Temptation and Choice Inconsistency

Samir Huseynov

Auburn University (Corresponding author, szh0158@auburn.edu)

Marco A. Palma

Texas A&M University

Ghufran Ahmad

Cardiff University

Rodolfo M. Nayga Jr.

Texas A&M University

Abstract

We examine food choice *inconsistency* vis-à-vis elicited Willingness-To-Pay (WTP) values using recent advancements in *menu-dependent preferences* and decision reference points. Previous theoretical and empirical studies predict the crucial impact of menu-induced visceral feelings and reference points on the inconsistency of choices. Our study employs a well-controlled and incentivized laboratory experiment to capture the role of normative utilities, temptation feelings, and decision reference points on the inconsistency of food choices. We document that 24% of food choices are inconsistent, and non-conventional menu-dependent preferences explain 40% of decisions. The salience of the external decision point reduces the frequency of temptation-driven preferences, albeit it does not affect food choice inconsistency. The intensity of temptation utility differences among food choice alternatives increases the decision inconsistency and reduces the welfare. Our study urges future studies to explicitly measure temptation feelings and model menu-dependent preferences when examining food choices, building policy insights, or empirical inferences based on elicited WTP measures. We also contribute to the theoretical menu-dependent preferences literature by showing its relevance and potential application venues in answering important empirical questions.

Keywords: Menu-Dependent Preferences, Set Betweenness, Willingness-To-Pay, Food Decisions.

JEL: C93, D03, I12.

1 Introduction

Standard economic models predict that a consumer will prefer the food product with a relatively higher *self-identified* Willingness-To-Pay (WTP) value to other available alternatives. Existing models do not account for the impact of choice-context-dependent visceral feelings (such as *temptation*) or attention-diluting effect of unavailable but salient items on food decisions.[1] What if these unaccounted factors are important and can lead to favoring an alternative that the very same decision-maker has previously assigned a lower decision value? These decision reversals or *inconsistencies* may invalidate behavioral models and consequently diminish the predictive power of WTP estimates.

A vast experimental literature on consumer decision-making relies on the stability of elicited Willingness-To-Pay measures. Preferring a higher valued alternative in terms of the self-revealed WTP scale—the canonical assumption of standard models—is at the core of studies estimating market shares of competing products, the adoption perspective of new technologies, the impact of label information on food choices, and policy implications.[2] Preferring a product with a lower self-identified WTP value, i.e., exhibiting *inconsistency*, has been regarded as random mistakes (Brouwer et al., 2010), and linked to choice complexity (Rigby et al., 2016). However, recent studies in the decision theory literature document systematic deviations from self-identified preferred alternatives in repeated and non-temporal choices (Agranov and Ortoleva, 2017; Fudenberg et al., 2018). A seminal paper by Gul and Pesendorfer (2001) shows that decision inconsistencies can be more prevalent in the food choice domain due to self-control issues. This new literature explains inconsistent choices with *menu-dependent* visceral feelings induced by the decision environment. Economic choices, inconsistent with respect to self-revealed product valuations, not only reduce the predictive importance of different WTP measures but also invalidate inferences built on them. The identification of underlying menu features leading to inconsistent choices can enable us to better understand the sources of this phenomenon and develop remedies to improve the external validity of WTP elicitation techniques. This article contributes to the decision inconsistency literature by modeling both canonical and non-canonical preference types in food decisions and develops an incentivized experimental procedure to showcase the role of temptation in inconsistent choices.[3]

[1]Facing unavailable menu options has especially become a widespread issue due to the COVID-19 and consequent disruptions in the food supply chain (Reardon et al., 2020).

[2]Examples include Alfnes and Rickertsen (2003), Carlsson et al. (2007), Pappalardo and Lusk (2016), and Costa-Font et al. (2008).

[3]Some recent studies also document decision inconsistencies/mistakes in different domains, such as risk-taking and

We follow a well-established experimental protocol to empirically document *decision inconsistency* and study its determinants.[4] Our design has two stages: 1) Independently eliciting WTP values for each food snack product using a novel incentivized procedure, 2) Pairing products in menus—in line with the model of Gul and Pesendorfer (2001)—to elicit *menu-dependent preferences*, temptation, and choice inconsistencies. In our design, the first stage is crucial in establishing *normative* decision values, and hence, identifying inconsistent food decisions vis-à-vis elicited WTP values in the second stage of the experiment.[5] Independently measuring individual WTP values also enables us to examine how the menu environment affects food decisions and causes behaviors contradicting self-revealed product valuations. Unlike some previous studies, we deliberately do not re-assign first-stage decision tasks in the second stage of the study to capture decision reversals (Agranov and Ortoleva, 2017). Repeating experimental tasks can document decision reversals, but this method is not sufficient to observe the process determinants of choice inconsistencies. Instead, we follow developments in the elicitation techniques of menu-dependent preferences to examine fundamental behavioral pillars of choice inconsistencies (Toussaert, 2018).

We manipulate decision reference points by introducing an unavailable menu item in treatment conditions. We also employ eye-tracking technology to measure visual attention to presented menus. Eye-tracking data help develop a *compliance* measure for each subject and examine the treatment effect without relying on the intent-to-treat approach. The *Conceptual Framework* section lays out *menu-dependent preferences* and associated models. We also explain the connections among recent menu-dependent decision models and how those models fit our research framework.

We conduct an incentivized laboratory study with 170 undergraduate students.[6] The experiment consists of two stages with an equal likelihood of being binding at the end of the study. In the first stage, participants are endowed with ten tokens, and each token gives subjects a 10% chance to receive the desired reward. In each task, participants are separately shown ten food snacks on their screens across ten decision tasks, and they are asked to allocate the available ten tokens between the displayed snack food or a one-dollar bill reward. For instance, if a subject allocates three tokens

lottery choices. However, those studies are silent on the process factors leading to behavioral inconsistencies (Nielsen and Rehbeck, 2020).

[4] Consult Fudenberg et al. (2018) for the appropriateness of similar research designs in examining the sources of economic choice inconsistencies.

[5] We define normative decision values using the modelling framework of Gul and Pesendorfer (2001) and explain details in the Experimental Procedures and Measures section.

[6] We used G*Power software to perform a power analysis. The results show that for capturing a mean difference in a between-subject design (the medium effect size; d=0.5), we need at least 51 subjects in each experimental condition to achieve the 80% power at the 10% statistical significance level.

to the displayed food snack, she opts in for a lottery, which has a 30% probability of yielding the food snack and a 70% likelihood of rewarding \$1.00. Corner allocations (i.e., allocating ten or zero tokens) guarantee the outcome with 100% probability (or 0% if no token is allocated) for the allocated alternative.[7] Participants are informed that if this stage is binding at the end of the study, one of the ten decisions will be randomly selected, and the associated lottery will be realized. This novel WTP elicitation mechanism helps measure participants' monetarily denominated preference strengths for each food product.

We construct 45 unique menus in the second stage using all the possible pairs from the ten food snacks presented during the first stage. Each menu consists of three sub-menus: two singleton sub-menus separately consisting of each alternative and one sub-menu that contains both products in the pair. Assume that the pair consists of Oreo Double Cookies and Ritz Bits. Then the associated second-stage menu will include A {Oreo Double Cookies}, B {Ritz Bits}, and C {Oreo Double Cookies and Ritz Bits} sub-menus. Having the sub-menu C and first-stage WTP values allow eliciting menu-dependent preferences (No Temptation, Overwhelming Temptation, Resistible Temptation, Flexibility, Guilt, and Indifferent) constructed by Gul and Pesendorfer (2001).[8] Food decisions not conforming to modeled menu-dependent preferences are inconsistent as those choice instances contradict elicited WTP values.[9] Subjects are asked to rank the displayed sub-menus in each menu and are informed that higher-ranked menus have a greater probability of being realized at the end of the experiment if the second stage and the menu are randomly selected to be binding. We employed probability structures used by Toussaert (2018) in converting ranking orders into menu-realization probabilities.[10] All food snacks were available at the lab. Subjects could only redeem their participation fees after completely eating their snack selection for the binding round.

Subjects are randomly assigned to treatment conditions (displaying an unavailable fourth sub-menu) and one control (only displaying available sub-menus). The purpose of the unavailable sub-menus was to exogenously change subjects' reference points in the menus and thus to observe the role of the external reference point in the inconsistency of food choices. Our treatment condition

[7] The primary logic behind this novel WTP elicitation technique is finding the p probability that satisfies $U(Food_i) = p * u(\$0.00) + (1 - p) * u(\$1.00)$ where $p =$ (number of tokens allocated to $Food_i$)/10. BDM and similar WTP elicitation techniques can be expressed as finding the lottery equivalent of the utility an individual derives from a product (see Healy (2018) for examples).

[8] We follow Kopylov (2012) and Gul and Pesendorfer (2001) in defining Guilt. Flexibility was constructed based on Gul and Pesendorfer (2001) and Toussaert (2018).

[9] We explain or approach in details in the Conceptual Framework section.

[10] The Experimental Procedures and Measures section describes ranking-based realization probabilities and other details of the experimental protocols.

is motivated by recent developments in the reference-point literature showing the importance of reference points in formation of the decision values of alternatives (Kıbrıs et al., 2018; Heydari, 2021; Caputo et al., 2020). Out-of-stock merchandises when shopping online or expectations of an introduction of a new product in the market serve as a motivation for our treatments. We conjecture that one of the sources of the choice inconsistency can be a change in the reference point of the decision context. Here we define the reference point as the salient item in the choice set in line with Kıbrıs et al. (2018) and Bhatia and Golman (2019). We display highly-tempting unavailable cakes that stand out from other available alternatives in the treatment conditions. Subjects made food decisions in the presence of eye trackers. Not all subjects pay the same visual attention to the unavailable cakes. We measure the average fixation time subjects spent on the cake stimuli to split our treatment sample into Low and High Compliance External Reference conditions. Measuring visual attention to the unavailable sub-menu enables us to link the reference point to the visual salience and scrutinize the effect of our treatment conditions via focusing on actually treated subjects. The *Experimental Procedures and Measures* section discusses our experimental design and protocols. This section also provides detailed information for our primary outcome measures: *Choice Consistency* and *Expected Loss*. We also spell out our hypotheses in this section.

The primary outcome of interest from our experiment is a binary inconsistency measure indicating whether the subject assigned a ranking order to sub-menus contradicting WTP estimates and associated menu-dependent preferences. We find that, on average, 24% of second-stage menu rankings are inconsistent with the elicited first-stage WTP measures. We also document that only 4% of decisions conform to standard No Temptation preferences in our sample, showing the importance of explicitly modeling and capturing menu-dependent preferences in WTP studies.

We do not detect any difference between the External Reference treatment conditions (External-Reference Low and High Compliance conditions) and the No Reference Control in terms of the proportion of inconsistent food choices. The null treatment effects indicate that the external reference point does not dilute the consistency of food choices. Facing a menu-ranking problem with one point higher temptation value difference (based on a Likert scale) between alternatives increases the percentage of inconsistent decisions by nine p.p. This result is in line with recent studies showing a higher temptation distance among food alternatives can reduce the proportion of normatively superior low-calorie choices (Huseynov et al., 2021).

We also construct a measure for the net expected WTP value loss (*expected loss*) from menu-rankings based on the probabilities assigned by identified preference structures (i.e., No Temptation, Resistible Temptation, Overwhelming Temptation, Guilt, Flexibility, and Indifference preferences). The difference between the net expected WTP value and the standard No Temptation preference baseline (provided that the decision-maker is consistent in the food decision) yields the net expected welfare loss for each decision due to exhibiting menu-dependent preferences. This measure enables us to estimate the welfare effect of being inconsistent in food choices. We do not detect an economically meaningful treatment effect on expected loss levels suggesting that external references also do not inflict welfare losses. Interestingly, we find that one point increase in the temptation value difference between alternatives also increases the net expected loss by around 40 cents. We conclude that a higher temptation value difference increases the proportion of inconsistent decisions and also elevates the welfare cost of making inconsistent decisions. The *Results* section discusses the further details of our findings.

In general, our findings show that the external reference point and its salience do not affect the food choice inconsistency and the magnitude of expected loss in net WTP choice values. In our sample, 24% of food decisions are inconsistent, and this proportion is very close to the recent studies using conventional discrete choice experiments with repeated tasks (Segovia and Palma, 2021; Fraser et al., 2021). We conclude that being consistent and adhering to self-identified WTP values significantly depend on the magnitude of temptation distance among food choice alternatives. Therefore, prospective WTP studies should account for choice-context-dependent visceral feelings in fine-tuning the predictions based on elicited consumer food valuations. While we do not detect the influence of external reference points on the consistency of food choices, we think that this result is suggestive. Future studies may explore other experimental design features to further examine the role of reference points on decisions.

Our findings are not exhaustive and do not explain *all* potential behavioral mechanisms triggering WTP inconsistencies. We view this study as an important first step in a new literature branch exploring behavioral underpinnings of decision inconsistency and its applications in the food choice domain. Our primary contribution includes offering an empirical framework supported by the novel set betweenness theoretical model to capture food choice inconsistencies and identifying the impact of two behavioral biases on decisions. We also show that food choices are intense with visceral feelings, and temptation can trigger food choice inconsistencies. We urge prospective studies to devise survey techniques for measuring *temptation* and accounting for its impact on the (in)consistency of WTP measures.

2 Conceptual Framework

2.1 Set Betweenness and Temptation

Let $A = \{a_1, a_2, \cdots, a_n\}$ be a finite set of food alternatives. Then, every $x \subseteq A$ is referred to as a menu of alternatives. Let $\succeq$ represent the preferences over the menus with $\succ$ and $\sim$ denoting the asymmetric and symmetric parts, respectively. Gul and Pesendorfer (2001) present a model that captures *temptation* faced in menu choices. Temptation is captured by the imposition of *set betweenness* which states the following:

$$x \succeq y \Longrightarrow x \succeq x \cup y \succeq y \tag{1}$$

Set betweenness admits the following preferences:

1. Standard No Temptation preferences: $x \sim x \cup y \succeq y$, and
2. Temptation: $x \succ x \cup y$,
   (a) Overwhelming Temptation (Temptation preferences): $x \cup y \sim y$, or
   (b) Resistible Temptation (Self-control preferences): $x \cup y \succ y$.

For standard No Temptation preferences, the agent only cares for the highest-valued alternative in each menu and, thus, she is indifferent between the menus x and $x \cup y$. The agent faces temptation from menu y when $x \succ x \cup y$ as she prefers facing menu x instead of menu $x \cup y$. This temptation is stated to be overwhelming (resistible) when $x \cup y \sim y$ ($x \cup y \succ y$) because the agent succumbs to (overcomes) temptation from the menu y. Under their axioms, Gul and Pesendorfer (2001) prove the following representation for menus $x \subseteq A$:

$$U_{GP}(x) = \max_{a \in x} \left[u(a) + v(a)\right] - \max_{b \in x} v(b) \tag{2}$$

In equation (2), $u(\cdot)$ represents the utility of a singleton alternative, i.e., the utility without exposure to multiple alternatives, which is referred to as the *normative utility.* On the other hand, $v(\cdot)$ represents the temptation of an alternative, which is pertinent when she faces at least two alternatives in the menu, and is referred to as the *temptation utility.*

As our experimental design focuses on binary menus, suppose that $\{a\} \succeq \{a, b\} \succeq \{b\}$ for alternatives $a, b \in A$. The representation in equation (2) admits standard preferences when $v(a) =$

$v(b)$. Overwhelming Temptation (i.e., Temptation preferences) arises when $v(b)-v(a) > u(a)-u(b)$, i.e., the gain in temptation utility by succumbing to temptation exceeds the gain in normative utility from committing to alternative a. Meanwhile, Resistible Temptation (i.e., Self-control preferences) arises when $v(b)-v(a) < u(a)-u(b)$, i.e., the gain in temptation utility by succumbing to temptation is not enough to overcome the gain in normative utility from staying committed to alternative a. Moreover, following Lemma 4 of Gul and Pesendorfer (2001), the temptation utilities can be calculated as $v(a) = U_{GP}(\{a, b\}) - U_{GP}(\{a\})$ and $v(b) = 0$, respectively.

Noor and Takeoka (2015) relax set betweenness to binary set betweenness which is defined as:

$$\{a\} \succeq \{b\} \Longrightarrow \{a\} \succeq \{a, b\} \succeq \{b\} \tag{3}$$

Based on their axioms, Noor and Takeoka (2015) prove the following utility representation for menu $x \subseteq A$:

$$U_{NT}(x) = \max_{a \in x} \left[u(a) - \psi \left(\max_{b \in x} v(b) \right) \left(\max_{b \in x} v(b) - v(a) \right) \right] \tag{4}$$

where $\psi(\cdot) > 0$ is a weakly increasing and continuous function. This representation allows the impact of self-control cost, $\max_{b \in x} v(b) - v(a)$, to vary with the menu by allowing $\psi(\cdot)$ to scale up or down. However, in the presence of only singleton and binary menus, binary set betweenness and set betweenness are identical. Therefore, in such a setting, menu-dependence of self-control acts is not pertinent (Noor and Takeoka, 2015).

2.2 Guilt and Temptation

Consider menus $x, y \subseteq A$. Two preference structures can arise when set betweenness is not satisfied; flexibility ($x \cup y \succ x, y$) and guilt ($x, y \succ x \cup y$).

Dekel et al. (2009) define positive and negative set betweenness by relaxing set betweenness, that permits guilt and flexibility, respectively. If $x \succeq y$, then positive set betweenness implies that $x \succeq x \cup y$ whereas negative set betweenness implies that $x \cup y \succeq y$. Positive set betweenness allows guilt as it does not restrict the preference relation between y and $x \cup y$. Similarly, negative set betweenness does not impose any preference relation between x and $x \cup y$, thus, allowing for flexibility. Figure 1 shows the relationship between positive set betweenness, negative set betweenness, and set betweenness. The set betweenness (SB) axiom defined by Gul and Pesendorfer (2001) is embedded in positive (SB+) and negative (SB-) set betweenness. In other words, set betweenness is the intersection of positive and negative set betweenness.

For positive set betweenness, Dekel et al. (2009) prove the no-uncertainty representation which is given as:

$$U_{\mathrm{NU}}(x)=\max_{a\in x}\left[u(a)+\sum_{j=1}^{J}\left\{v_j(a)-\max_{b\in x}v_j(b)\right\}\right] \tag{5}$$

where $j=1,\cdots,J$ represents the various temptation channels for the alternative. However, the representation in equation (5) permits guilt when there are at least two dimensions through which the alternatives can tempt the agent, i.e., $J\geq 2$, because equation (5) reduces to equation (2) when $J=1$. Moreover, for singleton and binary menus, guilt can arise only if, for $\{a\}\succeq\{b\}$, alternative b is not more tempting for all dimensions, as shown in Appendix A.1. Therefore, an agent faces guilt when she faces temptation from both of the alternatives in different dimensions. In such a scenario, she prefers singleton menus to the binary as regardless of what she chooses from the binary menu, she will face some degree of self-control costs.

For negative set betweenness, Dekel et al. (2009) prove the uncertain strength of temptation representation which is given as:

$$U_{\mathrm{US}}(x)=\sum_i q_i\max_{a\in x}\left[u(a)+\gamma_i\left\{v(a)-\max_{b\in x}v(b)\right\}\right] \tag{6}$$

where $q_i>0$ for each i with $\sum_i q_i=1$ and $\gamma_i\geq 0$ for all i. In this representation, γ_i represents the strength of temptation which occurs with a probability of q_i. When there is only one possible strength of temptation, $q_i=1$, equation (6) reduces to equation (4) and to equation (2) when additionally $\gamma_i=1$. Moreover, for singleton and binary menus, flexibility suggests that $\gamma_i[v(b)-v(a)]>u(a)-u(b)$ for some i (Appendix A.2). In simpler words, if the agent has a preference for flexibility, then she expects that there is at least one possible instance where she will succumb to temptation.

Kopylov (2012) provides another generalization of set betweenness which is referred to as perfectionist set betweenness. Perfectionist set betweenness is only imposed on those sets that have the same best normative alternative and is formally defined as:

$$x\succeq y\Longrightarrow\{a\}\succeq x\succeq x\cup y\succeq y \tag{7}$$

where a is the shared best normative alternative in menus x and y, i.e., $\{a\}\succeq\{b\}$ for each

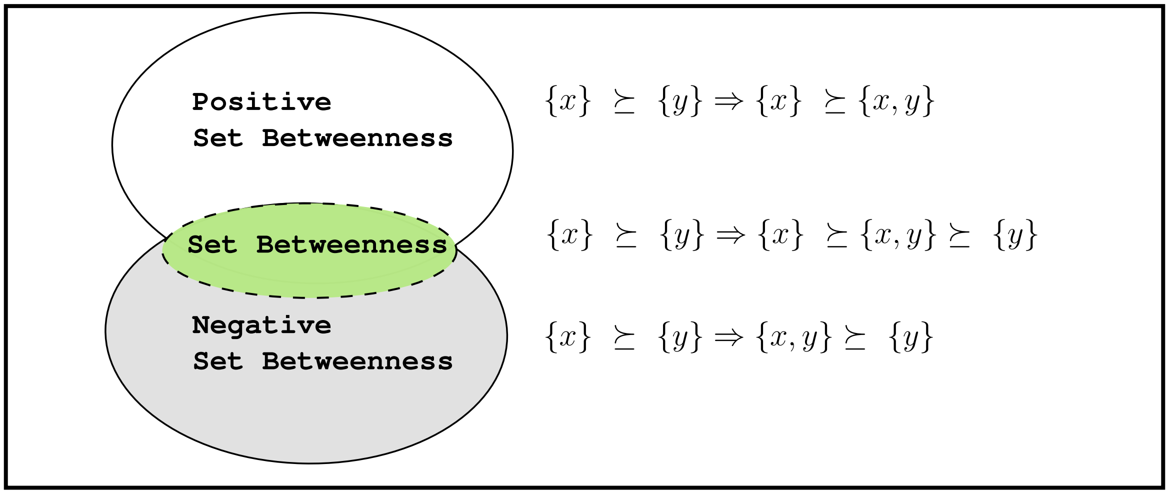


Figure 1: The relationship between Positive Set Betweenness, Negative Set Betweenness and Set Betweenness

$b \in x \cup y$. Under perfectionist set betweenness and other axioms, Kopylov (2012) proves the following representation:

$$U_{\mathrm{K}}(x) = \max_{a \in x} \left[u(a) - \left\{ \max_{b \in x} v(b) - v(a) \right\} \right] + \kappa \max_{c \in x} u(c) \tag{8}$$

where $\kappa > -1$. Note that equation (8) reduces to equation (2) when $\kappa = 0$. When $\kappa \neq 0$, flexibility and guilt can arise in menu preferences. Flexibility cannot arise when we are limited to singleton and binary menus (Appendix A.3). However, guilt can arise in this restricted setting. Suppose that $\{a\} \succeq \{b\}$. If $u(a) + v(a) \geq u(b) + v(b)$, then the menu preferences depict guilt if b is more tempting than a and κ is sufficiently small. Moreover, if $u(a) + v(a) < u(b) + v(b)$, then the menu preferences depict guilt if $u(a) > u(b)$ and $\kappa < 0$ (Appendix A.3).

2.3 Identifying Menu-Dependent Preference Types

We identify menu-dependent preference types based on the second-stage menu rankings and the first-stage WTP values. Table 1 shows menu-dependent preferences and associated menu rankings, WTP comparisons, and Set Betweenness types. We focus on Resistible Temptation (a decision-maker incurs a self-control cost but can resist the temptation when facing the combined sub-menu with both alternatives), Overwhelming Temptation (a decision-maker incurs a self-control cost but cannot resist temptation when facing the combined sub-menu with both alternatives), Guilt (a decision-maker avoids the combined sub-menu, hence dislikes trade-offs), and Flexibility (a decision-

Table 1: Menu Rankings and Menu-Dependent Preferences

WTP values	Ranking of Menus	Set Betweenness Types	Preference Types
$WTP(a)$≥WTP(b)	{a} = {a,b} > {b}	Set Betweenness: {a}≥ $\{a,b\} \geq \{b\}$	Standard No Temptation
$WTP(a)$≥WTP(b)	{a} > {a,b} > {b}	Set Betweenness: {a}≥ $\{a,b\} \geq \{b\}$	Resistible Temptation preferences
$WTP(a)$≥WTP(b)	{a} > {a,b} = {b}	Set Betweenness: {a}≥ $\{a,b\} \geq \{b\}$	Overwhelming Temptation preferences
$WTP(a)$≥WTP(b)	{a} > (or =) {b} > {a,b}	Positive Set Betweenness: {a}≥ $\{a,b\}$	Guilt preferences
$WTP(a)$≥WTP(b)	{a,b} > {a} > (or =) {b}	Negative Set Betweenness: {a,b} ≥ $\{b\}$	Flexibility preferences

maker always ranks the combined sub-menu first, hence prefers larger sets). In its turn, Standard No Temptation preferences can be identified when a decision-maker prefers one alternative and does not incur self-control costs when facing the combined sub-menu.[11] Flexibility (Guilt) arises when the negative (positive) set betweenness axiom holds but set betweenness does not hold: SB-/SB (SB+/SB). If menu orderings combined with the first stage preferences do not comply with SB, SB+, or SB-, we label the choice preference type as Non-SB or inconsistent since the choice pattern does not conform to our modeled preferences (including No Temptation preferences) and WTP values. For instance, choice inconsistency occurs when a decision-maker allocates four (three) tokens to product a (product b) in the first stage of the experiment and has the $Menu_b > Menu_a$ menu ordering (independent of the ranking of the combined sub-menu) in the second stage. Notice that in the case of the $Menu_b = Menu_a$ ordering, the decision-maker is still consistent, as the product with a higher token allocation is not ranked lower than the available alternative. Inconsistent decisions are in contradiction with both set betweenness and menu-dependent preferences (including the Standard No Temptation preferences).

3 Experimental Procedures and Measures

3.1 Protocols

A total of 170 undergraduate students of a university located in the South-Western United States participated in the study.[12] The experiment took place in the Fall of 2019.[13] Bulk recruitment

[11]The Indifference ranking preference arises when $\{a\} = \{a,b\} = \{b\}$.

[12]The raw total number of our recruited subjects was 218. A larger sample size enabled us to eliminate subjects who had zero WTP valuations for all food snack products or did not pay attention to product stimuli when completing menu rankings. Out of 218 subjects, 21 participants were eliminated from the analyses for not having non-zero token allocations to food snacks. We also excluded 27 subjects from our analysis sample for failing to sustain their attention in the second stage of the study.

[13]Experimental protocol and procedures had been approved prior to the implementation of experimental studies under the IRB2017-0011D application.

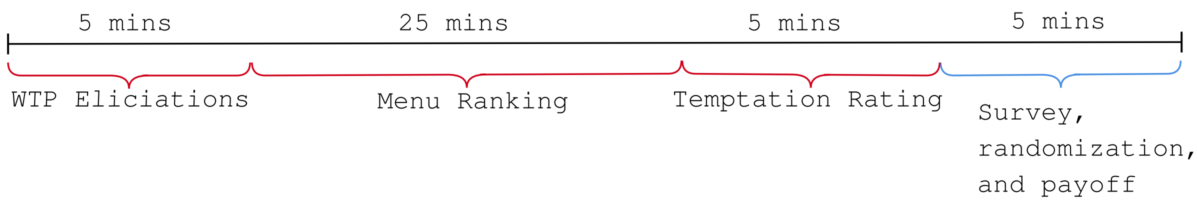


Figure 2: Timeline of the Experimental Procedures

emails were sent out to university-maintained email lists providing general information about the study. Students without known food allergies and restrictions were solicited to participate in the experiment involving real food choices. Participants were also informed that they would be eligible to earn a $15.00 participation fee for spending around 40 minutes in the laboratory following study protocols. The recruitment email also detailed that subjects could earn 1.00 USD or a snack food product in addition to the participation fee. Snack foods had to be consumed in the lab to be eligible to redeem the compensation.[14]

We conducted experimental sessions between 8.00 a.m. and 5.00 p.m., randomly assigning subjects to treatment and control conditions to offset the time-of-the-day effect. After arriving at the lab and signing the consent form, subjects were escorted to the experiment room. The session monitor read aloud general instructions and explained the procedures to determine the binding stage and decision. Participants were informed that the study would progress in two incentivized stages following a block with an additional task and a brief demographic survey. Instructions also conveyed that at the end of the study, using physical tokens and a bingo cage, one of the two stages of the experiment, and one decision would be randomly selected as binding.

Appendix B presents further details, including the list of snack food products and screenshots from Stage 2 of the experiment across treatment and control conditions. Stage 1 was identical for all experimental conditions. Subjects were independently shown snack products on the screen and asked to allocate endowed ten tokens between 1.00 USD and the snack product. The primary logic behind this WTP elicitation technique is identifying the p probability that satisfies $U(Food_i) = p_i * u(\$0.00) + (1 - p_i) * u(\$1.00)$ where $p =$ (number of tokens allocated to $Food_i$)/10. It has been shown that BDM and similar WTP elicitation techniques can be expressed as finding the lottery

[14] We selected snack food products available across campus food vending machines. We included a 1.00 USD bill as the alternative in the first stage of the study to prevent arbitrage opportunities as the approximate average magnitude of the prices of the selected snack products was around that amount. Using relatively well-known food snack products also helped us capture participants' product valuations by isolating potential issues stemming from being unaware of market prices.

equivalent of the utility an individual derives from a product (see Healy (2018) for an extensive discussion on this). For two food items, where $i \in \{1, 2\}$, $p_1 > p_2$ means $U(Food_1) > U(Food_2)$. Hence, $WTP_{Food_1} > WTP_{Food_2}$ when an individual allocates a higher amount of tokens to $Food_1$ compared to $Food_2$. We also assume that the utility function satisfies quasi-linearity properties.

Treatment conditions were introduced in Stage 2. Treatments included displaying unavailable sub-menu D that included a Nothing Bundt Cake with 580 or 710 calories. Appendix B shows that we do not detect any behavioral differences between treatment conditions that only differed across calorie contents (See Table B1 and B2).[15] Therefore, we combine treatment conditions and split the treated subjects into Low and High Compliance groups based on the total fixation time they spent fixating on sub-menu D.[16] Our result section is primarily based on comparing External-Reference Low Compliance, External-Reference High Compliance treatments, and the No Reference control conditions.

After the second stage of the study and before determining the binding stage and decision, subjects were separately shown each food snack on their screens and asked to indicate their temptation feelings to each item using 0 to 10 scale. Subjects concluded the study answering to demographic survey questions.

After the conclusion of each session, one subject volunteered to determine the binding stage and decision. The subject drew a token out of the bag containing two tokens, each separately indicating experimental stage numbers. Then depending on the binding stage, the subject spun the bingo cage with adequately numbered balls representing the decision tasks in that stage.[17] The randomly selected stage and decision task were binding for all subjects in the same session.

The session monitor approached each subject and conducted an additional randomization procedure after determining the binding decision. If the binding stage was the first stage of the experiment, subjects were asked to draw a token out of the bag containing ten tokens numbered from 1 to 10. For instance, if a subject allocated three tokens to the snack food product in the binding decision, drawing a token with the number of 1, 2, or 3 rewarded the subject with the product. Tokens with 4, 5, 6, 7, 8, 9, and 10 made the subjects eligible for an additional 1.00 USD reward.

Similar procedures were realized if the binding stage was the second stage of the experiment.

[15] The only statistical difference is detected in the proportion of male subjects across treatment conditions and we control for this variable in our analyses.

[16] We use the median value of the distribution to achieve close number of participants in each treatment condition.

[17] 10 (45) balls for the first (second) stage of the study.

Table 2: Menu Rankings and Assigned Probabilities

Ranking of sub-menus	Realization Probabilities
$Menu_1 > Menu_2 > Menu_3$	50%, 30%, 20%
$Menu_1 = Menu_2 > Menu_3$	40%, 40%, 20%
$Menu_1 > Menu_2 = Menu_3$	50%, 25%, 25%
$Menu_1 = Menu_2 = Menu_3$	33.3%, 33.3%, 33.3%

Each sub-menu was assigned a different realization probability depending on menu rankings following Toussaert (2018). Table 2 shows all possible menu rankings and assigned probabilities. For instance, the {a} > {a,b} > {b} menu ranking assigned 50%, 30%, and 20% realization probabilities for sub-menu $\{a\}$, sub-menu $\{a, b\}$ and sub-menu $\{b\}$, respectively. In this case, drawing a token numbered with 1, 2, 3, 4, or 5 rewarded the subject with sub-menu $\{a\}$. In the same fashion, tokens numbered 6, 7, or 8 (9 or 10) rewarded sub-menu $\{a, b\}$ (sub-menu $\{b\}$). After the completion of the randomization, each subject was sent to the lab's lobby to collect rewards.[18]

Assigned probabilities based on ranking orders are incentive-compatible when a decision-maker has strict ranking orderings (Toussaert, 2018). However, when a subject is indifferent between menus, she might prefer any probability assignment over menus. Toussaert (2018) also introduces a WTP assignment to revise ranking orders to link menu rankings to subjective WTP values. Therefore, in the primary analysis of this study, we only focus on choice instances where WTP differences between food alternatives are different than zero.

We elicit incentivized WTP values for each snack food product before menu orderings, and our primary research question is about studying the inconsistency between WTP measures and

[18]Only in two sessions (18 subjects in total), the binding sub-menu was $\{a, b\}$ (i.e., the combined menu), meaning subjects had to make one more choice to select their preferred snacks. Toussaert (2018) mainly focused on choices from the set betweenness in examining the role of menu-dependent preferences in temptation-intense decision menus. Therefore, Toussaert (2018) informed subjects that menu ranking induced realization probabilities (described in Table 2) would be used with the 50% likelihood. In the remaining cases, subjects would face the sub-menu c or $\{a, b\}$, independent of their menu rankings. We opted out from this method to make menu rankings shown in Table 2 salient to subjects. We also conjectured that having another lottery before the menu ranking induced realization probabilities could overwhelm study participants and they could rank menus without considering the information presented in Table 2. Subjects with the binding $\{a, b\}$ menu were sent to the lobby and were instructed to make their final selections there, not in the presence of the session monitor. Those participants made their final decisions in the lobby and claimed their preferred food snacks from the laboratory personnel who were in charge of monitoring the compliance (eating the chosen snack) and distributing study payoffs.

We did not record choices from the binding $\{a, b\}$ menu in the lobby, as the number of those cases were expected to be very small without yielding any statistically reliable conclusions. As mentioned above, Toussaert (2018) introduced another lottery to increase the number of cases with the binding $\{a, b\}$ sub-menu to be able to construct a reliable statistical inference.

ranking tasks. Eliciting subjects' WTP values for each food product enables us to explicitly capture inconsistent food choices based on second-stage menu rankings. Therefore, we focus on second-stage menus to examine the consistency of decisions where participants have non-zero WTP value differences between food products. The following subsection discusses the details of our food choice consistency measure.

3.2 Measuring Consistency

We measure $u(x)$ for $x \subseteq A$ using independent token allocations in the first stage of our experiment. When a food product is the only alternative in the menu, menu-dependent preferences and the associated utility function yield a normative utility. This can be easily illustrated using the utility function by Gul and Pesendorfer (2001) for *alternative* a in a singleton menu:

$$U_{GP}\left(x\right) = \max_{a \in x}\left[u\left(a\right) + v\left(a\right)\right] - \max_{a \in x} v\left(a\right) \tag{9}$$

Equation 9 yields $U_{GP}\left(x\right) = u\left(a\right)$ showing our first stage WTP measure is a proxy for normative utility of snacks. The primary intuition behind this assertion is that subjects can increase their chance of earning the snack product by increasing the number of token allocations for the food item. Therefore, $u\left(a\right)$ for each product is represented by the number of allocated tokens, hence with our WTP measure.

First stage $WTP(x)$ token allocations for $x \subseteq A$ represent the strength of subjects' normative utility preferences for each snack product. If $WTP(a) > WTP(b)$ for a subject, then we also expect to observe the singleton sub-menu containing *alternative* a to have a higher or the same ranking compared to the singleton sub-menu containing *alternative* b ($Menu_a \geq Menu_b$) in the second stage of the study. Since a higher ranking order increases the realization likelihood of a sub-menu, a consistent subject with a higher number of token allocations for product a compared to product b (i.e., $WTP(a) > WTP(b)$) is expected to exhibit $Menu_a \geq Menu_b$. Similarly, the subject is inconsistent when the $Menu_a < Menu_b$ ordering in the second stage contradicts with $WTP(a) > WTP(b)$ in the first stage.

If subjects have standard preferences, the existence of the combined sub-menu in the second stage does not introduce any additional information about preferences. If a consistent subject demonstrates a menu ranking ordering in the second stage adhering standard preferences, then

$WTP(a) > WTP(b)$ should lead to the $Menu_{a,b} = Menu_a > Menu_b$ ordering. This menu ordering in the second stage indicates that the consistent subject does not differentiate the combined sub-menu and the singleton sub-menu containing the alternative with a higher WTP value. Hence the subject does not have any visceral feelings or menu-dependent preferences for the alternative b.

The first stage $WTP(a) > WTP(b)$ and the second stage $Menu_{a,b} > Menu_b > Menu_a$ ordering indicate that the subject is inconsistent, while the decision-maker ranks the combined sub-menu with alternative a higher than other singleton sub-menus. In this case, the consistency measure still weakly stands, as a consistent decision-maker with $WTP(a) > WTP(b)$ should have exhibited the $Menu_a \geq Menu_b$ ordering no matter of the relative order of the combined sub-menu. It is also possible that with the $Menu_{a,b} > Menu_b > Menu_a$ ordering and $WTP(a) > WTP(b)$, a inconsistent subject still increases the chance of concluding the study with the binding combined sub-menu and choosing the alternative a. But at the same time, the decision-maker reduces the realization likelihood of $Menu_a$ with the product having a higher WTP value compared to the alternative b, which contradicts with the incentivized first stage outcomes.

If $WTP(a) = WTP(b)$ in the first stage of the study, then the decision-maker will be indifferent to any lottery over the alternatives (Toussaert, 2018). Therefore, our consistency measure only focuses on cases when $WTP(a) \neq WTP(b)$.

3.3 Measuring Expected Loss

Second stage decisions also generate expected WTP values based on ranking decisions since subjects also know how their second-stage menu orderings translate into realization probabilities. Being inconsistent in choices does not punish the decision-maker immediately but reduces the expected net WTP value of the decision. Moreover, the expected cost of being inconsistent depends on the self-assigned WTP values (i.e., normative utilities) to choice alternatives. Inconsistent decisions won't inflict a high expected welfare loss if the WTP value difference between alternatives is very small. But the expected cost of making inconsistent choices increases as the normative utility difference between menu alternatives becomes greater in magnitude.

If alternative a has a higher self-assigned WTP value than alternative b, then a consistent second-stage menu-ranking decision adhering to standard No Temptation or Resistible Temptation preferences yields the highest expected WTP value. However, the expected WTP value decreases when a subject shows overwhelming temptation preferences while even being consistent in menu orderings. Table 3 shows how we calculate expected WTP values based on second-stage menu rankings.

We assume that decision-makers can still choose one of the menu alternatives in combined sub-menus when they exhibit flexibility and guilt preferences. Therefore, we assume that their choice probabilities are 50% each. It is true that the singleton sub-menu for alternative a has a higher rank order compared to the singleton sub-menu for alternative b in the $\{a,b\} > \{a\} > \{b\}$ menu ordering. It may suggest that we should not assign equal realization likelihoods for the combined sub-menu alternatives for this menu ordering. However, a decision-maker could have substantially increased the realization probabilities of the alternative a by avoiding Flexibility preferences and adhering to standard No Temptation or resistible temptation preferences (See Table 3 for the relationship between expected net WTP values and menu-dependent preferences). The fact that the decision-maker preferred being flexible hints that her/his preferences for the alternative a have some uncertainties. Therefore, we assign the 50% realization probability for each combined sub-menu alternative.

We calculate the expected WTP value loss after constructing expected WTP values for each decision. Eliciting incentivized WTP values in the first stage of the study allows us to construct the highest possible expected WTP values. As shown in Table 3, standard or self-control preferences with consistent decisions yield the maximum expected outcome. The difference between the maximum possible expected outcome and the actual expected WTP value generates the expected WTP loss measure. By construction, the expected WTP loss measure gets USD values in the range of $[0, 0.6]$.

Table 3: Menu Rankings and Expected WTP values

Ranking of menus	Preference Types	Expected WTP formula	Expected WTP
{a} = {a,b} > {b}	No Temptation	0.4*$WTP(a)$+ 0.4*$WTP(a)$+ 0.2*$WTP(b)$	**0.8***$WTP(a)$ **+0.2***$WTP(b)$
{a} > {a,b} > {b}	Resistible Temptation	0.5*$WTP(a)$+ 0.3*$WTP(a)$+ 0.2*$WTP(b)$	**0.8***$WTP(a)$ **+0.2***$WTP(b)$
{a} > {a,b} = {b}	Overwhelming Temptation	0.5*$WTP(a)$+ 0.25*$WTP(b)$+ 0.25*$WTP(b)$	**0.5***$WTP(a)$ **+0.5***$WTP(b)$
{a} > {b} > {a,b}	Guilt	0.5*$WTP(a)$+ 0.3*$WTP(b)$+ 0.2*(0.5*$WTP(a)$+0.5*$WTP(b)$)	**0.6***$WTP(a)$ **+0.4***$WTP(b)$
{a} = {b} > {a,b}	Guilt	0.4*$WTP(a)$+ 0.4*$WTP(b)$+ 0.2*(0.5*$WTP(a)$+0.5*$WTP(b)$)	**0.5***$WTP(a)$ **+0.5***$WTP(b)$
{a,b} > {a} > {b}	Flexibility	0.5*(0.5*$WTP(a)$+0.5*$WTP(b)$)+ 0.3*$WTP(a)$+ 0.2*$WTP(b)$.	**0.55***$WTP(a)$ **+0.45***$WTP(b)$
{a,b} > {a} = {b}	Flexibility	0.5*(0.5*$WTP(a)$+0.5*$WTP(b)$)+ 0.25*$WTP(a)$+ 0.25*$WTP(b)$	**0.5***$WTP(a)$ **+0.5***$WTP(b)$
{a}= {a,b} = {b}	Indifference	0.33*$WTP(a)$+ 0.33*(0.5*$WTP(a)$+0.5*$WTP(b)$)+ 0.33*$WTP(b)$	**0.5***$WTP(a)$ **+0.5***$WTP(b)$

3.4 Hypotheses

We develop testable hypotheses based on our model and experimental study measures. Our hypotheses examine the role of external reference points, WTP (normative utility) and *temptation* (temptation utility) differences, and menu-dependent preferences in inconsistent food choice behavior.

Hypothesis 1

According to our model, Temptation preferences arise when $v(b) - v(a) > u(a) - u(b)$. In the same token, Self-control preferences emerge when $u(a) - u(b) > v(b) - v(a)$. It means that if normative utility differences of choice alternatives are less (more) than temptation utility differences, then the decision-maker is more likely to succumb to temptation (resist to temptation) and choose the choice option with a lower (higher) normative utility level (i.e., with a lower (higher) WTP value). Previous studies also show that an increase in the value difference between alternatives also makes choice options more distinguishable, hence improves the decision consistency (Smith and Krajbich, 2019; Huseynov and Palma, 2021). Since "choosing" corresponds to "assigning a higher ranking" in our study, we hypothesize that an increase in normative utility differences will also increase ranking orders of singleton menus with the highest WTP-valued food alternative.[19] Therefore, Hypothesis 1 can be formalized as follows:

H1: *An increase in WTP value differences will also reduce the proportion of inconsistent menu rankings in the second stage of the study.*

Hypothesis 2

Our second hypothesis closely follows Hypothesis 1. We conjecture that an increase in the temptation value difference between food products (when $v(b) - v(a) > u(a) - u(b)$) will lead to the importance of WTP values reduced. In this case, decision-makers won't adhere to WTP values and will show inconsistent choices more frequently due to the overwhelming effect of temptation.

H2: *An increase in Temptation value differences will also increase the proportion of inconsistent menu rankings in the second stage of the study.*

Hypothesis 3

[19] See Huseynov and Palma (2021) for the impact of value difference on food choice inconsistencies.

We define the choice reference point as the most salient alternative in the menu in line with Kıbrıs et al. (2018). Our treatments are constructed based on the visual salience of the sub-menu D across treated subjects. We conjecture that decision-makers will experience a higher intensity of temptation feelings when an unavailable delicious cake is salient in the food choice environment. According to our model, the $\psi(\cdot)$ representation can capture cases when the impact of temptation utility differences can change. We hypothesize that when there exists an unavailable but more appealing food alternative in menu ranking tasks, the impact of temptation utility differences $(v(b) - v(a))$ or the self-control cost will be higher compared to the No Reference control. This upward change in the impact of the self-control cost will have a higher magnitude when decision-makers exercise more visual attention to the unavailable alternative. Therefore, food menu rankings will be less consistent when the impact of the temptation utility differences is higher due to the external reference. Therefore, we hypothesize that our treatment conditions will increase (decrease) the number of Overwhelming Temptation (Resistible Temptation) preference cases and will reduce the proportion of consistent choices.

H3a: *Treatment conditions will increase (decrease) the proportion of Overwhelming Temptation (Resistible Temptation) preference types in food choices.*

H3b: *Treatment conditions will increase the number of inconsistent food choices.*

4 Results

4.1 Preamble A: Examining First-Stage Token Allocations

We start examining our study outcomes by focusing on first-stage token allocations. Appendix B describes overall token allocations to food snack products. The first-stage WTP elicitation task was identical for all treatment conditions, as the between-subject study treatments were introduced in the second stage of the experiment. Figure B1 shows that around 30% of token allocations assigned zero tokens to displayed food alternatives. Less than 5% of the token allocations yielded 10 tokens to the presented food snack products. Having a higher percentage of zero-token allocations might indicate that subjects overwhelmingly tried to earn additional monetary compensations instead of eating a snack product. Alternatively, no token allocations for food products can also hint not strong home-grown preferences towards available products.

A higher proportion of zero-token allocations can downward bias our WTP estimates indicating

Table 4: Sample Characteristics and Mechanisms for Treatment Conditions

Demographic Variables	N	No Reference N = 65	Low Compliance External Reference N = 53	High Compliance External Reference N = 52	p-value	p-value Corrected
Age	170	24.2 (7.2)	24.0 (8.5)	24.0 (7.3)	0.76	0.80
Male	170	36 (55%)	21 (40%)	20 (38%)	0.11	0.29
High Income (> 60,000 USD)	170	15 (23%)	15 (28%)	14 (27%)	0.80	0.80
BMI	165	24.3 (5.1)	23.8 (5.8)	23.9 (4.3)	0.63	0.80
Process Variables						
FixationCountMenua	170	40 (35)	26 (17)	52 (55)	0.00	0.00
FixationCountMenub	170	48 (44)	32 (29)	51 (43)	0.03	0.06
FixationCountMenuc	170	61 (42)	49 (32)	77 (72)	0.04	0.06
Token Allocations	170	2.97 (1.64)	2.76 (1.90)	2.16 (1.51)	0.02	0.11
Temptation Measure						
TemptationPostChoice	170	8.7 (3.0)	9.0 (3.8)	8.0 (3.5)	0.37	0.37

Statistics: Mean (SD). Tests: Kruskal-Wallis rank sum test; Benjamini & Hochberg correction for multiple testing.

weak preferences for snack products. A fee-chaser subject can assign zero tokens to all products while holding latent non-zero WTP values for snack products. Table B4 presents our regression analysis of the first-stage token allocations using OLS and censored Tobit regressions. On average, male participants allocate around five tokens more to food snacks compared to female subjects. Moreover, income is negatively correlated with token allocations to snack products. Subjects spending a greater fixation time and counts on product pictures also allocate a higher number of tokens to snack food items. Regression results reveal that high-income and female participants are exhibiting lower WTP values for snack products. Additionally, visual attention to product pictures also increases the number of allocated tokens to food products. Thus, we conclude that observed variations in token allocations across demographic and visual attention dimensions indicate that our WTP elicitation tasks captured variations in the preference strength towards snack products across gender, income, and visual attention dimensions. Since we randomly assign participants to experimental conditions after the first stage of the study, treatment effects and menu-dependent preferences can still manifest in the presence of potential downward bias of WTP estimates.

Table 4 displays a comparison of demographic profiles of participants across the experimental conditions. We do not detect any systematic differences among treatment conditions. Moreover, the TemptationPostChoice variable (the temptation feeling measure) does not vary because of external reference experimental conditions indicating that our treatment assignments have not generated a change in temptation feelings towards food snacks.

4.2 Preamble B: Comparing Second-Stage Rankings with the Simulated Benchmark

We conduct an additional robustness check analysis simulating a random menu ranking behavior for 1,000,000 decision-makers. This simulation enables us to test each subjects' data against the simulated random benchmark. Appendix C lays out details of our simulation analysis and our pseudo-code. All study participants pass our test and show menu ranking patterns statistically different than the simulated random benchmark. We conclude that the second stage menu rankings are not contaminated with a potential random behavior.

4.3 Result 1

WTP value differences do not affect food choice inconsistency. So, we do not confirm Hypothesis 1.

Table 5 shows Logit regression analyses of food decision inconsistency with different model specifications. The model specification in Column 2, where we control for treatment conditions and WTP value differences between food snack products, reveals that the prediction of Hypothesis 1 is not supported in our results. The WTP value difference has no impact on food choice inconsistency. Although the inclusion of temptation value differences in Column 4 detects a marginally significant impact of the determinant of the interest, this effect becomes null after the addition of individual demographic factors and the compliance measure for treatments.

Result 1 is not in line with previous findings in the literature (Huseynov and Palma, 2021). We conjecture that the impact of WTP value (i.e., normative utility) differences is not pronounced in the food choice environment where temptation and other visceral feelings are more active and can be measured. We think that modelling menu-dependent preferences and measuring temptation feelings offer a very granular picture of the food choice process, and therefore, explain most of the variation in our dependent variable. This might be a reason for not capturing the negative hypothesized effect of WTP value differences on the inconsistency of food decisions. Future studies might focus on the role of the normative utility in food choices when menu-dependent preferences and the temptation utility can be explicitly captured.

Table 5: Inconsistency

	(1)	(2)	(3)	(4)	(5)
(Intercept)	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
External-Reference Low Compliance	0.00	0.00	−0.00	−0.00	0.00
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
External-Reference High Compliance	0.00	0.00	−0.00	−0.00	0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)
WTP diff		−0.01	-0.06^{*}	−0.05	−0.05
		(0.03)	(0.04)	(0.04)	(0.04)
Temptation diff			0.09^{***}	0.09^{***}	0.09^{***}
			(0.03)	(0.03)	(0.03)
Male				0.01	0.02
				(0.02)	(0.02)
BMI				0.00	0.00
				(0.00)	(0.00)
Mean Fixation time on Menu D					−0.01
					(0.01)
Log Likelihood	−3359.60	−3359.53	−3354.29	−3257.22	−3256.74
Num. obs.	5195	5195	5195	5054	5054

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Note: The table shows the results of the Logit regression analyses for the food choice consistency across all experimental conditions. Standard errors are clustered at the subject level. Indifferent types were excluded from analyses becausethe *WTP diff* is zero in those cases.

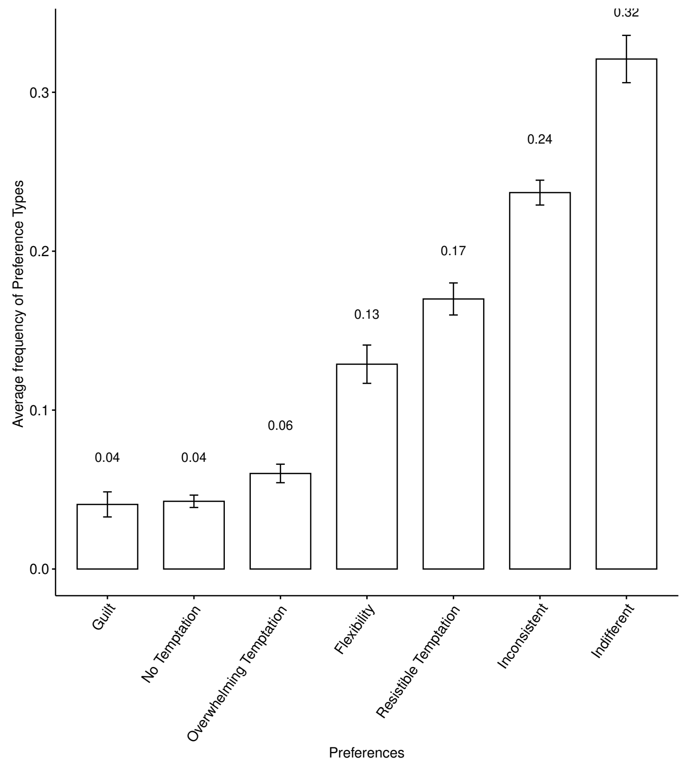
Table 6 shows the outcomes of Tobit regression analyses for Expected Loss. We employ the exact model specifications as in Table 5. Interestingly, both Table 5 and 6 yield overlapping results indicating the insignificance of WTP value differences in food choices. We conclude that normative value differences do not influence food choice consistency and do not lead to welfare-altering decisions in our sample.

4.4 Result 2

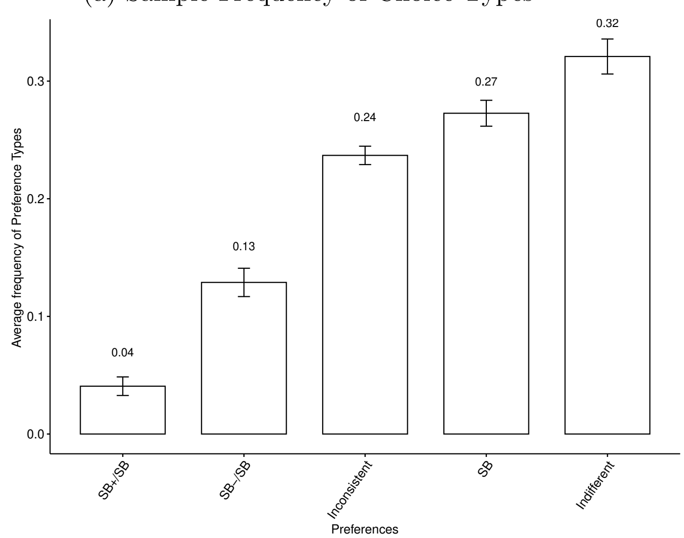
Temptation value differences strongly affect food choice inconsistency. Our results corroborate Hypothesis 2.

Table 5 Columns 3,4, and 5 display that temptation value (i.e., temptation utility) differences have a strong impact on food decision inconsistencies. One Likert-scale rating increase in temptation differences raises the proportion of inconsistent decisions by nine p.p. The inclusion of demographic profile and compliance measure variables do not alter our findings.

Table 6 Column 3,4, and 5 show that an upward change in temptation value differences also increases the magnitude of expected welfare loss in food choices. The combined results of Table 5 and 6 demonstrate that the impact of Temptation utility both in food choice consistency and expected welfare loss is more pronounced compared to normative utility.



(a) Sample Frequency of Choice Types



(b) Sample Frequency of Set-Betweenness Types

The figure shows observed proportions of modelled preferences in the entire study sample. Standard errors are clustered at the subject level. Wilcoxon test p-values ($^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$) are shown comparing the proportions of menu-dependent preferences to the proportion of the Standard preferences.

Figure 3: Menu Dependent Preferences (Second Stage of the Study)

Table 6: Expected Loss

	1	2	3	4	5
(Intercept)	-0.63^{***}	-0.61^{***}	-0.69^{***}	-0.74^{***}	-0.75^{***}
	(0.05)	(0.06)	(0.07)	(0.17)	(0.17)
External-Reference Low Compliance	0.00	0.00	-0.01	-0.02	0.01
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)
External-Reference High Compliance	0.00	0.00	-0.01	-0.00	0.12
	(0.07)	(0.07)	(0.07)	(0.07)	(0.14)
WTP diff		-0.05	-0.26^{*}	-0.24	-0.23
		(0.13)	(0.15)	(0.15)	(0.15)
Temptation diff			0.38^{***}	0.39^{***}	0.40^{***}
			(0.12)	(0.12)	(0.12)
Male				0.07	0.08
				(0.06)	(0.06)
BMI				0.00	0.00
				(0.01)	(0.01)
Mean Fixation time on Menu D					-0.05
					(0.05)
Log Likelihood	-3359.60	-3359.53	-3354.29	-3257.22	-3256.74
Num. obs.	5195	5195	5195	5054	5054

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: The table shows the results of the censored Tobit regression analyses for the expected WTP loss across all experimental conditions. Standard errors are clustered at the subject level.

4.5 Result 3A

A higher level of visual attention to External Reference points reduces the proportion of Overwhelming Temptation preferences and increases indifference among food snack products. However, External Reference Low-Compliance treatment condition reduces the frequency of Resistible Temptation preference types confirming our predictions. Our findings partially support the predictions of Hypothesis 3A.

We first focus on observed frequencies of standard and menu-dependent preferences in our study sample. Our research inquiry documents both conventional and non-conventional preference types in the food choice domain, displaying the importance of considering the peculiarities of decision environments and their impact on consumer behaviors. Figure 3 panel (a) shows average sample frequencies of modeled menu-dependent preference types. Specifically, we find that only 4% of second-stage menu rankings comply with standard No Temptation preferences indicating the absence of visceral feelings and self-control issues in food decisions. However, 40% of menu rankings are in line with menu-dependent non-conventional preferences indicating that most food choices are subject to a wide range of non-standard preferences associated with visceral feelings. The frequency of Flexibility preferences is 13%. The proportion of Resistible Temptation preferences is 17% and statistically different than the frequency of standard No Temptation preferences. Overwhelming Temptation and Guilt preferences constitute 6% and 4% of all food choices, respectively. Indifferent preferences explain 32% of menu rankings in the second stage of our experiment. Overall, the observed sample proportions of the modeled non-conventional preferences indicate the importance of explicitly measuring and documenting menu-dependent preferences during food decisions. Figure 3 panel (b) displays the proportion of preference types under the definition of set betweenness. Set betweenness and Indifferent types explain around 60% of second-stage food decisions showing the importance of menu-dependent models in understanding food choices.

Figure 4 compares the observed frequencies of preference types across experimental treatment conditions. Our analysis does not reveal treatment effects in the occurrences of standard No Temptation, Guilt, and Flexibility preferences across the study conditions. Interestingly, the share of Resistible Temptation preferences is 4% lower in the External-Reference Low Compliance treatment condition compared to the No Reference control ($Wilcoxon, z-score = 1.44, p = 0.08$). Moreover, the External-Reference High Compliance treatment condition reduces the proportion of Overwhelming Temptation preferences by 2% compared to the control No Reference condition ($Wilcoxon, z-score = 2.25, p = 0.01$). The share of Indifferent types also increases by 6% in the External-Reference High Compliance condition ($Wilcoxon, z-score = 1.39, p = 0.08$).

We also conduct a prediction exercise to understand the role of our primary independent variables in the formation of each menu difference type. We utilize dummy variables for each experimental condition and the differences in WTP and Temptation values between food choice alternatives in multinomial logit regressions to predict the probability of each preference type across treatments and the No Reference control. Figure 5 displays the mean values of fitted probabilities of menu-dependent preference types for each treatment condition. Our results are in line with the sample frequencies described in Figure 4. External reference points induce indifference among food choice alternatives and reduce the intensity of temptation-driven preferences.

We conclude that our results partially confirm our Hypothesis 3. The external reference point and its salience reduces the frequency of temptation driven preferences in the food choice domain. The outcomes of our analysis only confirms our prediction of the negative relationship between the external reference and Resistible Temptation preferences.

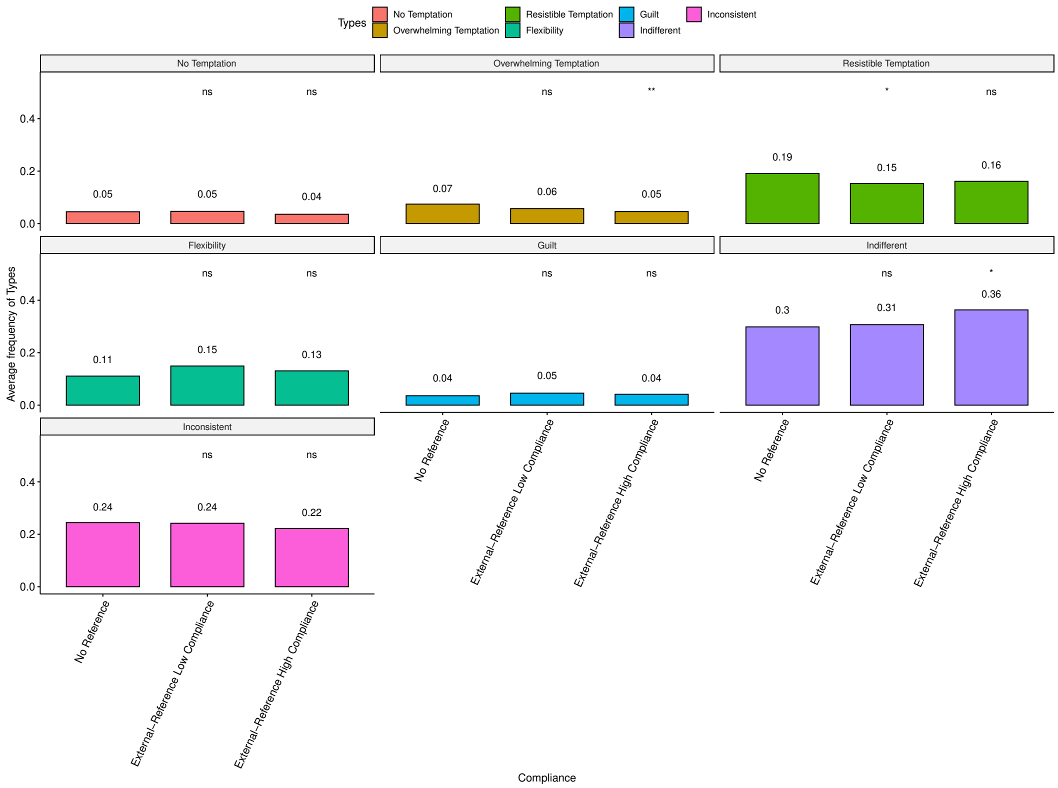


Figure 4: Frequency of Choice Types across Treatment Conditions

The figure shows observed proportions of modelled preferences across experimental conditions. Standard errors are clustered at the subject level. Wilcoxon test p-values ($^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$) are shown comparing the proportions of menu-dependent preferences to the proportion of the Standard No Temptation preferences.

4.6 Result 3B

External Reference points do not change food decision inconsistency. Our findings do not support the predictions of Hypothesis 3B.

Logit regression results in Table 5 and Tobit estimates in Table 6 unanimously show the null effect of treatment conditions. We do not detect any impact of the external reference treatment conditions both on our consistency measure and expected welfare loss dependent variables. This finding is robust to different model specifications.

Per our model and the constructed menu-ranking orderings, participants with Resistible and Overwhelming temptation preferences are conscious of self-control issues. Subjects with Overwhelming Temptation preferences predict choosing the self-identified normatively inferior but more tempting alternative in the combined menu. On the other hand, study participants with Resistible Temptation preferences expect to choose the self-identified normatively superior and less tempting alternative in the sub-menu C. Our findings shed light on the role of the external reference

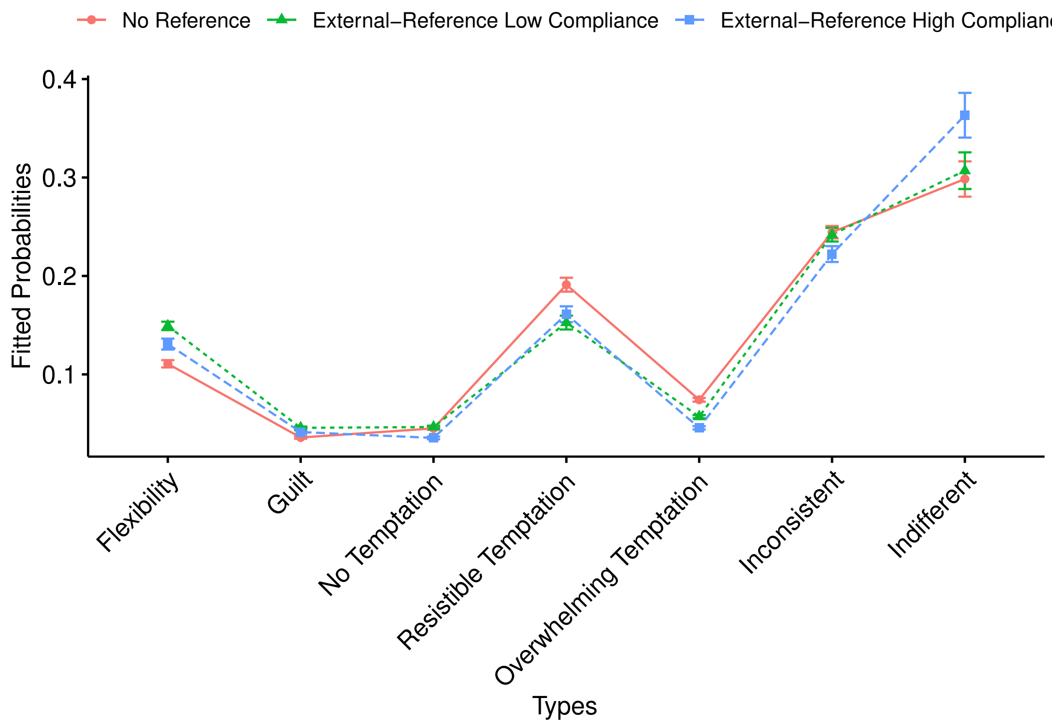


Figure 5: Predicting Menu-Dependent Preferences

The figure shows fitted probabilities of modelled preferences across experimental conditions.

point and the relative temptation to food choice alternatives. We conclude that the existence of an unavailable external reference affects the frequencies of Overwhelming Temptation preferences and not in the direction we hypothesized. A higher intensity of visual attention to the unavailable product reduces the proportion of Overwhelming Temptation types indicating that a salient external reference point decreases the severity of temptation feelings of the self-identified inferior alternative.

5 Related Literature and Policy Implications

Our paper relates to extensive literature examining the underpinnings of decision consistency. The first wave of economic studies scrutinized this phenomenon in the context of decision errors assuming that observed decision inconsistencies arise due to random mistakes of decision-makers (Manski, 1977; McFadden et al., 1973). Subsequent studies showed that economic agents' visceral (i.e., emotional) feelings and neuro-psychological profiles appear to be strong predictors of decision errors or time-inconsistent preferences. These advancements revealed the necessity of understanding the role of choice environments and behavioral urges in inconsistent economic decisions (Ben-Akiva et al., 2002; Frederick et al., 2002). The seminal paper by Gul and Pesendorfer (2001) presented the third explanation for inconsistent preferences showing decision contexts can also lead to inconsistent choices by imposing self-control costs. Increased decision-engineering and choice manipulation attempts in the marketplace and consequent self-control failures of consumers make the modeling framework developed by Gul and Pesendorfer (2001) more appealing and well-positioned to explain observed new empirical decision irregularities. This new strand of the literature also enables integrating decision errors, visceral feelings, and time-inconsistent decisions under a unified framework.

Decision consistency has also increasingly been scrutinized in experimental studies. For instance, Agranov and Ortoleva (2017) document that the 71% of choices are inconsistent when subjects make three repetitive decisions in a row among the same lottery alternatives. Before seeing the lotteries, subjects are informed that they will be shown the same lottery choice task three times to prevent any confusion about experimental procedures. Agranov and Ortoleva (2017) show that *random preferences* due to stochastic changes in utilities (Gul and Pesendorfer, 2006) and *bounded rationality* stemming from differential information accumulation paths (Ratcliff and McKoon, 2008) cannot explain their results as both behavioral models predict the same outcome when decisions are repeated in a row. Similar to our research design, Fudenberg et al. (2018) study decision inconsistency using a two-stage experimental setup where researchers can measure the number of subject-identified consistent choices. Participants independently rate alternatives in the first stage, revealing their idiosyncratic valuations. In the second stage of the experiment, they make choices in binary menus constructed from the stimuli rated in the first stage. Fudenberg et al. (2018) discuss the importance of scrutinizing *decision process*, and specifically, they show that the response time and the boundary of collected information impact choice consistency during the second-stage decisions. Bordalo et al. (2013) model the relationship between the choice context and decision reversals with *attribute salience*. However, their primary identification strategy relies on relative changes in attribute values that do not accommodate decision problems when attribute valuations do not change in repeated or two-stage settings. Primary models and approaches in this literature mostly operate based on structural assumptions, which are not straightforward to test in empirical studies. Therefore, we build on the other strand of this literature explaining choice inconsistency with *menu-dependent preferences* and simple choice statistics (Gul and Pesendorfer, 2001; Noor and Takeoka, 2010, 2015; Kopylov, 2012; Toussaert, 2018; Huseynov and Palma, 2021). Gul and Pesendorfer (2001) introduce *set-betweenness* (i.e., a set of assumptions on the relationship between singleton sub-menus and a combined sub-menu of all choice alternatives) to elicit menu-dependent preferences, temptation, and self-control costs. The primary advantage of this line of research is providing simple choice statistics to identify different decision types that can explain inconsistent outcomes.

Without explicitly studying the consistency of elicited WTP values, this new strand of literature brings new perspectives on the common sources of economic choice inconsistency, linking this phenomenon to peculiarities of the choice process and information collection (Fudenberg et al., 2018), visceral feelings (Ben-Akiva et al., 2002; Frederick et al., 2002), and menu-dependent preferences (Gul and Pesendorfer, 2001). Some empirical studies scrutinize correlates of consistency and inter-temporal stability of WTP estimates in different domains but do not explicitly test a decision-maker's commitment to self-identified WTP measures (Brouwer et al., 2017; Mørkbak and Olsen, 2015). Inconsistencies are frequent in the food decision domain favoring normatively inferior choices with severe health and welfare consequences (Ruhm, 2012). We examine the consistency of incentivized WTP values in repeated food decisions identifying its determinants and outcomes. Our contribution is explicitly testing the role of changes in decision *reference points*, *WTP differences*, and *temptation* in food choice consistency. Inferences based on empirical WTP studies should warrant further scrutiny if the food choice environment triggers inconsistent choices, contradicting elicited incentivized WTP measures. Modeling and empirically documenting decision environment factors leading to the inconsistency of food choices will improve prediction and policy design outcomes. By drawing on recent developments in the decision theory literature, our study contributes to the empirical WTP studies examining underpinnings of the consistency of food choices.

Our results contribute to the extensive literature investigating the determinants of low-calorie food choices, which aims to better inform health policies on designing effective nudges and incentive structures (Dolgopolova and Teuber, 2018; Jo et al., 2016; Wilson and Lusk, 2020). Jo et al. (2016) conducted an experiment with consumers using 173 food items, finding that taste is a significant factor driving willingness to pay (WTP) for food products. The authors also report that consumers react to provided health information by increasing their WTP for high-quality food alternatives, though the effect is sensitive to the decision environment. Our study expands existing knowledge, showing that the food choice environment and menu structure also play critical roles in shaping WTP values. Our results indicate that consumers may reduce their WTP for healthier food products when faced with tempting alternatives, making WTP estimates less reliable for policy insights.

Dolgopolova and Teuber (2018) discuss the current landscape of empirical evidence on WTP for low-calorie healthy products and emphasize policy challenges. They specifically address the mixed results in the relevant literature that offer uncertain policy insights. Our paper demonstrates that individually elicited WTP estimates might be inconsistent when consumers confront complex choice problems where tempting food alternatives are prominent. Consequently, our article cautions empirical studies that rely on comparing elicited WTP values for estimating market shares of food products or potential policy outcomes to consider the specificities of the decision environment. In particular, the inconsistency-inflating role of temptation and other menu-induced visceral feelings should be controlled for when choice alternatives involve food products.

The relevant literature also presents divergent views on the effectiveness of health policy measures in improving diet quality. Taxation and other external government measures have been shown to yield unintended consequences (Caputo and Just, 2022). Caputo and Just (2022) highlight the importance of "internal factors" (such as choice set restrictions) in enhancing the impact of policy initiatives. Our study offers a fresh perspective on this discussion by demonstrating how research can identify different consumer preference types, including self-restrained decision-makers who self-impose choice-set restrictions. Future studies can enhance the predictive power of survey instruments by employing menu-dependent preferences to identify consistent decision-makers for building robust policy inferences.

6 Discussions and Conclusion

Most economic activities involve repeated decisions over time, with immediate and (sometimes harmful) future consequences. According to conventional economic models, decision-makers have consistent preferences. Having consistent preferences and decisions ensures that agents will adhere to the initially identified optimal saving plans, effort levels, and healthy diets and won't deviate to not-welfare-maximizing alternatives. The primary implication of this modeling assumption in the food decision domain is that consumers make consistent food choices conforming to their self-revealed WTP valuations. Yet overwhelming evidence derived from experimental studies shows that economic agents frequently exhibit inconsistent decisions and prefer previously self-identified inferior alternatives (Agranov and Ortoleva, 2017; Fudenberg et al., 2018; Huseynov and Palma, 2021; Sadoff et al., 2020; Toussaert, 2019; Hsee and Zhang, 2010).

What is the relationship between the food decision context and WTP consistency? Can non-standard preferences explain deviations from choosing higher WTP-valued food alternatives? What are the welfare consequences of food choice inconsistencies? To answer our research questions, we construct a model based on menu-dependent preferences and visceral feelings to understand the impact of the food decision environment on choice (in)consistency (Gul and Pesendorfer, 2001; Ben-Akiva et al., 2002; Frederick et al., 2002; Agranov and Ortoleva, 2017). Our study investigates the consistency of food choices using preference elicitation techniques developed by Gul and Pesendorfer (2001) and Toussaert (2018). Our study design employs a controlled lab experiment using subjects' rankings of menus and self-identified WTP values for food products. We focus on choice inconsistency in non-temporal decisions as most empirical studies draw inferences with market and policy implications using lab-elicited incentivized and non-dynamic WTP values (Van Loo et al., 2015; Lusk and Hudson, 2004; Shogren et al., 2001). We also manipulate decision reference points to exogenously alter the salience of menu alternatives and detect changes in the proportion of consistent food choices. Our findings show that 24% of food decisions are inconsistent with the elicited incentivized WTP values. Additionally, higher *temptation* value differences between alternatives increase the inconsistency of food choices.

We also find that the external reference point and its salience do not alter the proportion of inconsistent food choices and the magnitude of expected loss. Our findings suggest that WTP studies are not vulnerable to changes in the decision environment via the reference point. We also show that menu-dependent preferences can explain 40% of food decisions. Our results reveal that the salience of the external decision reference point reduces the observed frequencies of temptation-related menu-dependent preferences and increases the indifference among food choice alternatives. While our model cannot offer insights on the behavioral channels yielding this result, we urge prospective studies to explore this research venue.

Our study is an important step towards understanding the (in)consistency of food choices. We offer a novel design where we measure consistency with incentivized WTP measures and unconventional menu ranking tasks. Our results show that non-standard choice models explain the majority of our data. Menu-dependent preferences may also have a heterogeneous impact on the consistency of food choices and welfare. Future studies should incorporate non-conventional models explaining visceral feelings and further investigate their role in economic decisions.

References

Agranov, M. and Ortoleva, P. (2017). Stochastic choice and preferences for randomization. *Journal of Political Economy*, 125(1):40–68.

Alfnes, F. and Rickertsen, K. (2003). European consumers' willingness to pay for us beef in experimental auction markets. *American Journal of Agricultural Economics*, 85(2):396–405.

Ben-Akiva, M., McFadden, D., Train, K., Walker, J., Bhat, C., Bierlaire, M., Bolduc, D., Boersch-Supan, A., Brownstone, D., Bunch, D. S., et al. (2002). Hybrid choice models: Progress and challenges. *Marketing Letters*, 13(3):163–175.

Bhatia, S. and Golman, R. (2019). Attention and reference dependence. *Decision*, 6(2):145.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and consumer choice. *Journal of Political Economy*, 121(5):803–843.

Brouwer, R., Dekker, T., Rolfe, J., and Windle, J. (2010). Choice certainty and consistency in repeated choice experiments. *Environmental and Resource Economics*, 46(1):93–109.

Brouwer, R., Logar, I., and Sheremet, O. (2017). Choice consistency and preference stability in test-retests of discrete choice experiment and open-ended willingness to pay elicitation formats. *Environmental and Resource Economics*, 68(3):729–751.

Caputo, V. and Just, D. (2022). The economics of food related policies: Considering public health and malnutrition. *Handbook of Agricultural Economics*, 6:5117.

Caputo, V., Lusk, J. L., and Nayga, R. M. (2020). Am i getting a good deal? reference-dependentdecision making when the reference price is uncertain. *American Journal of Agricultural Economics*, 102(1):132–153.

Carlsson, F., Frykblom, P., and Lagerkvist, C. J. (2007). Consumer benefits of labels and bans on gm foods—choice experiments with swedish consumers. *American Journal of Agricultural Economics*, 89(1):152–161.

Costa-Font, M., Gil, J. M., and Traill, W. B. (2008). Consumer acceptance, valuation of and attitudes towards genetically modified food: Review and implications for food policy. *Food policy*, 33(2):99–111.

Dekel, E., Lipman, B. L., and Rustichini, A. (2009). Temptation-driven preferences. *The Review of Economic Studies*, 76(3):937–971.

Dolgopolova, I. and Teuber, R. (2018). Consumers' willingness to pay for health benefits in food products: A meta-analysis. *Applied Economic Perspectives and Policy*, 40(2):333–352.

Fraser, I., Balcombe, K., Williams, L., and McSorley, E. (2021). Preference stability in discrete choice experiments. some evidence using eye-tracking. *Journal of Behavioral and Experimental Economics*, 94:101753.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401.

Fudenberg, D., Strack, P., and Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–84.

Gul, F. and Pesendorfer, W. (2001). Temptation and self-control. *Econometrica*, 69(6):1403–1435.

Gul, F. and Pesendorfer, W. (2006). Random expected utility. *Econometrica*, 74(1):121–146.

Healy, P. J. (2018). Explaining the bdm—or any random binary choice elicitation mechanism—to subjects. Technical report, mimeo.

Heydari, P. (2021). Luce arbitrates: Stochastic resolution of inner conflicts. *Games and Economic Behavior*, 126:33–74.

Hsee, C. K. and Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science*, 5(4):343–355.

Huseynov, S., Kassas, B., Segovia, M. S., and Palma, M. A. (2019). Incorporating biometric data in models of consumer choice. *Applied Economics*, 51(14):1514–1531.

Huseynov, S. and Palma, M. A. (2021). Food decision-making under time pressure. *Food Quality and Preference*, 88:104072.

Huseynov, S., Palma, M. A., and Ahmad, G. (2021). Does the magnitude of relative calorie distance affect food consumption? *Journal of Economic Behavior Organization*, 188:530–551.

Jo, J., Lusk, J. L., Muller, L., and Ruffieux, B. (2016). Value of parsimonious nutritional information in a framed field experiment. *Food Policy*, 63:124–133.

Kıbrıs, Ö., Masatlioglu, Y., and Suleymanov, E. (2018). A theory of reference point formation. *Unpublished Paper*.

Kopylov, I. (2012). Perfectionism and choice. *Econometrica*, 80(5):1819–1843.

Lusk, J. L. and Hudson, D. (2004). Willingness-to-pay estimates and their relevance to agribusiness decision making. *Applied Economic Perspectives and Policy*, 26(2):152–169.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229.

McFadden, D. et al. (1973). Conditional logit analysis of qualitative choice behavior.

Mørkbak, M. R. and Olsen, S. B. (2015). A within-sample investigation of test–retest reliability in choice experiment surveys with real economic incentives. *Australian Journal of Agricultural and*

Resource Economics, 59(3):375–392.

Nielsen, K. and Rehbeck, J. (2020). When choices are mistakes. *Available at SSRN 3481381.*

Noor, J. and Takeoka, N. (2010). Uphill self-control. *Theoretical Economics*, 5:127–158.

Noor, J. and Takeoka, N. (2015). Menu-dependent self-control. *Journal of Mathematical Economics*, 61:1–20.

Pappalardo, G. and Lusk, J. L. (2016). The role of beliefs in purchasing process of functional foods. *Food quality and preference*, 53:151–158.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, 20(4):873–922.

Reardon, T., Bellemare, M. F., Zilberman, D., et al. (2020). How covid-19 may disrupt food supply chains in developing countries. *IFPRI book chapters*, pages 78–80.

Rigby, D., Burton, M., and Pluske, J. (2016). Preference stability and choice consistency in discrete choice experiments. *Environmental and Resource Economics*, 65(2):441–461.

Ruhm, C. J. (2012). Understanding overeating and obesity. *Journal of Health economics*, 31(6):781–796.

Sadoff, S., Samek, A., and Sprenger, C. (2020). Dynamic inconsistency in food choice: Experimental evidence from two food deserts. *The Review of Economic Studies*, 87(4):1954–1988.

Segovia, M. S. and Palma, M. A. (2021). Testing the consistency of preferences in discrete choice experiments: an eye tracking study. *European Review of Agricultural Economics*, 48(3):624–664.

Shogren, J. F., Cho, S., Koo, C., List, J., Park, C., Polo, P., and Wilhelmi, R. (2001). Auction mechanisms and the measurement of wtp and wta. *Resource and Energy Economics*, 23(2):97–109.

Smith, S. M. and Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological science*, 30(1):116–128.

Toussaert, S. (2018). Eliciting temptation and self-control through menu choices: A lab experiment. *Econometrica*, 86(3):859–889.

Toussaert, S. (2019). Revealing temptation through menu choice: Field evidence. *Manuscript, University of Oxford.*

Van Loo, E. J., Caputo, V., Nayga Jr, R. M., Seo, H.-S., Zhang, B., and Verbeke, W. (2015). Sustainability labels on coffee: Consumer preferences, willingness-to-pay and visual attention to attributes. *Ecological Economics*, 118:215–225.

Wilson, L. and Lusk, J. L. (2020). Consumer willingness to pay for redundant food labels. *Food*

Policy, 97:101938.

Appendices

A Additional Proofs and Analyses of the Model

A.1 Guilt in no-uncertainty representation with singleton and binary menus

Consider alternatives $a, b \in A$ such that $\{a\} \succeq \{b\}$. Then, by positive set betweenness, we have $\{a\} \succeq \{a, b\}$. Note that $U_{\mathrm{NU}}(\{a\}) = u(a)$ and $U_{\mathrm{NU}}(\{b\}) = u(b)$. The menu preferences will satisfy guilt if $\{b\} \succ \{a, b\}$, i.e., we have the following:

$$\max\left\{u(a) + \sum_j v_j(a), u(b) + \sum_j v_j(b)\right\} - \sum_j \max_{c \in \{a,b\}} v_j(c) - u(b) < 0$$

This can be rewritten as:

$$\max\left\{u(a) + \sum_j v_j(a) - u(b) - \sum_j v_j(b), 0\right\} - \sum_j \left[\max_{c \in \{a,b\}} v_j(c) - v_j(b)\right] < 0$$

Since $\max\left\{u(a) + \sum_j v_j(a) - u(b) - \sum_j v_j(b), 0\right\} \geq 0$, the menu preferences will satisfy guilt only if $\sum_j \left[\max_{c \in \{a,b\}} v_j(c) - v_j(b)\right] > 0$. This is possible only if alternative b is not more tempting for all dimensions.

A.2 Flexibility in uncertain strength of temptation representation with singleton and binary menus

Consider the alternatives $a, b \in A$ such that $\{a\} \succeq \{b\}$. Then, by negative set betweenness, we have $\{a, b\} \succeq \{b\}$. Note that $U_{\mathrm{US}}(\{a\}) = u(a)$ and $U_{\mathrm{US}}(\{b\}) = u(b)$. Note that $U_{\mathrm{US}}(\{a, b\})$ can be written as:

$$U_{\mathrm{US}}(\{a, b\}) = \sum_i q_i \left[\max \{u(a) + \gamma_i v(a), u(b) + \gamma_i v(b)\} - \gamma_i \max_{c \in \{a,b\}} v(c) \right]$$

Since $U_{\mathrm{US}}(\{a\}) = u(a)$, $v(a) \leq \max_{c \in \{a,b\}} v(c)$, $\gamma_i \geq 0$, and $q_i > 0$, the menu preferences satisfy flexibility only if:

$$\sum_i q_i \max \{0, u(b) - u(a) + \gamma_i v(b) - \gamma_i v(a)\} > 0$$

The above condition suggests that $\gamma_i \left[v(b) - v(a)\right] > u(a) - u(b)$ for some i.

A.3 Flexibility and guilt in perfectionist set betweenness representation with singleton and binary menus

Consider the alternatives $a, b \in A$ such that $\{a\} \succeq \{b\}$. Following the representation in equation (8), we have $U_{\mathrm{K}}(\{a\}) = (1+\kappa)u(a)$ and $U_{\mathrm{K}}(\{b\}) = (1+\kappa)u(b)$. Since $\{a\} \succeq \{b\}$, we have $u(a) \geq u(b)$. Consider the following cases:

Case 1. $u(a) + v(a) \geq u(b) + v(b)$.

In this case, we can write $U_{\mathrm{K}}(\{a, b\})$ as follows:

$$U_{\mathrm{K}}(\{a, b\}) = (1+\kappa)u(a) + v(a) - \max_{c \in \{a,b\}} v(c)$$

Then, $U_{\mathrm{K}}(\{a, b\}) \leq U_{\mathrm{K}}(\{a\})$, i.e., $\{a\} \succeq \{a, b\}$, because $v(a) \leq \max_{c \in \{a,b\}} v(c)$. So, flexibility cannot arise in this case. For guilt to arise in the menu preferences, we must have the following:

$$(1+\kappa)[u(a) - u(b)] + v(a) - \max_{c \in \{a,b\}} v(c) < 0$$

Since $u(a) \geq u(b)$ and $\kappa > -1$, the above condition will be satisfied when $v(a) < v(b)$ and κ is sufficiently small.

Case 2. $u(a) + v(a) < u(b) + v(b)$.

Since $u(a) \geq u(b)$, we have $v(b) - v(a) > u(a) - u(b) \geq 0$, i.e., $v(b) > v(a)$. Then, $U_{\mathrm{K}}(\{a, b\})$ can be written as:

$$U_{\mathrm{K}}(\{a, b\}) = u(b) + \kappa u(a)$$

Then, $U_{\mathrm{K}}(\{a, b\}) \leq U_{\mathrm{K}}(\{a\})$, i.e., $\{a\} \succeq \{a, b\}$, because $u(a) \geq u(b)$. Therefore, flexibility cannot arise in this case either. Moreover, guilt arises in the menu preferences, $U_{\mathrm{K}}(\{a, b\}) < U_{\mathrm{K}}(\{b\})$, only if $u(a) > u(b)$ and $\kappa < 0$.

B Details of Experimental Procedures and Stimuli

We employed eye-tracking technology (Tobii x2-60) using iMotions software to measure subjects' eye movement and fixation patterns. This near-infrared technology relies on visible reflection on the cornea to infer gaze patterns, fixation time, and counts (Huseynov et al., 2019).

Table B1 depicts a statistical comparison of experimental treatment conditions based on the calorie content of unavailable cakes in Menu D. Table B2 also examines the same set of conditions across eye-tracking measures. We cannot detect any statistically different patterns among experimental conditions when treatments are determined by the calorie content of the cakes in the unavailable Menu D. Table B1, and B2 suggest that participants do not exhibit differential behavior to 580 and 710 calorie cakes. It is possible that the relative 160-calorie difference between cakes in the between-subject setting did not magnify any behavioral changes. If the effect of the unavailable menu D has been indistinctly aggregated on both cakes, then it can only be detected via the comparison of actually treated subjects who pay a higher level of visual attention to the cake to subjects paying a lower level of visual attention or no attention (participants in the Control condition) to the menu D.

We determine treatment conditions based on the median amount of the visual attention time (fixation time) on sub-menu D labeling the experimental treatment with subjects spending lower (higher) than the median level of the fixation attention as *External-Reference Low (High) Compliance*. Table 4 shows that the External-Reference Low Compliance condition is statistically different from the External-Reference High Compliance treatments across visual attention patterns to sub-menu A, B, and C. However, they are not statistically distinguishable based on observable demographic variables. Table 4 provides evidence that the level of visual fixation to Menu D alters the visual salience of sub-menu A, B, and C validating our treatment conditions based on the median split. We conclude that the existence of an unavailable sub-menu D changes the visual salience of the existing menu items.

We determine Area of Interest (AOI) to measure fixation time and counts for each sub-menu (sub-menu A, B, and C) and for the unavailable sub-menu D (See Figure B2). We do not define AOIs based on product pictures since the size of product pictures vary and it can mislead attention measures inflating fixation times and counts for products with larger images [20] Therefore, we draw the AOIs with the same size adjacent to the corresponding ranking scales for sub-menu A, B,

[20] We do not modify product pictures to capture their actual market demand and appealing.

and C in all food decision tasks in the second stage of our study. This approach ensures that we capture normalized fixation times and counts to each sub-menu, and our eye-tracking measures are comparable across second-stage tasks and treatment conditions. Although some subjects in the Control condition (who were not shown sub-menu D) exhibit very few fixation counts on the AOI determined for the unavailable food item, those visual patterns are substantially smaller than visual attention levels to existing items. Thus, we do not consider those noises in our analysis.

Table B1: Balance Statistics for Low- and High-Calorie references

Demographic Variables	N	No Reference N = 65	Low-Calorie Reference N = 68	High-Calorie Reference N = 37	p-value	p-value Corrected
Token Allocations	170	2.97 (1.64)	2.40 (1.73)	2.59 (1.75)	0.07	0.16
Age	170	24.2 (7.2)	24.0 (7.5)	23.9 (8.8)	0.52	0.76
Male	170	36 (55%)	32 (47%)	9 (24%)	0.01	0.05
High Income ($> 60,000$ USD)	170	15 (23%)	20 (29%)	9 (24%)	0.69	0.76
BMI	165	24.3 (5.1)	24.0 (5.4)	23.6 (4.4)	0.76	0.76

Statistics: Mean (SD); n (%). Tests: Kruskal-Wallis rank sum test; Pearson's Chi-squared test; Benjamini & Hochberg correction for multiple testing.

Table B2: Low- and High-Calorie references

Process Variables	N	No Reference N = 65	Low-Calorie Reference N = 68	High-Calorie Reference N = 37	p-value	p-value Corrected
FixationCountMenua	170	40 (35)	43 (50)	31 (24)	0.48	0.76
FixationCountMenub	170	48 (44)	44 (37)	38 (40)	0.36	0.76
FixationCountMenuc	170	61 (42)	66 (60)	58 (52)	0.76	0.76
Temptation Measure						
TemptationPostChoice	170	8.7 (3.0)	8.2 (3.8)	9.0 (3.4)	0.34	0.76

Statistics: Mean (SD). Tests: Kruskal-Wallis rank sum test; Benjamini & Hochberg correction for multiple testing.

Table B3: List if snack food products

Snack Product	Calorie Content
Debbie Oatmeal Creme pie	330 Calories
Whole Grain Peanut Butter Cracker	200 Calories
Ritz Cheese Sandwich Crackers	200 Calories
Oreo double cookies	210 Calories
Quaker Chewy Bar	140 Calories
Ritz Bits	150 Calories
Chips Ahoy cookies	140 Calories
Debbie Honey buns	230 Calories
Hostess Dark Chocolate Cupcakes	150 Calories
Hostess Twinkies	130 Calories

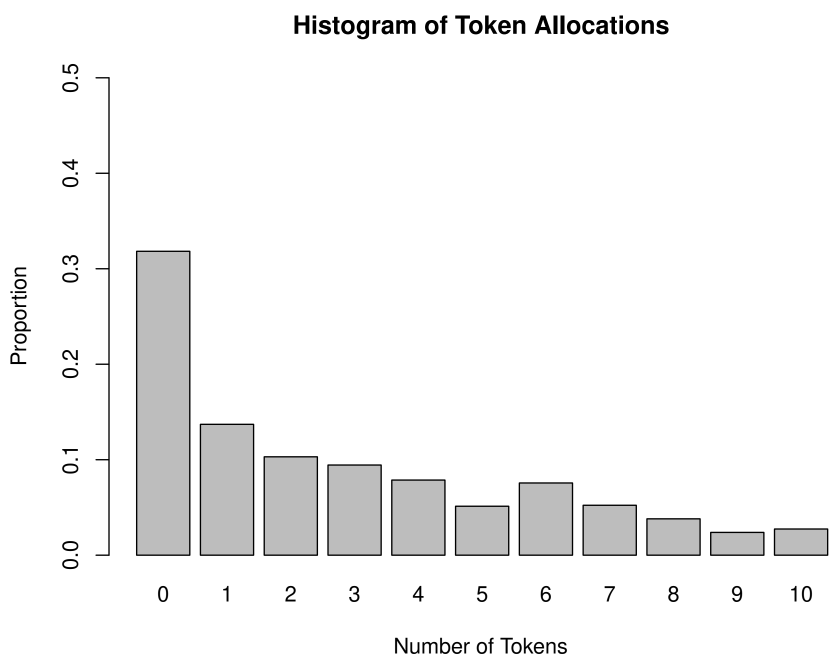
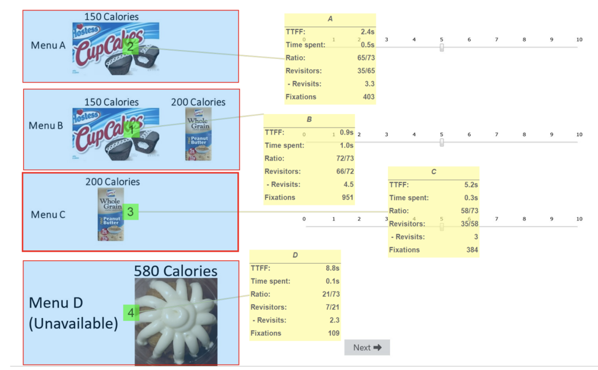
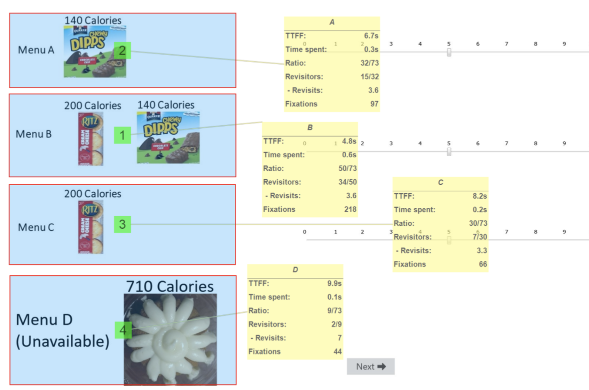


Figure B1: Histogram of first-stage token allocations for all participant across treatment conditions

(a) No Reference Experimental Condition



(b) External Reference Condition



(c) External Reference Condition

Figure B2: Experimental Stimuli (Second Stage of the Study)

Table B4: Token Allocations (First Stage)

	OLS	OLS	OLS	OLS	Tobit	Tobit
Intercept	28.29^{***}	-10.59	-14.18	-9.16	17.94^{***}	21.19^{***}
	(3.54)	(14.25)	(13.84)	(13.81)	(6.38)	(6.24)
Age	0.03	0.80^{***}	0.83^{***}	0.83^{***}	0.08	0.10
	(0.08)	(0.17)	(0.16)	(0.16)	(0.13)	(0.12)
Male	3.59^{**}	5.51	6.05	5.49	5.30^{**}	5.32^{**}
	(1.40)	(6.49)	(6.50)	(6.47)	(2.10)	(2.11)
Highincome ($> 60,000$ USD)	-5.50^{***}	-1.82	-1.83	-2.21	-8.69^{***}	-8.92^{***}
	(1.55)	(7.59)	(7.54)	(7.54)	(2.44)	(2.44)
BMI	-0.13	0.07	0.04	-0.11	-0.14	-0.19
	(0.13)	(0.29)	(0.29)	(0.28)	(0.20)	(0.20)
Number of Calories		-0.01	-0.01	-0.01	-0.01	-0.01
		(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Fixation Time (product picture)			1.66^{***}		1.85^{***}	
			(0.46)		(0.61)	
Fixation Time (calorie information)			-2.69		-3.00	
			(2.01)		(2.79)	
Fixation Counts (product picture)				0.31^{***}		0.21^{**}
				(0.06)		(0.08)
Fixation Counts (calorie information)				-0.56^{*}		-1.04^{**}
				(0.29)		(0.43)
Num. obs.	1650	1650	1650	1650	1650	1650

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

C Additional Analysis

Simulation Analysis of Preference Types

Having around 30% of first-stage decisions with zero token allocations to the presented food snack products leads to a concern that subjects may have randomly executed the second-stage menu ranking tasks. If participants do not value food snacks, they can also be reluctant to ranking sub-menus containing those products. We focus on menu ranking tasks without considering WTP values. The objective is to examine second-stage menu rankings and compare them with simulated random benchmarks. Therefore, we do not label inconsistent choices since we do not account for first stage Willingness-To-Pay values.

In the main analyses, we excluded subjects who did not pay visual attention to menus in the second stage of the study. However, a participant can still randomly act in spite of the fact s/he visually examines the product pictures. Therefore, in this subsection, we conduct further analysis to address the concern that the second stage results can be contaminated with random behavior patterns.

We execute a simulation analysis. In the simulation, we create 45 menu ranking decisions by randomly ranking singleton menus and the combined menu. For each sub-menu, random ranks are drawn from a uniform distribution of U(0,10). We repeat this routine 1,000,000 times. Therefore, we simulate random menu ranking decisions for 1,000,000 stochastic decision-makers. The pseudo code of the simulation can be described as follows:

Algorithm 1 Simulation

1: Randomly draw three sets of 45 integers from the uniform discreet distribution of U(0,10) with replacement for menua, menub, and menuc vectors. This enables us to mimic menu ratings in Stage 2 of the study, as subjects employed a 0-to-10 discreet scale for rating food alternatives.
2: Merge menua, menub, and menuc vectors to create 45 menu ranking decisions. Label these 45 decisions to represent one stochastic decision-maker.
3: Repeat Step 1 and 2 for 1,000,000 times.

After simulating 45 menu ranking decisions for 1,000,000 stochastic decision-makers, we define the preference type for each decision using the Ranking of Menus column in Table 1. Then, for each stochastic decision-maker, we calculate the percentage of each preference type among 45 menu ranking decisions.

Our simulation results show that, on average, only 8% of menu rankings are in line with standard

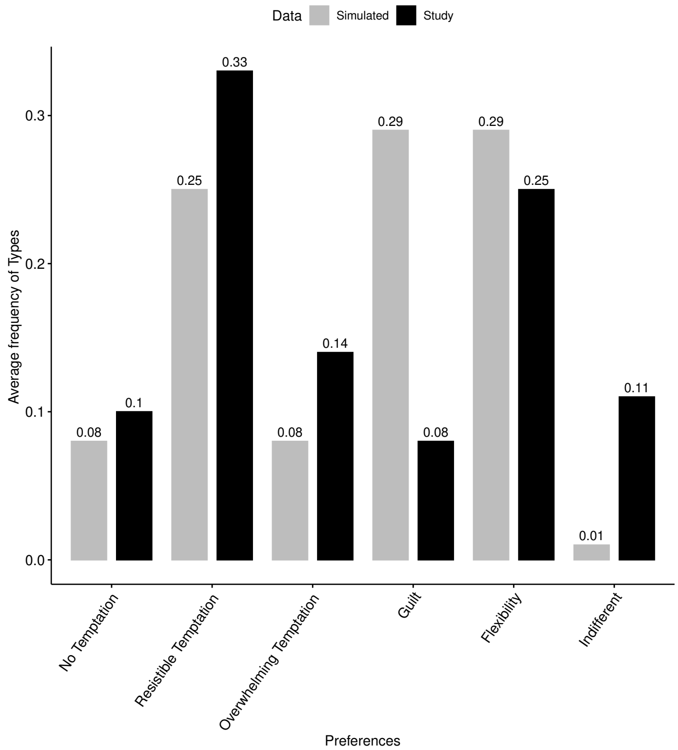


Figure C1: Frequency of Simulated Choice Types

The figure shows the observed proportions of modelled preferences in the entire study sample and the frequencies of simulated preferences. The proportion of the study preference types are not the same as reported in the main text, as we do not consider first-stage WTP values here and construct menu-dependent preferences solely based on menu rankings.

No Temptation preferences. Resistible Temptation preferences constitute 25% of menu rankings. The proportion of Overwhelming Temptation, Guilt and Flexibility preferences are 8%, 29%, and 29%, respectively. The proportion on Indifferent types is only 1% in our simulated sample. Figure C1 compares our simulation results to overall observed sample frequencies in our experiment.

In the second part of our analysis, using the ch-square test, we compare each subject's data from our study to our simulated random benchmark. All subjects pass our test and exhibit menu ranking patterns statistically different than our simulated random benchmark.