behavior. While positive feedback increases outcome expectancy and self-efficacy (Zajonc & Brickman 1969; Bandura & Cervone 1983; Fishbach et al. 2010), negative feedback can encourage greater effort by signaling that there is still much work to be done (Carver & Scheier 1981; Fishbach et al. 2010). Taken together, prior research suggests that highlighting consumers' negative streaks also has the potential to impact consumer behavior.

Therefore, we propose the following hypotheses regarding negative streaks:

Hypothesis 2A: Consumers will be more likely to end a negative streak when it is highlighted than when it is not.

Hypothesis 2B: Consumers will have shorter negative streaks when they are highlighted than when not.

2.2 How streak highlighting may encourage goal-inconsistent behavior.

Although a priori, we predicted that highlighting streaks of social media usage would motivate users to maintain positive streaks and break negative streaks, the specific ways in which consumers would change their behavior within our experimental context was an open question. We anticipated that highlighting streaks of "resisting" social media (positive streaks) will lead users to resist social media more frequently and achieve longer streaks of resisting. Conversely, we expect that highlighting streaks of "giving in" and browsing social media (negative streaks) will also lead users to resist social media more, and achieve shorter streaks of giving in. Yet in this context, streaks are *not* temporally bounded; rather than measuring behaviors at the hourly, daily, or even weekly level (similar to Apple Watch, Duolingo, and Peloton), *one sec* tracks and conveys patterns across each attempt at accessing social media. Therefore, there is some flexibility in how users can obtain streaks over time; one user may have a streak of resisting social media three times that spans several days and another may have the same streak across just a few minutes.

One possible outcome of our intervention is that users will not change their total number of attempts to open social media but will be more likely to actively resist following each attempt. For example, for a user who typically attempts to open their social media five times per day, highlighting positive streaks may shift them towards resisting three or more of those times in a row to build a streak, thereby lowering total visits to social media platforms. Conversely, highlighting negative streaks may shift them to resist on every third or fourth attempt to avoid or break a streak, thereby lowering total visits without changing the number of attempts. Given that our field experiment participants explicitly sought to reduce social media consumption, these patterns—leading to a reduction in overall social media visits—would be considered goal-consistent behavior.

Alternatively, highlighting and "gamifying" these patterns could instead lead to counterproductive, goal-inconsistent behavior. Similar to the effects observed when monetary incentives are introduced (Deci 1971; Kruglanski 1975; Gneezy 2011), highlighting positive streaks may crowd out motivation for the original goal to reduce consumption. As a result, users may change their behavior in service of the streak goal in ways that are incongruous with the original goal of reducing social media consumption. For example, users whose positive streaks are highlighted might intentionally build long positive streaks by attempting to open social media more frequently – say, three *additional* attempts without any intent to proceed onto the app – rather than reducing their consumption in the absolute sense. Conversely, users whose negative streaks are highlighted may break these streaks by resisting once only to immediately re-attempt to open the app and give in to social media moments later. Such behavioral patterns are consistent with emerging research on maladaptive gamification, or cases in which tracking and metrics counterproductively encourage consumers to maximize their performance within the game statistics, (Mogavi et al. 2022; Tseng et al. 2023; Kumar & Herger 2013).

In sum, our predictions in H1a-H2b involve users changing their behavior – namely actively resisting social media use more – in response to their streaks being highlighted. Yet this net increase in resisting social media may be tied to two competing strategies, both of which are supported, but not directly tested, by prior literature. Therefore, we propose the following two opposing hypotheses regarding how consumers may increase the frequency with which they resist when streaks are highlighted (versus not):

Hypothesis H3A (i.e., a goal-consistent behavior): Consumers will attempt to access social media *the same amount* (*or less)*, thereby reducing total consumption.

Hypothesis H3B (i.e., a goal-inconsistent behavior): Consumers will attempt to access social media *more frequently,* thereby *not* reducing total consumption.

2.3 Additional consequences of tracked streaks

Beyond immediate behavioral decisions, we can also make predictions about how tracking positive and negative streaks may influence consequential outcomes, such as brand attitudes, adoption, and disadoption. While both positive feedback on goal progress and negative feedback on goal failure can motivate goal pursuit (Fishbach et al 2010), people feel more motivated when they focus on the positive. Positive feedback (relative to negative feedback) increases feelings of competence and motivation (Ryan & Deci 2000; Fong et al 2019), particularly for novices (vs. experts; Fishbach et al 2010). Therefore, within the context we study, regardless of whether streaks reduce social media consumption, when positive streaks are highlighted, users will report more favorable attitudes toward one sec and be more likely to adopt (and less likely to disadopt) the tool. Specifically, we hypothesize:

Hypothesis 4A. Consumers will report more favorable attitudes toward firms that highlight positive streaks (versus not).

Hypothesis 4B. Consumers will disadopt technology less when consumers positive streaks are highlighted (versus not).

2.4 A note on highlighting both positive and negative streaks

As described in more detail below, our experiment features a "both" condition wherein users had both their positive streaks of resisting and their negative streaks of giving in and accessing social media highlighted. Prior literature makes conflicting predictions about how users might react to being shown both forms of feedback. Presenting users both positive and negative streaks may make it more difficult to interpret the messages, thereby reducing processing fluency and decreasing motivation (Song and Schwarz 2008; Iselin 1988). However, in certain circumstances, lowered fluency forces people to consume information more carefully and become less susceptible to automatic biases (Alter et al 2007), which in the case of this field study, may lead to more deliberate social media consumption. Given these possibilities, we focus on the effects of highlighting positive streaks (versus not) and highlighting negative streaks (versus not) and less so on the specific effects of the "both" condition in our main analyses, and report contrasts concerning the "both" condition in the Supplemental Materials.

2.5 Practical Application

Beyond our theoretical contributions, we test the effectiveness of streaks in a highly relevant context: social media consumption. Excessive social media use has been linked to increased risk of depression, sleep and eating disorders, and low self-esteem, particularly on young people (Department of Health and Human Services, 2023). Our intervention is especially appropriate to test within this context because average active users of social media (e.g. Instagram, TikTok, Facebook) open each of these apps over ten times a day (Datareportal.com, 2024). This frequent engagement creates an environment where many consumers frequently find themselves on both positive streaks of self-control (resisting the urge to use social media) and negative streaks (regularly giving in to the urge). By exploring how highlighting these behavioral patterns can aid in the prevalent goal of reducing social media consumption, we offer practical insights that can help consumers better manage their usage. Additionally, we provide a cautionary perspective for firms considering the use of behavioral tracking to improve consumer welfare, highlighting both the benefits and potential drawbacks.

Overview of Experiments

We test these hypotheses in two experiments. Experiment 1 is a six-week randomized controlled trial conducted in partnership with an app designed to help users reduce social media consumption. This field experiment demonstrates that when highlighted, consumers seek to build positive streaks (H1 A&B) and break negative streaks (H2 A&B). The results provide evidence that both positive and negative streaks can lead to maladaptive gamification (H3B). However, consumers rate the technology more favorably when they highlight positive streaks (versus not) and are less likely to disadopt it (H4 A&B).

Experiment 2 is a follow up laboratory study designed to provide additional support for the hypotheses that could be effectively tested in a lab setting (H1A, H2A, H4A). Specifically, experiment 2 finds further evidence that streak messaging motivates users to continue good streaks and break bad streaks, and that consumers view positive streak feedback more favorably, even if they are perceived as no more effective.

3. Field Experiment

For Experiment 1, we report the results of a six-week randomized controlled trial with 799 One Sec users assigned to four different streak conditions with a total of 347,801 observations. The post-test provides controlled, lab evidence supporting the mechanism of our field data. Preregistrations, data, materials, code and appendix materials are available on Open Science Framework (OSF) at https://osf.io/adpn7/?view_only=9c30a63af9e94e5eb7bd4b6bd89ccbcc. The University of Heidelberg IRB #XXXX approved this research.

Method

We aimed to recruit 1000 new one sec users with the Apple operating system (iOS) on a rolling basis in November-December 2023 via home screen and push notifications once they downloaded the app. Users who agreed to participate were redirected to a pre-survey with informed consent and, upon completion, were randomized into a treatment condition and redirected back to one sec. Users received no direct compensation for participating. However, they learned after signing up that they had access to features of the app that would normally only be available in the paid version (e.g., the ability to select multiple “target apps”).

Our preregistered inclusion criteria stipulated that we include all datapoints from any users who completed the pre-survey. Of the 1011 participants who completed the pre-survey, we received data from 799 users who ultimately installed one sec and opened a target app at least once. As preregistered robustness checks, we also ran our analyses on the subsamples of users who 1) had at least four observations and therefore could be exposed to our manipulation (n = 693), and 2) had observations in the sixth week of data collection (n = 433; see OSF for results). We define an observation as an instance where the user attempts to open a target app and thus encounters the one sec intervention (which, under the right circumstances of past behavior, leads them to see the manipulation).

Pre-survey

Participants reported on their current social media use and their perceptions of how they respond to positive and negative feedback. Participants selected from a list of problems they want to address with the app (i.e. excessive consumption, automatic consumption, feeling of no control, other) and from a list of goals they want to achieve with the app (less consumption, more awareness of consumption, consuming selectively, other). Participants reported how much time they spent on social media (in hours and minutes) and answered eight other questions (on seven-point scales) about their social media use and perceptions of positive feedback, including items like how frequently the use it and how frequently they would ideally use it. This survey was conducted prior to randomization and there were no differences in responses by condition ($Fs < 1.08$, $ps > .36$). We collected limited demographic or identifying data in accordance with *one* *sec*'s consumer privacy practices. Based on initial survey data, we can estimate that 48.95% of our participants were in the US and 37.38% in Europe.

Field Experiment.

For this six-week field experiment, we made two modifications to one sec's standard intervention. First, across all conditions, we modified the choice labels to the following: "Resist {target app}" (instead of "I don't want to open {target app}") and "Give in to {target app}" (instead of "Continue to {target app}". We modified these labels to ensure that each streak was clearly tied to a specific, easily understood action, and implemented this change for all users to preclude the possibility that any observed effects were driven by differences in language.

Second, and most critically, users were randomized into one of four conditions which differed in terms of what patterns of behavior were highlighted: *positive streaks*, where users were informed when they had resisted the target app three or more times in a row; *negative streaks*, where users were informed when they had given in to the target app three or more times in a row; *both*, where users were informed of both positive and negative streaks, and *control*, where users did not see any streak messaging. Random condition assignment occurred when users started the pre-survey. See Figure 3.1 for sample messages.

Figure 3.1. Sample interventions. These messages appeared after the user clicked a social media app (i.e., attempted).

Our key dependent variables were the number of attempts to open the target app, the number of resists per the number of attempts (i.e., the rate of resistance[2]), the total number of give-ins (visits to the target app), and the length of time between attempts.

Post-survey.

Six weeks after signing up for the experiment, users were invited to participate in a non-incentivized post-survey. They responded to the same measures about their current social media use as they did in the pre-survey. Additionally, they rated the perceived fun, frustration, effectiveness, and gamification ("*to what extent did using one sec feel like a game?*") associated with using the app. Finally, users in the *positive*, *negative*, and *both* conditions responded to

[2] We preregistered this variable as the rate of giving in (direct effect), but due to ease of narrative, we report this as rate of resistance (1 – rate of opening).

measures about how often they noticed the streak message, how it made them feel, how much it influenced them, and what they did after seeing it. All items were on 7-point scales.

Deviations from preregistration

There are two ways in which our analyses deviate from our preregistration that we would like to highlight, as it is a consistent deviation across several analyses. Specifically, while we report Tukey Honestly Significant Difference (HSD) tests (one of our primary preregistered analyses), we also report mixed-effects models (one of our secondary preregistered analyses). We do this because our dependent measures do not pass the Shapiro-Wilk test for normality of residuals nor the Bartlett's test for homogeneity of variances (see OSF), both of which are assumptions that must be validated to use the Tukey HSD or ANOVA; our secondary preregistered analyses do not rely on this assumption. Second, we aggregate averages per user at the daily level instead of the weekly level to best leverage the power of the available data and to make interpretation easier (e.g. daily attempts versus weekly attempts), but the significance of the results does not change. Full analysis of all our preregistered models is available on OSF.

Descriptive statistics

Overall, we collected 347,801 observations within six weeks of use from 799 users. All participants completed the pre-survey, 226 (28.29%) fully completed the post-survey, and another 68 (8.51%) provided partial responses. Social media was overwhelmingly the most common type of target app, with over 77% of all observations tied to an attempt to open social media. Users activated one sec on 2.04 target apps on average (median = 1, min = 1, max = 11). See Table 3.1 for a list of the most common target apps.

Table 3.1. Target apps in order of popularity among users.

Target App	App Category	Unique users	% of total users	Number of observations	% of total observations
Instagram	Social media	563	70.46%	193,703	55.69%
TikTok	Social media	169	21.15%	24,551	7.06%
Facebook	Social media	107	13.39%	22,305	6.41%
Youtube	Media content	167	20.90%	19,523	5.61%
Snapchat	Social media	61	7.63%	15,494	4.45%
Twitter	Social media	66	8.26%	12,392	3.56%
Whatsapp	Mobile Messaging	18	2.25%	12,076	3.47%
Other (e.g., Reddit, Safari, Hinge, Telegram, iMessage, Discord, Facebook Messenger, LinkedIn, Pinterest, Amazon)	Other	257	32.17%	47,757	13.73%

We assess the average participant in our study to be a frequent social media user based on their self-selection to use *one sec*, their self-reports in the pre-survey, and their observed data. In the pre-survey, participants could select one or more of the following goals: less consumption (88.23%), more awareness (62.58%), more selective consumption (51.06%), and other (5.63%). On average, participants reported in the pre-survey they spent 3.36 hours per day (SD=2.01) on social media, which is higher than the two and a half hour averages reported in the popular press (Datareportal.com, 2024). Across the six weeks, users attempted to open their target app every 2.66 hours on average, with a median of 19.8 minutes between attempts.

Results

Overall (that is, agnostic to condition assignment), using one sec helped users reduce their social media consumption. Replicating similar results to Grüning et al. (2023), users resisted 39.04% of their attempts across the six weeks and in the post-survey, they reported spending 66 minutes less (32.74%) per day on social media compared to the pre-survey.

Effects of highlighting positive streaks on behavior.

To test our first main hypotheses (H1A) that people will be more likely to maintain a positive streak when it is highlighted (versus not), we subset the data into observations on a positive streak and added a dummy variable for a) whether the user saw a streak message and b) whether the user was in the *both* condition, thereby controlling for how exposure to both types of messages impacted the message effect. When a user had a positive streak, they were more likely to continue that positive behavior when *one sec* highlighted that pattern, versus not. Specifically, a generalized linear mixed model (GLMM) with fixed effects for week and the presence of a positive streak and random effects controlling for individual and app heterogeneity found that highlighting positive streaks made users more likely to resist continuing to the app by 28.1%, $B=0.33$, $SE=0.09$, $z=3.65$, $p<0.001$, 95% CI [0.15, 0.51]. This effect is robust to using condition coding instead dummy variable. Regressing the log-odds of opening on condition maintaining the controls of week and individual and app heterogeneity, we find similar results. Users in the positive streak condition were more likely to resist than the control who did not see a message, $B=0.27$, $SE=0.11$, $z=2.51$, $p=0.012$, 95% CI [0.06, 0.48].

We found that highlighting positive streaks lead to longer positive streaks compared to when not highlighted (H1B). When regressing the positive streak length on condition, we find that users in the *positive streak* condition have longer resistance streaks than control ($B_{pos}=1.69$, $SE=0.16$, 95% CI [1.38, 2.01], $t(12463) = 10.57$, $p<0.001$)

Effects of highlighting negative streaks on behavior.

To test our second hypotheses (H2A) that people will be more likely to break a negative streak when it is highlighted (versus not), we subset the data into observations on a negative streak and added a dummy variable for a) whether the user saw a streak message and b) whether the user was in the *both* condition, thereby controlling for how exposure to both types of messages impacted the message effect. When a user had a negative streak, they were more likely to stop that negative behavior when *one sec* highlighted that pattern, versus not. A GLMM with fixed effects for week and the presence of a negative streak and random effects controlling for individual and app heterogeneity found that highlighting negative streaks also made users more likely to resist by 18.1%, B=0.20, SE=0.06, z=3.22, p=0.001, 95% CI [0.08, 0.31]. Once again, this effect is robust to using condition coding as well. Regressing the log-odds of opening on condition maintaining the controls of week and individual and app heterogeneity, we find similar results. Users in the negative streak condition were more likely to resist than the control who did not see a message, B=0.16, SE=0.07, z=2.29, p=0.022, 95% CI [0.02, 0.30].

We also found that highlighting negative streaks lead to shorter negative streaks compared to when not highlighted (H2B). When regressing the negative streak length on condition, we find that users in the *negative streak condition* have shorter open streaks than control (B_{neg} =-0.25, SE=0.07 95% CI [-0.39, -0.10], t(26233) = -3.37, p<0.001).

Evidence of goal-inconsistent behavior because of streak highlighting.

To test whether this increased likelihood of resisting upon seeing a streak message resulted in an overall decrease in social media use, we tested whether the streak conditions led to a reduction of social media usage by comparing total attempts, total opens, and the amount of time between opens across conditions. See Figure 3.2 to understand how these behaviors changed by condition across the 42 days in the study.

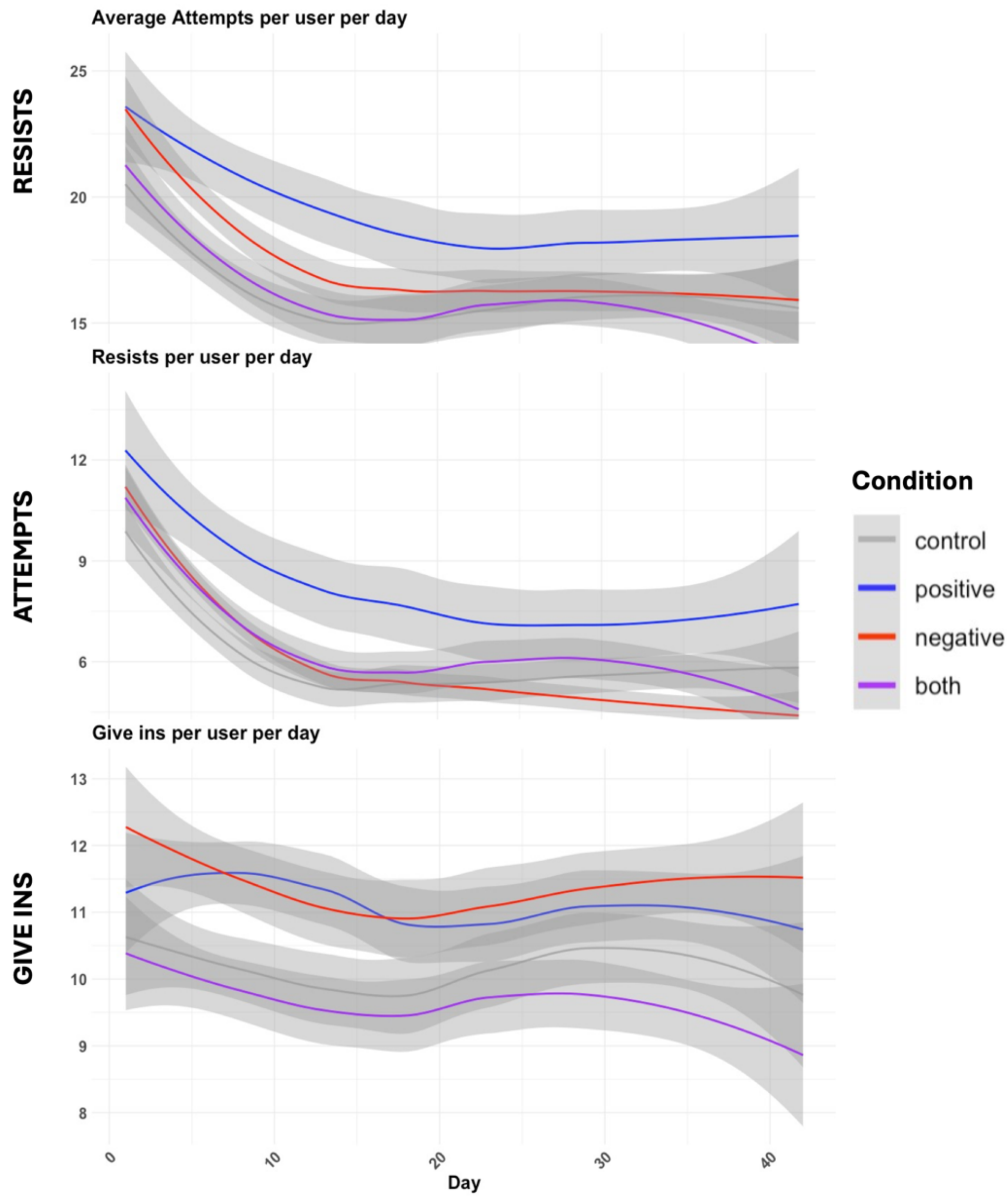

Figure 3.2. Average daily resists, attempts, and opens per user across the six-week field experiment by condition.

Attempts. In line with our goal inconsistent hypothesis (H3b), highlighting positive and negative streaks did *not* lead to fewer overall attempts to open social media usage. In fact, users in the *positive streak* condition had marginally *more* attempts across the six weeks than control. In our preregistered Tukey corrected pairwise comparison of attempts per user across the six weeks by condition, the *positive streak* condition (M_{pos} = 517.40, SD = 806.52) had on average 142.67 more attempts (95% CI [-8.96, 294.29], p =0.073) across six weeks to open the target app than control (M_{con} = 374.73, SD = 494.18). While the number of attempts in the negative condition (M_{neg} = 435.03, SD = 475.45) were not statistically different than control, directionally users in the negative condition also attempted 60.30 more times (95% CI [-90.36, 210.96], p =0.73) more than control. Moreso, on any given day within the six week trial, users' daily attempts were also higher in the positive and negative streak conditions than control. When regressing the total daily attempts on condition controlling for the number of days in the study (B_{day} =-0.11, SE=0.01, 95% CI [-0.13, -0.09], t(19897) = -9.36, $p<0.001$), participants in the *positive* condition attempted three times more per day on any given day of the trial than control (B_{pos} =3.23, SE=0.39, 95% CI [2.45, 4.00], t(19897) = 8.20, $p<0.001$) while participants in the *negative* condition attempted once more per day than the control (B_{neg} = 1.27, SE=0.40, 95% CI [0.49, 2.05], t(19897) =3.19, p=0.001).

In terms of heterogeneity, when we include random effects for participant ID, the effect on the positive condition remains significant, but the negative condition becomes insignificant. When controlling for both individual and app random effects, the positive condition becomes only marginally significant. See Table 3.2 for comparison of models.

Table 3.2. Heterogeneity of effects on number of attempts.

Comparison of Models for Attempts

	Standard LM			LM Controlling for Day			Mixed Model (UUID)			Mixed Model (UUID & App)		
Predictors	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*
(Intercept)	10.27	9.97 – 10.57	**<0.001**	11.28	10.90 – 11.66	**<0.001**	10.99	9.54 – 12.44	**<0.001**	7.44	5.73 – 9.15	**<0.001**
cond [positive]	2.09	1.68 – 2.51	**<0.001**	2.10	1.69 – 2.52	**<0.001**	2.27	0.22 – 4.32	**0.030**	1.91	-0.05 – 3.87	0.056
cond [negative]	1.01	0.59 – 1.43	**<0.001**	1.02	0.60 – 1.44	**<0.001**	1.23	-0.82 – 3.28	0.239	1.12	-0.84 – 3.07	0.262
cond [both]	0.22	-0.21 – 0.65	0.313	0.22	-0.21 – 0.64	0.326	0.15	-1.94 – 2.24	0.888	0.08	-1.91 – 2.08	0.935
day				-0.05	-0.06 – -0.04	**<0.001**	-0.06	-0.07 – -0.05	**<0.001**	-0.06	-0.07 – -0.05	**<0.001**
Random Effects												
σ^2							92.93			84.82		
τ_{00}							99.94_{UUID}			90.50_{UUID}		
										25.05_{app}		
ICC							0.52			0.58		
N							799_{UUID}			799_{UUID}		
										162_{app}		
Observations	31226			31226			31226			31226		
R^2 / R^2 adjusted	0.004 / 0.004			0.006 / 0.006			0.007 / 0.522			0.006 / 0.579		
AIC	250901.763			250830.858			232681.218			230061.726		

Give-ins. Users in the *positive streak* condition gave in to the target app more than users in any other condition. Although not significant, the Tukey HSD corrected pairwise comparison estimated that users in the *positive streak* condition opened the target app 62.71 times (95% CI [-16.00, 141.42], p =0.17) more across the six weeks than the control, while negative and both conditions were not significant, ps>.38. When regressing the total give-ins on condition controlling for the amount of days in the study (B_{day} =-0.01, SE=0.006, 95% CI [-0.02, -0.0006], t(19897) = -2.06, p=0.039), participants in the *positive* and *negative streak* conditions on average visited social media once more per day than control (B_{pos}=1.01, SE=0.20, 95% CI [0.61, 1.41], t(19897) = 4.95, p<0.001; B_{neg}= 1.23, SE=0.21, 95% CI [0.83, 1.63], t(19897) =5.98, p<0.001, respectively).

In terms of heterogeneity, when we include random effects for participant ID, the effect on the positive and negative condition becomes marginally significant. When controlling for both individual and app random effects, both effects become insignificant. See Table 3.3 for comparison of models.

Table 3.3. Heterogeneity of effects on number of give-ins.

Comparison of Models for Give-ins

	Standard LM			**LM Controlling for Day**			**Mixed Model (UUID)**			**Mixed Model (UUID & App)**		
Predictors	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*
(Intercept)	6.42	6.25 – 6.58	**<0.001**	6.39	6.19 – 6.60	**<0.001**	6.18	5.34 – 7.03	**<0.001**	4.12	3.12 – 5.12	**<0.001**
cond [positive]	0.68	0.45 – 0.91	**<0.001**	0.68	0.45 – 0.91	**<0.001**	1.15	-0.05 – 2.35	0.061	0.97	-0.18 – 2.12	0.100
cond [negative]	0.92	0.68 – 1.15	**<0.001**	0.92	0.68 – 1.15	**<0.001**	0.90	-0.30 – 2.10	0.142	0.81	-0.33 – 1.96	0.164
cond [both]	-0.17	-0.40 – 0.07	0.166	-0.17	-0.40 – 0.07	0.167	-0.21	-1.44 – 1.01	0.733	-0.22	-1.39 – 0.95	0.712
day				0.00	-0.01 – 0.01	0.718	-0.00	-0.01 – 0.01	0.936	-0.00	-0.01 – 0.00	0.740
Random Effects												
σ^2							29.33			26.76		
τ_{00}							34.63 $_{UUID}$			31.50 $_{UUID}$		
										8.44 $_{app}$		
ICC							0.54			0.60		
N							799 $_{UUID}$			799 $_{UUID}$		
										162 $_{app}$		
Observations	31226			31226			31226			31226		
R^2 / R^2 adjusted	0.004 / 0.004			0.004 / 0.004			0.005 / 0.544			0.004 / 0.600		
AIC	213700.671			213702.541			196736.510			194122.155		

Time between attempts. The streak conditions also impacted how often users attempted to go to the target app. Users in the *positive streak* condition waited on average 33 minutes fewer (B_{pos} =-0.56, SE=0.06, t(346170)=-9.76, p<0.001, 95% CI [-0.67, -0.45]) between attempts than in the control. Similarly, users in the *negative streak* condition waited on average 19.2 minutes fewer than control (B=-0.32, SE=0.06, t(346170)=-5.50, p<0.001, 95% CI [-0.44, -0.21]). When regressing the time between attempts on condition controlling for the week in the study, participants in the *positive* and *negative streak* conditions waited on average around 33 minutes (B_{pos} =-0.55, SE=0.06, 95% CI [-0.66, -0.44], t(346169) = -9.69, p<0.001) and 18.6 minutes (B_{neg} = -0.31, SE=0.06, 95% CI [-0.42, -0.19], t(346169) =-5.23, p<0.001) less between attempts than control, respectively.

Once again, in terms of heterogeneity, when we control for both individual and app random effects, both effects become insignificant. See Table 3.4 for comparison of models.

Table 3.4. Heterogeneity of effects on time between attempts.

Comparison of Models for Time between attempts

	Standard LM			**LM Controlling for Day**			**Mixed Model (UUID)**			**Mixed Model (UUID & App)**		
Predictors	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*	*Estimates*	*95% CI*	*p*
(Intercept)	2.93	2.84 – 3.01	**<0.001**	2.26	2.15 – 2.36	**<0.001**	5.16	3.77 – 6.54	**<0.001**	26.88	14.01 – 39.75	**<0.001**
cond [positive]	-0.56	-0.67 – -0.45	**<0.001**	-0.55	-0.67 – -0.44	**<0.001**	-0.64	-2.59 – 1.30	0.517	-0.06	-2.00 – 1.88	0.953
cond [negative]	-0.32	-0.44 – -0.21	**<0.001**	-0.31	-0.42 – -0.19	**<0.001**	-0.49	-2.45 – 1.47	0.622	-0.38	-2.33 – 1.57	0.704
cond [both]	-0.09	-0.21 – 0.03	0.138	-0.08	-0.20 – 0.04	0.194	-0.41	-2.41 – 1.58	0.684	-0.20	-2.18 – 1.78	0.840
day				0.04	0.03 – 0.04	**<0.001**	0.05	0.04 – 0.05	**<0.001**	0.05	0.04 – 0.05	**<0.001**
Random Effects												
σ^2							134.80			128.07		
τ_{00}							86.67 $_{\text{UUID}}$			85.32 $_{\text{UUID}}$		
										5806.95 $_{\text{app}}$		
ICC							0.39			0.98		
N							706 $_{\text{UUID}}$			706 $_{\text{UUID}}$		
										137 $_{\text{app}}$		
Observations	346174			346174			346174			346174		
R^2 / R^2 adjusted	0.000 / 0.000			0.002 / 0.002			0.002 / 0.392			0.000 / 0.979		
AIC	2705415.778			2704954.159			2683554.861			2666834.172		

Post-survey attitude measures. Finally, despite finding evidence of a goal-inconsistent behavior, we hypothesized that positive streaks would lead to more enjoyment (H4A) of and less disadoption (H4B). Only two measures in our post survey were significantly difference across conditions: fun($F(3, 236) = 2.81$, p=0.040) and frustrating ($F(3, 236) = 4.70$, p=.0033). Follow-up pairwise t-tests revealed that users in the positive streak condition found *one sec* less frustrating (M=3.86, SD=1.44) compared to the control (M=4.65, SD=1.64), t(118.18)=2.90, p=0.0044, or compared to the negative streak condition (M=4.75, SD=1.25), t(117.74)=3.67, p<0.001. Users in the positive streak condition also reported that *one sec* was more fun (M=3.83, SD=1.45) compared to control (M=3.12, SD=1.49), t(124.73)=2.78, p=0.0062, or compared to the negative streak condition (M=3.29, SD=1.49), t(108.57)=2.03, p=0.044. No other post survey measures were significant across conditions, (*F*s<2.33, ps>0.074).

Disadoption. Although we cannot directly observe participants discontinuing use of the technology, we find evidence suggesting that disadoption is lowest in the positive condition (H4B). We preregistered that we would proxy disadoption in two ways: 1) post-survey completion and 2) whether participants were still in the study after two weeks. Using a chi-squared test, there were no differences in survey completion between conditions (X-squared (3, N = 799) =4.42, p = 0.22). However, if we proxy disadoption by whether a user had any observations beyond the second week of the study, we find evidence of differential disadoption (X-squared (3, N = 799) =8.28, p = 0.041). In the pairwise chi-squared tests, we find less disadoption in the positive streak condition (25.6%) than in the control group (37.4%), p=0.014, but no statistical difference in disadoption between those conditions and the negative streak condition a (33.3%), ps>0.11. To further examine differential disadoption, if we specify disadoption by the number of active daily users, we can model disadoption using a regression predicting the number of unique users by condition controlling for each day in the study. Using this model, once again we find that we have ten more active users per day in the positive streak condition ($B = 10.21$, $SE = 2.05$, 95% CI [6.16, 14.26], $t(163) = 4.98$, $p<0.001$) and 5 more users per day in the negative streak condition ($B = 4.88$, $SE = 2.05$, 95% CI [0.83, 8.93], $t(163) = 2.38$, $p =0.018$) compared to control.

Discussion

Together, this evidence supports the broader claim that streaks are a powerful form of feedback. Highlighting a positive streak encourages people to continue that streak of desirable behavior (H1A), and as a result, positive streaks become longer when highlighted (H1B). Conversely, highlighting a negative streak motivates users to break that streak of undesirable behavior (H2A), and as a result, negative streaks before shorter when highlighted (H2B). Seeing any streak message, positive or negative, made users more likely to resist proceeding to social media (compared to those who were not reminded when they were on a similar streak).

However, most participants in this study had an overall goal of reducing social media consumption, not an intermediate goal of resisting more. Even though we find that streak messages lead people to resist more frequently, we find that highlighting streaks in this context led to behavior that was counterproductive toward the overall goal of consuming less (H3B). Rather than resisting more often while keeping total number of attempts unchanged (which would be goal-consistent), users on average *increased* the frequency of their attempts. Moreover, their increase in total number of resists did not match the increase in total attempts, which led to an overall *increase in giving in* (aka visiting social media more) across the positive and negative streak conditions compared to control.

One explanation for why participants in the streak conditions both resisted more and attempted more is that they were gaming the streaks. For example, participants whose streaks are highlighted might be more likely to attempt to open and then resist in quick succession (likely to build or regain a positive streak or remove a negative streak message). We find exploratory evidence that people may in fact be cheating the streaks. Using a binary GLM, we find that users in the *positive streak* condition were 65% more likely to have attempts within 30 seconds of one another than control (B_{pos}=0.51, SE=0.01, 95% CI [0.48, 0.53], z = 36.57, p<0.001, which could be consistent with attempts to game the streak.

Thus, it seems that at least some users "gamed" the system, changing their behavior to receive more positive or less negative feedback from *onesec*, while (perhaps inadvertently) increasing the amount they opened social media, a behavior completely counter to their stated goals. When we consider heterogeneity of effects across our participants, controlling for individual differences tends to reduce much of the differences between the control group and the streak conditions. This suggests that the streak intervention did not produce goal-inconsistent behavior for the entire participant pool. This heterogeneity is likely driven by different individual differences in people's desire to maximize the gamified aspects of the app and differences across the apps to which these interventions were applied.

In sum, this field experiment provides evidence that highlighting patterns of past behavior makes consumers want to build positive steaks and break negative streaks, and that consumers find the positive streak intervention to be more fun (H4A), leading to higher rates of retention (H4B). However, given the resulting differential attrition between conditions, we sought to replicate these results about consumer attitudes in a controlled, laboratory setting.

Experiment 2

We report the results of a controlled laboratory experiment in which participants evaluated the app *one sec* and made predictions about their behavior after viewing the streak interventions used in Experiment 1. Preregistrations, data, materials, coda and appendix materials are available at https://osf.io/adpn7/?view_only=9c30a63af9e94e5eb7bd4b6bd89ccbcc.

Method

Participants (n= 751, Prolific Workers, M_{age} = 38.18, 64.0% female) were assigned to three between subject message conditions (*positive*, *negative*, *control*) and two within-subject scenario conditions (history of giving in, history of resisting) in randomized order.

First, participants reflected on their current social media usage by answering questions about how much time they spend per day on social media and how problematic they find their consumption on a seven-point scale. Then participants read about the *one sec* app and watched a video of the baseline intervention.

Participants read that in this study, they would evaluate a new feature of the app. We varied this new feature across three between-subjects *messaging* conditions which aligned with the three types of messages in our field experiment: *positive streak*, *negative streak*, and *control*. Participants in all conditions first saw the control screen, and then participants in the positive and negative streak conditions saw what the respective streak screen would look like (see Figure 3.1

for the exact screenshots we used in the Experiment 2). Then each participant read two scenarios in a randomized order: "Imagine a day when you had tried to open Instagram about six times over the previous 24 hours. Lately, you feel like you have been resisting (giving in) more than you have been giving in (resisting)." Replicating our experimental conditions in Experiment 1, across both scenarios, participants in the control condition saw the control screenshot along with the scenario description. In the positive streak condition, participants saw the positive streak screenshot when they had a history of resisting, and the control screen when they had a history of giving in. In the negative streak condition, participants saw the negative streak screenshot when they had a history of giving in, and the control screen when they had a history of resisting.

Dependent Measures. We measured five variables consistent with our dependent measures in our field experiment. To capture behavioral responses to streak messaging, after each scenario, participants indicated which button they would click after seeing the message on a seven-point scale: 1 = Definitely "Give in", 7 = Definitely "Resist". To measure attitudes towards the app, participants rated the effectiveness of *one sec* at improving their social medial habits (1 = Extremely ineffective, 7 = Extremely effective), and the extent to which *one sec* seemed fun and frustrating (1 = Not at all, 7 = To a great extent). To measure adoption preferences, participants stated how likely they were to consider using *one sec* (1 = Extremely unlikely, 7 = Extremely likely).

Results.

Behavior. In the vignette where participants read that they had a history of resisting, highlighting the streak message made participants much more likely to resist compared to participants in the conditions whose streak was not highlighted. A one-way ANOVA revealed a significant effect of message condition on participants' behavior ($F(2, 748) = 16.26$, $p <0.001$).

Follow-up pairwise comparisons indicated that participants were much more likely to resist going on Instagram in the *positive* condition (M=4.83, SD=1.87) when the streak was highlighted compared to *negative* (M=3.90, SD=1.99), ($t(490.58) = 5.39, p < 0.001$) or compared to control (M=4.06, SD=2.02), ($t(501.72) = 4.48, p < 0.001$). The likelihood of resisting in the negative and positive conditions were not statistically different than the scale midpoint ($ps > 0.42$), meaning people were equally likely to give in or resist. However, in the positive condition when the streaks were highlighted, the likelihood of resisting was significantly higher than the midpoint, $t(251)=7.08$, $p<.001$, meaning participants were more likely to resist than give in.

In the vignette where participants read that they had a history of giving in, there was no significant effect of highlighting the negative streak. A one-way ANOVA revealed no significant effect of message condition on participants' behavior ($F(2, 748) = 1.30$, $p = 0.27$). Follow-up pairwise comparisons indicated that participants were equally likely ($ps > 0.14$) to resist going on Instagram in the *positive* (M=4.45, SD=1.91) *negative* (M=4.49, SD=2.03), and contol (M=4.23, SD=2.02) conditions. However, in this scenario, the likelihood of resisting in the control condition was not statistically different than the scale midpoint ($p=0.07$), meaning people were equally likely to give in or resist. However, participants were more likely to resist than give in across both the *positive* ($t(251)=3.79$, $p<.001$) and *negative* ($t(244)=3.85$, $p<.001$) conditions.

When comparing the effect of the message within subjects across scenario, participants in the *negative* condition were much more likely to resist when exposed to the negative streak screenshot (M=4.49, SD=2.03) compared to when exposed to the control screen(M=3.90, SD=1.99), ($t(487.91)=3.30$, $p=0.0010$). Similarly, participants in the *positive* condition were much more likely to resist when exposed to the positive streak screenshot (M=4.83, SD=1.87) compared to when exposed to the control screen (M=4.45, SD=1.91),(t(501.73)=2.24, p=0.026). Meanwhile, participants in the *control* condition were equally likely to resist across both scenarios, (p=0.33). See Figure 3.3.

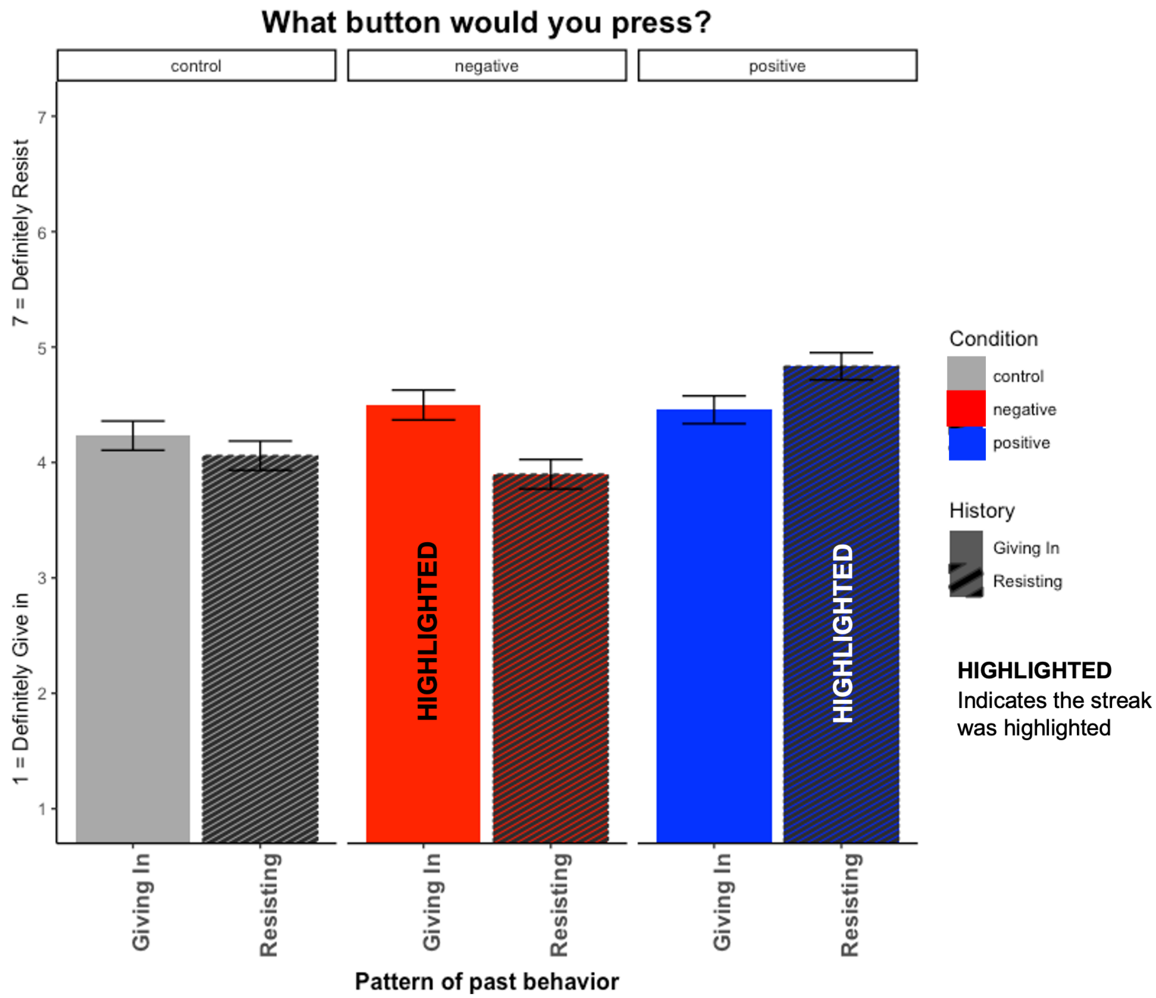

Figure 3.3. Within person effects of resisting based on streak highlighting in Experiment 2.

Fun. A one-way ANOVA revealed a marginal main effect of *messaging* condition on how fun *one sec* seemed (F(2, 748) = 2.77, p =0.064). Follow-up pairwise comparisons indicated that, in line with the results of Experiment 1, participants in the *positive streak* condition perceived the app as more fun (M=2.97, SD=1.70) than the *control* condition (M=2.62, SD=1.70), (t(503.9)=2.35, p =0.019) and directionally but not significantly more fun than in the *negative* condition (M=2.78, SD=1.77), (t(492.79)=1.21, p =0.23).

Effectiveness at changing social media habits. A one-way ANOVA revealed no significant effect of *messaging* condition on how effective *one sec* seemed (F(2, 748) = 2.33, p =0.36). Follow-up pairwise comparisons revealed that participants in the *positive* (M=4.54, SD=1.47), *negative* (M=4.45, SD=1.53), and *control* (M=4.35, SD=1.58) conditions rated the app equally effective (ps>0.15). Across all conditions, mean effectiveness was significantly greater than the midpoint (ts> 3.53, ps<.001).

Frustration. A one-way ANOVA revealed no significant effect of *messaging* condition on how frustrating *one sec* seemed (F(2, 748) = 0.37, p =0.69). Follow-up pairwise comparisons revealed that participants in the *positive* (M=4.02, SD=1.86), *negative* (M=4.20, SD=1.76), and *control* (M=4.11, SD=1.83) conditions rated the app equally frustrating (ps>0.39). By comparing the mean rating to the scale midpoint, only participants in the negative condition rated the app marginally more frustrating than not, (t(244)=1.81, p =0.071).

Likely to Use. A one-way ANOVA revealed no significant effect of *messaging* condition on how frustrating *one sec* seemed (F(2, 748) = 0.64, p =0.53). Follow-up pairwise comparisons revealed that participants in the *positive* (M=3.67, SD=2.03), *negative* (M=3.61, SD=2.11), and *control* (M=3.51, SD=2.03) conditions rated the app equally frustrating (ps>0.26). By comparing the mean likelihood to the scale midpoint, across all conditions, participants were not likely to consider using the app (ts> 2.60, ps<.01).

Discussion

The results of Experiment 2 confirm that many of the findings in the field hold in a more controlled, internally valid lab setting where people are predicting how they would respond to the intervention. That is, supporting H1A and H2A, we find evidence that participants seek to continue good steaks and break bad streaks when their streaks are highlighted. Supporting H4A, this study also suggests that consumers find the app more fun when it incorporates the positive streaks. While there was no significant difference across conditions on likelihood of using the app, how fun the app was a significant predictor of future use (B=0.72, SE=0.03, 95% CI [0.65,0.79], t(749)=20.60, p<.001). Therefore, we suspect that at scale, incorporating positive streaks would help the firm increase app adoption.

The fact that our main results hold in a population that did not self-select into using *one sec* like our participants in Experiment 1 speaks to the generalizability of the results. Participants in experiment 2 reported spending similar amounts of time on social media (M=3.28, SD=2.42) compared to participants in Experiment 1 (M=3.36, SD=2.01), t(1458.8)=0.72, p=.47, but reported finding their consumption much less problematic (M = 3.94, SD = 1.82) than in Experiment 1 (M=5.44, SD=1.32), t(1358.6)=18.52, p<.001.

General Discussion.

Our findings demonstrate that highlighting consumer streaks –both positive and negative – impacts behavior. Positive streak messages motivate users to maintain their streaks of desirable behavior, whereas negative streak messages increase the likelihood of breaking them. As a result, compared to when patterns of behavior were not highlighted, highlighting positive streaks led to longer positive streaks, whereas highlighting negative streaks led to slightly shorter negative streaks. However, streak highlighting seems to introduce new goals –maintaining or breaking the streak – that can compete with the original goal such as reducing social media consumption.

Theoretical Contribution

Our findings contribute to the understanding of how streaks influence behavior, revealing a double-edged sword: while they motivate users to engage in desired behaviors such as resisting social media, they can also encourage counterproductive behaviors such as being less deterred from attempting to open social media in the first place. Building on previous theoretical work on streaks and behavioral tracking (Silverman and Barasch 2023; Etkin 2016), this likely occurs because the pursuit of maintaining the streak or breaking the streak overshadows the original goal of reducing consumption. Because tracked streaks are highly evaluable (Hsee et al., 2003) highlighting streaks increases the importance of streak goals over their initial behavior modification goal. This also contributes to the emerging literature of maladaptive gamification and goal misalignment (Tseng et al 2023), whereby introducing gamified aspects to a platform can lead people to goal inconsistent behavior. Such insights underscore the complex relationship between behavioral incentives and digital consumption, suggesting that while positive reinforcement through streak highlighting is effective, it requires careful implementation to avoid unintended outcomes.

Marketing Implications

While these goal-inconsistent results would seem to suggest that firms should be wary of using streaks to help consumers reduce unwanted behaviors, it is important to note that consumers liked the positive streaks, and this increased liking correlated with less disadoption of *one sec* in the positive streak condition. One participant's comment on X (formerly Twitter) after the experiment concluded highlights this attachment: "What happened to the streak thingy??? I found it awesome. Please add it back [cry emoji]". We find that emphasis on positive streaks reduces disadoption, a central driver of firm profitability (Lehmann and Parker 2017).

Moreso, our heterogeneity analysis reveals that controlling for time in the study significantly predicted fewer attempts and longer time between attempts. Therefore, the longer people used One Sec, the more goal consistent they were regardless of condition. As indicated in Figure 3.2, although users opened their target app more in the two streak conditions than control, they still attempted and opened much less in the last week of the study than the first week across all conditions. In fact, participants across conditions resisted approximately 40% of their attempts over the six weeks and reported a 72-minute reduction (35.19%) in daily social media usage compared to their pre-survey behavior. Given the persistent goal-consistent effects of one sec across time, any intervention which keeps users from disadopting the app may in fact be better for the consumer by leading to less social media consumption over the long term.

How should firms balance these competing findings that consumers demonstrate a preference for positive streak highlighting even if it leads to less optimized behavior? Future analysis should consider the tradeoff between sustained user engagement and goal-consistent behavior. Firms should carefully weigh the tradeoff that a "fun" intervention such as streak highlighting may introduce between maladaptive consumer behavior against reduced disadoption that could ultimately benefit the firm and its customers in the long run.

How might our findings guide firms in implementing streaks within their own apps? Consider Starbucks. If Starbucks implemented a feature that let consumers set a goal to reduce sugar consumption, the app could remind customers of their streak—such as ordering three sugar-free drinks in a row—when they open the app to place an order. Our results predict that such an intervention would make consumers more likely to order the sugar free drink. Additionally, it might increase the frequency of app use, as consumers seek to build and maintain their streaks. Such an intervention could be a win-win for both the firm and the consumer if the consumer is in increasing consumption while substituting away from less healthy options.

Limitations and Future Directions

Our streak intervention is unique in three systematic ways that may impact the generalizability of the findings: the timing of the intervention, the structure of the streak, and the direction of the goal. Regarding timing, our streak manipulation occurred in the middle of the decision process, when a user has already succumbed to the temptation but has one more opportunity to opt in or out of the behavior. This decision interruption is a key feature of the *one sec* intervention but may not apply to other apps that consider implementing streak tracking. For example, a user tends to log into to a fitness app like Peloton after they have committed to working out, so streaks tend to be presented as a post-workout reward mechanism, not a pre-decisional tool. If they are merely contemplating working out or have decided not to workout, a fitness app would not have a natural interface with the consumer. Perhaps the most appropriate parallel to our intervention is for apps that use streaks in push notifications (e.g. *You are on a 3-day streak of working out!* Or *You have not worked out in three days*) where it might be more plausible to leverage the streak as a pre-decisional motivation tool. Future research should consider the timing of the streak message and its impact on behavior.

Regarding streak structure, our intervention used non-temporally bound streaks (i.e., based on sequential behavior) rather than a more traditional temporal streak (e.g., a daily or weekly streak). We suspect that daily streaks would lead to other patters of gamified behavior, notably a “what the hell” effect on days when a user gives-in early. Given that users in our study gave in, on average, ten times per day, we opted for the non-temporally bound streak, because few would have achieved a three-day positive streak. Our design highlighted small wins throughout the day. Alternatively, we could have used a combination of a cumulative and daily streak by having users set a daily limit of no more than *X* number of visits per day and counted a streak as the number of days in a row they met or exceeded that threshold. However, recent theoretical work on streaks suggests that cumulative streaks are more cognitively burdensome and lead to less commitment than daily streaks (Weather and Poehlman 2024). Additionally, this type of streak intervention may interact with other limit setting effects (Silverman, Etkin & Srna 2024). Future work on streaks should consider how temporally structuring streaks impacts consumer behavior.

As for the direction of the goal, our study population had an inhibition goal—reducing a negative behavior—whereas many apps encourage, acquisitional goals to do more of a good behavior (e.g. language lessons, exercise). Previous research on acquisitional versus inhibitional goals reveals that there are systematic differences in both the frequency of and response to feedback (Polivy and Herman 1985; Cochran and Tesser 1996; Soman and Cheema 2004). Tracking positive streaks may have reframed the inhibitional goal (give-in less) as an acquisitional goal (resist more), influencing consumer attitudes about the app and their progress. Open questions remain about how highlighting streaks in other contexts—differing in dimensions such as type or direction of behavior, temporal units, or periodicity—may influence whether streak highlighting promotes goal-consistent or goal-inconsistent behavior. Future work might consider both how positive and negative streaks impact behavior in naturalistic settings when the goal is acquisitional.

Our findings reveal heterogeneity in how users respond to streaks, as evidenced by our model comparisons. Not all users responded to streaks in a goal-inconsistent manner. In fact, we considered whether a subset of users may have their negative streak highlighted despite goal-consistent behavior. For example, a user deliberately limiting social media use to a one-to-one ratio of one attempt and one give in per day may still be told they are on a negative streak of giving in. We might expect backlash from these users who feel reprimanded for their goal-consistent, moderate behavior. However, we found such a behavioral pattern to be exceedingly rare. When we looked through our 347,801 observations, we only found 53 observations where there was a streak of three observations in a row, each at least 24 hours a part. Future work should explore individual differences between consumers—such as productivity orientation, need for structure and identity relevance—(Weathers & Poehlman 2024) to predict who may be more susceptible to gamified app features. Future research could also explore heterogeneity within users, when for example a consumer may be more likely to engaged in gamified, goal-inconsistent behavior (H3a) to repair a long broken streak which may evoke a strong feeling of failure.

How does streak highlighting interact with other important factors? Future research should explore aspects of tracked streaks which may moderate these effects such as whether the highlighting is public versus private, whether personalization is human-generated vs. technology-generated, or tracking over short versus longer periods of time. For example, behavioral tracking confers status (Huynh et al. 2024), so we suspect that sharing streaks publicly would strengthen the gamified incentives, amplifying goal-inconsistent behavior. Alternatively, human-generated (vs. technology-generated) feedback is seen as more judgmental and threatening, leading it to be more influential on people's judgments and behavior (Raveendhran & Fast 2021). Therefore, personalizing streak feedback by integrating an interactive artificial intelligence feature rather than an automatic tracking feature could amplify the effects. Finally, we look at streaks over long periods of time in the field, which is a contribution over recent streak work (Silverman & Barasch 2023). However, even longer temporal dynamics could be considered to see whether the effects of streak messaging increase or wear off over time. Moreover, an open question remains as to how consumers might react when such tracking is periodic or removed. The impacts of gamified rewards tend to drop off when rewards are removed (Blanchard & Palazzolo 2024). Future research should consider temporary (random or strategic) reminders of positive negative streaks, which may help users build good streaks and, more importantly, break unhealthy streaks without introducing goal-inconsistent incentives.

Conclusion.

Excessive social media use is one of many addictive behaviors that society currently grapples with. Overcoming these addictions requires individuals to exert sustained effort over a long period, and platforms may be able to assist by offering tools such as behavioral tracking to help users monitor their triggers, responses, and habits. Our results indicate that highlighting positive streaks can increase desirable behaviors, while highlighting negative streaks can motivate individuals to break detrimental patterns. However, both streaks can backfire when users focus on the streak itself and lose sight of their primary goals. Notably, if positive streaks can encourage prolonged engagement with self-monitoring tools, it may benefit consumers in the long run—even if this gamified approach promotes less-than-optimal behavior in the short term.

ACKNOWLEDGEMENTS

Chapter 3, in full, is currently being prepared for submission for publication of the material. Hillegass, Kathryn; Silverman, Jackie; Gershon, Rachel; Grüning, David; Riedel, Frederik; Wolf, Donatus. The dissertation author was the primary researcher and author of this material.

REFERENCES

Adams, G. S., Converse, B. A., Hales, A. H., & Klotz, L. E. (2021). People systematically overlook subtractive changes. Nature, 592(7853), 258-261.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. Journal of experimental psychology: General, 136(4), 569.

Alter, A. L., & Oppenheimer, D. M. (2006). From a fixation on sports to an exploration of mechanism: The past, present, and future of hot hand research. Thinking & Reasoning, 12(4), 431–444.

Anderson, I. A., & Wood, W. (2021). Habits and the electronic herd: The psychology behind social media's successes and failures. Consumer Psychology Review, 4(1), 83-99.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. Learning and motivation, 21(4), 434-455.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. Journal of personality and social psychology, 45(5), 1017.

Bayer, J. B., Anderson, I. A., & Tokunaga, R. S. (2022). Building and breaking social media habits. Current Opinion in Psychology, 45, 101303.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. Review of general psychology, 5(4), 323-370.
Hermsen, S., Frost, J., Renes, R. J., & Kerkhof, P. (2016). Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. Computers in Human Behavior, 57, 61-74.

Blanchard, S. J., & Palazzolo, M. (2024). Game Over? Assessing the Impact of Gamification Discontinuation on Mobile Banking Behaviors. Marketing Science.

Borovoi, L., Schmidtke, K., & Vlaev, I. (2020). The effects of feedback valance and progress monitoring on goal striving. Current Psychology, 1-18.

Bravata, D. M., Smith-Spangler, C., Sundaram, V., Gienger, A. L., Lin, N., Lewis, R., ... & Sirard, J. R. (2007). Using pedometers to increase physical activity and improve health: a systematic review. Jama, 298(19), 2296-2304.

Burke, L. E., Wang, J., & Sevick, M. A. (2011). Self-monitoring in weight loss: a systematic review of the literature. Journal of the American Dietetic Association, 111(1), 92-102.

Carlson, K. A., & Shu, S. B. (2007). The rule of three: How the third event signals the emergence of a streak. Organizational Behavior and Human Decision Processes, 104(1), 113-121.

Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. Journal of Experimental Social Psychology, 17(6), 545-568.

Cheng, C., Lau, Y. C., Chan, L., & Luk, J. W. (2021). Prevalence of social media addiction across 32 nations: Meta-analysis with subgroup analysis of classification schemes and cultural values. Addictive behaviors, 117, 106845.

Costantini, A. F., & Hoving, K. L. (1973). The effectiveness of reward and punishment contingencies on response inhibition. Journal of Experimental Child Psychology, 16(3), 484-494.

Coyne, S. M., Rogers, A. A., Zurcher, J. D., Stockdale, L., & Booth, M. (2020). Does time spent using social media impact mental health?: An eight year longitudinal study. Computers in human behavior, 104, 106160.

Csikszentmihalyi, M. (2002). Flow: The classic work on how to achieve happiness. Random House.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. Psychological inquiry, 11(4), 227-268.

Etkin, Jordan (2016), "The Hidden Cost of Personal Quantification," *Journal of Consumer Research*, 42(6), 967–84.

Fishbach, A., Eyal, T., & Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. Social and Personality Psychology Compass, 4(8), 517-530.

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. Educational Psychology Review, 31, 121-162

Frey, B. S., & Jegen, R. (2001). Motivation crowding theory. *Journal of economic surveys*, *15*(5), 589-611.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. Cognitive psychology, 17(3), 295-314.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. Journal of economic perspectives, 25(4), 191-210.

Goldsmith, K., Friedman, E. M., & Dhar, R. (2019). You don't blow your diet on Twinkies: Choice processes when choice options conflict with incidental goals. *Journal of the Association for Consumer Research*, *4*(1), 21-35.

Grüning, D. J., Riedel, F., & Lorenz-Spreen, P. (2023). Directing smartphone use through the self-nudge app one sec. Proceedings of the National Academy of Sciences, 120(8), e2213114120.

Haidt, J. (2024). The anxious generation: How the great rewiring of childhood is causing an epidemic of mental illness. Random House.

Hsee, C. K., Yu, F., Zhang, J., & Zhang, Y. (2003). Medium maximization. Journal of Consumer Research, 30(1), 1-14.

Hsee, C. K., Yang, Y., & Ruan, B. (2015). The mere-reaction effect: Even nonpositive and noninformative reactions can reinforce actions. *Journal of Consumer Research*, *42*(3), 420-434.

Huynh, Denny, Jordan Etkin, and Tanya Chartrand. "Behavioral Tracking as a Status Symbol." Working Paper.

Mogavi RH, Guo B, Zhang Y, Haq E-U, Hui P, Ma X (2022). When gamification spoils your learning: A qualitative case study of gamification misuse in a language-learning app.

Kruglanski, A. W., Riter, A., Amitai, A., Margolin, B. S., Shabtai, L., & Zaksh, D. (1975). Can money enhance intrinsic motivation? A test of the content-consequence hypothesis. Journal of Personality and Social psychology, 31(4), 744.

Kruglanski, A. W., & Szumowska, E. (2020). Habitual behavior is goal-driven. Perspectives on Psychological Science, 15(5), 1256-1271.

Kumar, J., Herger, M., Deterding, S., Schnaars, S., Landes, M., & Webb, E. (2013). Gamification@ work. In CHI'13 extended abstracts on human factors in computing systems (pp. 2427-2432).

Lee, D. S., Jiang, T., Crocker, J., & Way, B. M. (2022). Social media use and its link to physical health indicators. Cyberpsychology, Behavior, and Social Networking, 25(2), 87-93.

Lehmann, D. R., & Parker, J. R. (2017). Disadoption. Ams Review, 7, 36-51.

Lervik-Olsen, L., Andreassen, T. W., & Fennis, B. M. (2024). When enough is not enough: behavioral and motivational paths to compulsive social media consumption. European Journal of Marketing, 58(2), 519-547.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. American psychologist, 57(9), 705.

Mehr, Katie, Jackie Silverman, Marissa Sharif, Alixandra Barasch, and Katherine Milkman. "The Motivating Power of Streaks: Increasing Persistence Is as Easy as 1, 2, 3." Working paper.

Montag, C., Lachmann, B., Herrlich, M., & Zweig, K. (2019). Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. International journal of environmental research and public health, 16(14), 2612.

Myrseth, K. O. R., & Fishbach, A. (2009). Self-control: A function of knowing when and how to exercise restraint. Current Directions in Psychological Science, 18(4), 247-252.

National Alliance on Mental Illness. (n.d.). Social media and mental health. NAMI. Retrieved June 3, 2024, from https://www.nami.org/Your-Journey/Kids-Teens-and-Young-Adults/Teens/Social-Media-and-Mental-Health/.

Pantic, I. (2014). Online social networking and mental health. Cyberpsychology, Behavior, and Social Networking, 17(10), 652-657.

Pew Research Center. (2022, August 10). Teens, social media and technology 2022. Pew Research Center: Internet & Technology. https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

Raveendhran, R., & Fast, N. J. (2021). Humans judge, algorithms nudge: The psychology of behavior tracking acceptance. Organizational Behavior and Human Decision Processes, 164, 11-26.

Shaddy, F., Fishbach, A., & Simonson, I. (2021). Trade-offs in choice. *Annual Review of Psychology*, *72*, 181-206.

Silverman, Jackie and Alixandra Barasch (2023), "On or Off Track: How (Broken) Streaks Affect Consumer Decisions," *Journal of Consumer Research*, 49(6), 1095–1117.

Silverman, J., Barasch, A. P., & Small, D. A. (2023). Hot streak! Inferences and predictions about goal adherence. Organizational Behavior and Human Decision Processes, 179, 104281.

Silverman, Jackie, Jordan Etkin, and Shalena Srna. "Time Limits Act as Reference Points to Shape Time Spent." *Working paper (target: Psychological Science)*.

Soman, D., & Cheema, A. (2004). When goals are counterproductive: The effects of violation of a behavioral goal on subsequent performance. Journal of Consumer Research, 31(1), 52-62.

Song, H., & Schwarz, N. (2008). If it's hard to read, it's hard to do: Processing fluency affects effort prediction and motivation. Psychological science, 19(10), 986-988.

Thaler, R. H., & Shefrin, H. M. (1981). An economic theory of self-control. Journal of political Economy, 89(2), 392-406.

Tseng, S. L. A., Sun, H., Santhanam, R., Lu, S., & Thatcher, J. B. (2023). Rethinking Gamification Failure: A Model and Investigation of Gamified System Maladaptive Behaviors. Information Systems Research.

Vannucci, A., Flannery, K. M., & Ohannessian, C. M. (2017). Social media use and anxiety in emerging adults. Journal of affective disorders, 207, 163-166.

Verplanken, B., & Wood, W. (2006). Interventions to break and create consumer habits. Journal of public policy & marketing, 25(1), 90-103.

Verplanken, B., & Orbell, S. (2019). Habit and behavior change. Social psychology in action: Evidence-based interventions from theory to practice, 65-78.

Weathers, D., & Poehlman, T. A. (2024). Defining, and understanding commitment to, activity streaks. Journal of the Academy of Marketing Science, 52(2), 531-553.

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: thought, emotion, and action. Journal of personality and social psychology, 83(6), 1281.

Wrzesniewski, A., Schwartz, B., Cong, X., Kane, M., Omar, A., & Kolditz, T. (2014). Multiple types of motives don't multiply the motivation of West Point cadets. Proceedings of the National Academy of Sciences, 111(30), 10990-10995.

Zajonc, R. B., & Brickman, P. (1969). Expectancy and feedback as independent factors in task performance. Journal of Personality and Social Psychology, 11(2), 148.

Zimmermann, Laura. "Your screen-time app is keeping track": consumers are happy to monitor but unlikely to reduce smartphone usage." Journal of the Association for Consumer Research 6.3 (2021): 377-382.