A Behavioral Economic Approach to Quantifying Reinforcing Efficacy of Food

By

Rachel N.S. Foster
B.A., Gonzaga University, 2017

Submitted to the graduate degree program in the Department of Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Derek D. Reed, Ph.D., BCBA-D

Erin B. Rasmussen, Ph.D.

Vincent T. Francisco, Ph.D.

Date Defended: Friday, October 23, 2020

The thesis committee for Rachel N.S. Foster certifies that this is the approved version of the following thesis:

A Behavioral Economic Approach to Quantifying Reinforcing Efficacy of Food

Chair: Derek D. Reed

Date Approved: 23 October 2020

Abstract

In May 2018 the Patient Protection and Affordable Care Act mandated all restaurants with more than 20 locations in the U.S. provide calorie information on menus for food items served. Although policymakers have argued that adding calorie content to menu items is a useful tactic to decrease demand for high calorie foods, empirical studies assessing this initiative report mixed results. The current study evaluates the impact of high and low calorie contents on consumer demand for preferred sandwiches and snack foods, and further analyzes differences in demand between individuals of varying body mass index. Results indicate that at a macro level, demand between high calorie and low calorie sandwiches and snacks does not significantly differ. However, although not statistically different, researchers observe differences between high and low calorie sandwiches and snacks between individuals of differing BMI groups, specifically in the intensity of demand for snacks.

Keywords: behavioral economics, consumer demand, delay discounting, obesity, food, calories

Table of Contents

List of Figures

List of Tables

Introduction

Behavioral economics offers a unique approach in studying commodity valuation that is distinct from the field of traditional economics (Hursh, 1980; 1984). Broadly, behavioral economics is the quantification of consumer choice and decision-making using psychological principles and economic theory. Operant behavioral economics, a sub-division of behavioral economics largely rooted in behavior analysis, focuses on employing scientific methods to evaluate the relative reinforcing efficacy of commodities under various constrains, including increasing price, time, and response requirement (Hursh & Roma, 2016). Results from operant behavioral economic studies can be used to guide public policy decisions (Hursh & Roma, 2013; Kaplan, Gelino, et al., 2018; Reed et al., 2016) and understand choice and decision-making behaviors (Hursh & Roma, 2016). Additionally, operant behavioral economics provides a dynamic view of reward-seeking behavior within the larger scope of psychology and broader field of behavioral economics (Thaler & Sunstein, 2008). While both the broad field of behavioral economics at-large and operant behavioral economics as a subfield are highly impactful, researchers will be referring to operant behavioral economics for the methods used and implications of this study.

Recent research has shed light on the potential this science has to inform public health initiatives (Matjasko et al., 2016; Strickland et al., 2020). While operant behavioral economic's policy-relevancy has been influential in understanding and mitigating substance use (Bickel et al., 2014; McPherson et al., 2018), there are relatively fewer applications of this science to obesity prevention and treatment. Obesity is a leading cause of preventable death and is one such public health concern (Kushner, 2002); a behavioral economic approach to studying the multi-dimensional nature of obesity, how this relates to impulsivity, and its connection to consumer demand for food is imperative.

Recent data from the Center for Disease Control (CDC) indicates the nation's rates of severe obesity increased from 4.2% in 2000 to 9.2% in 2018, and in 2018, 42.4% of the U.S. population was obese (Hales et al., 2020). Although the CDC has yet to issue a current report of the national obesity rate for 2020, it is doubtful it has decreased 11.9% in two years, meeting the U.S. Department of Health and Human Service Healthy People Program 2020 goal of 30.5% (www.healthypeople.gov/2020). Needless to say, the national prevalence of obesity has become unacceptable (Bray et al., 2018).

Complications related to obesity cost the U.S. healthcare system over $140 billion annually, an enormous economic burden on insurance companies and taxpayers (CDC, 2020) Moreover, individuals with obesity-related health complications pay on average $1,429 in additional healthcare costs compared to healthy weight individuals, adding extra costs to overall healthcare bills for every American (CDC, 2020). Comorbidities such as heart disease, cancer, type II diabetes, and gallbladder disease result in $62.7 billion in doctors' visits and $39.3 billion annually for employers covering lost workdays (Runge, 2007).

Obesity increases the risk for developing comorbidities related to weight disorders, including cardiovascular disease, type II diabetes, osteoarthritis, kidney failure, various forms of cancer, and many other ailments leading to premature death (Field et al., 2001; Foxx-Orenstein, 2010). At least 5% of deaths in the US are attributable to obesity-related complications (Masters et al., 2013). In addition, obesity is an incredibly socially stigmatizing condition (Puhl & Brownell, 2003; Puhl & Latner, 2007), as individuals with overweight and obesity are more likely to experience bias, discrimination, and prejudice towards them (Puhl & Brownell, 2001; Puhl & Heuer, 2009).

Many people with obesity share common behavioral traits contributing to serious health-related complications (Goldschmidt et al., 2019). However, there are conflicting findings as to whether or not specific psychological traits such as depression and anxiety are significantly more prevalent in this population compared to the rest of society (Avila et al., 2015; Faith et al., 2002; Scott et al., 2008; Talen & Mann, 2009). Regardless, research seeks to identify common underlying personality factors and behavioral characteristics contributing to obesity and overweight.

One possible factor contributing to obesity is the presence of an addictive-like "personality."[1] Addictive-like personality traits are often linked to difficulty in delaying gratification or making impulsive choices (Murphy et al., 2014). Behavioral economists often quantify one's delay of gratification by measuring the rate at which an individual selects a larger, more delayed reward, versus a smaller, more immediate reward – this behavioral pattern is termed "delay discounting." As defined, delay discounting is the reduction in the present value of a reward as the delay to receiving that reward increases (Kirby et al., 1999). Delay discounting is most commonly referred to as an individual's *k* value, a free parameter derived from Mazur's 1987 hyperbolic discounting equation quantifying how steeply the value of a reinforcer declines as a function of the delay to its obtainment (Mazur, 1987; Odum, 2011a), where higher *k* values equate to steeper delay discounting, and thus colloquially referenced as greater impulsivity.

---

[1] From a behavior-analytic perspective, "personality" may be regarded as a temporally extended pattern of behavior. My use of the term may be operationally defined using Harzem's (1984) definition: "a particular cluster of individual differences or, in explicit behavioral terms, as a cluster of functional relations between (1) a set of variables and (2) the already established behavior patterns of an individual" (Harzem, 1984).

Recent research suggests *k* can be conceptualized as a personality trait, as it is relatively stable and generally reflective of how an individual may value the rate at which they receive a reward in the presence of a delay (Odum, 2011b). Additionally, steep delay discounting has been related to a variety of health-related concerns such as illegal substance use, smoking, heavy alcohol use, risky sex, and pathological gambling. (Amlung et al., 2019; Bickel et al., 2012; Daugherty & Brase, 2010). Conversely, shallow delay discounting has been associated with making proactive health-based decisions, such as having a mammogram, getting a Pap smear, getting a flu shot, and maintaining regular dental visits, suggesting an inverse relation between health conscientiousness and *k* (Daugherty & Brase, 2010). Therefore, *k* can be understood as a factor quantifying impulsivity, and a valuable way to measure individual valuation associated with many health-based outcomes.

Previous studies suggest there may be an underlying connection between obesity and delay discounting. Appelhans and colleagues (2011) found the interaction between monetary delay discounting rates and scores on the Power of Food Scale (PFS) related to palatable food intake for obese and overweight individuals in a bogus taste test experiment, such that higher PFS scores significantly predicted greater food intake for women ($p = 0.04$; Appelhans et al., 2011). Monetary delay discounting has also been shown to have a moderate, positive relation with BMI ($r_s = 0.308$; $p < .01$), as measured within a community sample (Jarmolowicz et al., 2014). Weller and colleagues (2008) found women with obesity more strongly discounted monetary rewards compared to healthy weight women. Researchers provided two seven-item monetary delay discounting tasks to a sample of overweight and obese people: a high-monetary delayed value ($50,000) and a low-monetary delayed value ($1,000). Results indicate BMI significantly mediated area under the curve (AUC) values for both the $50,000 ($p = .022$) and

$1,000 ($p$ = .01) rewards, and obese women significantly discounted both rewards greater than healthy weight controls (Weller et al., 2008).

Behavioral economists have also developed delay discounting tasks related to food reward sensitivity. Odum and Rainaud (2003) measured differences in discounting rates for alcohol, food, and money in healthy adults. Participants reported their preferred food and alcoholic beverage; amounts of these commodities were represented by portions of this food or beverage that could be purchased for $100 (Odum & Rainaud, 2003). Discounting was measured at 25 different amounts at seven different delays. For example, $100 of pizza ($10/pizza) expressed as 10 pizzas available immediately or 10 pizzas available after 1 week. Results indicate food and alcohol were discounted more steeply than money, illustrating a possible domain effect for these commodities. Researchers have coined the term “domain effect” to describe steeper discounting observed for primary reinforcers such as food, compared to generalized conditioned reinforcers such as money (Rasmussen et al., 2010).

Odum and colleagues (2020) explored both trait-like and state-like qualities of delay discounting. Their systematic review revealed that delay discounting for non-monetary rewards was significantly greater than discounting for money, indicating a state-like quality. Additionally, individuals who steeply discounted monetary rewards were more likely to discount non-monetary rewards as well, illustrating trait-like characteristics (Odum et al., 2020).

Specific discounting tasks for food, such as the Food Choice Questionnaire (FCQ), have been developed to quantify reward sensitivity to food (Hendrickson et al., 2015). In a validation study of the FCQ, Hendrickson and colleagues (2015) found women with higher percent body fat (PBF) discounted food more steeply compared to women with lower PBF; additionally, smaller bites of food were more steeply discounted compared to larger bites of food, a phenomenon researchers term a “magnitude effect.” Delay discounting as measured by the FCQ has shown to be subject to intervention; in one study, adults with high BFP and steep discounting for food demonstrated a decrease in discounting after participating in a mindfulness-based exercise (Hendrickson & Rasmussen, 2017). Thus, interventions designed to reduce impulsivity may be a viable tool in mediating weight-related disorders.

Although there is research to suggest a strong relation between discounting and obesity, findings from more recently published studies indicate otherwise (McClelland et al., 2016; Veillard & Vincent, 2020). Current research highlights a lack of connection between temporal discounting and body composition as measured by BMI, as well as a weak relation between temporal discounting and obesity. A meta-analysis by Amlung and colleagues (2016) appears to convincingly advocate for a robust relation between monetary and food delay discounting and obesity, noting a highly statistically significant, medium effect size across 39 studies ($d = 0.43$, $p < 10^{-14}$). However, authors noted large heterogeneity in methods used to measure and calculate delay discounting between studies (i.e. multi-item discounting assessments, MCQ, hyperbolic discounting function, area under the curve), largely limiting the generality of these results (Amlung et al., 2016). With a variety of unstandardized approaches in measuring discounting between published studies, researchers should question if overarching claims about these findings remain comparable.

A systematic review by McClelland and colleagues (2016) argues it may be more efficacious to study the relation between temporal discounting and eating disorders, as individuals with eating disorders may exhibit increased (bulimia nervosa, binge eating disorder), or decreased (anorexia nervosa) discounting compared to healthy weight controls, whereas there is much more limited evidence to support the notion that individuals with obesity exhibit increased discounting compared to controls. A study by Veillard and Vincent (2020) shows no relation between monetary discounting or weight loss rewards and BMI, and researchers were unable to find sufficient evidence suggesting temporal discounting is predictive of BMI. Researchers recruited participants interested in weight loss and used the 27-item Monetary Choice Questionnaire (MCQ; Kirby et al., 1999), as well as the Weight Loss Questionnaire (WCQ; Lim & Bruce, 2015), a discounting assessment adapted from the MCQ asking participants their preferences for smaller, immediate weight loss versus larger, later-delivered weight loss. Results of this study indicate that, not only was there a weak relation between BMI and monetary discounting (r = 0.059), but there was stronger evidence *against* there being a correlation (BF_{10} = 0.125); a weak relation was also observed between BMI and WCQ scores (r = 0.016), and there was stronger evidence against this relation as well, (BF_{10} = 0.081). While there is a much larger body of literature to support the relation between BMI and discounting, more recently published work denotes the importance of critically re-evaluating findings from these past studies, exploring the possibility this relation may be much less robust than previously established.

Related to food reward sensitivity, people with obesity may demonstrate greater demand for high energy dense (HED) foods, as illustrated by trends in consumer market research (Karnani et al., 2016). Obesity largely disadvantages individuals with low socioeconomic status (SES), as inexpensive foods tend to be less nutritious, contain more refined sugar, and are more processed (i.e. potato chips, soda) compared to better quality, organic, or fresh foods (i.e. fruits and vegetables, free-range eggs; Drewnowski & Specter, 2004; Lee et al., 2014). Many low-income communities, communities of color, and individuals living in rural areas experience a "grocery gap," a term used to describe when convenience stores and gas stations become more prevalent to consumers and replace grocery stores in their region. As a result, inexpensive, low-nutrient foods become the most easily accessible options, healthy foods become less available, and individuals in these communities become predisposed to developing weight-related comorbidities such as obesity and diabetes (Treuhaft & Karpyn, n.d.). As a result of these "food deserts," food insecurity—a situation in which families do not have sufficient access to safe, nutritious, foods in order to live a healthy and active life—emerges.

A recent meta-analysis notes many food insecure people suffer from depression, anxiety, and post-traumatic stress disorder, all common disorders known to majorly impact choice-making behaviors (Privitera et al., 2019; Tribble et al., 2020), including delay discounting (Amlung et al., 2016). Food insecurity and living in a chronically stressful environment influences demand and discounting for primary reinforcers such as food, as one's "temporal window", their perception of access to future reinforcers, narrows (Snider et al., 2016). Bickel and colleagues argue "poverty engenders greater discounting;" for these individuals, uncertainty about where their next meal is coming from may cause larger, later rewards to be devalued and immediate access to essential goods such as food to take priority (Bickel et al., 2016).

Economic disparity has led some researchers to consider the influence of taxes on HED foods and subsidies on low energy dense (LED) foods as one solution to increase demand for healthier foods (Epstein et al., 2015; Epstein, Dearing, Roba, et al., 2010). Epstein and colleagues (2010) designed an experimental marketplace to measure demand for foods with different calories for nutrients (CFN) densities; low CFN (LCFN) scores equate to healthier foods (fewer calories needed to obtain key nutrients), and high CFN (HCFN) scores equate to less healthy foods (more calories needed to obtain key nutrients; Drewnowski, 2005). Researchers found when subsidies on LCFN foods were implemented, mothers purchased *greater* quantities of HCFN foods; when taxes were placed on HCFN foods, demand for these foods decreased while demand for LCFN foods increased. Thus, placing subsidies on healthier foods did not result in a substitution effect for LCFN foods, but rather allowed participants to allocate a greater percentage of their funds to purchasing HCFN foods. However, taxes on HCFN foods resulted in a substitution effect; participants purchased more LCFN foods and decreased demand for HCFN foods (Hursh & Roma, 2013).

In a subsequent study by Epstein and colleagues (2015), researchers aimed to quantify overall calories of HCFN or LCFN foods purchased based on the effect of taxes and subsidies. Participants were women engaging in an online experimental marketplace. They found subsidies on LCFN foods (fruits, vegetables, water) led to increased calories purchased on LCFN foods, while taxes on HCFN foods (sweetened beverages, candy salty snacks) led to decreased calories purchased on HCFN foods. Additionally, subsidies on LCFN foods led to an overall increase in nutrient quality of foods purchased. Findings from these two studies suggest the current research on price differences and its impact on demand for foods is complex, as subsidizing healthier foods may lead to an increase in purchasing foods with higher nutritional quality but may also free up funds to purchase additional HED foods. A review of the literature suggests taxes and subsidies on foods may not have a significant impact on weight status or BMI for the general population, however, these price differences may have the most effect on weight-based outcomes on individuals with low SES (Powell & Chaloupka, 2009).

Regardless of current policies designed to increase the affordability of LED foods, it is important to consider how behavioral economics can play a role in quantifying the reinforcing efficacy of foods. Given that food is a primary reinforcer and people with obesity typically find food more reinforcing than other commodities, using behavioral economics to measure how individuals of varying weight status value food could be beneficial in informing healthcare policy. Additionally, one can assume the average American may exhibit greater demand for highly palatable foods compared to less palatable foods, regardless of their BMI.

Behavioral economic demand for commodities such as foods—or other consumable goods—is rooted in operant behavior analysis. Demand is behaviorally conceptualized and quantified through concepts such as elasticity, intensity, P_{max}, and O_{max}. “Elasticity” is the rate at which the demand curve changes from inelastic to elastic. “Intensity” (Q_0) is demand for the commodity when it is free ($0). Breakpoint is the first price at which the participant reports zero demand for the commodity. P_{max} is the price associated with the point on the demand curve that corresponds to a slope of -1. O_{max} is the maximum price a participant would be willing to pay for the commodity, calculated by multiplying P_{max} by the quantity of items consumed or likelihood of consumption of that product.

One study measured demand in women for preferred HED and LED snacks using a HPT with 19 successive prices (n = 191). Additionally, the researchers used an operant demand progressive fixed ratio schedule (4, 8, 16…), termed the “reinforcing value task” for these same HED and LED foods, instructing participants to “work” by using computer mouse clicks to allocate responses for HED or LED foods. Breakpoint for LED and HED foods was the primary outcome for the reinforcing value task. Results of this study indicate that, while there were some significant relations between demand parameters in the HPT and breakpoint in the reinforcing value task (elasticity and intensity), some HPT indices did not correlate (breakpoint and P_{max}). Researchers note differences in these parameters, specifically HPT breakpoint and relative reinforcing value breakpoint, may be due to the nature of the different tasks, “real” versus hypothetical responding, the effort required to engage in either task (several mouse clicks in the reinforcing value task), and ability to choose reinforcers in the reinforcing value task versus purchase one in the HPT (Epstein, Paluch, et al., 2018). Findings suggest when evaluating the reinforcing efficacy of food, hypothetical versus actual responses may not adequality match.

Although the relation between hypothetical and “real life” (i.e., in non-experimental natural settings) purchase tasks has been well-established (Amlung et al., 2012; J. G. Murphy et al., 2009), it may be difficult for someone to accurately tact their likelihood of purchasing and consuming foods in a HPT. Specifically, HED foods are engineered using well-designed formulas to enhance palatability and make people crave more of these foods (Fazzino et al., 2019). Enhanced palatability makes it difficult for people to stop eating HED foods, causing them to consume much more than they originally planned at that time, possibly confounding say-do correspondence.

In another study by Epstein and colleagues, researchers used HPTs for preferred HED and LED snacks to evaluate latent factor structure, specifically, *persistence* and *amplitude*. These factors illustrate the components of reinforcement of a commodity across a demand curve. Persistence typically represents the constellation elasticity, breakpoint, P_{max}, and O_{max}, while amplitude is typically defined by intensity. Researchers were able to replicate this same factor structure across both the LED and HED tasks, indicating consistent reinforcing factors between HED and LED foods (Epstein, Stein, et al., 2018). Although repeated measures ANOVAs indicated statistically significant differences in all demand indices for both LED versus HED foods, indices only differed by about a $1 amount. Additionally, individuals were *more price sensitive to HED foods*, as intensity, breakpoint, P_{max}, and O_{max} were all greater for LED foods, while elasticity was greater for HED foods. The only factor significantly (yet weakly) related to BMI was intensity for HED foods ($p < 0.05$, $r = 0.18$) suggesting HPTs for food do not adequately measure the complex nature of obesity but may be useful in determining the value of food as a reinforcer when provided for free. Researchers acknowledge the reinforcing value of food is only one aspect of the multi-dimensional nature of obesity; lack of exercise as well as the nutritional quality of the foods consumed are two other major contributors not captured in these methods.

There have been various attempts to decrease consumption of HED foods in an effort to combat obesity rates in the USA. One such initiative has been to include calorie information on menus, a public policy implemented to prevent obesity. In May 2018, as mandated by section 4205 of the Patient Protection and Affordable Care Act, all restaurants in the USA with more than 20 locations were required to display caloric information next to food items served (Patient Protection and Affordable Care Act, 2010). While this initiative was intended to guide consumers to make healthier choices, most studies indicate little to no reduction in calories, attending to calories, or modifications of food orders following the addition of calories on menus, a massive “failure” of calorie labeling (Carroll, 2015).

Of the studies that have been successful in changing behavior after the addition of caloric information, many of them did not take place in naturalistic settings, such as a restaurant or university dining halls, places where this mandate is actually implemented (Kiszko et al., 2014). While an experimentally controlled setting is considered the gold standard approach in measuring the effect of an independent variable, researchers should question whether doing this kind of research in a lab is the best way to evaluate the impact of an extensive public policy.

One study that successfully reduced average caloric intake at a chain restaurant on a college campus displayed calories on the left-hand side of the menu, directly before the name of the food item (Dallas et al., 2019). Researchers reasoned that, since calories are typically displayed to the right of menu items, displaying calories to the left forces consumers to first read the calories and then the food item. This approach worked well, as participants ordered significantly fewer calories when they were displayed on the left (*M* = 654.53, *SD* = 390.45) compared to no calorie information (*M* = 914.34, *SD* = 560.94), or when calories were displayed to the right of the food item (*M* = 865.41, *SD* = 517.26). Interestingly, this effect maintained when used with a sample of Hebrew-speaking Israelis; in this condition, calories were placed to the right-hand side of the menu item, since Hebrew is read right-to-left. Results of this study indicate perhaps the position of this stimuli is crucial. This leads to the empirical question to whether stimulus salience is of most importance in order for consumers to attend to calories.

A study by Roberto and colleagues (2010) successfully decreased consumption of high-calorie foods when caloric information was provided to participants in an experimental study diner. Participants were randomized into three conditions: a menu with no calories (no calorie labels), a menu with caloric information provided per item (calorie labels), and a menu with caloric information provided per item and the following statement, "the recommended daily caloric intake for an average adult is 2000 calories" (calorie labels plus information). There was a significant reduction in calories ordered in both the calorie labels condition and the calorie labels plus information condition as determined by post-hoc LSD tests, (P = .03; d = 0.32; P = 0.3, d= 0.31, respectively). Interestingly, results of a dietary recall assessment measuring calories consumed throughout the day after the experimental meal indicate participants in the calorie labels condition consumed significantly more calories throughout the rest of the day (M= 294 ± 387) compared to the no calorie labels condition (M = 179 ± 310) or the calorie labels plus information condition (M = 177 ± 309) and were more likely to have had an evening snack (70%) compared to the no calorie labels condition (57%) or the calorie labels plus information condition (46%; Roberto et al., 2010). Thus, perhaps caloric information is best attended to when placed in context, such as knowing an appropriate daily range.

In an experimental study by Pulos & Leng (2010), researchers evaluated the overall nutritional content of entrées sold at six restaurants in Pierce County, Washington, for one month before and after providing nutrition information (fat, (g), sodium (mg), and carbohydrates (g)) on menus. Results indicate a moderate effect, with an average decrease in 15 fewer calories, 1.5 fewer g of fat, and 45 fewer mg of sodium in entrées purchased. Interestingly, although 71% of consumers noticed the nutritional information added to the menus, only 20.4% of consumers reported to modify their orders based on caloric content, and 16% based on fat content. This suggests adding nutrition information to menus, particularly calories, may be a useful tactic for a subset of consumers. However, there needs to be more research on specific traits of people who are more likely to attend to caloric content and adjust their behavior based on this stimulus (Pulos & Leng, 2011).

A review of the literature in 2014 on the effects of calorie labeling suggests caloric content does not greatly impact demand for most consumers. Researchers reviewed 31 studies published between 2007 to 2013 and concluded that, while caloric content influenced consumer demand in some studies, for most naturalistic settings, providing caloric content had minimal significant impact on consumer demand (Kiszko et al., 2014). While this review highlighted several studies noting a positive impact on caloric information provided to participants, the overwhelming majority of studies highlights a lack of significant effects of caloric content on consumption. Studies included in this review indicated the proportion of individuals who were impacted by caloric information included residents of wealthier neighborhoods, individuals between 18-24 years of age (Dumanovsky et al., 2011), and women (Krieger et al., 2013). In

2015, Long and colleagues conducted a systematic review and meta-analysis on the impact of caloric contents on menus in restaurants. Results indicated of the 19 studies included in this review, calorie labeling was associated with an average calorie reduction of only 18.13 kilocalories (kg); additionally, controlled studies in restaurants demonstrated only a 7.63 kg reduction after displaying calories on menus. These results support findings suggesting including calories on menus may not have a significant effect on consumer demand, and might only slightly reduce caloric consumption at best (Long et al., 2015).

There is a host of empirical data to support the notion that caloric information may not significantly impact demand for the average consumer. However, a push to provide nutrition information for entrées at restaurants still remains widely implemented, possibly because this is one of the most cost-effective approaches in an attempt to tackle such a seemingly uncontrollable obesity epidemic. Moreover, few studies have compared the impact of high versus low caloric contents on demand for the same product, a primary aim of the current study.

The purpose of the current study is to evaluate the impact of explicit and implicit caloric information on demand for highly preferred sandwiches and snacks, respectively. Additionally, researchers sought to measure the relation between behavioral economic demand indices, delay discounting, BMI, and reinforcing value of food as measured by several clinical scales. More broadly, this study uses behavioral economics to quantify demand for a variety of foods in an attempt to better understand the multifaceted nature of obesity and reinforcing efficacy of food.

Method

All study procedures were approved by the Human Subjects Committee of the institutional review board, KU IRB #STUDY2065. Researchers recruited participants via Amazon Mechanical Turk (MTurk), a crowdsourced web-based platform incentivizing Workers, to complete Human Intelligences Tasks (HITs), such as surveys. MTurk has gained prominence over the last several years as a cost-effective and efficient way to collect crowdsourced data for behavioral science and psychology research (Strickland & Stoops, 2019). Additionally, some research indicates that results generated from experiments using MTurk data are comparable to data collected from national samples (Coppock, 2019), and are as reliable as data collected from in-person samples (Buhrmester et al. 2011). In the current study, inclusion criteria for MTurk workers were that they are located in the United States and have a HIT approval rate greater than 95%. Only MTurk workers who met this inclusion criteria were allowed to access the survey. Participants were paid $0.60 upon survey completion. Survey questions consisted of a purchase task assessing likelihood of purchasing a preferred sandwich, a quantity purchase task for preferred snack foods, a demographics questionnaire, a monetary discounting questionnaire, a food discounting questionnaire, three clinical scales assessing relative reinforcing efficacy of food and sensitivity to food reward, and questions related to dieting history.

Likelihood of Preferred Sandwich Purchase

Participants completed a likelihood of purchase task for sandwiches (Roma et al., 2016). Pictures of sandwiches used in this study were identical to those used in Roma et al. (2016), however, the survey included a question prompting participants to "click on the burger/sandwich that most closely resembles" their ideal sandwich, assessing sandwich preference prior to the administration of the purchase task questions. The image of their most preferred sandwich was piped into each question related to the sandwich purchase task. Additionally, we added labels to each sandwich picture ("cheeseburger", "grilled chicken sandwich", "turkey club", or "veggie sub") to help participants accurately identify pictures. Although the vignette presented in this study was identical to the vignette used in Roma et al., 2016, we added one assumption: *your* *preferred dressings, veggies, and/or meat are included in this sandwich.* This assumption was added to control for any ingredients depicted in sandwich pictures that may not be preferred for an individual. Additionally, this assumption allowed participants to imagine their preferred ingredients for each sandwich were included. Lastly, participants were required to pass three verification questions at the end of the vignette, ensuring they had understood the details included. If a verification question was answered incorrectly, participants were prompted to re-answer the question until correct. The following 17 prices were presented in ascending order to assess demand: $0.00, $0.05, $0.10, $0.25, $0.50, $1, $2, $3, $5, $10, $20, $40, $60, $80, $100, $250, $500. Researchers opted to use the price sequence containing 17 prices from Roma et al., 2016, as is in-line with recommendations from Roma and colleagues. Additionally, 17 prices is an "ideal" density to best measure sensitivity to prices, as well as provide optimal demand curve fits (Kaplan, Foster, et al., 2018). All prices were presented on the same page. This purchase task will be referred to as the "Standard T1" task throughout the remainder of this document.

Quantity Purchase Task for Snack Foods

Participants completed a purchase task for preferred snack foods (Epstein, Dearing, & Roba, 2010; Epstein, Paluch, et al., 2018). Participants selected one preferred low energy density (LED) and one preferred high energy density (HED) snack food from a list. LED snacks included apples, bananas, mandarin oranges, low-fat strawberry yogurt, celery and dip, carrots and dip, applesauce, red seedless grapes, and pineapple chunks (Figures

Figure 1). HED snacks included nacho cheese Doritos, milk chocolate M&M's, Chips Ahoy! cookies, Reese's peanut butter cups, Hershey's chocolate, mini Oreos, Pringles chips, and Little Debbie zebra cakes (Figure 2).

After selecting these items, participants were shown pictures of 30g of each item and read the statement, "you chose [preferred snack food] as your most preferred snack food. To ensure you are paying attention, click on the image of [preferred snack food]." This question was included to preserve data quality; researchers did not retain data for subsequent analyses for participants who did not click on the correct picture corresponding to the snack food they selected. Additionally, this question was added to ensure participants could accurately tact the snack food item they selected. Although pictures had not been implemented in previous administrations of this purchase task in previous studies, researchers sought to improve upon earlier methods by providing a visual depiction of 30g of each food; pictures served as a visual aid for participants who may not be able to accurately estimate a 30g portion of these snack food items. Participants completed separate purchase tasks for both LED and HED snacks. Participants were asked how many portions of their preferred snack food they would consume if they were priced at each of the following 19 values: $0, $0.01, $0.05, $0.13, $0.25, $0.50, $1, $2, $3, $4, $5, $6, $11, $35, $70, $140, $280, $560, $1120 (Epstein, Paluch, et al., 2018). Prices were presented in an ascending order and all prices were presented on one page.

Demographics

The demographics survey included questions regarding sex, age (years), race, height (feet and inches), weight (lbs.), primary language, and annual income (Appendix).

Caloric Manipulations

After completing all above questionnaires, participants were randomized into three conditions: a control group, in which participants were administered the original likelihood of sandwich purchase task again (no caloric information provided, "referred to as Standard T2"), a "low calorie" group, in which they were provided a relatively low caloric density for their preferred sandwich, and a "high calorie" group, in which they were provided a relatively high caloric density for their preferred sandwich. In the control condition, participants were provided with the statement, "*below is the same purchase task you were previously administered. Please re-read the scenario and answer the following questions*." For the low calorie and high calorie groups, the vignette in this purchase task was identical to that used in the first likelihood of sandwich purchase task, however, the statement "when you look at the menu to order the sandwich/burger, you notice for the first time that it contains ____ calories," with the corresponding caloric information, was added. Participants were also provided with an image of a menu with their preferred sandwich and corresponding caloric information located on the right side of the sandwich. Participants had one verification question for this purchase task, prompting them to verify the caloric content of the sandwich. They did not have to complete the same verification questions previously administered for the Standard T1 likelihood of sandwich purchase.

Hypothetical caloric densities were determined in a preliminary study with a different MTurk sample (*N* = 200); participants (n = 50 per sandwich) were randomly shown a picture of one of the four sandwiches and asked to estimate how many calories were in that sandwich. Researchers used the ROUT method for identifying and excluding outliers in GraphPad Prism® (n = 5 outliers in total,

Table **3**). Mean caloric densities were divided by two for each sandwich to obtain "low caloric" densities and multiplied by two to obtain "high caloric" densities used in this study (

Table **4**).

Monetary Choice Questionnaire

A 27-item Monetary Choice Questionnaire was used to assess rates of delay discounting for money (Kirby et al., 1999). In this task, participants are asked in 27 distinct trials to choose between a smaller sum of money to be received immediately versus a larger sum money to be received after a delay. From these choices, a delay discounting rate (*k*) is quantified based on the pattern of responding across these 27 trials. Researchers used the 27-item Monetary Choice Questionnaire automated scorer (Kaplan et al., 2016a).

Food Choice Questionnaire

A 27-item Food Choice Questionnaire (FCQ) was used to assess rates of delay discounting for food (Hendrickson et al., 2015). Participants were told to imagine an image of a 5/8 inch white cube as a bite of their favorite food and were asked choose between eating fewer bites of this food immediately versus a eating a greater number of bites after a delay (e.g., “4 bites now” versus “8 bites 5 hours from now”). The FCQ was developed from and modeled after the MCQ and *k* values are calculated using the same methods (Hendrickson et al., 2015). Researchers adapted a 27-item MCQ automated scorer to calculate discounting for food by changing the reward values and *k* values at each indifference point per trial, such that an overall *k* value was obtained per participant.

Clinical Scales

The Yale Food Addiction Scale (YFAS), Three Factor Eating Questionnaire (TFEQ), and the Power of Food Scale (PFS) have been implemented in various studies and with diverse populations to measure sensitivity to food reward, cognitive and behavioral effects of food consumption, and psychological impact of being in an environment abundant with highly palatable foods (Cappelleri, Bushmakin, Gerber, Leidy, Sexton, Karlsson, et al., 2009; Gearhardt et al., 2009; Stunkard & Messick, 1985). Both the YFAS and PFS scales have adequate internal consistency and test-retest reliability (Davis et al., 2011; Eichen et al., 2013; Gearhardt et al., 2009; Lowe et al., 2009; Mitchell et al., 2016; Murphy et al., 2014). The TFEQ was developed in 1985; since then, its original factor structure had failed to be replicated (Ganley, 1988; Karlsson et al., 2000; Mazzeo et al., 2003). However, more recent modifications to the TFEQ, such as the 18-item TFEQ, the 21-item TFEQ, and the 18-item TFEQ 2.0, have been developed and validated, supporting the TFEQ as a psychometrically sound tool (Cappelleri, Bushmakin, Gerber, Leidy, Sexton, Lowe, et al., 2009).

Studies on of survey-based research indicate longer questionnaires tend to increase response burden and result in survey fatigue, or negligent responding to questions (Porter et al., 2004). Given this consideration to try to keep the survey at a minimal response effort, the more recently developed, brief versions of these scales were used in this study (see Table 1 for clinical scale outcomes; see Table 2 for clinical scale outcomes broken down by BMI).

mYale Food Addiction Scale 2.0

The Yale Food Addiction Scale (YFAS) has been widely used to assess symptoms of addiction towards foods with high sugar and high fat contents (Gearhardt et al., 2009). Symptoms of food addiction as measured by the YFAS are based on the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2000) substance dependence criteria (SUD), namely, measuring behavior towards food and food consumption as indicative of tolerance and withdrawal. As the initial scale was developed in 2009, it has since been adapted to measure symptoms of food dependence within a variety of populations, in both clinical and non-clinical samples (Brunault et al., 2014; Clark & Saules, 2013; Gearhardt et al., 2013; Li et al., 2017; Meule & Gearhardt, 2014). A more recently-developed version of the

YFAS, the mYFAS 2.0, was used in this study (Schulte & Gearhardt, 2017). The YFAS 2.0 was developed to reflect changes to substance use dependence criteria included in the DSM-V (Gearhardt et al., 2016). The mYFAS 2.0 was developed as a brief version of the YFAS 2.0, to be used when a more concise measure is suitable for a study, and with the intention of measuring specific addictive symptoms towards food and eating (Schulte & Gearhardt, 2017). The mYFAS 2.0 has strong internal reliability (Kuder-Richardson α = 0.86), similar to the internal reliability of the full YFAS 2.0 (α = 0.97). There are two scoring options for the mYFAS 2.0: a symptom count score and a diagnostic score. The symptom count score is the total number of items an individual endorses related to DSM-V SUD criteria (scale items 1-11). The diagnostic score is calculated when an individual endorses at least two SUD criteria and also endorse items relating to either impairment or distress, "clinical significance" (scale items 12-13). There are three diagnostic categories: mild (endorsing 2-3 SUD criteria and clinical significance), moderate (endorsing 4-5 SUD criteria and clinical significance), or severe (endorsing 6+ SUD criteria and clinical significance). Researchers in the present study used both scoring options to quantify food addiction.

The 18-item Three Factor Eating Questionnaire, Version 2.0

The Three Factor Eating Questionnaire (TFEQ) is a psychometric tool developed to measure properties of eating behavior related to dietary restraint, disinhibition, and hunger (Stunkard & Messick, 1985). The original scale is comprised of 51 questions, however, briefer versions with strong internal consistency have been developed (Karlsson et al., 2000; Tholin et al., 2005). Researchers used Version 2 of the 18-item Three Factor Eating Questionnaire (TFEQ R-18 V2), (Cappelleri, Bushmakin, Gerber, Leidy, Sexton, Lowe, et al., 2009). The TFEQ-R18 V2 was developed after re-testing the factor structure of the TFEQ 21-item scale. The TFEQ-

R18 V2 improved upon the TFEQ-R21 by removing three items relating to the cognitive restraint subscale, thereby increasing the reliability and robustness of the scale's factor structure. The TFEQ-R18 V2 examines the factors of uncontrolled eating (UE), cognitive restraint (CR), and emotional eating (EE). Additionally, the TFEQ-R18 V2 has exceptional internal consistency (Cronbach's α = 0.89 for UE domain, 0.78 for CR domain, and 0.94 for EE domain). The TFEQ-R18 V2 is the most recently developed modification to TFEQ and is optimal for large-*n* studies containing many survey measures. Specific questions relate to the score for each factor subscale. Endorsing values of 1-2 scores as 1; 3-4 as 2; 5-6 as 3, and 7-8 as 4. The CR subscale contains six items, UE contains nine items, and EE contains six items. Higher values indicate greater scores in any one factor.

Power of Food Scale

The Power of Food Scale is a 15-item clinical scale developed to measure "hedonic hunger," eating beyond one's daily caloric need, as is common for individuals living in food-abundant environments (Mitchell et al., 2016). The PFS is best used to assess desire to eat hyper-palatable foods and measure engagement in loss-of-control eating (Espel-Huynh et al., 2018). In non-clinical samples, high PFS scores have been associated with more frequent daily snacking and predictive of snacking in the presence of food-related cues (Schüz et al., 2015). Additionally, when used with clinical samples, higher scores have been shown to predict binge eating frequency in women with bulimia nervosa (Witt & Lowe, 2014). Although PFS scores have been associated with appetite for and energy intake of palatable foods, recent studies indicate PFS scores do not consistently predict changes in BMI reflecting overweight/obesity onset, nor dieting history (Lipsky et al., 2019). Since its development in 2009, its factor structure has been successfully replicated and has shown to have adequate test-retest reliability, and internal consistency (Cronbach's α = 0.81-0.91) rendering the original scale as an acceptable tool (Cappelleri, Bushmakin, Gerber, Leidy, Sexton, Karlsson, et al., 2009; Lowe et al., 2009). Thus, researchers used the original 15-item PFS to assess relations between PFS scores and demand for high energy density snack foods in this study (Cappelleri, Bushmakin, Gerber, Leidy, Sexton, Karlsson, et al., 2009). The PFS scores are comprised of scores in three subscales: Food Available, Food Present, and Food Tasted, as well as an aggregate score. The score for each subscale is the average of all items scored in each domain. Scores range from 1 (*I don't agree*) – 5 (*I strongly agree*). The aggregate score is the mean of the three subscales.

Data Quality and Exclusions

Researchers completed a power analysis using *G*Power* in order to determine the minimum sample size necessary to detect a significant effect of the independent variable (caloric manipulations) on the dependent variable (participants' demand per each condition), (Faul et al., 2007). Results of a repeated measures, between factors analysis of variance (ANOVA) indicate in order for the study to be adequately powered at the 0.80 significance level, the necessary sample size is 116. Additionally, in order for the study to be powered at the 0.95 significance level, the necessary sample size is 176. After data exclusions, the sample size of this study (N = 172) met the minimum sample size necessary to detect a significant at 0.80 power.

Researchers used Qualtrics® online survey platform to host all survey questions administered to MTurk participants. Data were then exported to Microsoft® Office Excel for organization and data synthesis. Participants self-reported their height (ft and in) and weight (lbs) during the survey; researchers converted feet to inches and calculated BMI using the following equation:

$$BMI = (Weight[lbs]/Height[in]^2) * 703 \tag{1}$$

According to the World Health Organization, a BMI below 18.5 is considered "underweight;" 18.5 – 24.9 is "normal;" 25.0 – 29.0 is "overweight;" and ≥ 30.0 is "obese" (WHO, n.d.). While individuals with eating disorders may have a BMI that falls within any of these ranges, the severity factor for adults with anorexia nervosa is typically determined by BMI (Beumont et al., 1988; Garber et al., 2019). Individuals with "mild" anorexia have a BMI ≤ 17 (American Psychiatric Association, 2013). In order to ensure demand for food at the group level was not impacted by individuals who may fall into this diagnostic category, researchers excluded participants with a BMI ≤ 17 (n = 40). Post-exclusions for BMI, only two participants' BMI was in the "underweight" range; their data were grouped with those of participants within the "normal" BMI range for subsequent analyses (n = 76).

To determine deviations from systematic purchase task data, researchers used procedures recommended by Stein, Koffarnus, Snider, Quisenberry, and Bickel (2015). Stein and colleagues outline best practices for identifying nonsystematic purchase task data, flagging data violating the criteria of bounce, trend, and reversals from zero (Stein et al., 2015). The trend criterion identifies data with nonnegligible decreases in consumption as prices increase. The bounce criterion flags price-to-price increases in consumption exceeding 25% of consumption at free. Reversals from zero are flagged when a participant reports zero consumption at one price followed by reported consumption of that good at a higher price.

Stein and colleagues note excluding nonsystematic data has utility in most cases, however, using the nonsystematic data criteria instead as a descriptor of data quality is an appropriate alternative. For instance, authors note exclusion of data violating the trend criteria is typically recommended in most cases, as trend is based on the law of demand stating increases in price of a good results in reduction of consumption of that good (Stigler, 1954). However, the criteria of bounce and reversals from zero are more susceptible to influences from extraneous variables such as environmental factors or demographic differences. Thus, researchers indicate an alternative to data exclusion based on these two factors is to selectively evaluate data on a case-by-case basis that violate these two assumptions. Given there are no empirical standards for using the Stein criteria with likelihood of purchase tasks, as well as the novelty of research published using such tasks (Naudé et al., 2019; Reed et al., 2016a; Roma et al., 2016), researchers used the Stein criteria as a descriptor of data quality rather than a standard for data exclusion.

Researchers used the criteria of trend by excluding data from any participant who reported demand at a later price that was greater than the intensity of their demand (demand for the product when it is free at $0.00; n = 102 total cases across all conditions). These exclusions were applied for both the likelihood of sandwich purchase task and the snack quantity demand task. Additionally, we excluded data from participants who reported intensity greater than or less than 3.29 standard deviations above the mean intensity for the snack demand purchase task (n = 15 total cases). Researchers also flagged participants who reported intensity at three standard deviations above or below the mean intensity for the sandwich purchase task; these data were further examined and excluded (n= 11 total cases). Whereas some researchers would opt to Winsorize data above or below 3.29 standard deviations from the mean (Tabachnick and Fidell, 2013; Tukey & McLaughlin, 1963), researchers in this study aimed to fit a curve to data specifically reported by participants, such that all data used for analyses are representative of participants' actual responses.

Data Analyses

Researchers used a freely-available template on GraphPad Prism® (GraphPad Software, www.graphpad.com) to fit the exponential model of demand (Hursh & Silberberg, 2008) to the purchase task data at the group level:

$$\log Q = logQ_0 + k(e^{-\alpha Q_0 C} - 1) \tag{2}$$

Researchers compared R^2 from the exponential model to the exponentiated model of demand (Koffarnus et al., 2015) to determine the equation that best fit both the sandwich and snack demand data:

$$Q = Q_0 * 10^{k(e^{-\alpha} Q_0 P - 1)} \tag{3}$$

In both equations, Q is the number of portions of a snack food item purchased or likelihood of purchasing a sandwich at each price (i.e. P) and Q_0 is demand for the products when they are free (0 converted to 0.01 to fit a curve in log-log space using exponential demand model). The parameter k is the range of consumption across all prices. For the snack demand purchase task, k was calculated by log (mean max/mean min) + 0.5; for likelihood of sandwich purchase, k was calculated by log (mean max/mean min). The same k values were used across both equations. The parameter α is the rate of change in elasticity across the demand curve. P_{max} is the point of unit elasticity, depicted where the slope of the demand curve equals -1. P_{max} was calculated using a slope-based approach, by which researchers used curve-derived α, k, and Q_0 to generate 50,000 price points (112,000 for snack demand purchase task) between each price point and found the exact price where the slope of the curve reached -1.

Researchers fit curves to demand at the group level for all purchase tasks used in this study – Standard T1, Standard T2, Low Calorie, High Calorie, LED snack, and HED snack.

Additionally, researchers fit curves to demand at the group-level based on BMI to better depict potential differences in demand between individuals in different BMI categories.

Post-exclusions, researchers calculated individual empirical (“observed”) demand indices for both the sandwich and snack demand purchase tasks (Foster et al., 2020a, 2020b). Empirical demand indices are distinct from curve-derived demand indices because they are a model-free measure of each demand parameter. Breakpoint 1 is defined as the first price associated with zero reported consumption of the commodity. Empirical P_{max} is the price associated with maximum reported consumption of the commodity. Empirical O_{max} is understood as two different outcomes for both the likelihood of purchase task and the snack demand purchase task. For likelihood of purchase tasks, this is viewed as expected revenue from the view of producer; for quantity purchase tasks, this is recognized as the overall maximum amount an individual would pay for the good (Naudé et al., 2019; Roma et al., 2016).

In order to evaluate significant differences between demand curves, researchers implemented Akaike’s Information Criterion (AIC) in GraphPad Prism. AIC is a model selection tool, estimating the quality of both models (in this case, whether or not demand curves are the same between conditions), and provides the best model fit, indicating the probability that alpha is the same between curves (Posada & Buckley, 2004). Researchers report AIC for all likelihood of purchase demand curves and snack purchase demand curves; researchers report the number of participants in each condition and across each BMI category that meet an AIC criterion for the probability that alpha is the same between both curves (conditions).

Researchers used SPSS Statistics Version 26 and GraphPad Prism for all analyses between observed demand indices, demographics variables, and clinical scale outcomes. Additionally, researchers implemented both the Shapiro-Wilk and the D’Agostino-Pearson tests to determine data normality and tested for significant differences at the individual level between observed demand indices for Standard T1 and Standard T2, Standard T1 and Low Calorie, Standard T1 and High Calorie, Low Calorie and High Calorie, HED and LED snacks, and delay discounting *k* values. Results of data normality tests led researchers to use the nonparametric Wilcoxon matched-pairs signed rank test to determine significant differences between these conditions. Researchers also examined Spearman's rank order correlation coefficients to determine the strength of correlations between demand indices in each task, BMI, and clinical scale outcomes.

Results

Exclusions

Initially, researchers recruited 506 participants; of those, 168 were excluded for providing incomplete datasets (n = 338 remaining). Next, researchers excluded participants who did not pass the attention verification question and/or the snack demand verification questions for both HED or LED snacks (n = 82 excluded in total; n = 21 who did not pass validation question alone; n = 61 who did not pass snack demand verification questions alone). Of the 61 participants excluded based on snack demand verification, 19 of them failed verification questions on *both* the HED and LED snack purchase tasks. Thirty-seven participants failed verification on the HED task alone, and five participants failed verification on the LED task alone. A total of 256 participants remained after exclusions based on data verification checks.

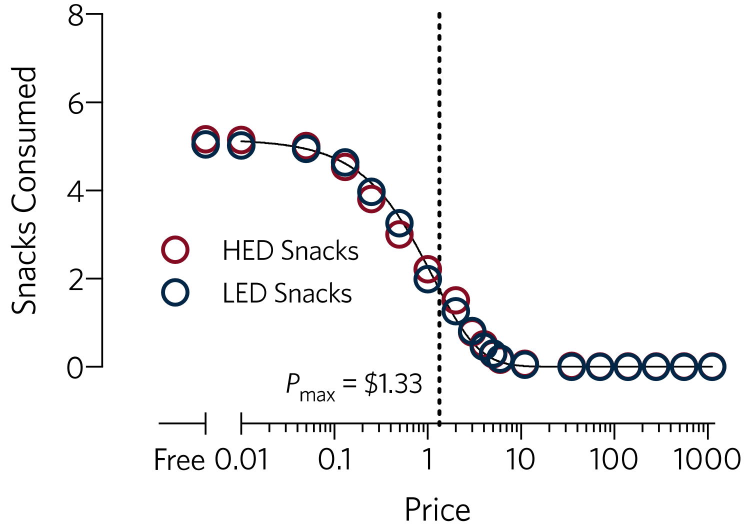
Forty participants provided height and weight data yielding a BMI ≤ 17; (n = 27 of these did not pass the verification question(s)); a total of 13 were excluded for providing abhorrent height and weight data alone, yielding a BMI ≤ 17. After exclusions for BMI, a total of 243 participants remained.

Of the 243 participants remaining, 71 were excluded based on the researcher's demand data exclusion criterion, leaving a total of 172. As researchers used the Stein et al. (2015) nonsystematic data criterion solely as a data descriptor, Table 5 shows a breakdown of participants' data flagged as nonsystematic using the nonsystematic purchase task identification tool. Of these 71 participants excluded, demand reported at a price that was greater demand at Q_0 occurred 116 times (Standard T1 n = 39; LED snack demand n = 14; HED snack demand n = 17; Standard T2 n = 9; Low Calorie condition n = 14; High Calorie condition n = 23). Additionally, of the 71 participants excluded, intensity reported 3.29 standard deviations above or below the mean intensity occurred 26 times (Standard T1 n = 4; LED snack demand n = 7; HED snack demand n = 8; Standard T2 n = 3; Low Calorie condition n = 1; High Calorie condition n = 3).

Post-exclusions, bananas were the most highly preferred LED snack food selected (n = 29), and Reese's Peanut Butter Cups were the most highly preferred HED snack food selected (n = 49). Applesauce was the least highly preferred LED snack food selected (n = 7), and mini Oreos were the least highly preferred HED snack food selected (n = 5). Although bananas and Reese's Peanut Butter Cups were chosen as the most highly preferred snack foods, Q_0 demand was highest for pineapple chunks and Pringle's chips (see Figure 3 and Figure 4 illustrating demand for individual LED and HED snacks, respectively).

The final sample of 172 consisted of 69 (40.1%) participants randomized into the Standard T2 condition, 41 (23.8%) participants randomized into the Low Calorie condition, and 62 (36.0%) randomized into the High Calorie condition (see Figure 5 for a flowchart outlining data exclusions and Table 6 outlining participant demographics).

Clinical Scale Scores



Snacks Consumed
8
6
4
2
0
HED Snacks
LED Snacks
P_{max} = $1.33
Free
0.01
0.1
1
10
100
1000
Price

Figure 12

Normal BMI – Standard T1 vs Standard T2

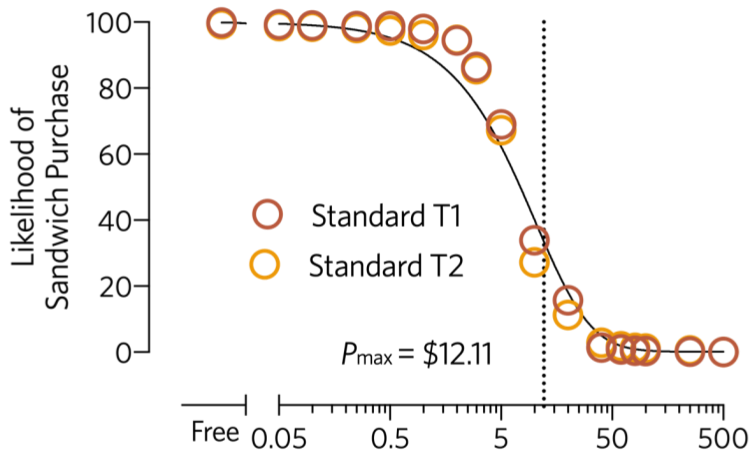


Figure 13

Normal BMI – Standard T1 vs High Calories

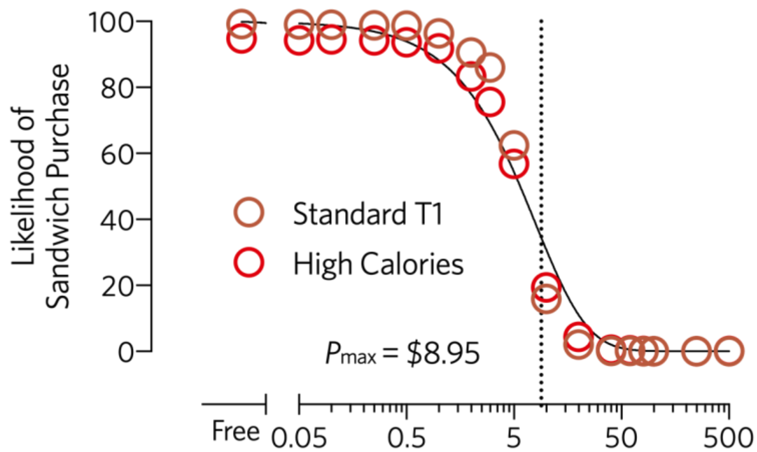


Figure 14

Normal BMI – Standard T1 vs Low Calories

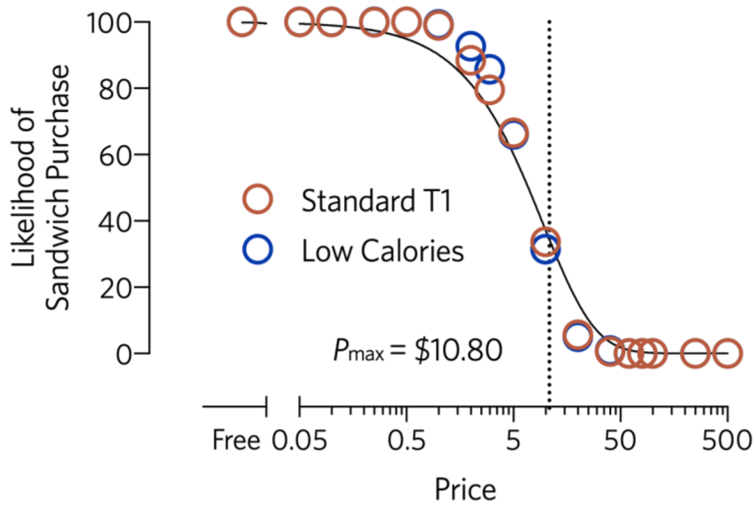


Figure 15

Overweight BMI – Standard T1 vs Standard T2

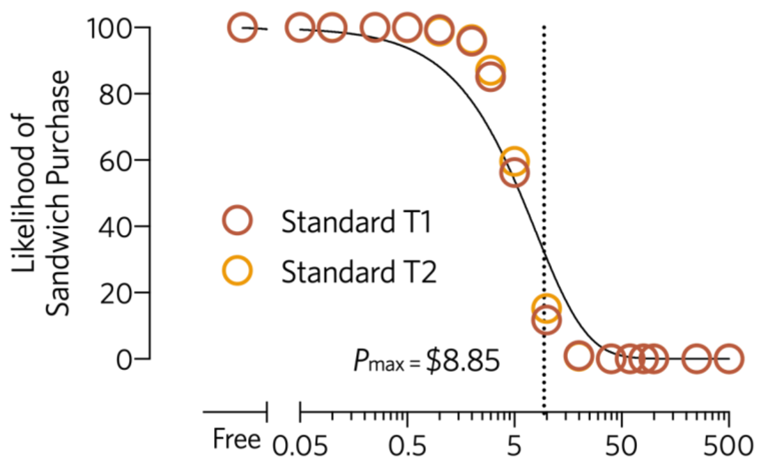


Figure 16

Overweight BMI – Standard T1 vs High Calories

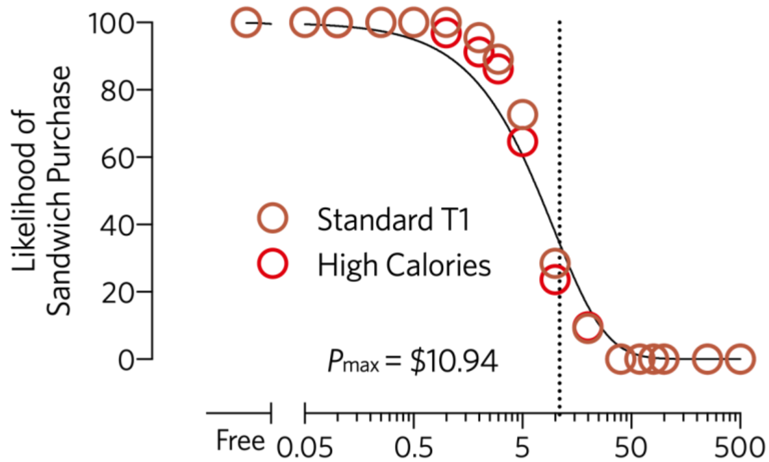


Figure 17

Overweight BMI – Standard T1 vs Low Calories

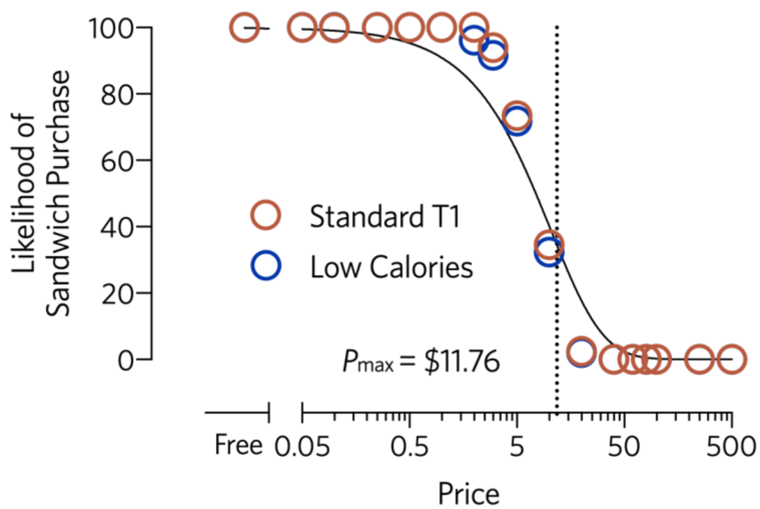


Figure 18

Obese BMI – Standard T1 vs Standard T2

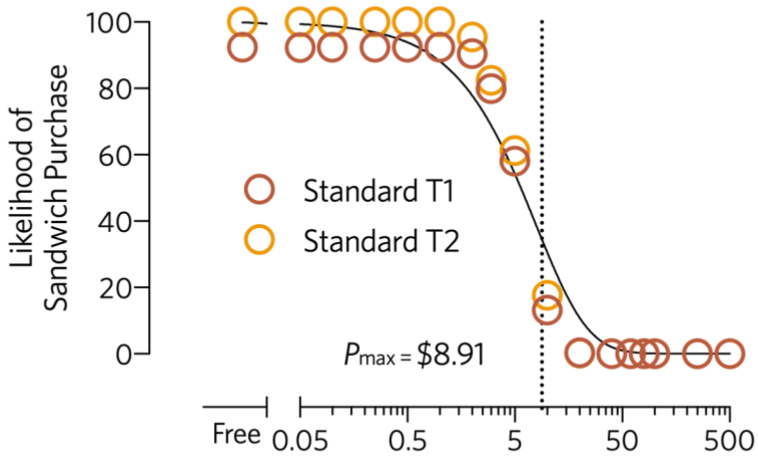


Figure 19

Obese BMI – Standard T1 vs High Calories

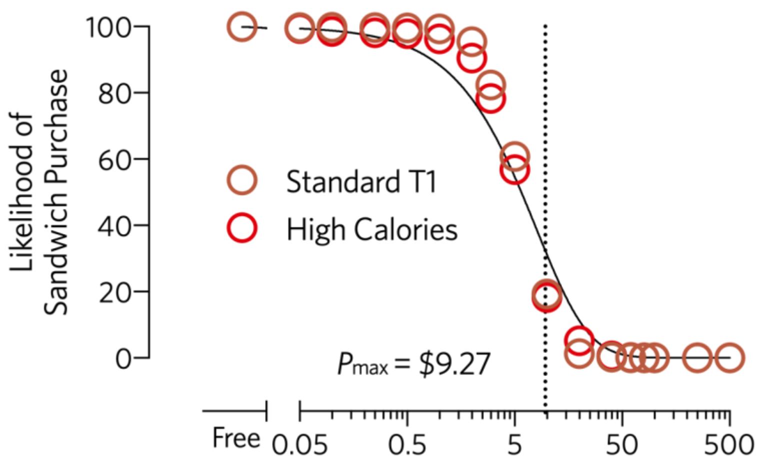


Figure 20

Obese BMI – Standard T1 vs Low Calories

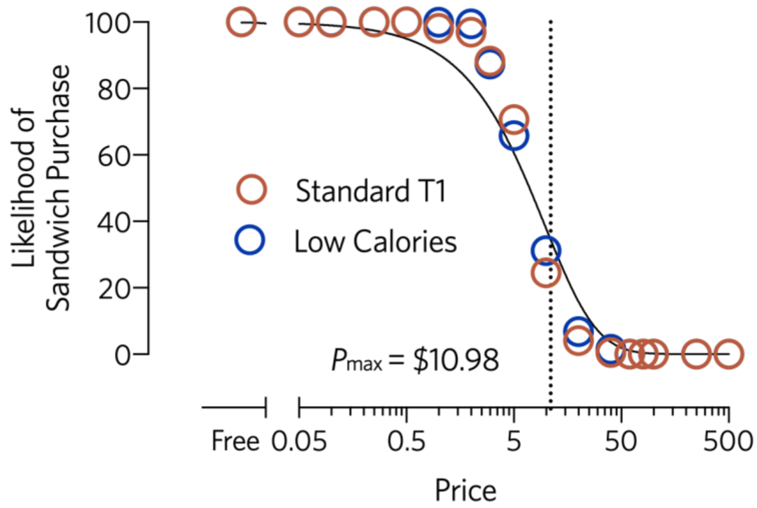


Figure 21

Normal BMI – LED vs HED Snacks

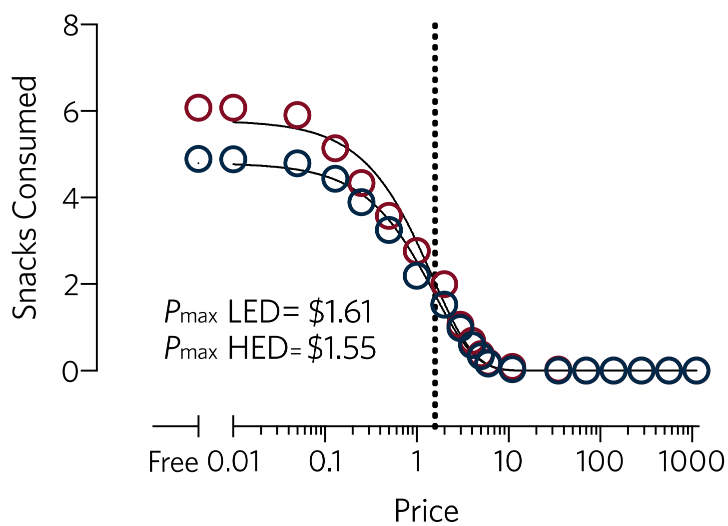


Figure 22

Overweight BMI – LED vs HED Snacks

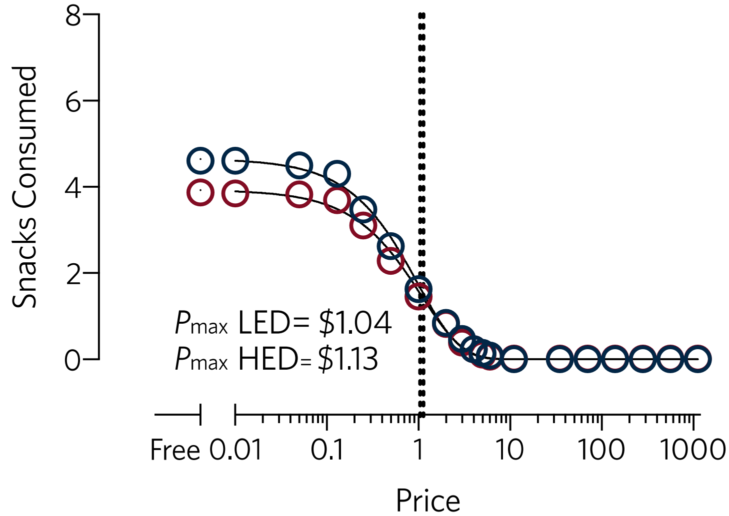
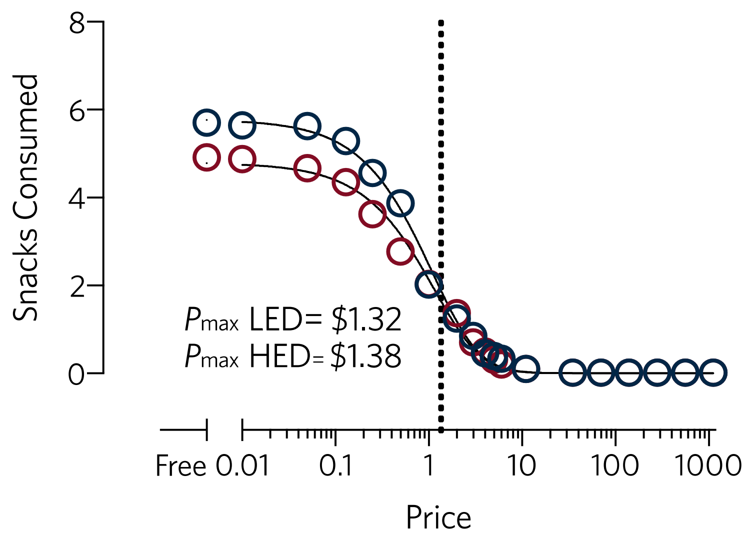


Figure 23

Obese BMI – LED vs HED Snacks



Tables

Table 1 outlines scores on clinical scales for all participants. The majority of participants (n = 119) scored 0 for the Symptom Count on the mYFAS 2.0 (69.2%); (*M* = 1.10 ± 2.07). Out of a total possible score of 11, this indicates that most participants did not endorse traits relating to addiction of foods with high sugar and high fat contents. Of the eight participants that met the criterion for a clinical diagnostic score, four of them scored “severe,” two scored “moderate,” and two scored “mild”. Thus, 4.7% of the overall sample fell within the range of the scale to merit a diagnostic criterion of food addiction. Internal consistency for the YFAS 2.0 scale was high (Cronbach’s α = .93), comparable to the Kuder-Richardson α = .86 in the original development of the mYFAS 2.0 (Schulte & Gearhardt, 2017).

Results of the TFEQ-R18 V2 indicate most participants scored between the 20-40 overall score range (n = 90; 52.33%) and 41-60 range (n = 76, 44.19%); (*M* = 39.51 ± 11.10) out of a total possible score of 72. Mean subscale scores are as follows: UE (*M* = 20.05 ± 5.94); CR (*M* = 6.72 ± 2.49); EE (*M* = 12.74 ± 5.10). Total possible scores for the subscales are as follows: UE = 36; CR = 12; EE = 24. Broadly, results of the TFEQ-R18 V2 highlight that most participants’ scores fell within the midrange of scores relating to specific eating behaviors represented by the subscales. Internal consistency for the TFEQ-R18 V2 was high (Cronbach’s α = 0.92). Internal consistency for the subscales was also acceptable (UE α = 0.87; CR α = 0.83; EE α = 0.93). Cronbach’s α values for TFEQ-R18 V2 responses in this study are comparable with Cronbach’s α values from the original validation of this questionnaire (UE domain = 0.89, CR domain = 0.78, and EE domain = 0.94; Cappelleri et al., 2009).

The POF Scale adequately measured the psychological effects of living in food-abundant environments and appetite for highly palatable foods in this sample. Most participants scored between the 21- 30 range (n = 48; 27.91%) or in the 31- 40 range (22.09%) for the aggregate score. Mean subscale scores are as follows: Food Available (*M* = 12.45 ± 6.61); Food Present (*M* = 10.34 ± 4.57); Food Tasted (*M* = 13.41 ± 5.06). Average aggregate score was 36.20 ± 14.86. Total possible scores are as follows: Food Available = 30; Food Present = 20; Food Tasted = 25; aggregate score = 75. Internal consistency for the POF scale overall was excellent in this sample (Cronbach's α = 0.95). Internal consistency for the subscales was also acceptable (Food Available = 0.93; Food Present = 0.92; Food Tasted = 0.85). Participants' scores on these scales indicate that that this sample is within the midrange of measurement for hedonic hunger and loss-of-control eating. These values are better than Cronbach's α values calculated from the validation of the 15-item POF scale (Food Available = 0.87; Food Present = 0.87; Food Tasted = 0.81; aggregate score = 0.90; Cappelleri et al., 2009).

Delay Discounting

Figure 6 illustrates differences in discounting between the two tasks, with steeper discounting demonstrated in the FCQ compared to the MCQ. The natural log of the average (*M* ± *SD*) score on the MCQ was -4.42 ± 1.95; the natural log of the average score on the FCQ was -1.60 ± 1.22. Thus, researchers observed a domain effect between money and food, such that participants chose the smaller, sooner reward for food more often than the smaller, sooner, reward for money. Consistency scores as determined by data consistencies < 75% were calculated using the automated scorers for both discounting tasks (Kaplan et al., 2016b). Consistency scores define the degree to which a participant's responses to the delay discounting questions are consistent with each other. They are calculated by summing how often a participant choose the smaller, sooner reward (0) versus the larger, later reward (1) both before and after their individual *k* value and is then divided by the number of questions in the discounting questionnaire (n = 27 for both the FCQ and MCQ). A larger value indicates a more consistent pattern of responding. For this sample, the average consistency score for the both the FCQ and MCQ was high (*M* = 90.98% ± 10.35; *M* = 95.56% ± 5.53, respectively). Nineteen participants had a consistency score lower than 75% in the FCQ; two had a consistency score lower than 75% in the MCQ. These two participants also had consistency scores < 75% in the FCQ.

Demand Curves

Since R^2 values between the exponential and exponentiated demand equations were comparable, researchers opted to use the exponential demand equation for all subsequent demand curve analyses in this experiment. See **Error! Reference source not found.** for a comparison of R^2 values and AIC criterion between the exponential and exponentiated demand equations. Additionally, for this specific experiment the exponentiated demand equation was also the most conservative approach in data interpretation, as AIC model selection more often reported a greater probability alpha was the same between demand conditions, allowing researchers to be less likely to report a Type I error (**Error! Reference source not found.**). Lastly, the exponentiated demand equation better depicts differences in demand for commodities towards the lower end of the price ranges, a range in which differences in demand were most apparent for this study. The AIC model selection tool provides the best fit model comparing two curves. This section outlines both individual P_{max} values for each condition, as well as shared P_{max} values between curves. Figure 7 illustrates demand for both sandwiches and snacks categorized by BMI.

Group Level

Figure 8 depicts demand for sandwiches in the Standard T2 condition. Results of this condition indicate that the likelihood of purchase task has solid test-retest reliability; AIC model selection indicated a 76.97% probability that alpha is the same for both data sets. P_{max} for Standard T1 was \$10.74, and P_{max} for Standard T2 was also \$10.74. Shared P_{max} for this condition was \$10.74, indicating that P_{max} was equal for individuals in both conditions for which caloric information was not provided. AIC model selection at the individual level for all participants in the Standard T2 group indicated an average of 73.96% probability that alpha is the same for both data sets.

Figure 9 depicts demand for sandwiches in the High Calorie condition. AIC model selection indicated a 71.82% probability that alpha is the same for both data sets. P_{max} for sandwiches in the Standard T1 condition is \$10.23; P_{max} for sandwiches in the High Calorie condition is \$9.30. Shared P_{max} for this condition is \$9.77, thus, demand for high calorie sandwiches is \$0.97 lower compared to demand for sandwiches in the Standard T1 condition. AIC model selection at the individual level for all participants in the High Calorie group indicated an average of 64.34% probability that alpha is the same for both data sets.

Figure 10 depicts demand for sandwiches in the Low Calorie condition. AIC model selection indicated a 76.92% probability that alpha is the same for both data sets. P_{max} for sandwiches in the Standard T1 condition is \$11.82; l P_{max} for sandwiches in the Low Calorie condition is \$11.94. Shared P_{max} is \$11.88, indicating that demand for low calorie sandwiches is \$1.14 higher compared to demand for sandwiches in the Standard T1 condition. AIC model selection at the individual level for all participants in the Low Calorie group indicated an average of 78.83% probability that alpha is the same for both data sets.

Figure 11 depicts demand for both high calorie snacks and low calorie snacks. At the group level, AIC model selection indicated a 75.41% probability that alpha is the same for both data sets. Demand was nearly identical for both high calorie and low calorie snacks, as P_{max} for low calorie snacks was \$1.32 and P_{max} for high calorie snacks was \$1.36, only a \$0.04 difference. Shared P_{max} is \$1.33. AIC model selection at the individual level for all participants indicated an average of 18.17% probability that alpha is the same for both data sets.

Analyzed by BMI

Sandwich Demand

Normal Weight

Researchers further analyzed demand for both the sandwiches and snack foods as separated by BMI categories. Figures 12 – 14 depict demand for sandwiches for normal weight participants (n = 76). For normal weight participants randomized into the Standard T2 control condition (n = 32), AIC model selection indicated a 68.14% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$12.78; Standard T2 P_{max} was \$11.64. Shared P_{max} for normal weight individuals randomized into this condition was \$12.11 (Figure 12).

For normal weight participants randomized into the High Calorie condition (n = 30), AIC model selection indicated a 64.5% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$9.20; P_{max} for the High Calorie condition was \$8.70. P_{max} shared between both curves was \$8.95, \$3.16 less than demand for sandwiches in the Standard T2 condition (Figure 13).

For normal weight participants randomized into the Low Calorie condition (n = 16), AIC model selection indicated a 85.84% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$10.96; P_{max} for the Low Calorie condition was \$10.88. Shared P_{max} between curves was \$10.80, \$1.31 lower than for sandwiches in the Standard T2 condition (Figure 14).

Overweight

Figures 15 – 17 illustrate demand for sandwiches for participants with overweight (n = 46) For participants with overweight randomized into the Standard T2 condition (n = 24), AIC model selection indicated an 84.17% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$8.61; Standard T2 P_{max} was \$9.11. P_{max} shared between both curves was \$8.85 (Figure 15).

For participants with overweight randomized into the High Calorie condition (n = 11), AIC model selection indicated an 80.56% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$11.58; P_{max} for the High Calorie condition was \$10.32. Shared P_{max} between both data sets was \$10.95. Interestingly, demand actually *increased* for participants with overweight in the High Calorie condition compared to those randomized into the Standard T2 condition by \$2.09. However, as compared within-subject, demand decreased \$0.63 from Standard T1 to the High Calorie condition (Figure 16).

For participants with overweight randomized into the Low Calorie condition (n = 11), AIC model selection indicated a 87.31% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$12.03; P_{max} for the Low Calorie condition was \$11.49. Shared P_{max} was \$11.76. Demand for sandwiches in the Low Calorie condition was also greater compared to demand for sandwiches in the Standard T1 condition by \$2.91; demand for sandwiches was greatest in this condition for participants with overweight (Figure 17).

Obese

Figures 18 – 20 depict demand for sandwiches for participants with obesity (n = 48). For participants with obesity randomized into the Standard T2 condition (n = 13), AIC model selection indicated an 82.26% probability that alpha was the same for both data sets. Standard T1

P_{max} was \$8.79; Standard T2 P_{max} was \$9.13. P_{max} shared between both conditions was \$8.91 (Figure 18).

For participants randomized into the High Calorie condition (n = 21), AIC model selection indicated an 82.65% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$9.27; P_{max} for participants randomized into the High Calorie condition was \$8.81. P_{max} shared between conditions was \$9.27. Thus, demand for sandwiches increased in the High Calorie condition compared to the Standard T2 condition by \$0.36, however, within-subjects, demand decreased \$0.46 from the Standard T1 condition (Figure 19).

For participants randomized into the Low Calorie condition (n = 14), AIC model selection criterion indicated a 86.91% probability that alpha was the same for both data sets. Standard T1 P_{max} was \$10.74; P_{max} for participants in the Low Calorie condition was \$11.25. P_{max} shared between conditions was \$10.98. Participants with obesity had the greatest demand for sandwiches in the Low Calorie condition; participants' demand in this condition was greater by \$2.07 compared to Standard T1 and greater by \$1.71 compared to the High Calorie condition. This suggests that although demand decreases in the High Calorie condition, demand is greatest when low calorie information is provided, over and above no calorie information or high calorie information (Figure 20).

Snack Demand

When comparing snack demand across BMI categories, AIC criterion indicated differences in demand between HED and LED snacks across all BMI groups. However, the greatest differences in demand were observed at Q_0 between BMI status. Individuals in the normal BMI range reported on average greater consumption of high calorie snacks at free (6 portions), compared to overweight participants (4 portions) or participants with obesity (5 portions).

For normal weight participants, AIC criterion indicated a 99.97% probability that alpha was different between LED and HED snacks. P_{max} for LED snacks was \$1.61; P_{max} for HED snacks was \$1.55 (Figure 21).

For participants with overweight, AIC criterion indicated a 98.65% probability that alpha was different between LED and HED snacks. P_{max} for LED snacks was \$1.04; P_{max} for HED snacks was \$1.13 (Figure 22).

For participants with obesity, AIC criterion indicated a 93.39% probability that alpha was different between LED and HED snacks. P_{max} for LED snacks was \$1.32; P_{max} for HED snacks was \$1.38 (Figure 23).

Statistical Analyses

Tests for Significant Differences

Researchers first used both the D'Agostino-Pearson and Shapiro-Wilk normality tests to determine how closely observed demand indices and delay discounting values resemble a Gaussian distribution and proceeded to test for significant differences based on these results. Observed demand indices at the individual level were compared from demand indices at Standard T1 to each condition. All observed demand indices (breakpoint 1, intensity, empirical O_{max}, empirical P_{max}) per condition, as well as MCQ and FCQ values, were non-normally distributed as evaluated by both D'Agostino-Pearson and Shapiro-Wilk normality tests.

As some researchers would opt to transform non-normal data to better resemble a normal distribution, this is typically necessary when subsequent analyses used (i.e. regressions) assume a normal distribution. Additionally, some statisticians argue that transforming data is unnecessary when parametric tests can adequately describe outcomes, or that it over-complicates data interpretation (Norris & Aroian, 2004). Regardless, researchers performed square root transformations on all observed demand indices and MCQ and FCQ values in order to determine if the transformed data more closely approached normality; in almost all cases, D'Agostino-Pearson and Shapiro-Wilk normality tests still declared non-normality of the data. Empirical O_{max} for the Low Calorie condition was the only condition in which square root transformation normalized the distribution; however, a t-test indicated a non-significant effect, $t(40) = 0.78$, $p = 0.44$. When employing analyses to check for significant differences between groups, normal distributions are not a necessary assumption and nonparametric tests are adequate; thus, researchers implemented the Wilcoxon matched-pairs signed rank test on untransformed data to compare differences in indices and discounting values between dependent conditions. Additionally, the normality assumption is relaxed when sample size is large (N > 30); in this case, all conditions have a large sample size.

Standard T1 versus Standard T2

A Shapiro-Wilk test indicated a significant departure from normality for breakpoint 1 values in the Standard T1 condition, $W(69) = 2.03$, $p < .001$, as well as the Standard T2 condition, $W(69) = .221$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 144.6$, $p < .001$; $K^2 = 128.7$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between breakpoint 1 values between Standard T1 and Standard T2, $Z = -1.060$, $p = .289$.

A Shapiro-Wilk test indicated data were non-normally distributed for intensity values for the Standard T1 condition, $W(69) = .107$, $p < .001$, as well as the Standard T2 condition, $W(69) = 1.63$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 =$

150.5, $p < .001$; $K^2 = 138.2$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between intensity values between Standard T1 and Standard T2, $Z = -3.65$, $p = .715$.

A Shapiro-Wilk test indicated data were non-normally distributed for empirical O_{max} values for the Standard T1 condition, $W(69) = 2.67$, $p < .001$, as well as the Standard T2 condition, $W(69) = .333$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 58.59$, $p < .001$; $K^2 = 64.42$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical O_{max} values between Standard T1 and Standard T2, $Z = -.463$, $p = .644$.

A Shapiro-Wilk test indicated data were non-normally distributed for empirical P_{max} values for the Standard T1 condition, $W(69) = .216$, $p < .001$, as well as the Standard T2 condition, $W(69) = .227$, $p < 0.001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 144.9$, $p < .001$; $K^2 = 134.2$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical P_{max} values between Standard T1 and Standard T2, $Z = -.816$, $p = .415$.

Standard T1 versus Low Calorie Condition

A Shapiro-Wilk test indicated a significant departure from normality for breakpoint 1 values in the Standard T1 condition, $W(41) = 0.733$, $p < .001$, as well as the Low Calorie condition, $W(41) = 0.742$, $p < 0.001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 29.99$, $p < .001$; $K^2 = 26.91$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between breakpoint 1 values between Standard T1 and the Low Calorie condition, $Z = -.494$, $p = .622$.

All intensity values reported for both the Standard T1 condition and the Low Calorie condition were identical (100% likelihood of purchase). As such, researchers were unable to compute normality tests for these data. A Wilcoxon signed-ranks test indicated that there were no differences between intensity values between Standard T1 and the Low Calorie condition, $Z = 0.00$, $p = 1.00$.

A Shapiro-Wilk test indicate data were non-normally distributed for empirical O_{max} values for the Standard T1 condition, $W(41) = 0.961$, $p < 0.001$, as well as the Low Calorie condition, $W(41) = 0.948$, $p < .001$. Results of a D’Agostino-Pearson normality test also indicate nonnormality, $K^2 = 3.686$, $p < .001$; $K^2 = 3.605$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical O_{max} values between Standard T1 and the Low Calorie condition, $Z = -.754$, $p = .451$.

A Shapiro-Wilk test indicate data were non-normally distributed for empirical P_{max} values for the Standard T1 condition, $W(41) = 0.889$, $p < .001$, as well as the Low Calorie condition, $W(41) = 0.887$, $p < .001$. Results of a D’Agostino-Pearson normality test also indicate nonnormality, $K^2 = 17.91$, $p < .001$; $K^2 = 15.92$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical P_{max} values between Standard T1 and the Low Calorie condition, $Z = -.158$, $p = .874$.

Standard T1 versus High Calorie Condition

A Shapiro-Wilk test indicated a significant departure from normality for breakpoint 1 values in the Standard T1 condition, $W(62) = 0.605$, $p < .001$, as well as the High Calorie condition, $W(62) = 0.508$, $p < .001$. Results of a D’Agostino-Pearson normality test also indicate nonnormality, $K^2 = 66.39$, $p < .001$; $K^2 = 78.62$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between breakpoint 1 values between Standard T1 and the High Calorie condition, $Z = -1.739$, $p = .082$.

A Shapiro-Wilk test indicate data were non-normally distributed for intensity values for the Standard T1 condition, $W(62) = 0.185$, $p < .001$, as well as the High Calorie condition, $W(62) = 0.332$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 105.6$, $p < .001$; $K^2 = 70.47$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference in intensity values between Standard T1 and the High Calorie condition, $Z = -1.782$, $p = .075$.

A Shapiro-Wilk test indicate data were non-normally distributed for empirical O_{max} values for the Standard T1 condition, $W(62) = 0.821$, $p < .001$, as well as the High Calorie condition, $W(62) = 0.838$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 3.686$, $p < .001$; $K^2 = 3.605$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical O_{max} values between Standard T1 and the High Calorie condition, $Z = -.208$, $p = .836$.

A Shapiro-Wilk test indicate data were non-normally distributed for empirical P_{max} values for the Standard T1 condition, $W(62) = 0.174$, $p < .001$, as well as the High Calorie condition, $W(62) = 0.151$, $p < .001$. Results of a D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 134.8$, $p < .001$; $K^2 = 135.8$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was insufficient evidence to suggest a difference between empirical P_{max} values between Standard T1 and the High Calorie condition, $Z = -.274$, $p = .784$.

Delay Discounting Values

A Shapiro-Wilk test indicate data were non-normally distributed for MCQ values, $W(172) = .956$, $p < .001$, as well as FCQ values, $W(172) = .890$, $p < .001$. Results of a

D'Agostino-Pearson normality test also indicate nonnormality, $K^2 = 10.63$, $p = .005$; $K^2 = 50.69$, $p < .001$. A Wilcoxon signed-ranks test indicated that there was sufficient evidence to suggest a difference between MCQ and FCQ values for all participants, $Z = -.11.170$, $p < .001$.

Correlations

Researchers implemented a Spearman's rank order correlation coefficients analysis between select demand indices, clinical scale scores, discounting values, and BMI (**Error! Reference source not found.**). Results of this analysis highlight that the only variables significantly related to BMI were the clinical scale outcomes. A weak, positive correlation was present between BMI and the POF Scale, $r_s = .262$, $p = .001$, as was present between BMI and the mYFAS 2.0, $r_s = .227$ $p = .003$, and BMI and the TFEQ-R18 V2 total score, $r_s = .296$ $p = .000$. Additionally, clinical scales did not significantly correlate with any other variable, but were significantly related to each other. Researchers observed a moderately positive relation between POF and mYFAS 2.0 scores, $r_s = .574$, $p = .000$, and between POF and TFEQ-R18 V2 scores, $r_s = .621$, $p = .000$. Additionally, a moderate, positive relation between mYFAS 2.0 and TFEQ-R18 V2 scores emerged, $r_s = .536$, $p = .000$.

Researchers observed significant, weak positive relations between discounting scores on the MCQ and FCQ, $r_s = .290$, $p = .000$. In addition, there was a weak, positive relation between FCQ scores and LED O_{max} values, $r_s = .154$, $p = .044$. Researchers found weak, positive relations between MCQ values and empirical O_{max} for HED and LED snacks, $r_s = .228$, $p = .003$; $r_s = .213$, $p = .005$, respectively.

Empirical O_{max} values significantly correlated with one another across the snack demand purchase tasks (both HED and LED) and the sandwich demand purchase tasks (T1 and T2). In addition, researchers included intensity for HED and LED snacks in the correlation analysis, as these were the only variables significantly different as determined by AIC when comparing demand for snacks across BMI groups. However, intensity for HED and LED snacks only significantly correlated with empirical O_{max} values for HED and LED snacks, as well as with each other.

Discussion

This study demonstrates the utility of using behavioral economic metrics to quantify relative reinforcing efficacy of food and demand for preferred foods, as well as measure changes in demand due to stimuli (e.g. calories). This study also represents the multi-faceted etiology of obesity, highlighting that adding calorie contents to menus as a public health initiative may not significantly shift consumer demand to thwart obesity. Results of this study are consistent with the larger body of literature indicating that, at a macro level, calorie content on menus does not change demand for preferred foods for the majority of consumers. Researchers observed statistically nonsignificant differences in demand between LED and HED snacks, and in demand between Standard T1 and the other three conditions. More research is needed to determine what scientists and policymakers can do to decrease demand for high energy dense foods within the obese and overweight population.

Correlations between variables highlighted the fact that BMI and clinical scales were significantly related, but that there were no relations between BMI and empirical demand indices, or between BMI and delay discounting values. Additionally, clinical scales correlated with one another, however, they did not correlate with empirical demand indices. Although there were statistically significant interactions between FCQ and MCQ scores and demand indices, these correlations were weak. Nonetheless, future research should further examine the relations between delay discounting for money and food, and how this relates to overall food expenditure.

When interpreting significance of results, it is important to note differences between clinical and statistical significance and why this is imperative for this particular study. Clinical significance refers to the influence of a study's findings on clinical practice, regardless of if statistically significant effects are present (Ranganathan et al., 2015). For this study, researchers interpret clinical significance to mean the significance this research has on both public policy and obesity prevention at both the group and individual level. Although AIC criterion indicated more than 70% probability there were no differences between curves is most conditions for the exponentiated demand equation (**Error! Reference source not found.**) and nonparametric tests indicated nonsignificant effects for all conditions, differences in demand by even $1.00 can impact national revenue on sandwich purchases. National reports indicate that nearly one half (47%) of adults ages 20+ in the U.S. eat at least one sandwich per day, and that the majority of sandwiches and/or sandwich ingredients are purchased from a store (58%). Additionally, individuals who reported consuming sandwiches daily had a greater average daily caloric intake compared to those who did not, and sandwiches contributed about 12% to total energy intake. Sandwiches used in the current study are representative of those most purchased by Americans: cold cuts are the most commonly consumed (27%), followed by burgers (17%), and poultry (12%; Sebastian et al. 2015). Given the prevalence of sandwich consumption for Americans, it is clinically significant to evaluate the impact of a reduction of even a few dollars on national revenue, as well as determine how the nutritional content of sandwiches from popular restaurants and chains can be modified to create healthier options.

Future research should examine the implications of adding caloric information to the left-hand side of the menu – appearing before the menu item itself – to determine if this can cause a significant decrease in demand, as shown in Dallas et al. (2019). Additionally, future research should examine if other kinds of nutritional information (e.g. g of saturated fat, g of added sugar) can shift demand more than calorie contents on menus does. One small-scale study (N = 16) demonstrated that when nutritional labels were added to menu items in a college food court, most students reported attending to calories, however, for some students, interest in the ingredients listed modulated whether or not changes in ordering were made. For example, one student noted that they were willing to eat an entrée containing 2,000 calories as long as the calories were “worthwhile” (i.e. coming from healthier food sources such as protein and vegetables) compared to a quesadilla (Kolodinsky et al., 2008). In a large-scale study (N = 1,817) by Christoph and colleagues (2018), researchers found that, of the young adults who reported attending to nutrition facts (31.4%), most attended to sugars (74.1%), total calories (72.9%), serving size (67.9%), and the ingredient list (65.8%) compared to other nutrition facts listed (Christoph et al., 2018). Thus, individuals who read nutrition labels may mostly attend to calories, however, there are other nutrition components that may change an individual’s preference for a certain food item.

Another approach is to research the effects of stimulus saliency on demand for preferred foods. For example, studies suggest that using a color-coded traffic light guide to categorize food into different groups based on their nutritional content increases healthy food choices compared to traditional numerical guidelines (Rramani et al., 2020). The positive effects of the “stoplight guide” in informing food choices has been influential for both typically developing individuals (Cecchini & Warin, 2016) and individuals with intellectual and developmental disabilities (Saunders et al., 2011). Thus, perhaps the saliency of the stimulus chosen to communicate nutritional content of foods is important in helping people attend to nutrition quality.

Although the current study has clear implications for public health and policy research, the researchers would like to acknowledge a few limitations. First, researchers did not assess nutritional literacy within this sample. Nutritional literacy refers to how well an individual can understand and interpret nutrition facts, as well as apply this information to their own metabolic and caloric needs (Silk et al., 2008). Without information on nutritional literacy, it is largely unknown as to if participants were aware that caloric information provided per sandwich was relatively "high" or "low" for their own daily intake. Additionally, participants may not have been aware of the specific differences in nutritional content between the HED and LED snack foods, or of nutritional differences between each snack in both categories. Conversely, perhaps some participants answered the behavioral economic demand questions in a specific way *because* of their understanding of the nutritional variations between these products.

Second, due to the nature of data collection on MTurk, researchers were only able to collect BMI information from self-reported height and weight for each participant. While self-report answers to hypothetical demand assays have been validated to translate well to demand in naturalistic settings (Amlung et al., 2012; J. G. Murphy et al., 2009; Wilson et al., 2016), accuracy of self-report for height and weight is not as high. Research suggests adults tend to underreport weight and BMI, and overreport height (Gorber et al., 2007; Taylor et al., 2006), and that women tend to underreport weight more often than men (Flegal et al., 2019). Another limitation to this study is that more participants reported being in the normal weight range compared to overweight and obese; a greater number of participants in the latter two ranges would have allowed for more equal comparisons across BMI ranges.

Regardless of whether or not participants accurately reported these anthropometrics, BMI is an inadequate indicator of overall health and wellbeing because it fails measure differences in mass attributed to fat versus bone and muscle, and cannot identify weight attributed to fat mass in different parts of the body (Burkhauser & Cawley, 2008; Nuttall, 2015). Additionally, BMI fails to account for differences in body composition due to race, ethnicity, and sex (Ahima & Lazar, 2013). While BMI is not an optimal measure of overall health, it is a quick, easy, and cost-effective way to determine weight status that largely informs public health and epidemiological research. Additionally, obese BMI has been significantly predictive of elevated mortality risk and other comorbidities, while underweight BMI is one useful metric to measure facets of anorexia nervosa (Gorwood et al., 2019; Gutin, 2018). Future research may examine percent body fat (PBF) as a predictor of differences in demand for preferred foods. High PBF has been predictive of more impulsive choices for both food and money in delay discounting tasks (Hendrickson & Rasmussen, 2017). Additionally, percent body fat is a better measure of overall fat in the body compared to the estimation provided by BMI, and more accurately measures the effects of energy intake on adiposity (Tucker et al., 1997).

Another limitation of this study was that participants were not representative of a clinical weight management sample, meaning that researchers did not know enough about the individual characteristics of each participant that may have been contributing to their demand and discounting. Such questions include whether or not these participants were actively trying to lose or gain weight during this study, if they were currently adhering to special diets or had food allergies or restrictions (e.g. lactose intolerance, celiac disease), if they had undergone medical procedures that would limit their food intake abilities such as bariatric surgery or were on diet pills, if they had inhibited sensory abilities (taste, smell), and if cultural considerations played a part in their food choices. Additionally, while researchers were able to quantify facets of eating disordered behaviors with the clinical scales used, there is no way to determine if these participants had clinical eating disorders of any kind such as food addiction, anorexia nervosa, or bulimia nervosa. An individual can have a normal BMI and still exhibit thought patterns and behaviors indicative of a clinical diagnosis and necessitating an intervention (Dingemans & van Furth, 2012; Goldschmidt et al., 2011), a factor this study did not capture. While this is not an exhaustive list of considerations, there are many aspects of health and behavior that can impact this kind of decision-making.

Lastly, researchers did not assess state-based hunger for participants and did not ask when they last ate. Current hunger at the time of completing the purchase tasks and choice of preferred snacks, as well as answering questions to the clinical scale indices and FCQ, could potentially have an impact on responding. Contrarily, a participant who was satiated or full while they were completing the task may have reported lower levels of demand for foods than they would have otherwise, and may have reported choosing the larger, later food reward in the FCQ task more often.

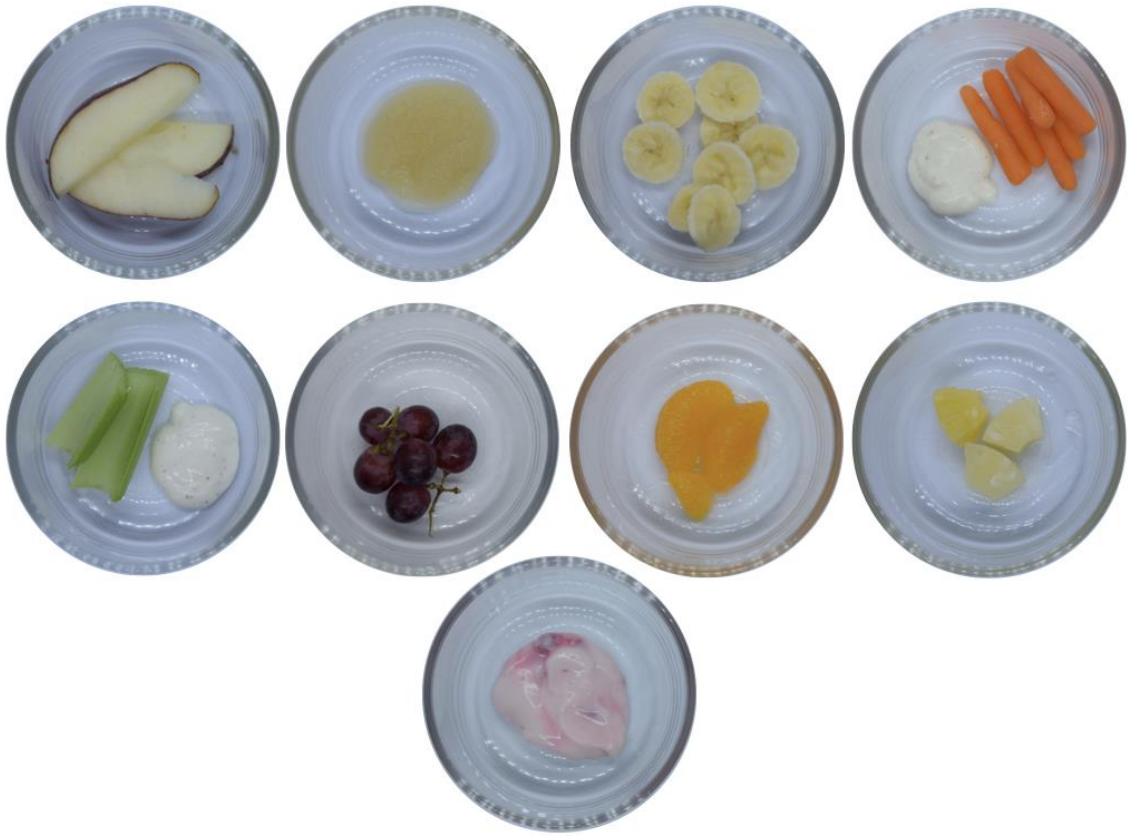
Combining behavioral economic methods with science from other disciplines can help researchers understand how to best combat, treat, and prevent obesity. Future research regarding the impact of nutrition labeling on consumer demand may consider the use of eye-tracking devices and neuroimaging techniques to determine what kind of nutritional information individuals are most attending to when reading food labels, and which parts of the brain are more activated for individuals of varying weight status or percent body fat. Some neuroscience research suggests that caloric information can alter brain response by reducing reward center activation and heightening control system activation, and that this effect is stronger in individuals with more experience attending to caloric information, such as dieters (Courtney et al., 2018). Additionally, some research suggests that hyper-responsivity of reward valuation regions in the brain is linked to overeating (Ng et al., 2011). A study by Grabenhorst and colleagues (2013) highlights the effect of food labels on amygdala activation in the brain, noting that the amygdala plays a key role in food choice and behavioral shifts towards healthier choices (Grabenhorst et al., 2013). Thus, there are many facets of behavior—both overt and covert—that influence food choice and demand, and combining methodology from multiple scientific disciplines may be advantageous in aiding both obesity prevention and understanding the complexities of obesity.

The purpose of the current study is to use behavioral economics to quantify relative reinforcing efficacy of food and better understand the impact of calorie labeling on demand, both at a macro-level and as separated by individuals of varying BMI status. Results indicate that calorie content does not significantly impact demand for the majority of consumers, however, some interesting differences are observed between individuals of normal, overweight, and obese BMI, both when explicit and implicit calorie contents are implemented. Future research should address the impact of other nutritional information on demand, the effects of stimulus saliency, determine if differences in demand are observed for individuals of varying percent body fat, and evaluate how neurological research can be integrated with behavioral science to aid obesity prevention. Findings from this work can inform public policy and healthcare initiatives at both a national and global level.

Figures

Figure 1

Images Depicting 30g of Each LED Snack Food Item



Note. From left: apples, applesauce, bananas, carrots with dip, celery with dip, red seedless grapes, mandarin oranges, pineapple chunks, and low-fat strawberry yogurt

Figure 2

Images Depicting 30g of Each HED Snack Food Item

Note. From left: Chips Ahoy! cookies, Hershey's chocolate, Pringles chips, Reese's peanut butter cups, Little Debbie zebra cakes, mini Oreos, milk chocolate M&M's, and nacho cheese aDoritos

Figure 3

Individual Level LED Snack Demand

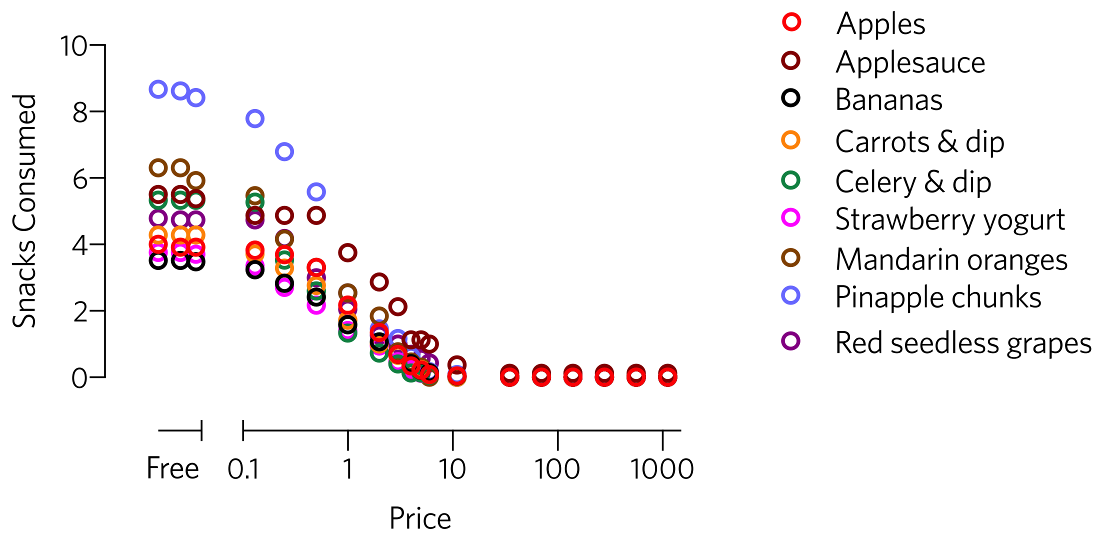


Figure 4

Individual Level HED Snack Demand

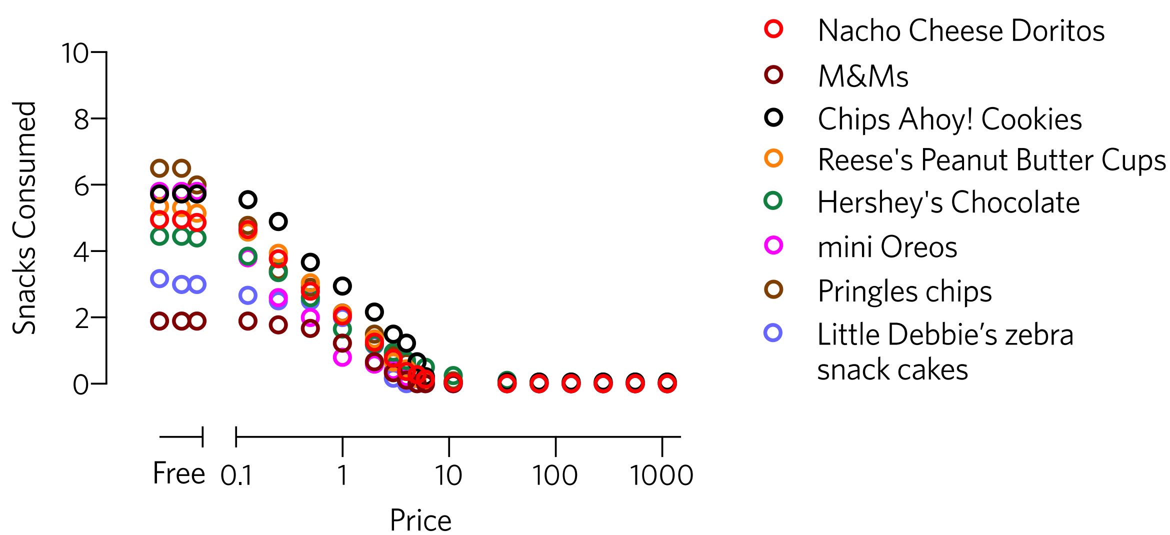


Figure 5

Data Exclusion Flowchart

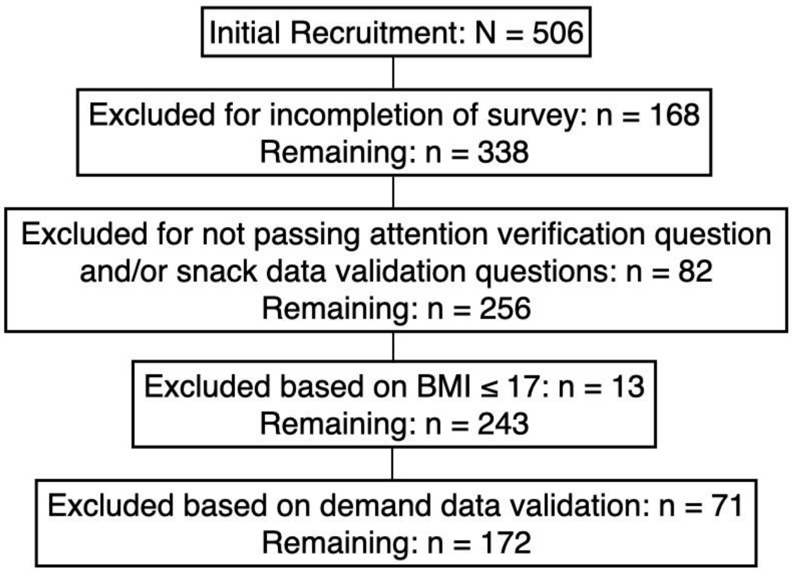


Figure 6

Natural log-transformed Discounting Values

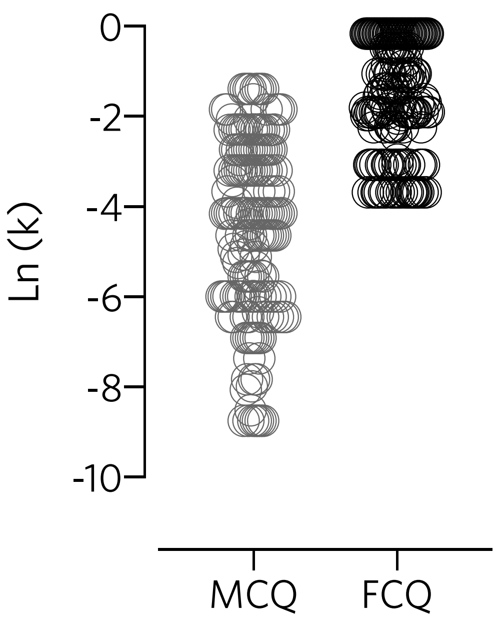
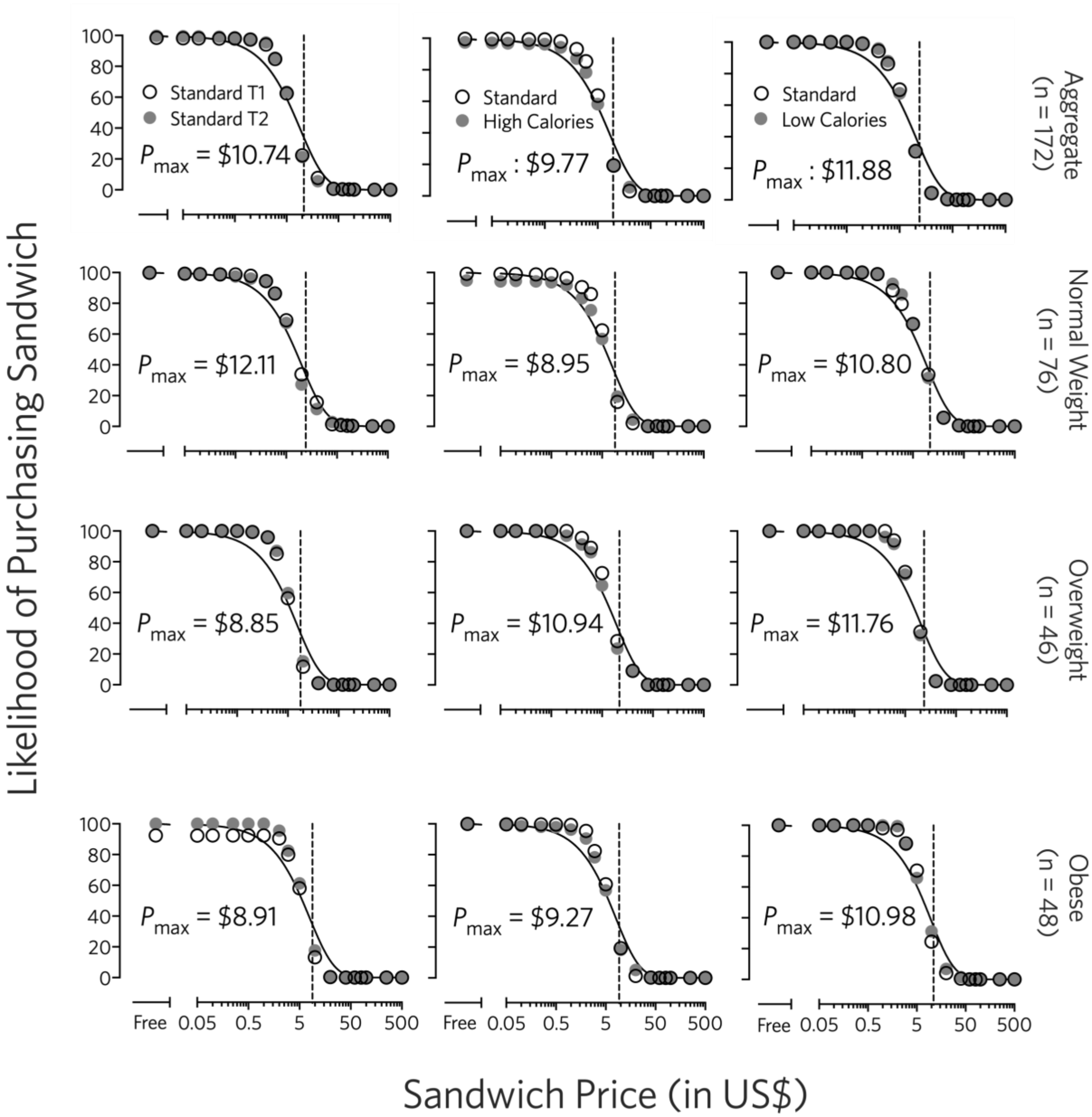


Figure 7

Demand Curves



Note. Mean likelihood of sandwich purchase (y-axis) depicted as a function of price (x-axis; note log scaling) for aggregate (top row), normal weight (second row), overweight (third row), and obese (fourth row) participants. Curve-derived Pmax is included on each plot. Each pair of dataset points in each plot is best described by 1 non-linear (exponentiated) demand curve.

Figure 8

Standard T1 vs Standard T2

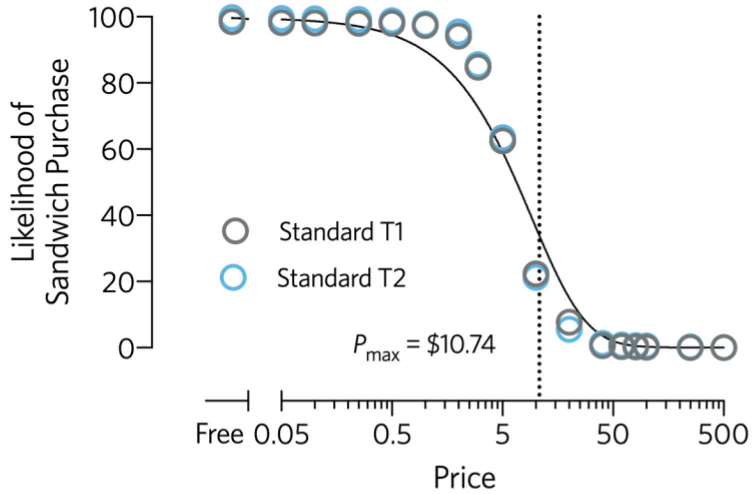


Figure 9

Standard T1 vs High Calories

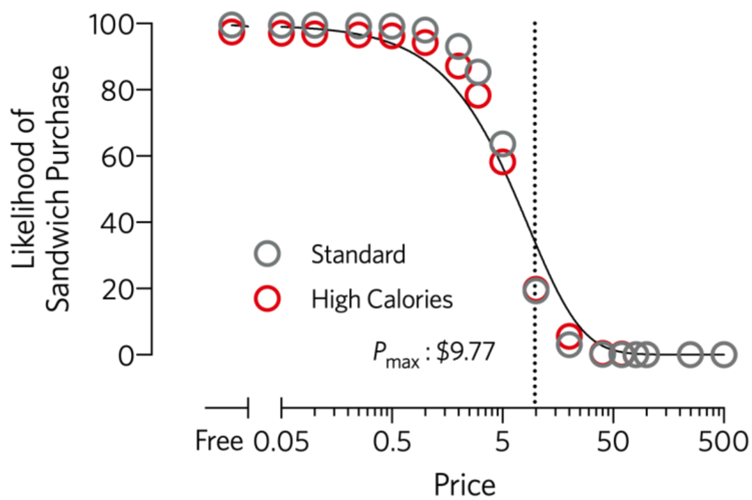


Figure 10

Standard T1 vs Low Calories

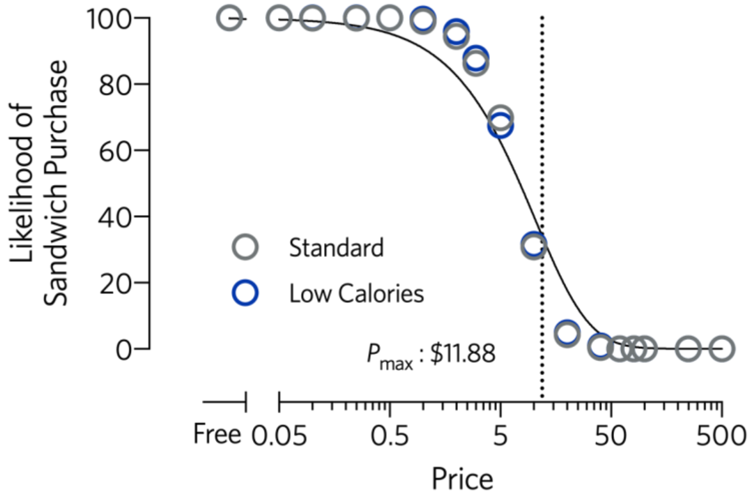


Figure 11

HED vs LED Snacks

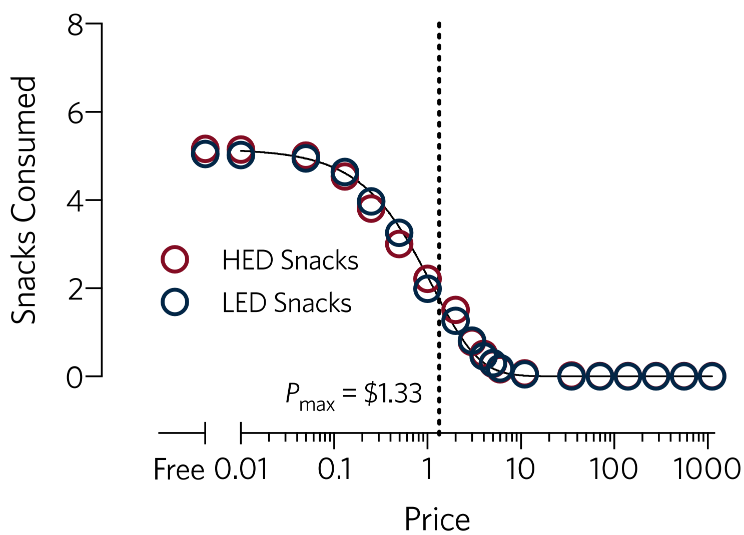


Figure 12

Normal BMI – Standard T1 vs Standard T2

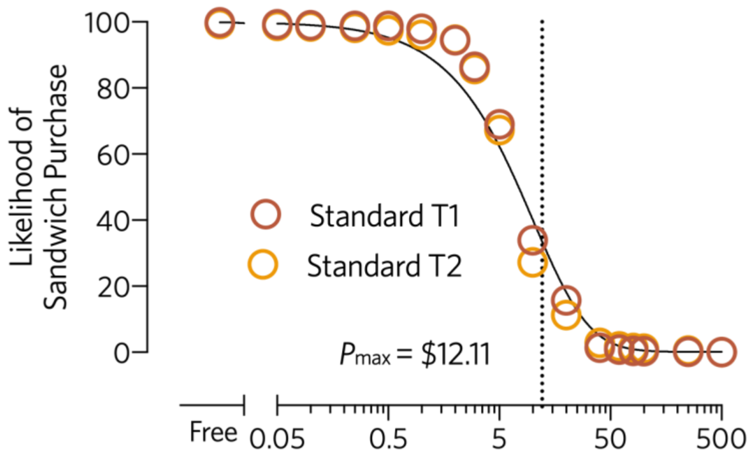


Figure 13

Normal BMI – Standard T1 vs High Calories

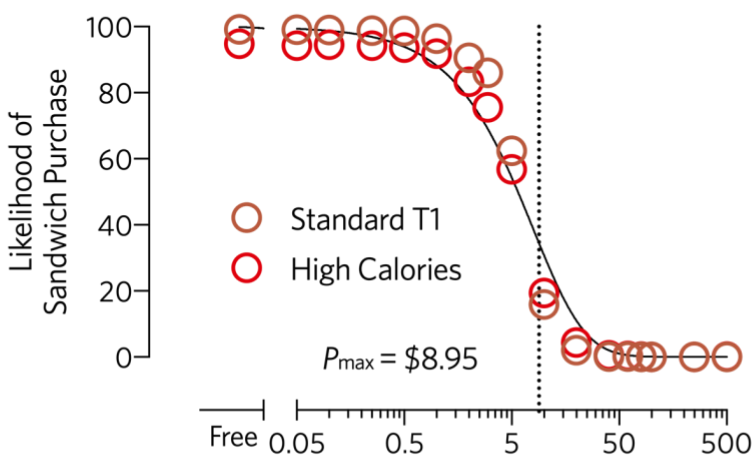


Figure 14

Normal BMI – Standard T1 vs Low Calories

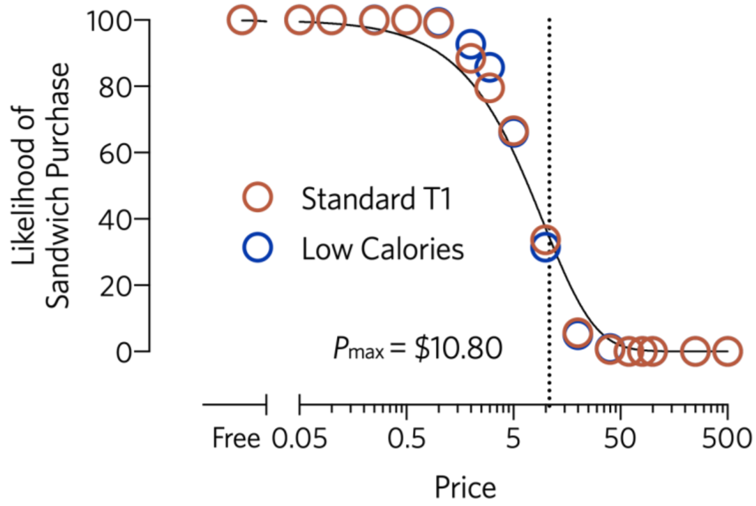


Figure 15

Overweight BMI – Standard T1 vs Standard T2

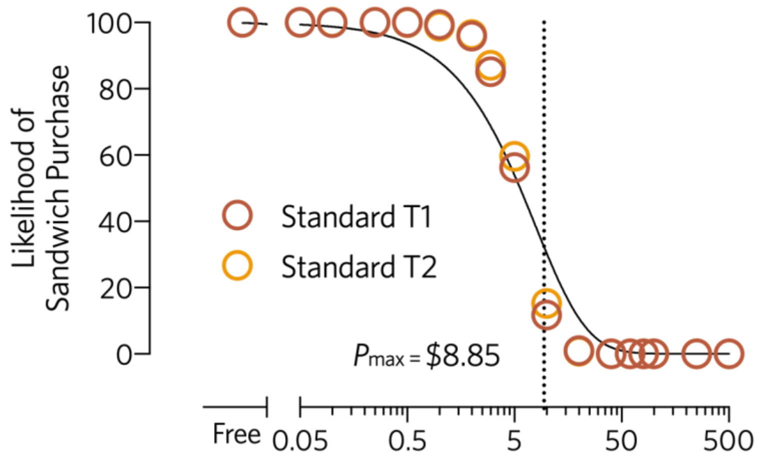


Figure 16

Overweight BMI – Standard T1 vs High Calories

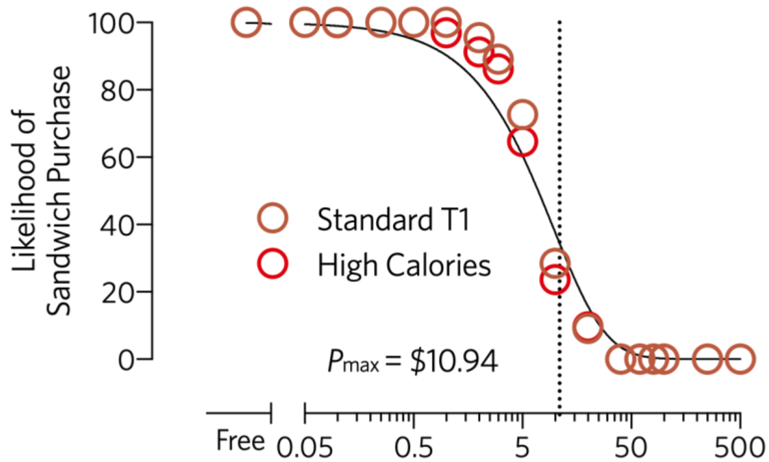


Figure 17

Overweight BMI – Standard T1 vs Low Calories

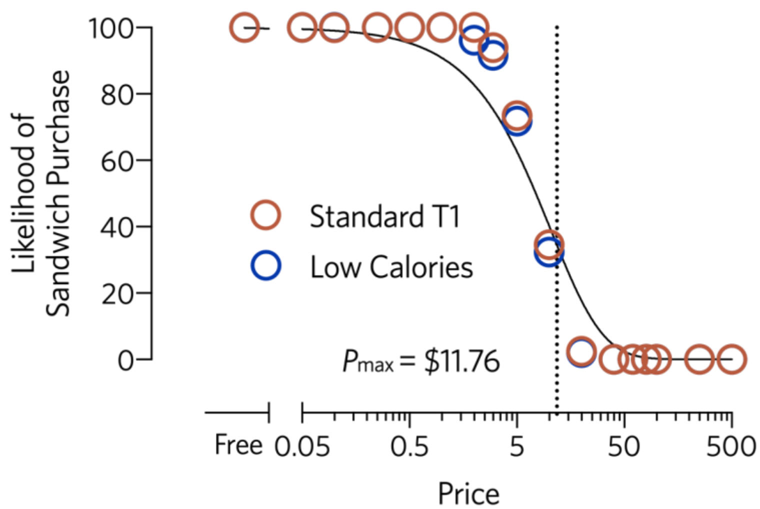


Figure 18

Obese BMI – Standard T1 vs Standard T2

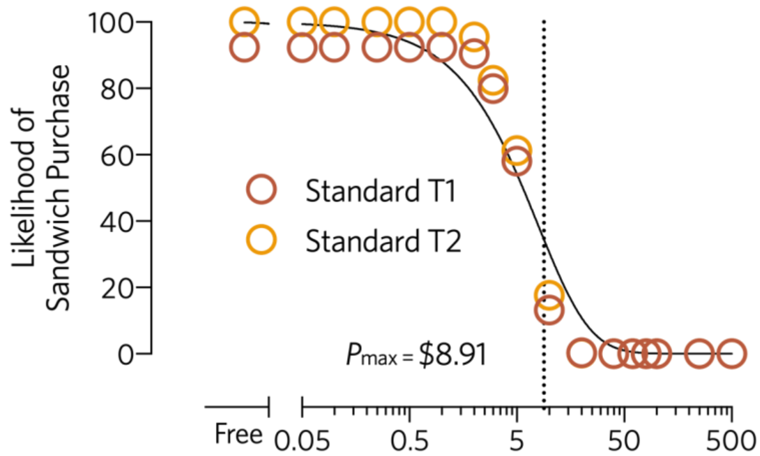


Figure 19

Obese BMI – Standard T1 vs High Calories

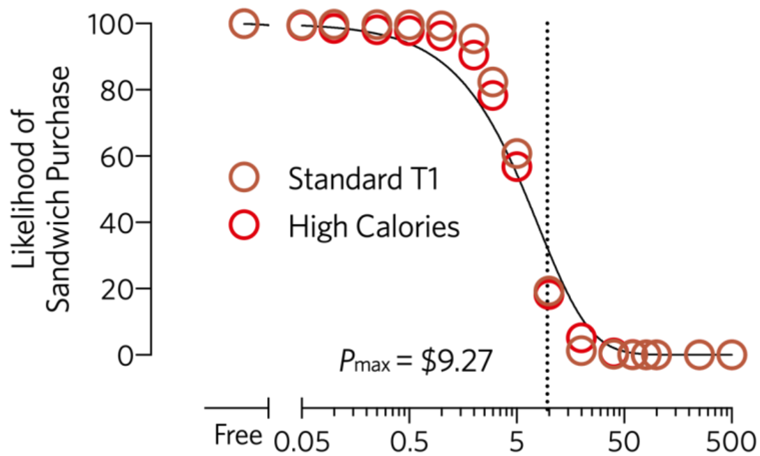


Figure 20

Obese BMI – Standard T1 vs Low Calories

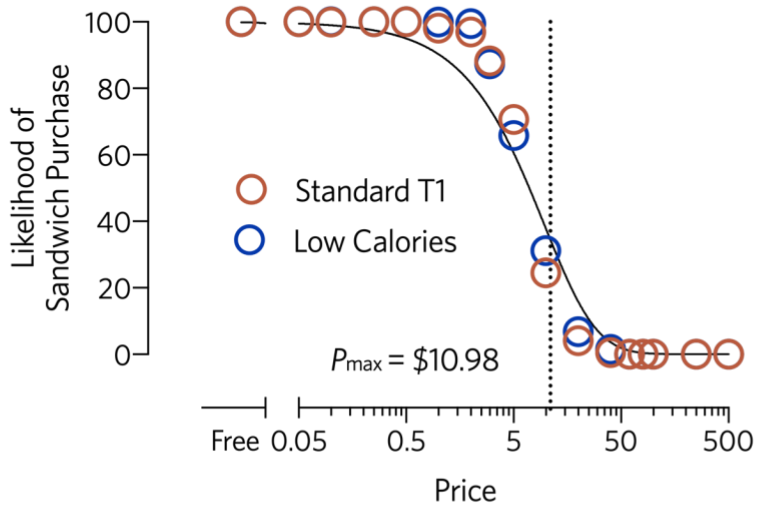


Figure 21

Normal BMI – LED vs HED Snacks

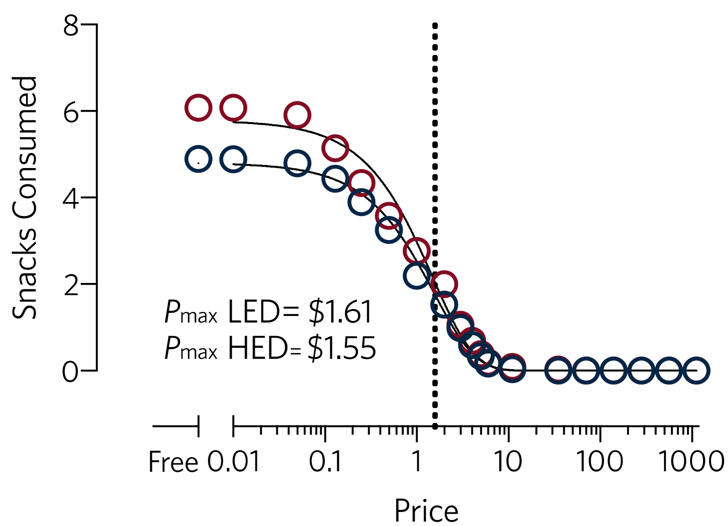


Figure 22

Overweight BMI – LED vs HED Snacks

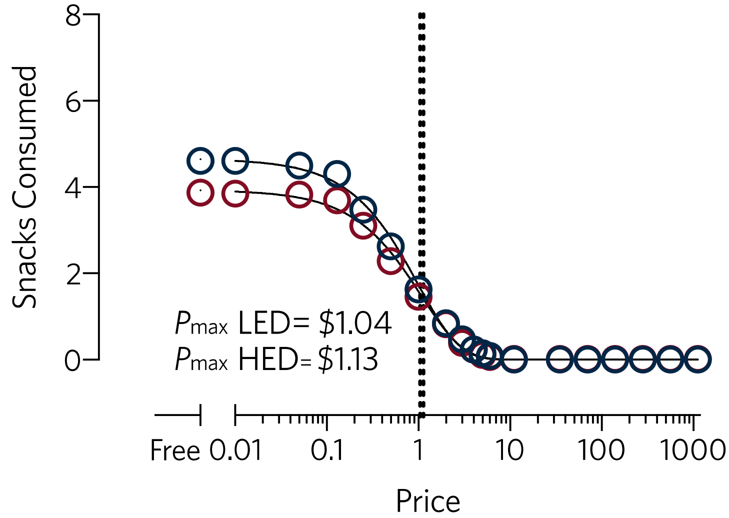
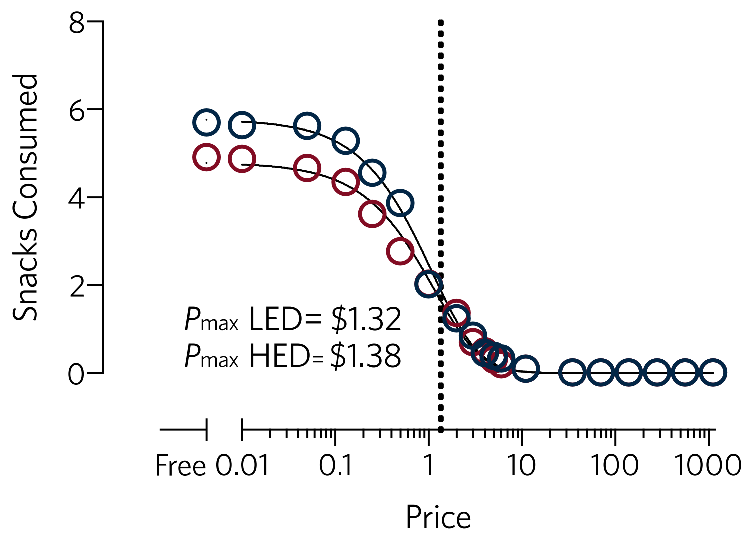


Figure 23

Obese BMI – LED vs HED Snacks



Tables

Table 1

Clinical Scale Scores

Clinical Scales	No. (%) of 172 Participants	
mYFAS 2.0		
Symptom Count Score		
0	119	(69.2%)
1	15	(8.72%)
2	5	(2.90%)
3	7	(4.07%)
4	9	(5.23%)
5	5	(2.91%)
6	7	(4.07%)
7	1	(0.58%)
8	3	(1.74%)
9	1	(0.58%)
Diagnostic Score		
Mild	2	(1.16%)
Moderate	2	(1.16%)
Severe	4	(2.33%)
TFEQ-R18 V2		
Uncontrolled Eating		
0-10	4	(2.33%)
11-20	92	(53.49%)
21-30	64	(37.21)
30+	12	(6.98%)
Cognitive Restraint		
0-5	54	(31.40%)
6-10	107	(62.21%)
11+	11	(6.40%)
Emotional Eating		
6-10	64	(37.21%)
11-20	98	(56.98%)
21+	10	(5.81%)
Overall Score		
<20	2	(1.16%)
20-40	90	(52.33%)

41-60	76	(44.19%)
60+	4	(2.33%)
POF Scale		
Food Available		
0-10	95	(55.23%)
11-20	47	(27.33%)
21-30	30	(17.44%)
Food Present		
0-10	95	(55.23%)
11-20	77	(44.77%)
Food Tasted		
0-10	54	(31.40%)
11-20	102	(59.30%)
21-25	16	(9.30%)
Aggregate Score		
15-20	25	(14.53%)
21-30	48	(27.91%)
31-40	38	(22.09%)
41-50	25	(14.53%)
51-60	23	(13.37%)
61+	13	(7.56%)

Table 2

BMI - Clinical Scale Scores

Scale	Cronbach's α	Clinical Scale Scores ($M \pm SD$)		
		Normal BMI (n = 76)	Overweight BMI (n = 46)	Obese BMI (n = 48)
TFEQ Overall Score	$\alpha = .92$	36.96 ± 10.36	38.43 ± 11.35	44.46 ± 10.53
TFEQ UE	$\alpha = .87$	19.04 ± 5.68	19.30 ± 5.83	22.38 ± 6.00
TFEQ CR	$\alpha = .83$	6.42 ± 2.75	6.83 ± 2.27	7.08 ± 2.23
TFEQ EE	$\alpha = .93$	11.50 ± 4.68	12.30 ± 4.95	15.00 ± 5.06
mYFAS 2.0 symptom	$\alpha = .93$	0.66 ± 1.55	0.82 ± 1.67	1.98 ± 2.67
mYFAS 2.0 diagnosis threshold and category (n of sample)	N/A	n = 2 (1 Moderate, 1 Severe)	n = 0	n = 6 (2 Mild, 1 Moderate, 3 Severe)
POF Food Available	$\alpha = .93$	11.19 ± 6.36	11.46 ± 5.88	15.44 ± 6.84
POF Food Present	$\alpha = .92$	9.64 ± 4.63	9.56 ± 4.43	12.23 ± 4.17
POF Food Tasted	$\alpha = .85$	13.13 ± 5.46	12.89 ± 4.01	14.38 ± 5.26
POF Aggregate	$\alpha = .95$	33.96 ± 14.97	33.91 ± 12.94	42.04 ± 15.10

Note: TFEQ = Three Factor Eating Questionnaire – R18 V2; UE = Uncontrolled Eating; CR = Cognitive Restraint; EE = Emotional Eating; mYFAS 2.0 = modified Yale Food Addiction Scale 2.0; POF = Power of Food Scale

Table 3

Mean Caloric Estimates per Sandwich

Sandwich	Calories *M*	*SD*
Cheeseburger	539.2	302.6
Grilled chicken sandwich	486.1	214.8
Veggie sub	566.8	210.6
Turkey club	482.5	247.7

Table 4

Caloric Densities Provided per Sandwich

Sandwich	Low Calorie	High Calorie
Cheeseburger	270	1,078
Grilled chicken sandwich	243	972
Veggie sub	283	1,134
Turkey club	241	965

Table 5

Nonsystematic Purchase Task Identification Tool

Nonsystematic Purchase Task Identification Tool (N = 243)				
Purchase Task	n	Trend	Bounce	Reversals from Zero
		n of cases per condition (% of condition)		
Standard T1	n = 243	16 (6.58%)	5 (2.06%)	9 (3.70%)
Standard T2	n = 85	3 (3.5%)	0 (0.0%)	4 (4.71%)
High Calorie	n = 94	7 (7.45%)	3 (3.19%)	6 (6.38%)
Low Calorie	n = 64	9 (10.59%)	3 (3.53%)	3 (3.53%)
HED Snack Demand	n = 243	14 (5.76%)	14 (5.76%)	0 (0.0%)
LED Snack Demand	n = 243	15 (6.17%)	13 (5.35%)	2 (0.82%)

Table 6

Participant Demographics

Characteristic	No. (%) of 172 Participants
Age, years	
20-30	50 (29.1%)
31-40	66 (38.4%)
41-50	25 (14.5%)
50+	31 (18.0%)
Ethnicity	
White/Caucasian	134 (77.9%)
African American	15 (8.72%)
Asian	11 (6.40%)
Hispanic	9 (5.23%)
Native American	1 (0.58%)
Other	2 (1.16%)
Female	86 (50.0%)
Primary language	
English	172 (100%)
BMI	
Underweight (17.8-18.4)	2 (1.16%)
Normal (18.5-24.9)	76 (44.2%)
Overweight (25.0-29.0)	46 (26.7%)
Obese (30.0+)	48 (27.9%)
Preferred sandwich	
Cheeseburger	95 (55.2%)
Grilled chicken sandwich	26 (15.1%)
Turkey club	15 (8.72%)
Veggie sub	36 (20.9%)
Average income	
Under $10,000	20 (11.6%)
$10,000-$19,999	16 (9.30%)
$20,000-$29,999	25 (14.5%)
$30,000-$39,999	32 (18.6%)
$40,000-$49,999	16 (9.30%)
$50,000-$74,999	21 (12.2%)
$75,000-$99,999	20 (11.6%)
$100,000-$150,000	12 (6.98%)
$150,000+	4 (2.33%)
Unreported	6 (3.49%)

Table 7

Comparisons between Exponential and Exponentiated Demand Equations

Comparisons of R^2 & AIC for Exponential and Exponentiated Demand Equations								
			Exponential			Exponentiated		
	k value	Demand Curve Analysis	R^2		AIC Model Selection (probability alpha is same for both data sets, %)	R^2		AIC Model Selection (probability alpha is same for both data sets, %)
Sandwich Demand Purchase Task	3.5	Standard T1 vs. Standard T2	Standard T1: 0.9905	Standard T2: 0.9550	(7.465%)	Standard T1: 0.9803	Standard T2: 0.9972	(76.97%)
	3.8	Standard T1 vs. High Calories	Standard T1: 0.9858	High Calories: 0.9866	(15.52%)	Standard T1: 0.9757	High Calories: 0.9876	(71.82%)
	3.6	Standard T1 vs. Low Calories	Standard T1: 0.9652	Low Calories: 0.9727	(75.69%)	Standard T1: 0.9804	Low Calories: 0.9798	(76.92%)
	3.6	High Calories vs. Low Calories	High Calories: 0.9868	Low Calories: 0.9727	(77.26%)	High Calories: 0.9878	Low Calories: 0.9798	(30.81%)
Snack Demand Purchase Task	3.4	LED snack demand vs. HED snack demand	LED snack: 0.9589	HED snack: 0.8289	(73.13%)	LED snack: 0.9962	HED snack: 0.9924	(75.41%)

Note: Researchers used the same *k* values for both equations in order to make equal comparisons of goodness-of-fit. Refer to methods section for a description of how *k* was calculated.

Table 8

Spearman's Rank Order Correlations

	Relations between BMI, Clinical Scales, and Demand Indices											
	1	2	3	4	5	6	7	8	9	10	11	12
1 HED O_{max}												
2 LED O_{max}	.604**											
3 T1 O_{max}	.321**	.310**										
4 T2 O_{max}	.246**	.259**	.739**									
5 BMI	-.064	-.121	-.075	-.024								
6 ln*k* MCQ	.228**	.213**	.085	.074	-.075							
7 ln*k* FCQ	.071	.154*	-.040	-.067	-.084	.290**						
8 POF	.071	-.023	.063	.020	.262**	.046	-.041					
9 mYFAS 2.0	.084	-.052	-.033	-.020	.227**	.049	.038	.574**				
10 TFEQ-R18 V2	-.039	-.029	.061	-.040	.296**	.012	.061	.621**	.536**			
11 HED Q_0	.329**	.170*	.000	-.016	-.083	.093	.002	-.043	-.039	-.092		
12 LED Q_0	.221**	.353**	.016	-.017	.034	.070	.019	.103	.109	.054	.548**	

**Correlation significant at 0.01 level; *Correlation significant at the 0.05 level
T2 indicates "time two" of likelihood of sandwich purchase task administration; these scores are representative of values across all three conditions.

References

Ahima, R. S., & Lazar, M. A. (2013). The health risk of obesity - Better metrics imperative. *Science*, *341*(6148), 856–858. https://doi.org/10.1126/science.1241244

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

Amlung, M., Petker, T., Jackson, J., Balodis, I., & Mackillop, J. (2016). Steep discounting of delayed monetary and food rewards in obesity: A meta-analysis. *Psychological Medicine*, *46*, 2423–2434. https://doi.org/10.1017/S0033291716000866

Amlung, M. T., Acker, J., Stojek, M. K., Murphy, J. G., & MacKillop, J. (2012). Is talk "cheap"? An initial investigation of the equivalence of alcohol purchase task performance for hypothetical and actual rewards. *Alcoholism: Clinical and Experimental Research*, *36*(4), 716–724. https://doi.org/10.1111/j.1530-0277.2011.01656.x

Amlung, M., Marsden, E., Holshausen, K., Morris, V., Patel, H., Vedelago, L., Naish, K. R., Reed, D. D., & McCabe, R. E. (2019). Delay discounting as a transdiagnostic process in psychiatric disorders: A meta-analysis. *JAMA Psychiatry*, *76*(11), 1176–1186. https://doi.org/10.1001/jamapsychiatry.2019.2102

Appelhans, B. M., Woolf, K., Pagoto, S. L., Schneider, K. L., Whited, M. C., & Liebman, R. (2011). Inhibiting food reward: Delay discounting, food reward sensitivity, and palatable food intake in overweight and obese women. *Obesity*, *19*(11), 2175–2182. https://doi.org/10.1038/oby.2011.57

Avila, C., Holloway, A. C., Hahn, M. K., Morrison, K. M., Restivo, M., Anglin, R., & Taylor, V.

H. (2015). An overview of links between obesity and mental health. *Current Obesity Reports*, *4*(3), 303–310. https://doi.org/10.1007/s13679-015-0164-9

Beumont, P., Al-Alami, M., & Touyz, S. (1988). Relevance of a standard measurement of undernutrition to the diagnosis of anorexia nervosa: Use of Quetelet's Body Mass Index (BMI). *International Journal of Eating Disorders*, *7*(3), 399–405. https://doi.org/10.1002/1098-108X(198805)7:3<399::AID-EAT2260070312>3.0.CO;2-R

Bickel, W. K., Wilson, G. A., Chen, C., Koffarnus, M. N., & Franck, C. T. (2016). Stuck in time: Negative income shock constricts the temporal window of valuation spanning the future and the past. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0163051

Bickel, W. K., Jarmolowicz, D. P., Mueller, E. T., Koffarnus, M. N., & Gatchalian, K. M. (2012). Excessive discounting of delayed reinforcers as a trans-disease process contributing to addiction and other disease-related vulnerabilities: Emerging evidence. *Pharmacology and Therapeutics*, *134*(3), 287–297. https://doi.org/10.1016/j.pharmthera.2012.02.004

Bickel, W. K., Johnson, M. W., Koffarnus, M. N., MacKillop, J., & Murphy, J. G. (2014). The behavioral economics of substance use disorders: Reinforcement pathologies and their repair. *Annual Review of Clinical Psychology*, *10*, 641–677. https://doi.org/10.1146/annurev-clinpsy-032813-153724

Bray, G. A., Heisel, W. E., Afshin, A., Jensen, M. D., Dietz, W. H., Long, M., Kushner, R. F., Daniels, S. R., Wadden, T. A., Tsai, A. G., Hu, F. B., Jakicic, J. M., Ryan, D. H., Wolfe, B. M., & Inge, T. H. (2018). The science of obesity management: An endocrine society scientific statement. *Endocrine Reviews*, *39*(2), 79–132. https://doi.org/10.1210/er.2017-00253

Brunault, P., Ballon, N., Gaillard, P., Réveillère, C., & Courtois, R. (2014). Validation of the

French version of the yale food addiction scale: an examination of its factor structure, reliability, and construct validity in a nonclinical sample. *Canadian Journal of Psychiatry*, *59*(5), 276–284. https://doi.org/10.1177/070674371405900507

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5. https://doi.org/10.1177/1745691610393980

Burkhauser, R. V., & Cawley, J. (2008). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics*, *27*(2), 519–529. https://doi.org/10.1016/j.jhealeco.2007.05.005

Cappelleri, J. C., Bushmakin, A. G., Gerber, R. A., Leidy, N. K., Sexton, C. C., Karlsson, J., & Lowe, M. R. (2009). Evaluating the Power of Food Scale in obese subjects and a general sample of individuals: development and measurement properties. *International Journal of Obesity*, *33*, 913–922. https://doi.org/10.1038/ijo.2009.107

Cappelleri, J. C., Bushmakin, A. G., Gerber, R. A., Leidy, N. K., Sexton, C. C., Lowe, M. R., & Karlsson, J. (2009). Psychometric analysis of the Three-Factor Eating Questionnaire-R21: results from a large diverse sample of obese and non-obese participants. *International Journal of Obesity*, *33*, 611–620. https://doi.org/10.1038/ijo.2009.74

Carroll, A. E. (2015, November 30). *The failure of calorie counts on menus.* The New York Times. https://www.nytimes.com/2015/12/01/upshot/more-menus-have-calorie-labeling-but-obesity-rate-remains-high.html?_r=1.

Center for Disease Control and Prevention (2020, June 29). *Overweight & Obesity: Adult Obesity Facts.* https://www.cdc.gov/obesity/data/adult.html

Cecchini, M., & Warin, L. (2016). Impact of food labelling systems on food choices and eating behaviours: a systematic review and meta-analysis of randomized studies. *Obesity reviews: an official journal of the International Association for the Study of Obesity*, *17*(3), 201–210. https://doi.org/10.1111/obr.12364

Christoph, M. J., Larson, N., Laska, M. N., & Neumark-Sztainer, D. (2018). Nutrition facts panels: Who uses them, what do they use, and how does use relate to dietary intake? *Journal of the Academy of Nutrition and Dietetics*, *118*(2), 217–228. https://doi.org/10.1016/j.jand.2017.10.014

Clark, S. M., & Saules, K. K. (2013). Validation of the Yale Food Addiction Scale among a weight-loss surgery population. *Eating Behaviors*, *14*(2), 216–219. https://doi.org/10.1016/j.eatbeh.2013.01.002

Coppock, A. (2019). Generalizing from survey experiments conducted on mechanical turk: A replication approach. *Political Science Research and Methods, 7*(3), 613-628. https://doi.org/10.1017/psrm.2018.10

Courtney, A. L., PeConga, E. K., Wagner, D. D., & Rapuano, K. M. (2018). Calorie information and dieting status modulate reward and control activation during the evaluation of food images. *PLoS ONE*, *13*(11), e0204744. https://doi.org/10.1371/journal.pone.0204744

Dallas, S. K., Liu, P. J., & Ubel, P. A. (2019). Don't count calorie labeling out: calorie counts on the left side of menu items lead to lower calorie food choices. *Journal of Consumer Psychology*, *29*(1), 60–69. https://doi.org/10.1002/jcpy.1053

Daugherty, J. R., & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences*, *48*(2), 202–207. https://doi.org/10.1016/j.paid.2009.10.007

Davis, C., Curtis, C., Levitan, R. D., Carter, J. C., Kaplan, A. S., & Kennedy, J. L. (2011).

Evidence that “food addiction” is a valid phenotype of obesity. *Appetite*, *57*(3), 711–717. https://doi.org/10.1016/j.appet.2011.08.017

Dingemans, A. E., & van Furth, E. F. (2012). Binge eating disorder psychopathology in normal weight and obese individuals. *International Journal of Eating Disorders*, *45*(1), 135–138. https://doi.org/10.1002/eat.20905

Drewnowski, A. (2005). Concept of a nutritious food: toward a nutrient density score. *The American Journal of Clinical Nutrition*, *82*(4), 721–732. https://doi.org/https://doi.org/10.1093/ajcn/82.4.721

Drewnowski, A., & Specter, S. E. (2004). Poverty and obesity: The role of energy density and energy costs. In *American Journal of Clinical Nutrition* (Vol. 79, Issue 1, pp. 6–16). https://doi.org/10.1093/ajcn/79.1.6

Dumanovsky, T., Huang, C. Y., Nonas, C. A., Matte, T. D., Bassett, M. T., & Silver, L. D. (2011). Changes in energy content of lunchtime purchases from fast food restaurants after introduction of calorie labelling: Cross sectional customer surveys. *BMJ*, *343*(d4464). https://doi.org/10.1136/bmj.d4464

Eichen, D. M., Lent, M. R., Goldbacher, E., & Foster, G. D. (2013). Exploration of “food addiction” in overweight and obese treatment-seeking adults. *Appetite*, *67*, 22–24. https://doi.org/10.1016/j.appet.2013.03.008

Epstein, L. H., Dearing, K. K., & Roba, L. G. (2010). A questionnaire approach to measuring the relative reinforcing efficacy of snack foods. *Eating Behaviors*, *11*(2), 67–73. https://doi.org/10.1016/J.EATBEH.2009.09.006

Epstein, L. H., Dearing, K. K., Roba, L. G., & Finkelstein, E. (2010). The influence of taxes and subsidies on energy purchased in an experimental purchasing study. *Psychological Science*, *21*(3), 406–414. https://doi.org/10.1177/0956797610361446

Epstein, L. H., Finkelstein, E., Raynor, H., Nederkoorn, C., Fletcher, K. D., Jankowiak, N., & Paluch, R. A. (2015). Experimental analysis of the effect of taxes and subsides on calories purchased in an on-line supermarket. *Appetite*, *95*, 245–251. https://doi.org/10.1016/j.appet.2015.06.020

Epstein, L. H., Paluch, R. A., Carr, K. A., Temple, J. L., Bickel, W. K., & MacKillop, J. (2018). Reinforcing value and hypothetical behavioral economic demand for food and their relation to BMI. *Eating Behaviors*, *29*, 120–127. https://doi.org/10.1016/J.EATBEH.2018.03.008

Epstein, L. H., Stein, J. S., Paluch, R. A., MacKillop, J., & Bickel, W. K. (2018). Binary components of food reinforcement: Amplitude and persistence. *Appetite*, *120*, 67–74. https://doi.org/10.1016/j.appet.2017.08.023

Espel-Huynh, H. M., Muratore, A. F., & Lowe, M. R. (2018). A narrative review of the construct of hedonic hunger and its measurement by the Power of Food Scale. *Obesity Science and Practice*, *4*(3), 238–249. https://doi.org/10.1002/osp4.161

Faith, M. S., Matz, P. E., & Jorge, M. A. (2002). Obesity - Depression associations in the population. *Journal of Psychosomatic Research*, *53*(4), 935–942. https://doi.org/10.1016/S0022-3999(02)00308-2

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191. https://doi.org/10.3758/BF03193146

Fazzino, T. L., Rohde, K., & Sullivan, D. K. (2019). Hyper-palatable foods: Development of a quantitative definition and application to the US food system database. *Obesity*, *27*(11), 1761–1768. https://doi.org/10.1002/oby.22639

Field, A. E., Coakley, E. H., Must, A., Spadano, J. L., Laird, N., Dietz, W. H., Rimm, E., & Colditz, G. A. (2001). Impact of overweight on the risk of developing common chronic diseases during a 10-year period. *Archives of Internal Medicine*, *161*(13), 1581–1586. https://doi.org/10.1001/archinte.161.13.1581

Flegal, K. M., Ogden, C. L., Fryar, C., Afful, J., Klein, R., & Huang, D. T. (2019). Comparisons of self-reported and measured height and weight, BMI, and obesity prevalence from national surveys: 1999-2016. *Obesity, 27*(10), 1711–1719. https://doi.org/10.1002/oby.22591

Foster, R. N. S., Fincham, T. A., Becirevic, A., & Reed, D. D. (2020). *Likelihood of Purchase Calculator* [spreadsheet application]. https://osf.io/abys9/ doi:10.17605/OSF.IO/ABYS9

Foster, R. N. S., Fincham, T. A., Becirevic, A., & Reed, D. D. (2020). *Observed Demand Calculator* (Version 2.0) [spreadsheet application] https://osf.io/gpvw6/ doi: 10.17605/OSF.IO/GPVW6

Foxx-Orenstein, A. E. (2010). Gastrointestinal symptoms and diseases related to obesity: An overview. *Gastroenterology Clinics of North America*, *39*, 23–37. https://doi.org/10.1016/j.gtc.2009.12.006

Ganley, R. M. (1988). Emotional eating and how it relates to dietary restraint, disinhibition, and perceived hunger. *International Journal of Eating Disorders*, *7*(5), 635–647. https://doi.org/10.1002/1098-108X(198809)7:5<635::AID-EAT2260070507>3.0.CO;2-K

Garber, A. K., Cheng, J., Accurso, E. C., Adams, S. H., Buckelew, S. M., Kapphahn, C. J., Kreiter, A., Le Grange, D., Machen, V. I., Moscicki, A.-B., Saffran, K., Sy, A. F., Wilson, L., & Golden, N. H. (2019). Weight loss and illness severity in adolescents with atypical anorexia nervosa. *Pediatrics*, *144*(6), e20192339. https://doi.org/10.1542/peds.2019-2339

Gearhardt, A. N., Corbin, W. R., & Brownell, K. D. (2009). Preliminary validation of the Yale Food Addiction Scale. *Appetite*, *52*(2), 430–436. https://doi.org/10.1016/J.APPET.2008.12.003

Gearhardt, A. N., Corbin, W. R., & Brownell, K. D. (2016). Development of the Yale Food Addiction Scale Version 2.0. *Psychology of Addictive Behaviors*, *30*(1), 113–121. https://doi.org/10.1037/adb0000136

Gearhardt, A. N., Roberto, C. A., Seamans, M. J., Corbin, W. R., & Brownell, K. D. (2013). Preliminary validation of the Yale Food Addiction Scale for children. *Eating Behaviors*, *14*(4), 508–512. https://doi.org/10.1016/j.eatbeh.2013.07.002

Goldschmidt, A. B., Le Grange, D., Powers, P., Crow, S. J., Hill, L. L., Peterson, C. B., Crosby, R. D., & Mitchell, J. E. (2011). Eating disorder symptomatology in normal-weight vs. obese individuals with binge eating disorder. *Obesity*, *19*(7), 1515–1518. https://doi.org/10.1038/oby.2011.24

Goldschmidt, A. B., Smith, K. E., Lavender, J. M., Engel, S. G., & Haedt-Matt, A. (2019). Trait-level facets of impulsivity and momentary, naturalistic eating behavior in children and adolescents with overweight/obesity. *Journal of Psychiatric Research*, *110*, 24–30. https://doi.org/10.1016/j.jpsychires.2018.12.018

Gorber, S. C., Tremblay, M., Moher, D., & Gorber, B. (2007). A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review. *Obesity Reviews*, *8*(4), 307–326. https://doi.org/10.1111/j.1467-789X.2007.00347.x

Gorwood, P., Duriez, P., Lengvenyte, A., Guillaume, S., & Criquillion, S. (2019). Clinical insight in anorexia nervosa: Associated and predictive factors. *Psychiatry Research*, *281*, 112561. https://doi.org/10.1016/j.psychres.2019.112561

Grabenhorst, F., Schulte, F. P., Maderwald, S., & Brand, M. (2013). Food labels promote healthy choices by a decision bias in the amygdala. *NeuroImage*, *74*(1), 152–163. https://doi.org/10.1016/j.neuroimage.2013.02.012

Gutin, I. (2018). In BMI we trust: Reframing the body mass index as a measure of health. *Social Theory & Health*, *16*, 256–271. https://doi.org/10.1057/s41285-017-0055-0

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2020, February). *Prevalence of obesity and severe obesity among adults: United States, 2017-2018.* Center for Disease Control and Prevention. https://www.cdc.gov/nchs/products/databriefs/db360.htm

Harzem, P. (1984). Experimental analysis of individual differences and personality. *Journal of the Experimental Analysis of Behavior*, *42*(3), 385–395. https://doi.org/10.1901/jeab.1984.42-385

Hendrickson, K. L., & Rasmussen, E. B. (2017). Mindful eating reduces impulsive food choice in adolescents and adults. *Health Psychology*, *36*(3), 226–235. https://doi.org/10.1037/hea0000440

Hendrickson, K. L., Rasmussen, E. B., & Lawyer, S. R. (2015). Measurement and validation of measures for impulsive food choice across obese and healthy-weight individuals. *Appetite*, *90*, 254–263. https://doi.org/10.1016/J.APPET.2015.03.015

Hursh, S. R. (1980). Economic concepts for the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, *34*(2), 219–238. https://doi.org/10.1901/jeab.1980.34-219

Hursh, S. R. (1984). Behavioral economics. *Journal of the Experimental Analysis of Behavior*, *42*(3), 435–452. https://doi.org/10.1901/jeab.1984.42-435

Hursh, S. R., & Roma, P. G. (2013). Behavioral economics and empirical public policy. *Journal* *of the Experimental Analysis of Behavior*, *99*(1), 98–124. https://doi.org/10.1002/jeab.7

Hursh, S. R., & Roma, P. G. (2016). Behavioral economics and the analysis of consumption and choice. *Managerial and Decision Economics, 37*(4–5), 224–238. https://doi.org/10.1002/mde.2724

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review*, *115*(1), 186–198. https://doi.org/10.1037/0033-295X.115.1.186

Jarmolowicz, D. P., Cherry, J. B. C., Reed, D. D., Bruce, J. M., Crespi, J. M., Lusk, J. L., & Bruce, A. S. (2014). Robust relation between temporal discounting rates and body mass. *Appetite*, *78*, 63–67. https://doi.org/10.1016/j.appet.2014.02.013

Kaplan, B. A., Amlung, M., Reed, D. D., Jarmolowicz, D. P., Mckerchar, T. L., & Lemley, S. M. (2016b). Automating scoring of delay discounting for the 21- and 27-item monetary choice questionnaires. *The Behavior Analyst*, *39*(2), 293–304. https://doi.org/10.1007/s40614-016-0070-9

Kaplan, B. A., Foster, R. N. S., Reed, D. D., Amlung, M., Murphy, J. G., & MacKillop, J. (2018). Understanding alcohol motivation using the alcohol purchase task: A methodological systematic review. *Drug and Alcohol Dependence, 191*, 117–140. https://doi.org/10.1016/j.drugalcdep.2018.06.029

Kaplan, B. A., Gelino, B. W., & Reed, D. D. (2018). A behavioral economic approach to green consumerism: Demand for reusable shopping bags. *Behavior and Social Issues, 27*, 20–30. https://doi.org/10.5210/bsi.v27i0.8003

Karlsson, J., Persson, L.-O., Sjöström, L., & Sullivan, M. (2000). Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEQ) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. *International Journal of Obesity*, *24*, 1715–1725. https://doi.org/10.1038/sj.ijo.0801442

Karnani, A., McFerran, B., & Mukhopadhyay, A. (2016). The obesity crisis as market failure: An analysis of systemic causes and corrective mechanisms. *Journal of the Association for Consumer Research*, *1*(3), 445–470. https://doi.org/10.1086/686244

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78–87. https://doi.org/10.1037/0096-3445.128.1.78

Kiszko, K. M., Martinez, O. D., Abrams, C., & Elbel, B. (2014). The influence of calorie labeling on food orders and consumption: A review of the literature. *Journal of Community Health*, *39*(6), 1248–1269. https://doi.org/10.1007/s10900-014-9876-0

Koffarnus, M. N., Franck, C. T., Stein, J. S., & Bickel, W. K. (2015). A modified exponential behavioral economic demand model to better describe consumption data. *Experimental and Clinical Psychopharmacology*, *23*(6), 504–512. https://doi.org/10.1037/pha0000045

Kolodinsky, J., Green, J., Michahelles, M., & Harvey-Berino, J. R. (2008). The use of nutritional labels by college students in a food-court setting. *Journal of American College Health*, *57*(3), 297–302. https://doi.org/10.3200/JACH.57.3.297-302

Krieger, J. W., Chan, N. L., Saelens, B. E., Ta, M. L., Solet, D., & Fleming, D. W. (2013). Menu labeling regulations and calories purchased at chain restaurants. *American Journal of Preventive Medicine*, *44*(6), 595–604. https://doi.org/10.1016/j.amepre.2013.01.031

Kushner, R. F. (2002). Medical management of obesity. *Seminars in Gastrointestinal Diesease*, *13*(3), 123-132.

Lee, H., Andrew, M., Gebremariam, A., Lumeng, J. C., & Lee, J. M. (2014). Longitudinal associations between poverty and obesity from birth through adolescence. *American*

Journal of Public Health, *104*, e70-6. https://doi.org/10.2105/AJPH.2013.301806

Li, W.-Q., McGeary, J. E., Cho, E., Flint, A., Wu, S., Ascherio, A., Rimm, E., Field, A., & Qureshi, A. A. (2017). Indoor tanning bed use and risk of food addiction based on the modified Yale Food Addiction Scale. *Journal of Biomedical Research, 31*(1), 31–39. https://doi.org/10.7555/JBR.31.20160098

Lim, S. L., & Bruce, A. S. (2015). Can't wait to lose weight? Characterizing temporal discounting parameters for weight-loss. *Appetite*, *85*, 8–13. https://doi.org/10.1016/j.appet.2014.11.001

Lipsky, L. M., Nansel, T. R., Haynie, D. L., Liu, D., Eisenberg Colman, M. H., & Simons-Morton, B. (2019). Lack of prospective relationships of the Power of Food Scale with Body Mass Index and dieting over 2 years in U.S. emerging adults. *Eating Behaviors*, *34*, 101302. https://doi.org/10.1016/j.eatbeh.2019.101302

Long, M. W., Tobias, D. K., Cradock, A. L., Batchelder, H., & Gortmaker, S. L. (2015). Systematic review and meta-analysis of the impact of restaurant menu calorie labeling. *American Journal of Public Health*, *105*(5), e11–e24.

Lowe, M. R., Butryn, M. L., Didie, E. R., Annunziato, R. A., Thomas, J. G., Crerand, C. E., Ochner, C. N., Coletta, M. C., Bellace, D., Wallaert, M., & Halford, J. (2009). The Power of Food Scale. A new measure of the psychological influence of the food environment. *Appetite*, *53*(1), 114–118. https://doi.org/10.1016/j.appet.2009.05.016

Masters, R. K., Reither, E. N., Powers, D. A., Yang, Y. C., Burger, A. E., & Link, B. G. (2013). The impact of obesity on US mortality levels: The importance of age and cohort factors in population estimates. *American Journal of Public Health*, *103*(10), 1895–1901. https://doi.org/10.2105/AJPH.2013.301379

Matjasko, J. L., Cawley, J. H., Baker-Goering, M. M., & Yokum, D. V. (2016). Applying behavioral economics to public health policy: Illustrative examples and promising directions. *American Journal of Preventive Medicine*, *50*(5), S13–S19. https://doi.org/10.1016/j.amepre.2016.02.007

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 5. The effect of delay and of intervening events on reinforcement value* (pp. 55-73). Hillsdale, NJ: Erlbaum.

Mazzeo, S. E., Aggen, S. H., Anderson, C., Tozzi, F., & Bulik, C. M. (2003). Investigating the structure of the eating inventory (three-factor eating questionnaire): A confirmatory approach. *International Journal of Eating Disorders*, *34*(2), 255–264. https://doi.org/10.1002/eat.10180

McClelland, J., Dalton, B., Kekic, M., Bartholdy, S., Campbell, I. C., & Schmidt, U. (2016). A systematic review of temporal discounting in eating disorders and obesity: Behavioural and neuroimaging findings. In *Neuroscience and Biobehavioral Reviews* (Vol. 71, pp. 506–528). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2016.09.024

McPherson, S., Burduli, E., Smith, C., Herron, J., Oluwoye, O., Hirchak, K., Orr, M., McDonell, M., & Roll, J. (2018). A review of contingency management for the treatment of substance-use disorders: adaptation for underserved populations, use of experimental technologies, and personalized optimization strategies. *Substance Abuse and Rehabilitation*, *9*, 43–57. https://doi.org/10.2147/sar.s138439

Meule, A., & Gearhardt, A. N. (2014). Five years of the Yale Food Addiction Scale: Taking stock and moving forward. *Current Addiction Reports, 1*, 193–205.

https://doi.org/10.1007/s40429-014-0021-z

Mitchell, T. B., Cushing, C. C., & Amaro, C. M. (2016). Psychometric properties of the Power of Food Scale in a community sample of preadolescents and adolescents. *Journal of Child and Family Studies*, *25,* 2733–2739. https://doi.org/10.1007/s10826-016-0444-3

Murphy, C. M., Stojek, M. K., & MacKillop, J. (2014). Interrelationships among impulsive personality traits, food addiction, and Body Mass Index. *Appetite*, *73*, 45–50. https://doi.org/10.1016/j.appet.2013.10.008

Murphy, J. G., MacKillop, J., Skidmore, J. R., & Pederson, A. A. (2009). Reliability and validity of a demand curve measure of alcohol reinforcement. *Experimental and Clinical Psychopharmacology*, *17*(6), 396–404. https://doi.org/10.1037/a0017684

Naudé, G. P., Foster, R. N. S., Bartley, M., Martinetti, M. P., Ayers, L. O., & Reed, D. D. (2019). Predicting adverse consequences of alcohol consumption in underage college students using a novel Fake ID Purchase Task. *Experimental and Clinical Psychopharmacology*. Advance online publication. https://doi.org/10.1037/pha0000345

Ng, J., Stice, E., Yokum, S., & Bohon, C. (2011). An fMRI study of obesity, food reward, and perceived caloric density. Does a low-fat label make food less appealing? *Appetite*, *57*(1), 65–72. https://doi.org/10.1016/j.appet.2011.03.017

Norris, A. E., & Aroian, K. J. (2004). To transform or not transform skewed data for psychometric analysis: That is the question! *Nursing Research*, *53*(1), 67–71.

Nuttall, F. Q. (2015). Body mass index: Obesity, BMI, and health: A critical review. *Nutrition Today*, *50*(3), 117–128. https://doi.org/10.1097/NT.0000000000000092

Odum, A. L. (2011a). Delay discounting: I’m a k, you’re a k. *Journal of the Experimental Analysis of Behavior*, *96*(3), 427–439. https://doi.org/10.1901/jeab.2011.96-423

Odum, A. L. (2011b). Delay discounting: Trait variable? *Behavioural Processes*, *87*(1), 1–9. https://doi.org/10.1016/j.beproc.2011.02.007

Odum, A. L., Becker, R. J., Haynes, J. M., Galizio, A., Frye, C. C. J., Downey, H., Friedel, J. E., & Perez, D. M. (2020). Delay discounting of different outcomes: Review and theory. *Journal of the Experimental Analysis of Behavior*, *113*(3), 657–679. https://doi.org/10.1002/jeab.589

Odum, A. L., & Rainaud, C. P. (2003). Discounting of delayed hypothetical money, alcohol, and food. *Behavioural Processes*, *64*(3), 305–313. https://doi.org/10.1016/S0376-6357(03)00145-1

Patient Protection and Affordable Care Act. (2010). 42 U.S.C. § 18001 et seq.

Porter, S. R., Whitcomb, M. E., & Weitzer, W. H. (2004). Multiple surveys of students and survey fatigue. *New Directions for Institutional Research*, *2004*(121), 63–73. https://doi.org/10.1002/ir.101

Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology*, *53*(5), 793–808. https://doi.org/10.1080/10635150490522304

Powell, L. M., & Chaloupka, F. J. (2009). Food prices and obesity: Evidence and policy implications for taxes and subsidies. *Milbank Quarterly*, *87*(1), 229–257. https://doi.org/10.1111/j.1468-0009.2009.00554.x

Privitera, G. J., King-Shepard, Q. W., Cuifolo, K. N., & Doraiswamy, P. M. (2019). Differential food intake and food choice by depression and body mass index levels following a mood manipulation in a buffet-style setting. *Journal of Health Psychology*, *24*(2), 199–208. https://doi.org/10.1177/1359105316650508

Puhl, R. M., & Brownell, K. D. (2001). Bias, discrimination, and obesity. *Obesity Research*, *9*(12), 788–805. https://doi.org/10.1038/oby.2001.108

Puhl, R. M., & Brownell, K. D. (2003). Psychosocial origins of obesity stigma: toward changing a powerful and pervasive bias. *Obesity Reviews*, *4*(4), 213–227. https://doi.org/10.1046/j.1467-789X.2003.00122.x

Puhl, R. M., & Heuer, C. A. (2009). The stigma of obesity: A review and update. *Obesity*, *17*(5), 941–964. https://doi.org/10.1038/oby.2008.636

Puhl, R. M., & Latner, J. D. (2007). Stigma, obesity, and the health of the nation's children. *Psychological Bulletin*, *133*(4), 557–580. https://doi.org/10.1037/0033-2909.133.4.557

Pulos, E., & Leng, K. (2011). Evaluation of a voluntary menu-labeling program in full-service restaurants. *American Journal of Public Health, 100*(6), 1035-1039. https://doi.org/10.2105/AJPH.2009.174839

Ranganathan, P., Pramesh, C. S., & Buyse, M. (2015). Common pitfalls in statistical analysis: Clinical versus statistical significance. *Perspectives in Clinical Research*, *6*(3), 169-170. https://doi.org/10.4103/2229-3485.159943

Rasmussen, E. B., Lawyer, S. R., & Reilly, W. (2010). Percent body fat is related to delay and probability discounting for food in humans. *Behavioural Processes*, *83*(1), 23–30. https://doi.org/10.1016/j.beproc.2009.09.001

Reed, D. D., Kaplan, B. A., Becirevic, A., Roma, P. G., & Hursh, S. R. (2016). Toward quantifying the abuse liability of ultraviolet tanning: A behavioral economic approach to tanning addiction. *Journal of the Experimental Analysis of Behavior*, *106*(1), 93–106. https://doi.org/10.1002/jeab.216

Roberto, C. A., Larsen, P. D., Agnew, H., Baik, J., & Brownell, K. D. (2010). Evaluating the impact of menu labeling on food choices and intake. *American Journal of Public Health*, *100*, 312–318. https://doi.org/10.2105/AJPH.2009.160226

Roma, P. G., Hursh, S. R., & Hudja, S. (2016). Hypothetical purchase task questionnaires for behavioral economic assessments of value and motivation. *Managerial and Decision Economics*, *37*(4–5), 306–323. https://doi.org/10.1002/mde.2718

Rramani, Q., Krajbich, I., Enax, L., Brustkern, L., & Weber, B. (2020). Salient nutrition labels shift peoples' attention to healthy foods and exert more influence on their choices. *Nutrition Research*, *80*, 106–116. https://doi.org/10.1016/j.nutres.2020.06.013

Runge, C. F. (2007). Economic consequences of the obese. *Perspectives in Diabetes*, *56*(11), 2668–2672. https://doi.org/10.2337/db07-0633

Saunders, R. R., Saunders, M. D., Donnelly, J. E., Smith, B. K., Sullivan, D. K., Guilford, B., & Rondon, M. F. (2011). Evaluation of an approach to weight loss in adults with intellectual or developmental disabilities. *Intellectual and Developmental Disabilities*, *49*(2), 103–112. https://doi.org/10.1352/1934-9556-49.2.103

Schulte, E. M., & Gearhardt, A. N. (2017). Development of the modified yale food addiction scale version 2.0. *European Eating Disorders Review*, *25*(4), 302–308. https://doi.org/10.1002/erv.2515

Schüz, B., Schüz, N., & Ferguson, S. G. (2015). It's the power of food: individual differences in food cue responsiveness and snacking in everyday life. *International Journal of Behavioral Nutrition and Physical Activity*, *12*(1), 149. https://doi.org/10.1186/s12966-015-0312-3

Scott, K. M., Bruffaerts, R., Simon, G. E., Alonso, J., Angermeyer, M., De Girolamo, G., Demyttenaere, K., Gasquet, I., Haro, J. M., Karam, E., Kessler, R. C., Levinson, D., Medina Mora, M. E., Oakley Browne, M. A., Ormel, J., Villa, J. P., Uda, H., & Von Korff, M.

(2008). Obesity and mental disorders in the general population: Results from the world mental health surveys. *International Journal of Obesity*, *32*(1), 192–200. https://doi.org/10.1038/sj.ijo.0803701

Sebastian, R. S., Wilkinson Enns C., Goldman, J. D., Hoy, M.K., & Moshfegh, A. J., 2015. *Sandwich Consumption by Adults in the U.S.*: *What we Eat in America, NHANES 2009-2012.* Beltsville Human Nutrition Research Center: U.S. Department of Agriculture, Food Surveys Research Group. https://www.ars.usda.gov/ARSUserFiles/80400530/pdf/DBrief/14_sandwich_consumption_0912.pdf

Silk, K. J., Sherry, J., Winn, B., Keesecker, N., Horodynski, M. A., & Sayir, A. (2008). Increasing nutrition literacy: Testing the effectiveness of print, web site, and game modalities. *Journal of Nutrition Education and Behavior*, *40*(1), 3-10. https://doi.org/10.1016/j.jneb.2007.08.012

Snider, S. E., LaConte, S. M., & Bickel, W. K. (2016). Episodic future thinking: expansion of the temporal window in individuals with alcohol dependence. *Alcoholism: Clinical and Experimental Research*, *40*(7), 1558–1566. https://doi.org/10.1111/acer.13112

Stein, J. S., Koffarnus, M. N., Snider, S. E., Quisenberry, A. J., & Bickel, W. K. (2015). Identification and management of nonsystematic purchase task data: Toward best practice. *Experimental and Clinical Psychopharmacology*, *23*(5), 377–386. https://doi.org/10.1037/pha0000020

Stigler, G. J. (1954). The early history of empirical studies of consumer behavior. *Journal of Political Economy*, *62*(2), 95–113. https://doi.org/10.1086/257495

Strickland, J. C., Marks, K. R., & Bolin, B. L. (2020). The condom purchase task: A hypothetical demand method for evaluating sexual health decision-making. *Journal of the Experimental Analysis of Behavior*, *113*(2), 435–448. https://doi.org/10.1002/jeab.585

Strickand, J. C., & Stoops, W. W. (2019). The use of crowdsourcing in addiction science research: Amazon Mechanical Turk. *Experimental and Clinical Psychopharmacology, 27*(1), 1-18. https://dx.doi.org/10.1037/pha0000235

Stunkard, A. J., & Messick, S. (1985). The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. *Journal of Psychosomatic Research*, *29*(1), 71–83. https://doi.org/10.1016/0022-3999(85)90010-8

Tabachnick, B. G. & Fidell, L. S. (2013). *Using multivariate statistics (6th ed.)*. Boston: Pearson.

Talen, M. R., & Mann, M. M. (2009). Obesity and mental health. *Primary Care - Clinics in Office Practice*, *36*(2), 287–305. https://doi.org/10.1016/j.pop.2009.01.012

Taylor, A. W., Dal Grande, E., Gill, T. K., Chittleborough, C. R., Wilson, D. H., Adams, R. J., Grant, J. F., Phillips, P., Appleton, S., & Ruffin, R. E. (2006). How valid are self-reported height and weight? A comparison between CATI self-report and clinic measurements using a large cohort study. *Australian and New Zealand Journal of Public Health*, *30*(3), 238–246. https://doi.org/10.1111/j.1467-842X.2006.tb00864.x

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Yale University Press.

Tholin, S., Rasmussen, F., Tynelius, P., & Karlsson, J. (2005). Genetic and environmental influences on eating behavior: the Swedish Young Male Twins Study. *The American Journal of Clinical Nutrition*, *81*(3), 564–569. https://doi.org/10.1093/ajcn/81.3.564

Treuhaft, S., & Karpyn, A. (n.d.). *The grocery gap: who has access to healthy food and why it matters*. Retrieved April 28, 2020, from http://thefoodtrust.org/uploads/media_items/grocerygap.original.pdf

Tribble, A. G., Maxfield, A., Hadley, C., & Goodman, M. (2020). Food insecurity and mental health: A meta-analysis. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3520061

Tukey, J. W., & McLaughlin, D. H. (1983). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/winsorization 1. *Sankhyā: The Indian Journal of Statistics, Series A, 25*(3), 331-352.

Tucker, L. A., Seljaas, G. T., & Hager, R. L. (1997). Body fat percentage of children varies according to their diet composition. *Journal of the American Dietetic Association*, *97*(9), 981–986. https://doi.org/10.1016/S0002-8223(97)00237-X

Veillard, M. L., & Vincent, B. T. (2020). Temporal discounting does not influence body mass index. *Physiology and Behavior*, *221*, 112893. https://doi.org/10.1016/j.physbeh.2020.112893

Weller, R. E., Cook, E. W., Avsar, K. B., & Cox, J. E. (2008). Obese women show greater delay discounting than healthy-weight women. *Appetite*, *51*(3), 563–569. https://doi.org/10.1016/j.appet.2008.04.010

World Health Organization (n.d.). *Body mass index - BMI*. https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi

Wilson, A. G., Franck, C. T., Koffarnus, M. N., & Bickel, W. K. (2016). Behavioral economics of cigarette purchase tasks: Within-subject comparison of real, potentially real, and hypothetical cigarettes. *Nicotine and Tobacco Research*, *18*(5), 524–530. https://doi.org/10.1093/ntr/ntv154

Witt, A. A., & Lowe, M. R. (2014). Hedonic hunger and binge eating among women with eating disorders. *International Journal of Eating Disorders*, *47*(3), 273–280.

https://doi.org/10.1002/eat.22171

Appendix A

Demographics Questionnaire

1) What is your sex?
- Female
- Male

2) What is your current age? ______

3) What is your race?
- White/Caucasian
- African American
- Hispanic
- Asian
- Native American
- Pacific Islander
- Other

4) What is your current height?
- Feet ______
- Inches _____

5) What is your current weight? (lbs) ______

6) What is your primary language?

English	Dutch
Spanish	Japanese
Chinese	Hebrew
French	Swedish
German	Other (specify) ______

7) Please indicate your current annual income in U.S. dollars. Do not include financial aid. In other words, indicate your current annual income from jobs in which you get paid.
- Rather not say
- Under $10,000
- $10,000 - $19,999
- $20,000 - $29,999
- $30,000 - $39,999
- $40,000 - $49,999
- $50,000 - $74,999
- $75,000 - $99,999
- $100,000 - $150,000
- Over $150,000