Rationalizing self-defeating behaviors: theory and evidence*

Lars J. Lefgren† Olga B. Stoddard‡ John E. Stovall§

10 September 2020

Abstract

Why do individuals engage in self-harm? Why do they experience the apathy of depression or inaction when trapped by multiple competing problems? We propose a framework for explaining these and other related behaviors based on the insight that individuals can only experience a limited number of latent stimuli to which they are exposed. We conduct an experiment to test this model and find that more than two thirds of the subjects behave consistent with our theoretical framework.

*We thank Matthew Ross for his excellent research assistance and Gordon Dahl, Jeff Denning, Erik Snowberg, Andrew Oswald, Daniel Sgroi, and Eugenio Proto for their valuable feedback on earlier drafts of the paper. We express our appreciation to audience members at Brigham Young University, the Singapore Economic Association Conference, the Institute for Social Research Conference at Jinan University, the University of Nottingham, the 14th Meeting of the Society for Social Choice and Welfare, the 2018 North American Summer Meeting of the Econometric Society, the 2018 Foundations of Utility and Risk Conference, the 2018 Economic Science Association World Meeting, the 2019 Utah Experimental Conference, and the University of Rochester. We thank the Brigham Young University College of Family, Home, and Social Science for generous research support. John Stovall acknowledges the financial support for travel provided by the David M. Kennedy Center for International Studies at Brigham Young University.

†Department of Economics, Brigham Young University and National Bureau of Economic Research; email: lars_lefgren@byu.edu

‡Department of Economics, Brigham Young University; email: olga.stoddard@byu.edu

§Department of Economics, Brigham Young University; email: johnestovall@byu.edu

1 Introduction

Economists have long been interested in why individuals engage in self-defeating behaviors. Such behaviors include seemingly myopic financial decisions, self-harm, substance abuse, and apparent apathy in response to the challenges of life. Behavioral explanations for these pervasive issues may include lack of self-control (Strotz, 1955), hyperbolic discounting (Thaler, 1981; Rachlin et al., 1991), and the cognitive strains of poverty (Ridley et al., 2020; Dean et al., 2019; Kremer et al., 2019).

We propose another framework for explaining these behaviors based on the insight that individuals can only experience a limited number of latent stimuli to which they are exposed. This is consistent with research in psychology suggesting that individuals can only attend to a very limited number of stimuli (Cherry, 1953; Treisman and Gelade, 1980; McCaul and Haugtvedt, 1982; Borkovec and Roemer, 1995). Our model can explain a variety of seemingly dysfunctional behaviors including self-harm, the apathy of depression, and the inaction of individuals trapped by multiple competing problems. In addition, we provide an empirical test of the model in the lab and find that it can explain behavior of 68 percent of our subjects.

To understand the intuition underlying our model, consider an individual who is exposed to many latent stimuli but prone to experience only the most salient one. Suppose an individual is watching a movie, with a utility measure of 10, while experiencing a headache, with a utility measure of -2. Suppose further that the movie and headache each have a subjective measure of salience, and that the individual only experiences the utility of the most salient stimulus.[1] If the salience of the movie exceeds that of the headache, the individual does not notice the headache in the background and enjoys a utility of 10 from watching the movie. On the other hand, if the salience of the headache exceeds that of the movie, then the individual cannot

[1]This corresponds to our main model. We also present a more general model that allows the individual to experience the utility of a less salient stimulus, but doing so requires a 'concentration cost.'

pay attention to the movie and instead experiences the utility level of the headache, -2. What are the behavioral implications of such a decision maker?

This framework has the power to explain a variety of interesting economic behaviors and phenomena. Naturally, this model predicts that individuals and firms will often bundle unpleasant stimuli with a more pleasant distraction (Filcheck et al., 2005; Milkman et al., 2014; Al-Khotani et al., 2016). Hence, individuals may listen to music while exercising or watch TV while prone in the dentist's chair. More significantly, however, the model rationalizes a number of seemingly dysfunctional behaviors as well.

First, consider health-related behaviors such as self-harm or substance abuse. In our model, individuals may engage in such activities even if they are not intrinsically pleasurable in isolation. Suppose an individual is exposed to an emotionally painful event, perhaps due to poor mental health or a difficult interpersonal conflict. In this case, an individual may engage in cutting, not because the stimulus is pleasurable but rather because it has higher utility than the emotional suffering and is more salient. In this sense, it acts as an effective, if unpleasant, distraction from an even more painful stimulus. While the individual would prefer a pleasant distraction, there may not exist any pleasant experiences that are sufficiently salient to distract from the emotional suffering. In this sense, the self-harming behavior serves as a feasible optimal distraction.

Second, our model explains the behavior of individuals experiencing depression. We model depression as a very salient negative stimulus. Because depression is so salient, it crowds out the utility associated with activities that would normally be considered pleasant. Consequently, depressed individuals have little motivation to engage in a variety of activities associated with a functional life. Indeed, to the extent that the only stimuli sufficiently salient to break through depression are either risky or harmful, our model also explains why depressed individuals are at risk of

illicit drug use, overeating, and self-harm.

Third, our model predicts that individuals experiencing multiple problems have a diminished incentive to fix any single problem. This is because the benefit of removing one negative stimulus is negligible if the individual is also experiencing another more salient negative stimulus. Even the benefit of solving the problem associated with the most salient negative stimulus is limited by the fact that removing the stimulus will simply bring another problem to the fore of the individual's attention. In this manner, individuals may appear apathetic about improving their situation in the presence of multiple problems. This can explain why stressed individuals act in a manner that seems self-defeating or why the poor make seemingly irrational decisions (Schilbach et al., 2016; Haushofer and Fehr, 2014).

We present empirical predictions of our model and test them in a laboratory setting. The key prediction is that an individual would always rank experiencing two stimuli simultaneously between the two stimuli separately. In particular, when the stimuli are undesirable, then the individual would rank two bad things as weakly better than just the worst of the two. We examine this in the case of subjects who are asked their willingness to endure listening to a painfully loud fire alarm, put a hand in ice cold water, or do both at the same time. Strikingly, we find that fully 68 percent of individuals weakly prefer to endure both negative stimuli compared to enduring just the most painful one. Of these, 13 percent have a strict preference. This is strong evidence suggesting our model is empirically relevant in explaining human behavior.

Our work relates to a growing body of research in health economics on rational addiction (Becker and Murphy, 1988; Darden, 2017; Darden and Papageorge, 2018) and studies in behavioral economics exploring the relationship between cognitive function and poverty-induced stress. Ridley et al. (2020); Kremer et al. (2019) and Dean et al. (2019) document robust causal links between poverty, psychological well-being, and

economic behaviors. They describe a feedback loop in which poverty reinforces itself through reduced cognitive function. We contribute to this literature by presenting an additional explanation for why the poor and those in otherwise difficult circumstances often engage in seemingly irrational behaviors. We also contribute to prior empirical work in economics which has explored the relationship between self-harm behaviors and employment, income, and education (Hansen and Lang, 2011; Marcotte, 2003; Rodriguez Andres, 2006) as well as depression and mental health in contexts like fertility (Wilson, 2019), unemployment (Hakulinen et al., 2019; Tefft, 2011; Marcus, 2013; Green, 2011), crime (Mahuteau and Zhu, 2015), and human capital accumulation (Patton et al., 2016).

This paper also contributes to a long-standing literature in psychology, cognitive science and decision theory, which we discuss in detail in Section 3 below.

2 Examples

Before proceeding to a formal development of our model, we present several simple examples that demonstrate the relevance of our idea. These examples explain a number of behaviors of interest to economists, clinical psychologists, and policy makers.

For these examples, we assume that each stimulus is indexed by a measure of utility and a measure of salience. Like utility, salience is subjective and thus derived from preferences. When exposed to a set of stimuli, an individual experiences the utility of the stimulus within this set that has the highest salience. That is, for stimulus x, let $u(x)$ denote the individual's utility of x and let $s(x)$ denote the individual's salience of x. If $X = \{x, y, z, \ldots\}$ is the individual's set of stimuli, then the utility experienced by the individual under X is

$$U(X) = \max_{x \in X} u(x) \text{ subject to } s(x) \geq s(y) \text{ for all } y \in X. \tag{1}$$

Underlying our model is the assumption that an individual is aware of all stimuli to which she is exposed, with their corresponding measures of utility and salience.

Returning to the headache and movie example of the introduction, we have $u(\text{movie}) = 10$ and $u(\text{headache}) = -2$. If we had $s(\text{movie}) = 5$ and $s(\text{headache}) = 10$, then the individual would experience the utility $U(\{\text{movie}, \text{headache}\}) = -2$ since the headache is more salient.

While the above utility function takes the same form as one attributed to Strotz (1955) in the temptation literature, we reinterpret the framework to generate important insights in very different domains. We discuss the relationship between our work and Strotz (1955) further in section 3 following the presentation of our examples.

2.1 Self-injury

> *I used self-injury as a coping mechanism to help me overcome the emotional stress that I was incapable of dealing with in any other way. Self-injury was a means of escape, a way to relieve the numbness, and an expression of the pain within me.*
>
> –Giblin (2006), *Hailey's Story.*

Extensive prior research in psychology has explored the factors that may lead people to engage in self-defeating behaviors. These include creating boundaries (Suyemoto, 1998), replacing suicidal behaviors (Firestone and Seiden, 1990), stopping or eliciting dissociation (Herpertz, 1995; Himber, 1994; Miller and Bashkin, 1974), controlling sexuality (Friedman et al., 1972) and externalizing emotions (e.g., Friedman et al., 1972; Herpertz, 1995; Himber, 1994). Theoretically, Nock and Prinstein (2004)'s four-function model suggests a framework for understanding motivations for self-harm. Consistent with this framework, the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, or DSM-5, states that "most commonly, the

purpose is to reduce negative emotions, such as tension, anxiety, and self-reproach, and/or to resolve an interpersonal difficulty." Consequently, those who engage in nonsuicidal self-injury (NSSI) will often report an immediate sensation of relief that occurs during the process.

Nock (2010) provides a comprehensive review of research on NSSI in psychology. While there is no consensus on its exact prevalence, researchers estimate that between 10 and 30 percent of adolescents in the general population engage in self-harming behaviors. Most common among these behaviors are found to be cutting, burning, and overdosing on medications (Doyle et al., 2015). Sociodemographic and psychological factors, including exposure to self-harming friends or family members, dysfunctional family relationships, and sexual orientation are found to be the strongest correlates (Doyle et al., 2015; Swannell et al., 2014; Somer et al., 2015; Kharsati and Bhola, 2016).

To explain self-injury in the context of our framework, suppose there are three possible stimuli with the following utilities and salience.

Stimulus	u	s
g	1	0
b	-2	3
h	-1	4

We think of g as a good stimulus, b as a bad stimulus (e.g. being in an abusive relationship), and h as self-harm (i.e. a negative but salient stimulus). The utility an individual receives from a set of stimuli is given by U in Equation 1.

Suppose that life can either be going well or poorly for the decision maker. If life is going well, then he experiences the set $\{g\}$. If life is going poorly, then he experiences the set $\{g, b\}$. Suppose further that the decision maker has the ability to add h to any set of experiences. If life is going well, then he would rather not have h

present. I.e.

$$U(\{g\}) > U(\{g, h\}).$$

However, when life is going poorly, then this preference switches and the individual does want stimulus h present. I.e.

$$U(\{g, b, h\}) > U(\{g, b\}).$$

Thus when things are going well, the decision maker will not engage in self-harm. However, if life takes a turn for the worse, then the decision maker will engage in self-harm in order to distract himself from the bad stimulus.

This example is of relevance to economists considering the causes of a variety of destructive behaviors. For example, Carpenter et al. (2017) find that substance abuse is more common among individuals facing economic hardship. Our model predicts that this would be the case. Additionally, it suggests that efforts to treat substance abuse may be of limited effectiveness without either alleviating the underlying stress that made it optimal in the first place or providing alternative, less harmful, strategies for distraction. More broadly, Cawley and Ruhm (2011) explores the economic models for risky behaviors and evidence for such models. Our framework provides an additional lens through which to analyze such behaviors.[2]

Similarly, by understanding some risky behaviors as an optimized response to a set of stimuli, clinicians may have better insights into how to help individuals in such conditions. For example, McCart et al. (2014) outlines a set of therapy strategies for youth struggling with substance abuse. Our framework might refine contingency contracting approaches described in this paper in which therapists work with clients to discuss alternative rewards that can compete with the risky behaviors.

[2]While our model provides one explanation for understanding why individuals might find it optimal to engage in substance abuse, as a static model it cannot speak to addiction, which is fundamentally a dynamic phenomenon.

2.2 Depression

> *To have depression is to have no motivation; No motivation to wake up, no motivation to socialize, no motivation to live. It is a crushing weight that you just need some support to lift.*
>
> –Anonymous (2014).

Major Depressive Disorder is the most common mood disorder in the US affecting over 16 million adults. The DSM-5 characterizes depressive disorders by "sad, empty, or irritable mood, accompanied by somatic and cognitive changes that significantly affect the individual's capacity to function." Risk factors for depressive disorders include both genetic and environmental factors. Additionally, bereavement and other severe life challenges can induce symptoms of a depressive disorder without meeting the criteria for such a diagnosis. Regardless of the causes of depressive disorders and symptoms, our framework provides insight into the resulting behaviors.

Regarding major depressive episodes, the DSM-5 reports, "The mood in a major depressive episode is often described by the person as depressed, sad, hopeless, discouraged, or 'down in the dumps'." However, on occasion individuals "complain of feeling 'blah,' having no feelings, or feeling anxious." From these descriptions, it seems reasonable to model the phenomenon of depression as an extremely salient, and generally negative, stimulus that crowds out other feelings. In the context of our theoretical framework, the existence of such a powerful stimulus provides a compelling explanation for a variety of observed behaviors of individuals suffering from depression.

First, the DSM-5 states, "Loss of interest or pleasure is nearly always present, at least to some degree. Individuals may report feeling less interested in hobbies, 'not caring anymore,' or not feeling any enjoyment in activities that were previously considered pleasurable. In some individuals, there is a significant reduction from previous

levels of sexual interest or desire." To see how this relates to our model, consider the following example. Assume three possible stimuli: the absence of depression, g; depression, d; and a pleasurable activity, a. The utility and salience of these measures are given in the table below.

Stimulus	u	s
g	1	0
d	-10	10
a	5	5

An individual will enjoy the pleasurable activity in the absence of depression, since $U(\{g, a\}) > U(\{g\})$. However, when depression is within the set of stimuli, an individual will not find that the pleasurable activity increases utility since it is not sufficiently salient to be enjoyable, yielding $U(\{d, a\}) = U(\{d\})$. Thus for sufficiently salient depression, individuals would become indifferent to many activities they would otherwise find enjoyable.[3]

A similar example can explain the difficulty that, according to the DSM-5, depressed individuals exhibit in thinking, concentration, and decision making. Note that depression need not be associated with very low utility or extreme sadness in order to bring about these changes in behavior. Indeed, what is most significant is not the utility associated with the stimulus of depression but rather its overwhelming salience which numbs an individual to other stimuli, both pleasant and unpleasant.

Second, the DSM-5 indicates that depression is often comorbid with substance-related disorders and that while some depressed individuals display a lack of interest in food, others report increased appetite and weight gain. Similarly, prior research has found a significant relationship between depression and unhealthy behaviors such

[3]An alternative way to model depression is to think of it as a disorder in which the salience of negative stimuli is increased relative to that of positive stimuli. Such a model would similarly predict diminished interest in formerly pleasant activities and also predict that individuals dwell on negative stimuli.

as smoking, obesity, physical inactivity, heavy drinking, workforce productivity, educational attainment, and suicide in various cross-sections of adults in the US (Strine et al., 2008; Beck et al., 2011; Berndt et al., 1998; Kessler, 2012). Zetterqvist (2015) reports that among individuals engaging in NSSI, 70 to 80 percent are depressed as well. These behaviors associated with depression are consistent with our explanation of self-harm in the prior section. Indeed, depressed individuals are likely to engage in any behavior which is more pleasant than depression and sufficiently salient to increase their utility.

Our model predicts that depression is likely to compete for attention with tasks associated with productivity in the labor market and household. Consistent with this prediction, several studies have found that depressive symptoms are negatively associated with economic productivity (Peng et al., 2016), employment (Frasquilho et al., 2016), income (Lund et al., 2010) and food insecurity (Noonan et al., 2016). In addition to its adverse effects on cognitive function, Ridley et al. (2020) find that depression may also distort beliefs individuals hold about themselves or the world. This results in depressed individuals being more likely to remember negative stimuli (Gotlib and Joormann, 2010) and updating their beliefs more pessimistically (Korn et al., 2014). Our framework complements these studies and suggests that the economic benefits of effective treatments for depression are likely to be substantial.

2.3 The Trap of Competing Problems

> *Shawn, an office manager in Cleveland, was struggling to make ends meet. He was late on a bunch of bills. His credit cards were maxed out. His paycheck ran out quickly. As he said, "There is always more month than money." Every phone call made him tense: another creditor calling to "remind" him? Being out of money was also affecting his personal life. And there was no end in sight. He had bought a Blu-ray player on credit,*

with no payments for the first six months. That was five months ago. How would he pay this extra bill next month?

–Mullainathan and Shafir (2013)

Researchers have long puzzled over why individuals in difficult situations fail to undertake action to improve their situation, or even engage in behavior that would seem to exacerbate the difficulties they already face.[4] Going back to Strotz (1955), economists have also considered the possibility that individuals have limited self-control and hence are subject to temptation that they may wish to avoid. Furthermore, Mani et al. (2013) show that individuals in poverty demonstrate reduced cognitive function that prevents them from making optimal financial decisions. The evidence for these theories is compelling. Our framework, however, presents an additional explanation for such behaviors.

In particular, when individuals face a large number of problems or negative stimuli, the effect on their realized utility of eliminating one negative stimulus may be quite small. An individual who receives utility only from the most salient stimulus has no incentive to remove a negative stimulus that is insufficiently salient to be felt. Even when the experienced negative stimulus is undesirable, the benefit of eliminating it may be limited by the fact that another negative, if slightly less salient, stimulus will simply be brought to the fore of the individual's attention. This intuition is similar to that of a competing risks model in epidemiology. The life-saving benefit of curing one illness is limited by the health risks posed by a second. For example, reducing the health risks of heart disease by dieting are negligible for an individual diagnosed

[4]Long-standing work on ego depletion by Baumeister et al. (1998), Muraven et al. (1998), Muraven and Baumeister (2000), and Schmeichel and Vohs (2009), for example, relates one's self-control to a muscle that grows tired with repeated use. Self-control governs thoughts, feelings, physical endurance, and task persistence. They find that while human behavior is governed by automatic and controlled processes, it is possible to override these processes at a significant cost to one's self-control resources. In situations where people face constant stress such as enduring multiple negative experiences and depression, they find themselves in a chronic state of self-control depletion, which keeps them from making the changes in their life necessary to improve their conditions.

with terminal cancer.

Consider the following example that illustrates this intuition. Suppose that there are three possible stimuli with the following utilities and salience.

Stimulus	u	s
g	1	0
b	-2	3
w	-3	4

We think of g as a good stimulus, b as a bad stimulus (e.g. being unemployed), and w as a worse stimulus (e.g. marital problems).

Suppose the decision maker is experiencing the set $\{g, b, w\}$, but has the ability to remove b from this set. If he does this, he will experience utility $U(\{g, w\}) = -3$. However, $U(\{g, b, w\}) = -3$. Thus the decision maker is not willing to pay any cost to remove only b from this stimulus set.

Even the willingness to remove the worse stimulus, w, is limited by the existence of the bad stimulus, b. An individual experiencing the set $\{g, w\}$ would be willing to pay a utility cost up to 4 to eliminate w from the stimulus set. However, if the decision maker is experiencing the set $\{g, b, w\}$, he would only be willing to pay a utility cost of 1 or less to eliminate w.

This example also highlights the indifference of individuals experiencing negative stimuli to the addition of other negative stimuli. To an individual with the preferences we describe, $U(\{g, b, w\}) = U(\{g, w\})$ and $U(\{b, w\}) = U(\{w\})$. Hence the addition of the bad stimulus, b, to the set that already includes the worse stimulus, w, has no effect on realized utility.

Of course the individual would prefer to have no negative stimuli in his stimulus set. However, this example demonstrates that the existence of multiple problems limits the willingness to eliminate any one problem. Consequently, individuals with a variety of problems may not find it optimal to fix any of them depending on their

constraints. They may also find it suboptimal to prevent the occurrence of new problems. Hence, behavior that may seem irrational, impulsive, or demonstrating poor cognitive function may instead reflect the complementarity of negative stimuli arising from the fact that once a person has problems, adding more problems may not change experienced utility.

Our model suggests that clinicians may want to consider the full portfolio of an individual's challenges when providing treatment on how to improve a client's situation. In particular, an individual may not find it meaningful to improve in one domain unless he or she is better able to cope in another, possibly seemingly unrelated, domain. Hence, our model suggests a broader and more holistic approach to helping individuals facing multiple problems.

This model may also be of relevance in health economics. Specifically, if individuals from low-income backgrounds or disadvantaged minority groups face a variety of negative stimuli in their life, it may not seem optimal to engage in behaviors to improve health. Pampel et al. (2010) find that low-income individuals engage in fewer behaviors to improve health than high-income individuals. Similarly, Escarce et al. (1993) find that elderly African Americans are substantially less likely to utilize medical services than whites despite having the same access to Medicare. Future researchers may wish to examine specifically whether exposure to multiple problems reduces willingness of individuals to invest in their health.

We recognize that this trap of competing problems example presupposes an individual is aware of all stimuli to which she is latently exposed, even though she experiences the utility of only the most salient one. This means that she is aware both of the utility and salience of stimuli that are in the background of her attention. People's ability to respond to surveys (such as the General Social Survey) about a variety of problems and anxieties suggests that individuals are aware of many challenges to which they are exposed. It also seems likely that individuals who distract

themselves know the consequences of removing the distraction—indeed the decision to distract is based on personal experience regarding the utility and salience of the distraction and the stimulus being avoided. We are not aware, however, of evidence of broader metacognition of the full set of stimuli to which an individual is exposed. Such research would provide important insight regarding the existence and nature of the trap of competing problems.

3 Theory

These preceding examples highlight the economic relevance of our model. In this section, we provide a brief technical treatment of the Strotz (1955) utility representation and also consider alternative utility representations. This formalization provides empirical predictions which are testable in the lab.

3.1 Model

Let A denote the (finite) set of possible stimuli, and let $\mathcal{P}(A)$ denote the set of all subsets of A. A decision maker will experience a set of stimuli $X \in \mathcal{P}(A)$, and has a preference relation $\succeq$ over $\mathcal{P}(A)$. We define $\succ$ and $\sim$ in the usual way. We say U *represents* $\succeq$ if $U(X) \geq U(Y)$ if and only if $X \succeq Y$. Note that we include the empty set in our domain. For all of the following representations and without loss of generality, we set the utility of the empty set equal to zero.

Definition. We say $\succeq$ has a *Strotz representation* if there exist real-valued functions u and s such that

$$U^S(X) = \max_{x \in X} u(x) \text{ subject to } s(x) \geq s(y) \text{ for all } y \in X,$$

represents $\succeq$. If $\succeq$ has a Strotz representation, then we say that $\succeq$ is a *Strotz prefer-*

ence.

As previously discussed, the interpretation is that s is the decision maker's subjective measure of salience of stimuli while u is the decision maker's true utility. The decision maker experiences the utility of only the most salient stimulus.

Strotz's original model was one of changing tastes, and he considered how a self-aware individual might behave in such a situation. He proposed that the decision maker would prefer commitment. However in the absence of commitment opportunities, Strotz proposed the "strategy of consistent planning": Today's self chooses a plan of consumption that tomorrow's self will actually implement. In this interpretation of U^S, s represents tomorrow's preferences while u represents today's.

One possible objection to the Strotz representation is that it is overly restrictive in the sense that the decision maker can only pay attention the most salient stimulus. Could one not ignore salient stimuli through effort and concentration? The next representation addresses this issue.

Gul and Pesendorfer (2001) (henceforth GP) extended Strotz's analysis to a setting of temptation and self-control in a paper that kicked off a large decision theoretic literature on temptation.[5] They introduced the following representation.

Definition. We say $\succeq$ has a *GP representation* if there exist real-valued functions u and s such that

$$U^{GP}(X) = \max_{x \in X} [u(x) + s(x)] - \max_{x \in X} s(x),$$

represents $\succeq$. If $\succeq$ has a GP representation, then we say that $\succeq$ is a *GP preference*.

One way to think about the Strotz representation is as a limiting case of the GP

[5]See Lipman and Pesendorfer (2013) for a survey of this literature. Also, note that Gul and Pesendorfer's (2001) set up is different than our finite one. Specifically, their primitive is a preference relation over non-empty compact subsets of lotteries over a compact metric space. See Gul and Pesendorfer (2005) for an in depth analysis of these preferences in a finite setting.

representation. To see this, consider the GP representation

$$U_{\alpha}^{GP}(X) = \max_{x \in X} [u(x) + \alpha s(x)] - \max_{x \in X} \alpha s(x), \;\; \alpha > 0,$$

and note that $U_{\alpha}^{GP} \to U^S$ as $\alpha \to \infty$.

The GP representation can be applied in our setting of sensory limitations. The functions u and s are the true utility and salience respectfully, as before. However as opposed to the Strotz representation, a decision maker can experience the utility of a stimulus other than the most salient one, but doing so requires exerting some mental effort. To see this, rewrite U^{GP} is

$$U^{GP}(X) = \max_{x \in X} [u(x) - c_s(x, X)],$$

where $c_s(x, X) = \max_{y \in X} s(y) - s(x)$. The function $c_s(x, X)$ represents the cost of concentrating on $x \in X$. Thus a decision maker with GP preferences evaluates sets of stimuli according to the utility function u net concentration costs.[6]

To better illustrate the intuition of the GP representation, it is helpful to go back to the motivating example from our introduction. Recall that the utility of the movie is 10 while that of the headache is -2. Further suppose that the salience of the movie is 5, while that of the headache is 10. Under the Strotz representation, an individual cannot enjoy the movie because the headache is more salient. Under the GP framework, however, an individual could pay a concentration cost (measured in utility) of 5, which is the difference in salience between the two stimuli, to experience the utility of the movie. In this case, the individual can still focus on the movie

[6]The decision theoretic literature on temptation provides other possible models to extend to our setting of sensory limitations. For example, representations could be adapted to allow for uncertainty about salience (Stovall, 2010) or to allow for multiple measures of salience (Dekel et al., 2009). However, to keep our paper focused, we do not introduce these representations. In addition, our experiment does not provide enough structure to differentiate these models from the GP representation.

and benefit from watching it, but utility is lower since the concentration cost must be deducted from the utility of the less salient stimulus. Essentially, in the GP representation an individual can still focus on only one stimulus but has the ability to change their focus from one stimulus to another, possibly at a utility cost.

It is straightforward to show that both the Strotz and GP preferences satisfy:

Set Betweenness. If $X \succeq Y$, then $X \succeq X \cup Y \succeq Y$.

However only the Strotz representation satisfies:

No Compromise. For all X and Y, either $X \sim X \cup Y$ or $Y \sim X \cup Y$.

Recognizing these differences will allow us to test in an experimental setting whether a decision maker is consistent with the Strotz and GP preferences. It will also allow us to potentially differentiate between the two models.[7]

One obvious alternative model to compare ours to is an additive model. After all, if X represents the set of stimuli experienced by the individual, perhaps the utility from X is simply the sum of utilities from each stimulus:

$$U^A(X) = \sum_{x \in X} u(x).$$

The key behavioral property of the additive model is that any single stimulus is either always positive, always negative, or always neutral.

Definition. We say stimulus x is *universally positive* if for every $X \not\ni x$ we have $X \cup \{x\} \succ X$. We say x is *universally negative* if for every $X \not\ni x$ we have $X \succ X \cup \{x\}$. We say x is *universally neutral* if for every $X \not\ni x$ we have $X \cup \{x\} \sim X$.

[7] Given the limited data we can observe in an experimental setting, we can only potentially differentiate GP preferences from Strotz preferences; we cannot differentiate Strotz preferences from GP preferences. This is because any preference consistent with No Compromise is also consistent with Set Betweenness, while there are preferences consistent with Set Betweenness that are inconsistent with No Compromise (e.g. $X \succ X \cup Y \succ Y$).

Additivity. For every $x \in A$, x is exactly one of the following: universally positive, universally negative, universally neutral.

More generally, we will refer to the additive model to be not just those preferences that can be represented by the functional form above, but to be all preferences satisfying the axiom Additivity. Note that Set Betweenness and Additivity are generally opposed, as the following lemma shows.

Lemma 1. *Suppose $\succeq$ satisfies $\{w\} \succ \{x\} \succ \{y\} \succ \{z\}$. Then $\succeq$ cannot satisfy Additivity and Set Betweenness.*

Proof. Suppose $\succeq$ satisfies Set Betweenness. Then we have $\{x\} \succeq \{x, y\} \succeq \{y\}$, with one of these strict.

Case 1 – $\{x\} \succ \{x, y\}$. Then y is not universally positive or universally neutral. However since $\{y\} \succ \{z\}$, Set Betweenness implies $\{y, z\} \succeq \{z\}$. But this implies that y cannot be universally negative.

Case 2 – $\{x, y\} \succ \{y\}$. Then x is not universally negative or universally neutral. However since $\{w\} \succ \{x\}$, Set Betweenness implies $\{w\} \succeq \{w, x\}$. But this implies that x cannot be universally positive. □

Thus under Strotz and GP preferences, a stimulus is not universally positive or negative — it depends on the set of stimuli to which it is added.

3.2 Prior Literature Motivating our Model

We conclude this section with a discussion on how our model fits in with other theoretical work in economics and psychology. Though our model borrows the utility representations from Strotz's and GP's work in the temptation literature, more broadly our model belongs to the growing literature on preference over sets (e.g. Kreps, 1979; Bossert et al., 2000; Dekel et al., 2001). However, this literature generally views a set of alternatives as representing an opportunity set from which the decision maker will

ultimately consume a single alternative in an unmodeled future period, whereas in our model the decision maker consumes (or experiences) all alternatives in the set. There has been some work that considers preferences over sets in which the alternatives are not mutually exclusive. But this has usually been in the context of group choice, such as purely hedonic games (Bogomolnaia and Jackson, 2002), voting (Barberà et al., 1991), and matching (Roth, 1985). See Barberà et al. (2004) for a comprehensive review of the literature on ranking sets.

Salience plays an important role in the work of Bordalo et al. (2012, 2013b,a, 2015), Kőszegi and Szeidl (2012) and Ellis and Masatlioglu (2017). However, while our work and theirs do share the term 'salience', the implementation of the idea is different. In their frameworks, salience is an objective attribute of a product that attracts the focus of an individual. In addition, this literature focuses on choice from sets, and thus does not address the inability of individuals to experience multiple stimuli at once. Consequently, those concepts of salience have difficulty explaining the seemingly dysfunctional behaviors we examine. They also cannot explain the behavior of subjects in our experiment.

This paper also builds upon several important literatures in psychology and neuroscience. Going back to Cherry (1953) and Broadbent (1958), psychologists have outlined models of attention in which an individual is incapable of attending to all stimuli to which she is exposed. Early experiments were in the domain of listening in which individuals attempted to attend to one specific voice when exposed to multiple sounds. Significantly for the purpose of our model, Cherry (1953) found that when individuals were exposed to a different voice in each ear, they could focus on one or the other but could not follow both at once. Indeed, subjects had no recollection of what was said by the non-focal voice. Subsequent to these early experiments, similar behavior was observed with regards to vision (Treisman and Gelade, 1980), pain (McCaul and Haugtvedt, 1982), and even anxiety (Borkovec and Roemer, 1995).

This inability to attend to multiple stimuli at once is consistent with the model we present. However, the simplest version of our model implies that while individuals may have control over which stimuli they expose themselves to, they attend only to the most salient stimulus to which they are exposed. This is, of course, a simplification. Individuals exhibit an ability to choose which stimulus they wish to focus their attention on (Cherry, 1953). They also exhibit trouble, however, focusing on the preferred stimulus when exposed to distracting stimuli that are particularly salient (Lavie, 2010). Additionally, concentrating on a preferred stimulus in the presence of salient distractions is mentally taxing and requires substantial effort (Boksem et al., 2005).[8] This motivates the GP representation of our model which allows individuals to focus on a less salient stimulus with some cost to utility. The behavioral implications in this extension are similar, however, to our simple case.

4 Laboratory Experiment

4.1 Experimental Protocol

We conducted a laboratory experiment to test the key predictions of our model, No Compromise and Set Betweenness. The experiment was administered to each subject individually in a private room. Upon signing the consent form, the subject was asked to practice each of four unpleasant tasks for 30 seconds. The tasks were: 1) listening to a loud (85 decibel) fire alarm sound through headphones, 2) holding a hand in ice cold water, 3) doing 1 and 2 simultaneously, and 4) holding two hands in ice cold water.[9] To avoid potential order effects, subjects practiced each task twice in random

[8] Psychologists have demonstrated that attention control is a cognitive activity governed by the prefrontal cortex (Kane and Engle, 2002). Individuals with attention deficit hyperactivity disorder (ADHD) (Swanson et al., 2003), schizophrenia (Everett et al., 1989), and other mental disorders have substantially greater difficulty focusing attention on non-salient stimuli.

[9] While unpleasant, these tasks were in compliance with Occupational Safety and Health Administration regulations and not deemed dangerous by a physician. The Institutional Review Board of Brigham Young University approved the protocol. Kahneman et al. (1997) and others have tested

order. For the purposes of the current study, we focus on the first three tasks, though in the results section we briefly discuss results from the fourth task.

Participants were then asked to write down the minimum amount (from \$0 to \$15, in fifty-cent increments) that they would be willing to be paid to complete each of the four tasks for two minutes. They were told that after they write down the amount for each task, the experimenter would randomly select one of the four tasks as well as one of ten threshold amounts, ranging from \$0 to \$15, from an envelope. If the chosen threshold amount exceeds the minimum amount listed by the subject for a particular task, the subject would receive that amount after she completes the task for two minutes. If the threshold amount is less than the minimum amount listed by the subject for a particular task, the subject would not have an opportunity to complete the task, but would need to sit in the room for the remaining two minutes until they can be paid a show-up fee of \$2. This protocol ensured that the amounts listed by the subjects were incentive compatible and strategy-proof. From these amounts, we construct subjects' preferences over the tasks.

To ensure the subjects' understanding of the instructions, before writing down the amounts, they were asked to answer multiple comprehension questions correctly to proceed. At the end of the experiment, subjects completed a post-experimental questionnaire intended to collect their demographic information, including gender, age, university status, GPA, and study major. An average experimental session lasted approximately 15 minutes and subjects earned \$6 on average, in addition to the show up fee.

Subjects in our experiment were students at Brigham Young University recruited through email advertisements. The exact content of the recruitment email and detailed experimental instructions are included in the appendix.

Relating this experiment to our proposed model, let ℓ and h denote listening to

individuals' responses to painful stimuli by having subjects place their hand in cold water.

the loud siren for two minutes and submerging one hand in ice water for two minutes respectively. How a subject ranks $\emptyset$, $\{\ell\}$, $\{h\}$, and $\{\ell, h\}$ (as revealed by his/her reservation payments) will determine whether he or she is consistent with a given model. For example, the ordering

$$\emptyset \succ \{\ell\} \succ \{\ell, h\} \succ \{h\}$$

is consistent with Set Betweenness, but not No Compromise or Additivity. Hence this ordering is consistent with GP preferences, but not Strotz or additive preferences. As another example, the ordering

$$\emptyset \sim \{\ell\} \succ \{\ell, h\} \sim \{h\}$$

is consistent with No Compromise, Set Betweenness, and Additivity. Hence it is consistent with Strotz, GP, and additive preferences.

4.2 Results

We collected data from 65 subjects. From this sample, we drop 5 subjects who stopped the protocol prior to completion. This leaves us with an analysis sample of 60 subjects who completed the protocol. Table 1 shows summary statistics for these subjects. Consistent with the fact that we recruited subjects in a university setting, the average age is approximately 22 years old. Only about a third of our subjects are female and a majority are white. After practicing with all of the tasks, students reported the minimum amount they would be willing to accept to perform each of the tasks, which we refer to as a reservation payment. The average reservation payments for the various tasks range from \$2.50 to \$5.28.

We now turn our attention to the theoretical predictions of our model. Table 2 shows the fraction of subjects exhibiting each preference ordering across sets of stim-

uli. Note that only the preference orderings actually exhibited by subjects are shown in the table. Table 3 shows more concisely the fraction of subjects exhibiting preferences consistent with each utility representation. Note that the sum of these fractions exceeds one since some of these preferences are consistent with multiple models. Examining the table, we see that 63 percent of subjects exhibit preferences consistent with the Strotz representation in that the reservation payment is exactly equal to the reservation payment of one of the stimuli in isolation. 68 percent exhibit preferences consistent with a GP representation in that the reservation payment for both stimuli lies in the closed interval between the reservation payments of the two individual stimuli. Only 40 percent of subjects exhibit additive preferences in which an additional stimulus increases the reservation payment. One individual (2 percent of subjects) demonstrates inconsistent preferences in the sense that the subject reports no disutility from hearing the siren yet the reservation payment for the siren and hand exceeds that of the hand alone. Collectively, the experimental evidence strongly suggests that most individuals only have a limited ability to experience multiple latent stimuli at the same time. Strikingly, the simple Strotz model in which an individual is able to experience the utility of only a single stimulus is sufficient to explain a majority of subjects' decisions in this setting.

Note that while the experimental evidence is consistent with our model for a majority of subjects, the siren serves as an effective distraction for the hand in cold water for only 7 percent of subjects. In order for the loud siren to be an effective distraction, it must be the case that the measures of salience and disutility to be discordant across the two stimuli. In other words, the siren needs to be less painful and more salient than the hand in cold water. It may not be surprising that only a minority of subjects experience the stimuli in this fashion. It seems plausible that often the most painful stimulus will also be the most salient. However, in non-experimental settings individuals have a potentially broad set of stimuli to choose

from as distractions including watching TV, working in the office, or substance abuse. There is no reason to think that one person's optimal distraction would necessarily be optimal for a different individual.

To represent our results another way, Figure 1 shows the empirical CDF of subjects' reservation payment for one hand in cold water and listening to a siren divided by the maximum reservation payment of the two stimuli in isolation. In our sample, all subjects for whom this ratio is equal to or below one exhibit preferences consistent with the GP and possibly Strotz representations.[10] We see that fully two thirds of individuals reveal a reservation payment for the two stimuli less than or equal to the greater of the two reservation payments corresponding to the individual stimuli. There is a discontinuous increase in the density at a value of 1 demonstrating that for many subjects the disutility of two stimuli is exactly equal to the disutility of the most uncomfortable individual stimulus. The behavior of a majority of subjects stands in stark contrast to the prediction of any additive model in which each additional stimulus should increase the subject's reported reservation payment.

One might be concerned that our results are driven in part by individuals who do not find the stimuli unpleasant. The results are virtually identical if we exclude the 6 observations who report a reservation payment of 0 for one or more stimuli.

As mentioned in our description of the protocol, we also elicited from subjects their reservation payment for placing two hands in cold water. In this setting 35 percent of subjects indicate the same reservation payment for placing two hands in water as for placing one hand in water. 2 percent (one subject) indicate a lower reservation payment for two hands than for one. 63 percent require a reservation payment for two hands that exceeds the reservation payment for putting a single hand into cold water. If we view each hand as a separate stimulus, 35 percent of subjects in this setting

[10] It is possible for preferences to be inconsistent with the Strotz or GP preferences and still have this ratio be less than one: $\{\ell, h\} \succ \{\ell\} \succ \{h\}$ is one such example. However none of our subjects exhibited such preferences.

demonstrate preferences consistent with the GP and Strotz representations. To the extent that placing two hands in cold water represents an increased intensity of the same stimulus relative to placing one hand in cold water, this part of the protocol provides a less informative test of the theory.

5 Future Research Directions

Our model provides a simple framework for economists to understand a variety of interesting behaviors relevant for health and economics. Additionally, we present empirical evidence suggesting this model is important for explaining human behavior. However, the simplicity of the model and the lab setting of the empirical evidence provide room for future researchers. In this section we outline possible future directions both in terms of modelling and empirical analyses.

5.1 Model Extensions

Psychologists, including Muraven et al. (1998), Muraven and Baumeister (2000), and Schmeichel and Vohs (2009), have established that individuals possess a limited capacity for self-control in a variety of domains including emotional regulation. In a dynamic extension of the model, it would be interesting to explore the implications of a limited attention resource that could be allocated to focus on stimuli which are not the most salient. For example, an individual could choose whether to allocate this limited attention resource to focus on work instead of a headache. However, such a decision could make it infeasible to later attend to family responsibilities that are less salient than the ongoing health challenge.

We currently don't model how stimuli are either produced or alleviated. Instead, we outline a structure in which individuals have preferences over sets of stimuli in which the choice of available sets is outside of the model. As researchers further inves-

tigate specific applications, it will be helpful to place more structure on the process of how stimuli come into being. Such extensions might also include investigations in which stimuli can be moderated on the intensive margin as well as the extensive margin, which is the focus of our current analysis.

In our study, we assume that individuals can only experience a single stimulus. This assumption may work well in some settings, such as our motivating examples or the setting we test in our experiment. However, it is certainly the case that in other settings multiple stimuli can be experienced jointly. For example, music in conjunction with fine dining may be more pleasant than either in isolation. Future researchers could aid in the refinement of the model through empirical studies that determine the settings in which stimuli can be experienced jointly and settings in which one stimulus acts to displace another.

Our model is agnostic with regards to which stimuli are most salient and why. A better understanding of the determinants of salience would allow researchers to place additional structure on the model that could increase both its predictive power and policy relevance. Such an understanding could allow researchers to predict the specific circumstances under which individuals engage in harmful distractions or become paralyzed by the existence of multiple negative stimuli. It might also allow researchers to design more effective interventions to help individuals coping with such problem behaviors. We encourage researchers to engage in the empirical research required to better understand the determinants of salience.

On a related note, salience is a fixed characteristic of a stimulus in our model. However, psychologists (Higgins, 1996) have long known that priming individuals about characteristics of a situation or even their own identity can increase the salience of particular dimensions of a problem, situation, or identity. For example, Papoiu et al. (2011) find that when individuals observe someone scratching their skin, they are more likely to experience itching sensations themselves. More significantly, counsellors

will work with clients to reframe the way in which they perceive the life situations or stimuli to which they are exposed. This implicitly involves increasing the salience of some stimuli to improve the client's level of function and well-being. Similarly, brain chemistry may affect relative salience of positive and negative stimuli. Consequently, salience of specific stimuli may also be affected by psychotropic medications such as those designed to alleviate clinical depression. It seems likely that extensions of the model to endogenize salience would be a fruitful research direction.

5.2 Further Tests of the Model

While our experiment represents a convincing test of our model, future researchers should consider additional lab and quasi-experiments as well as observational analyses more closely aligned to the economic and clinical phenomena that we used to motivate the model. For each of our motivating examples, we outline possible protocols, quasi-experiments, and data sources.

Thus far, analyses of NSSI have been primarily observational using convenience samples not generally available to the public. Experimental tests to understand the reasons for NSSI are challenging due to the need to observe strict ethical and safety standards.[11] Our model explains NSSI as an effort of individuals to distract themselves from painful psychological processes by exposing themselves to a less painful but more salient physical stimulus. The key testable implications of our model in this setting are that 1) the incidence of NSSI will be higher with an exogenous in-

[11] Fox et al. (2017) provide experimental evidence on why NSSI engagement may make individuals feel better. Examining a population of individuals with a history of self-harm, the researchers induced a negative mood by asking the subjects to spend five minutes writing about the most significant time "in which they failed or let themselves down in their life." Control subjects were exposed to no additional stimulus while other subjects were also exposed to a physically painful stimulus. One might predict that, according to our model, individuals exposed to a physically painful stimulus would experience improved mood relative to the control group. However, this presupposes that the priming of negative experiences induced greater distress that was both more painful and less salient than the physically painful stimulus. If this condition doesn't hold, there is no reason to expect mood or perceived utility to improve by the addition of a painful physical stimulus to a painful psychological stimulus.

crease in stressful events and 2) increasing the availability of non-harmful distractions can improve individuals' ability to deal with stressful situations without engaging in NSSI.

Testing this first prediction can be effectively done providing that one can track NSSI behavior of individuals and identify exogenous sources of psychological stress. For example, one could follow a set of youth who found school to be a very stressful setting. Following Jacob and Lefgren (2003), one could observe whether NSSI was less common during teacher in-service days when school was out of session to otherwise similar days when school was in session. This would directly test the prediction of our model that NSSI occurs as an endogenous response to negative stimuli.

The second implication could be tested experimentally in a clinical setting. In the control group, individuals would receive appropriate treatment for NSSI (Turner et al., 2014). In treated group, this treatment would be supplemented with joint explorations between the care provider and the subjects regarding distractions which were less harmful but also sufficiently salient to provide an effective distraction to emotional distress. Our model suggests that such treatments could reduce the incidence of NSSI.

Considering our example of depression, we predict that interventions that reduce the severity of depression should increase engagement with activities that are insufficiently salient to be enjoyed in the depressed state. One of the sub-components of the commonly used Hamilton Depression Rating Scale (HDRS-6) asks whether individuals have exhibited "loss of interest in activity, hobbies or work." Our model predicts that interventions that reduce the incidence of depression should have the effect of improving this sub-component of the HDRS-6. Consistent with our theory, experimental studies of selective serotonin reuptake inhibitors (SSRI), a class of drugs used to treat depression, have shown improvement in this domain (Hieronymus et al., 2019). Data from many clinical trials registered at the Food and Drug Administration (FDA) can be requested at vivli.org. Additional tests of our theory could focus on

this measure of depression using new or existing datasets.

The key implication of the "trap of competing problems" example is that multiple problems lead to inaction regarding any problem. In a lab setting, one can experimentally adjust the number of negative stimuli to which an individual is exposed. Then one can see whether individuals exposed to multiple negative stimuli are less likely to engage in costly actions to reduce any of them. Alternatively, consider a pool of subjects struggling to deal with multiple problems. Our model suggests that the exogenous provision of help with one problem should induce complementary efforts on the part of subjects to reduce the severity of additional problems.

In observational settings, a test of the "trap of competing problems" could leverage information from the American Time Use Survey (ATUS). In particular, one could observe the extent to which individuals engage in distracting behavior (e.g. television watching or video games) relative to productive behavior. However, this would need to be coupled with information on the number of multiple stresses to which an individual is exposed. For example, our model would predict that some individuals who exogenously lose their job may engage in less home production and instead engage in distracting activities. This would be because the joint worries of job loss and home concerns would reduce willingness to invest in either one. Fortunately, the ATUS samples subjects from the Current Population Survey (CPS), which includes information on employment status.

6 Conclusion

We reinterpret the Strotz and GP utility representations to provide insight into how people behave when they have only a limited ability to experience competing latent stimuli. One key insight from our model regards the complementarity of negative stimuli. In particular, once one is experiencing a salient and negative stimulus the

utility cost of additional negative stimuli might be quite small. This can lead to a variety of seemingly dysfunctional behaviors that are nevertheless consistent with utility maximization.

In particular, our theoretical framework provides explanations for phenomena including destructive distractions, an unwillingness to ameliorate negative situations in one's life, and the apathy of severe depression. In each of these examples, actions that would seem to objectively improve an individual's life fail to be optimal if the benefits are insufficiently salient to rise above the pain of other life circumstances. Furthermore, individuals may engage in behavior that seems destructive if doing so provides a salient distraction from a greater pain. Understanding how these behaviors are optimal from a utility maximization perspective may provide researchers, policy makers, and clinicians with insights regarding how to better help individuals in these conditions. In particular, policies to improve the circumstances of such individuals should take into account the full set of latent circumstances and stimuli to which an individual is exposed.

While our theory has a strong intuitive appeal, we also present convincing empirical evidence that our theoretical framework predicts behavior in a setting with both real payoffs and consequences. We find that over two thirds of individuals exhibit preferences consistent with our framework in an experiment in which subjects are exposed to single and multiple painful stimuli. In particular, after experiencing the stimuli together and in isolation, 68 percent of individuals experience the same or *less* disutility from two painful stimuli than one of the stimuli in isolation. This suggests that, at least in this experimental context, our framework explains the majority of peoples' choices in a way that runs counter to what an additive model would predict.

Tables and Figures

Table 1: Summary Statistics

Variable	Mean
Age	22.18
	(2.21)
Female	0.35
	(0.48)
GPA	3.68
	(0.34)
White	0.83
	(0.38)
Asian	0.13
	(0.34)
Reservation Payment to Perform Task	
Siren	\$2.50
	(2.21)
One Hand in Water	\$4.12
	(2.90)
Siren and One Hand	\$4.57
in Water	(3.17)
Two Hands in Water	\$5.28
	(3.56)
Observations	60

Notes: Standard deviations are in parentheses.

Table 2: Preference Ordering of Various Stimuli

Preference Ranking	Fraction	Consistent Representations
$\emptyset \succ \{\ell\} \succ \{\ell, h\} \sim \{h\}$	0.27	Strotz, GP
	(0.06)	
$\emptyset \succ \{h\} \succ \{\ell, h\} \sim \{\ell\}$	0.02	Strotz, GP
	(0.02)	
$\emptyset \succ \{\ell\} \sim \{\ell, h\} \succ \{h\}$	0.05	Strotz, GP
	(0.03)	
$\emptyset \succ \{\ell\} \sim \{\ell, h\} \sim \{h\}$	0.20	Strotz, GP
	(0.06)	
$\emptyset \sim \{\ell\} \succ \{\ell, h\} \sim \{h\}$	0.07	Strotz, GP, Additive
	(0.03)	
$\emptyset \sim \{\ell\} \sim \{\ell, h\} \sim \{h\}$	0.03	Strotz, GP, Additive
	(0.02)	
$\emptyset \sim \{\ell\} \succ \{\ell, h\} \succ \{h\}$	0.03	GP
	(0.02)	
$\emptyset \succ \{\ell\} \succ \{\ell, h\} \succ \{h\}$	0.02	GP
	(0.02)	
$\emptyset \succ \{\ell\} \succ \{h\} \succ \{\ell, h\}$	0.23	Additive
	(0.06)	
$\emptyset \succ \{h\} \sim \{\ell\} \succ \{\ell, h\}$	0.05	Additive
	(0.03)	
$\emptyset \succ \{h\} \succ \{\ell\} \succ \{\ell, h\}$	0.02	Additive
	(0.02)	
$\emptyset \sim \{\ell\} \succ \{h\} \succ \{\ell, h\}$	0.02	Inconsistent
	(0.02)	
Observations	60	

Notes: Robust standard errors are in parentheses.

Table 3: Fraction of Subject Preferences Consistent with Each Utility Representation

Utility Representation	Fraction
Strotz	0.63
	(0.06)
GP	0.68
	(0.06)
Additive	0.40
	(0.06)
Inconsistent	0.02
	(0.02)
Observations	60

Notes: Robust standard errors are in parentheses.

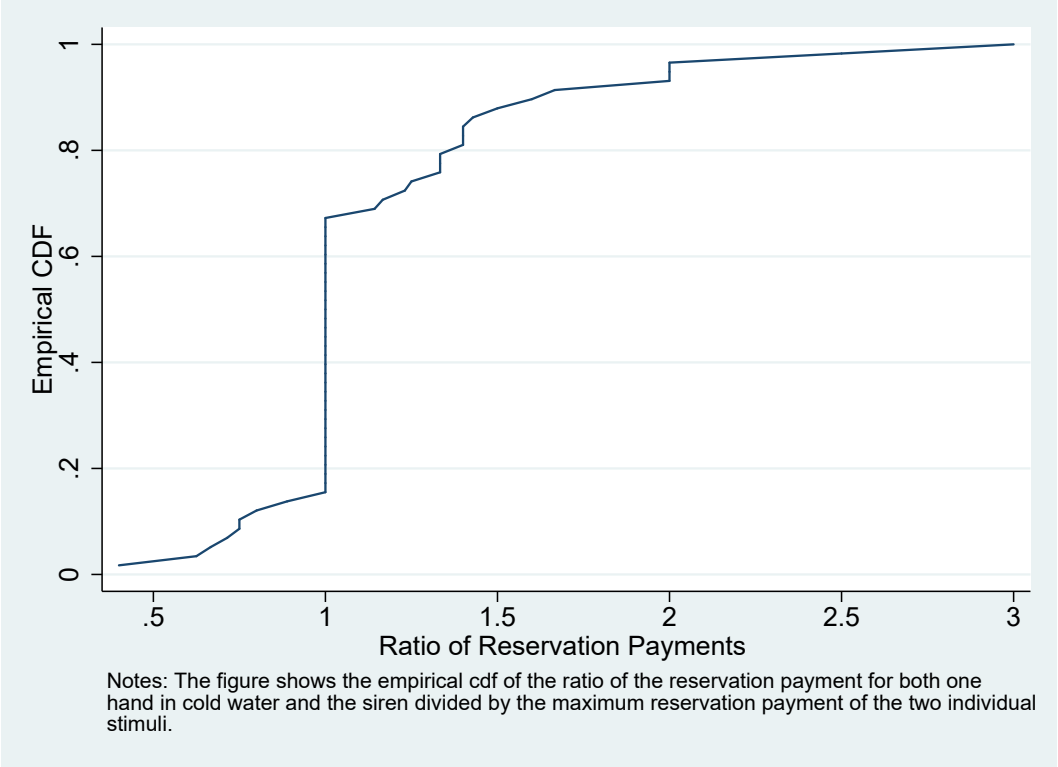

Figure 1: Are Reservation Payments Higher for Two Negative Stimuli than for One?

Appendix - Experimental Protocol

A Email announcement

New time-slots are available for the research study "Economics of decision making" next week. If you are interested in participating, please sign up for an available session of "Economics study on decision making." Please show up on time to your scheduled session in room 340 TNRB (Behavioral Lab).

*The link that will be provided (http://byu-marriott.sona-systems.com/Default.aspx?ReturnUrl=%2f) which will direct the students to the SONA recruitment site where they can sign up for a session. They will see the following details about the project before they sign up.

Study name: Economics experiment on decision making.
Brief abstract: In this study you will participate in a task that will involve listening to a loud (85 dB) sound as well as holding your hands submerged in cold water. You will also complete a short demographic questionnaire. Please read the eligibility requirements carefully before you sign up to participate.
Eligibility requirements and risks: To participate in the study, you need to have good hearing and be willing to have your hands submerged in uncomfortably cold water for up to 5 minutes. If you have poor hearing and/or use a hearing device, you cannot participate in the experiment.

Also, be aware that there are certain medical pre-existing conditions, such as circulatory, rheumatological, and autoimmune disorders, which could cause longer-term symptoms. If you have any of these conditions, you will not be able to participate in this experiment.
Duration: 15 minutes
Pay: $6 on average, depending on your decisions. The exact compensation amounts

range from $2 and $15, depending on your decisions.

If you have any questions, please contact the primary researcher, Olga Stoddard, Ph.D. at olga.stoddard@byu.edu

B Participant Instructions

Welcome and thank you for your participation in this study. Please follow along as we read through the instructions. If you have any questions, please feel free to ask these questions as we move forward.

If you complete the study, you will receive at least $2 for your participation. The exact amount that you will be paid at the end of the study depends on your decisions, as I will explain shortly.

There are four tasks that you will need to perform during this experiment. To begin, we would like you to experience and practice each task for 30 seconds, two times (for a total of eight practice runs). We would like you to:

Treatment 1:	The subject listens to a loud siren-like sound (at 85 decibels) through headphones.
Treatment 2:	The subject submerges one hand in ice-cold water (41° F) up to the wrist.
Treatment 3:	The subject submerges both hands in ice-cold water up to. the wrists.
Treatment 4:	The subject submerges one hand in ice-cold water up to the wrist while listening to a loud siren-like sound through headphones.

Next, you are asked to write down the minimum amount (in 50 cent increments, up to $15) that you would be willing to accept to do one of these tasks for 2 minutes.

You will need to write down the minimum amount for each task, but only one of them will actually be selected for you to do. These four envelopes contain all four tasks. After you complete this sheet, you will select an envelope containing one of the four tasks.

Also, prior to the study, we have randomly chosen 10 different threshold amounts between \$0 and \$15 and have placed each one in an envelope. After you indicate the minimum amount you are willing to accept to do the task, you will be asked to choose one of the envelopes at random reveal the threshold amount. If the minimum amount you are willing to accept for the chosen task is less than the threshold amount, then you will be paid the threshold amount once you complete the task for 2 minutes. If the amount you are willing to accept is more than the threshold amount, then you will not have an opportunity to complete the task, and will be paid a \$2 show up fee to sit in the room for the remaining 2 minutes.

Note that it is in your best interest to state the true minimum amount that you are willing to accept since you can never receive less than that amount, but you can end up with a lot more.

To make sure that you understand the instructions, consider two hypothetical examples:

Example 1: Suppose that the minimum amount that you listed to do the chosen task is \$5. We draw the threshold amount of \$7.

Question: Will you have to do the task? How much will you earn?

Example 2: Suppose that the minimum amount that you listed to do the chosen task is \$5. We draw the threshold amount of \$3.

Question: Will you have to do the task? How much will you earn?

C In-Experiment Questionnaire

Participant ID #:

What is the minimum amount (in 50 cent increments, up to $15) that we would have to pay you to do one of these tasks for 2 minutes?

Task	Amount ($0 - $15)
Listen to the loud siren sound through the headphones	
Hold one hand submerged in ice cold water	
Hold two hands submerged in ice cold water	
Hold one hand submerged in ice cold water while also listening to loud siren through the headphones	

D Post-Experimental Exit Questionnaire

Participant ID #:

1. What is your age? ___
2. What is your gender? ___ M(=1) F(=2)
3. What is your ethnicity? ___

   (0=Caucasian, 1=Asian, 2=Hispanic, 3=African American, 4=Other)
4. What is your major at BYU? ___

   (1=Business, 2=Economics, 3=English, 4=Science, 5= Sociology,

   6=Mathematics, 7=Other)
5. What is your GPA? ___ (0=2.0, 100=4.0)
6. What is your marital status? ___ (1=single, 2=married)
7. How easy were these instructions? ___ (0=easy, 100=Hard)
8. Did you like the experiment? ___ Yes(=1) No(=2)
9. Do you have any comments regarding this experiment?

References

Al-Khotani, Amal, Lanre A'aziz Bello, and Nikolaos Christidis (2016), "Effects of audiovisual distraction on children's behaviour during dental treatment: A randomized controlled clinical trial." *Acta Odontologica Scandinavica*, 74, 494 – 501.

American Psychiatric Association (2013), *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Anonymous (2014), "What does depression feel like?" URL `https://www.7cups.com/qa-anxiety-16/what-does-depression-feel-like-14`.

Barberà, Salvador, Walter Bossert, and Prasanta K. Pattanaik (2004), "Ranking sets of objects." In *Handbook of Utility Theory* (Salvador Barberà, Peter Hammond, and Christian Seidl, eds.), volume 2, 893–977, Springer.

Barberà, Salvador, Hugo Sonnenschein, and Lin Zhou (1991), "Voting by committees." *Econometrica*, 595–609.

Baumeister, R., E. Bratslavsky, M. Muraven, and D. Tice (1998), "Ego depletion: Is the active self a limited resource?" *Journal of Personality and Social Psychology*, 74, 1252–1265.

Beck, Arne, A. Lauren Crain, Leif I. Solberg, Jürgen Unützer, Russell E. Glasgow, Michael V. Maciosek, and Robin Whitebird (2011), "Severity of depression and magnitude of productivity loss." *The Annals of Family Medicine*, 9, 305–311.

Becker, Gary S. and Kevin M. Murphy (1988), "A theory of rational addiction." *Journal of political Economy*, 96, 675–700.

Berndt, Ernst R., Stan N. Finkelstein, Paul E. Greenberg, Robert H. Howland, Alison Keith, A. John Rush, James Russell, and Martin B. Keller (1998), "Workplace

performance effects from chronic depression and its treatment." *Journal of Health Economics*, 17, 511–535.

Bogomolnaia, Anna and Matthew O. Jackson (2002), "The stability of hedonic coalition structures." *Games and Economic Behavior*, 38, 201–230.

Boksem, A.S., F. Meijman, and M. Lorist (2005), "Effects of mental fatigue on attention: An ERP study." *Cognitive Brain Research*, 25, 107–116.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2012), "Salience theory of choice under risk." *The Quarterly Journal of Economics*, 127, 1243–1285.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013a), "Salience and asset prices." *American Economic Review*, 103, 623–28.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2013b), "Salience and consumer choice." *Journal of Political Economy*, 121, 803–843.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer (2015), "Competition for attention." *The Review of Economic Studies*, 83, 481–513.

Borkovec, T.D. and L. Roemer (1995), "Perceived functions of worry among generalized anxiety disorder subjects: Distraction from more emotionally distressing topics?" *Journal of Behavior Therapy and Experimental Psychiatry*, 26, 25–30.

Bossert, Walter, Prasanta K. Pattanaik, and Yongsheng Xu (2000), "Choice under complete uncertainty: Axiomatic characterizations of some decision rules." *Economic Theory*, 16, 295–312.

Broadbent, Donald E. (1958), *Perception and Communication*. Pergamon Press, London.

Carpenter, Christopher S., Chandler B. McClellan, and Daniel I. Rees (2017), "Economic conditions, illicit drug use, and substance use disorders in the United States." *Journal of Health Economics*, 52, 63 – 73.

Cawley, John and Christopher J. Ruhm (2011), "The economics of risky health behaviors." In *Handbook of Health Economics* (Mark V. Pauly, Thomas G. Mcguire, and Pedro P. Barros, eds.), volume 2 of *Handbook of Health Economics*, chapter 3, 95 – 199, Elsevier.

Cherry, E. Colin (1953), "Some experiments on the recognition of speech, with one and with two ears." *Journal of the Acoustical Society of America*, 25, 975—-979.

Darden, Michael (2017), "Smoking, expectations, and health: A dynamic stochastic model of lifetime smoking behavior." *Journal of Political Economy*, 125, 1465–1522.

Darden, Michael E. and Nicholas W. Papageorge (2018), "Rational self-medication." NBER Working Paper No. 25371.

Dean, Emma Boswell, Frank Schilbach, and Heather Schofield (2019), "Poverty and cognitive function." In *The Economics of Poverty Traps* (Christopher B. Barrett, Michael R. Carter, and Jean-Paul Chavas, eds.), National Bureau of Economic Research Conference Report, University of Chicago Press.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2001), "Representing preferences with a unique subjective state space." *Econometrica*, 69, 891–934.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini (2009), "Temptation-driven preferences." *Review of Economic Studies*, 76, 937–971.

Doyle, Louise, Margaret P. Treacy, and Ann Sheridan (2015), "Self-harm in young people: Prevalence, associated factors, and help-seeking in school-going adolescents." *International Journal of Mental Health Nursing*, 24, 485–494.

Ellis, Andrew and Yusufcan Masatlioglu (2017), "A regional approach to salience." Working paper.

Escarce, José J., Kenneth R. Epstein, David C. Colby, and J. Sanford Schwartz (1993), "Racial differences in the elderly's use of medical procedures and diagnostic tests." *American Journal of Public Health*, 83, 948–954.

Everett, J., L. Laplante, and J. Thomas (1989), "The selective attention deficit in schizophrenia: Limited resources or cognitive fatigue?" *Journal of Nervous and Mental Disease*, 177, 735–738.

Filcheck, Holly A., Keith D. Allen, Hilary Ogren, James Brant Darby, Brian Holstein, and Steve Hupp (2005), "The use of choice-based distraction to decrease the distress of children at the dentist." *Child & Family Behavior Therapy*, 26, 59–68.

Firestone, R. W. and R. H. Seiden (1990), "Suicide and the continuum of self-destructive behavior." *Journal of American College Health*, 38, 207–2013.

Fox, K.R., K.E. Toole, J.C. Franklin, and J.M. Hooley (2017), "Why does nonsuicidal self-injury improve mood? A preliminary test of three hypotheses." *Clinical Psychological Science*, 5, 111–121.

Frasquilho, D., M. MAtos, F. Salonna, D. Guereiro, C. Storti, and adn Caldas de Almeida J. Gapar, T. (2016), "Mental health outcomes in times of economic recession: a systematic literature review." *BMC Public Health*, 16, 115–126.

Friedman, M., E. Laufer, M. Laufer, and M. Wohl (1972), "Attempted suicide and self-mutilation in adolescence: Some observations from a psychoanalytic research project." *International Journal of Psychoanalysis*, 53, 179–183.

Giblin, Hailey (2006), *Hailey's Story*. John Blake.

Gotlib, I. and J. Joormann (2010), "Congnition and depression: current status and future directions." *Annual Review of Clinical Psychology*, 6, 285–312.

Green, Francis (2011), "Unpacking the misery multiplier: How employability modifies the impacts of unemployment and job insecurity on life satisfaction and mental health." *Journal of Health Economics*, 30, 265–276.

Gul, Faruk and Wolfgang Pesendorfer (2001), "Temptation and self-control." *Econometrica*, 69, 1403–1435.

Gul, Faruk and Wolfgang Pesendorfer (2005), "The simple theory of temptation and self-control." Mimeo.

Hakulinen, C., M. Elovainio, M. Arffman, S. Lumme, S. Pirkola, I. Keskimaki, K. Manderbacka, and P. Bockerman (2019), "Mental disorders and long-term labour market outcomes: nationwide cohort study of 2055720 individuals." *Acta Psychiatrics Scandinavia*, 140, 371–381.

Hansen, Ben and Matthew Lang (2011), "Back to school blues: Seasonality of youth suicide and the academic calendar." *Economics of Education Review*, 30, 850–861.

Haushofer, Johannes and Ernst Fehr (2014), "On the psychology of poverty." *Science*, 344, 862–867.

Herpertz, S. (1995), "Self-injurious behavior: Psychopathological and nosologial characteristics in subtypes of self-injurers." *Acta Psychiatrica Scandinavica*, 91, 57–68.

Hieronymus, Fredrik, Alexander Lisinski, Staffan Nilsson, and Elias Eriksson (2019), "Influence of baseline severity on the effects of ssris in depression: An item-based, patient-level post-hoc analysis." *Lancet Psychiatry*, 6, 745–752.

Higgins, E. Tory (1996), "Knowledge activation: Accessibility, applicability, and

salience." In *Social Psychology: Handbook of Basic Principles* (E. Tory Higgins and Arie W. Kruglanski, eds.), Guilford Press.

Himber, J. (1994), "Blood rituals: Self-cutting in female psychiatric inpatients." *Psychotherapy*, 31, 620–631.

Jacob, Brian A. and Lars Lefgren (2003), "Are idle hands the devil's workshop? Incapacitation, concentration, and juvenile crime." *American Economic Review*, 93, 1560–1577.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin (1997), "Back to Bentham? Explorations of experienced utility." *The Quarterly Journal of Economics*, 112, 375–406.

Kane, M.J. and R.W. Engle (2002), "The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective." *Psychonomic Bulletin & Review*, 9, 637–671.

Kessler, Ronald C. (2012), "The costs of depression." *Psychiatric Clinics*, 35, 1–14.

Kharsati, Naphisabet and Poornima Bhola (2016), "Self-injurious behavior, emotion regulation, and attachment styles among college students in India." *Industrial Psychiatry Journal*, 25, 23–28.

Korn, C., T. Sharot, H. Walter, H. Heekeren, and R. Dolan (2014), "Depression is related to an absence of optimistically biased belief updating about future life events." *Psychology Medicine*, 44, 579–592.

Kőszegi, Botond and Adam Szeidl (2012), "A model of focusing in economic choice." *The Quarterly Journal of Economics*, 128, 53–104.

Kremer, Michael, Gautam Rao, and Frank Schilbach (2019), "Behavioral development economics." In *Handbook of Behavioral Economics - Foundations and Applications*

2 (B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds.), North-Holland.

Kreps, David M. (1979), "A representation theorem for 'preference for flexibility'." *Econometrica*, 47, 565–578.

Lavie, N. (2010), "Attention, distraction, and cognitive control under load." *Current Directions in Psychological Science*, 19, 143–148.

Lipman, Barton L. and Wolfgang Pesendorfer (2013), "Temptation." In *Advances in Economics and Econometrics: Tenth World Congress, Volume 1, Economic Theory* (Daron Acemoglu, Manuel Arellano, and Eddie Dekel, eds.), Econometric Society Monographs, Cambridge University Press.

Lund, C., A. Breen, A. Flisher, R. Kakuma, J. Corrrigall, J. Joska, L. Swartz, and V. Patel (2010), "Poverty and mental disorders in low and middle income countries: A systematic review." *Social Science Medicine*, 71, 517–528.

Mahuteau, Stephane and Rong Zhu (2015), "Crime victimisation and subjective well-being: Panel evidence from Australia." *Health Economics*, 25, 1448–1463.

Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao (2013), "Poverty impedes cognitive function." *Science*, 341, 976–980.

Marcotte, Dave (2003), "The economics of suicide, revisited." *Southern Economic Journal*, 69, 628–643.

Marcus, Jan (2013), "The effect of unemployment on the mental health of spouses – Evidence from plant closures in Germany." *Journal of Health Economics*, 32, 546–558.

McCart, Michael R., Ashli J. Sheidow, and Elizabeth J. Letourneau (2014), "Risk

reduction therapy for adolescents: Targeting substance use and HIV/STI-risk behaviors." *Cognitive and behavioral practice*, 21, 161–175.

McCaul, K.D. and C. Haugtvedt (1982), "Attention, distraction, and cold-pressor pain." *Journal of Personality and Social Psychology*, 43, 154–162.

Milkman, Katherine L., Julia A. Minson, and Kevin G. M. Volpp (2014), "Holding the Hunger Games hostage at the gym: An evaluation of temptation bundling." *Management Science*, 60, 283 – 299.

Miller, F. and E. A. Bashkin (1974), "Depersonalization and self-mutilation." *The Psychoanalytic Quarterly*, 43, 638–649.

Mullainathan, Sendhil and Eldar Shafir (2013), *Scarcity: Why Having Too Little Means So Much*. Times Books.

Muraven, Mark and Roy F. Baumeister (2000), "Self-regulation and depletion of limited resources: Does self-control resemble a muscle?" *Psychological Bulletin*, 126, 247–259.

Muraven, Mark, Dianne M. Tice, and Roy F. Baumeister (1998), "Self-control as a limited resource: Regulatory depletion patterns." *Journal of Personality and Social Psychology*, 74, 774–789.

Nock, M. K. and M. J. Prinstein (2004), "A functional approach to the assessment of self-mutilative behavior." *Journal of Consulting and Clinical Psychology*, 72, 885–890.

Nock, Matthew K. (2010), "Self-injury." *Annual Review of Clinical Psychology*, 6, 339–363.

Noonan, Kelly, Hope Corman, and Nancy E. Reichman (2016), "Effects of maternal depression on family food insecurity." *Economics & Human Biology*, 22, 201 – 215.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney (2010), "Socioeconomic disparities in health behaviors." *Annual Review of Sociology*, 36, 349–370.

Papoiu, A.D.P., H. Wang, R.C. Coghill, Y-H. Chan, and G. Yosipovitch (2011), "Contagious itch in humans: A study of visual 'transmission' of itch in atopic dermatitis and healthy subjects." *British Journal of Dermatology*, 164, 1299–1303.

Patton, G., S. Sawyer, J. Santelli, D. Ross, R. Afifi, N. Allen, M. Arora, P. Azzopardi, W. Baldwin, C. Bonell, R. Kakuma, and E. Kennedy (2016), "Our future: a lancet commission on adolescent health and wellbeing." *Lancet*, 287, 2423–2478.

Peng, Lizhong, Chad D. Meyerhoefer, and Samuel H. Zuvekas (2016), "The short-term effect of depressive symptoms on labor market outcomes." *Health Economics*, 25, 1223–1238.

Rachlin, Howard, Andres Raineri, and David Cross (1991), "Subjective probability and delay." *Journal of the Experimental Analysis of Behavior*, 55, 233–244.

Ridley, M., G Rao, F. Shilbach, and V. Patel (2020), "Poverty, depression, and anxiety: Causal evidence and mechanisms."

Rodriguez Andres, Antonio (2006), "Income inequality, unemployment, and suicide: A panel data analysis of 15 european countries." *Applied Economics*, 439–451.

Roth, Alvin E. (1985), "The college admissions problem is not equivalent to the marriage problem." *Journal of Economic Theory*, 36, 277–288.

Schilbach, Frank, Heather Schofield, and Sendhil Mullainathan (2016), "The psychological lives of the poor." *American Economic Review*, 106, 435–40.

Schmeichel, Brandon J. and Kathleen Vohs (2009), "Self-affirmation and self-control: Affirming core values counteracts ego depletion." *Journal of Personality and Social Psychology*, 96, 770–782.

Somer, Oya, Tezan Bildik, Bürge Kabukçu-Başay, Duygu Güngör, Ömer Başay, and Richard F. Farmer (2015), "Prevalence of non-suicidal self-injury and distinct groups of self-injurers in a community sample of adolescents." *Social Psychiatry and Psychiatric Epidemiology*, 50, 1163–1171.

Stovall, John E. (2010), "Multiple temptations." *Econometrica*, 78, 349–376.

Strine, Tara W., Ali H. Mokdad, Shanta R. Dube, Lina S. Balluz, Olinda Gonzalez, Joyce T. Berry, Ron Manderscheid, and Kurt Kroenke (2008), "The association of depression and anxiety with obesity and unhealthy behaviors among community-dwelling us adults." *General Hospital Psychiatry*, 30, 127–137.

Strotz, Robert H. (1955), "Myopia and inconsistency in dynamic utility maximization." *The Review of Economic Studies*, 23, 165–180.

Suyemoto, Karen L. (1998), "The functions of self-mutilation." *Clinical Psychology Review*, 18, 531–554.

Swannell, Sarah V., Graham E. Martin, Andrew Page, Penelope Hasking, and Nathan J. St. John (2014), "Prevalence of nonsuicidal self-injury in nonclinical samples: Systematic review, meta-analysis and meta-regression." *Suicide and Life-Threatening Behavior*, 44, 273–303.

Swanson, J.M., N.D. Volkow, J. Newcorn, B.J. Casey, R. Moyzis, D. Grandy, and M. Posner (2003), "Attention deficit hyperactivity disorder."

Tefft, Nathan (2011), "Insights on unemployment, unemployment insurance, and mental health." *Journal of Health Economics*, 30, 258–264.

Thaler, Richard (1981), "Some empirical evidence on dynamic inconsistency." *Economics Letters*, 8, 201–207.

Treisman, A.M. and G. Gelade (1980), "A feature-integration theory of attention." *Cognitive Psychology*, 12, 97–136.

Turner, Brianna J., Sara B. Austin, and Alexander L. Chapman (2014), "Treating nonsuicidal self-injury: A systematic review of psychological and pharmacological interventions." *Canadian Journal of Psychiatry*, 59, 576–585.

Wilson, Sarah (2019), "The other costs of children: Motherhood, substance use, and depression." *Working paper*.

Zetterqvist, Maria (2015), "The DSM-5 diagnosis of nonsuicidal self-injury disorder: A review of the empirical literature." *Child and Adolescent Psychiatry and Mental Health*, 9.